

UNICA

UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

Ph.D. DEGREE IN
Physics

Cycle XXXVIII

TITLE OF THE Ph.D. THESIS

Measurement of the Electroweak W Boson Production via Vector-Boson Fusion in Proton-Proton Collisions at √s = 13 TeV with the CMS Detector

Scientific Disciplinary Sector(s)

FIS-01

Ph.D. Student:	Andrea Claudio Maria Bulla
Supervisor	Prof. Pierluigi Bortignon
Co-Supervisor	Dott.ssa Mia Tosi

Final exam. Academic Year 2024/2025
Thesis defence session: February 2026

Abstract

This work reports a precise measurement of the electroweak production of a W boson in association with two jets ($pp \to Wjj$) at $\sqrt{s} = 13$ TeV, using the full Run II dataset recorded by the CMS experiment at the CERN Large Hadron Collider (LHC), corresponding to an integrated luminosity of 138 fb^{-1}. The process, characterised by the vector boson fusion (VBF) topology, provides a sensitive probe of the gauge structure of the Standard Model (SM) and of possible deviations from it.

The first part of the work establishes the theoretical and statistical foundations of the measurement, introducing the mathematical structure of the SM and its Effective Field Theory (EFT) extension, together with the likelihood-based methods used for parameter inference. A brief description of the LHC and the CMS detector follows, outlining the experimental environment that enables precision measurements of electroweak processes.

The analysis targets single-lepton final states ($W \to \ell\nu jj$, $\ell = e, \mu$) consistent with the VBF topology. Particular emphasis is placed on the reliability of simulated samples and on the data-driven corrections applied to improve their modelling. A deep neural network classifier enhances the separation between electroweak and QCD components, and the signal strength is extracted with a simultaneous binned profile-likelihood fit, yielding an expected total uncertainty of 10%.

Fiducial and differential measurements are also performed within a restricted particle-level phase space to reduce model dependence. Unfolded cross sections are extracted as functions of several generator-level observables, and an innovative unfolding strategy based on multiclass neural networks is compared to a traditional approach relying on reconstructed observables only.

Finally, the measurement is interpreted in the context of the Standard Model Effective Field Theory (SMEFT), providing sensitivity to higher-dimensional operators that modify electroweak interactions. An independent EFT interpretation of the Wjj process is followed by a global CMS combination including all major bosonic channels, in which each input uses fully simulated EFT signal samples with their associated experimental and theoretical uncertainties. This combined result offers a coherent picture of the electroweak sector within the EFT framework and delivers globally consistent constraints on potential deviations from the SM.

Contents

1 Mathematical foundations **1**
- 1.1 Group Theory . . . 1
- 1.2 Poincaré group . . . 2
- 1.3 The Standard Model Formalism . . . 4
- 1.4 Euler-Lagrange Equation . . . 5
- 1.5 Noether's theorem . . . 5
- 1.6 Gauge transformations . . . 6
- 1.7 The Standard Model of Particle Physics . . . 8
  - 1.7.1 Quantum Electrodynamics . . . 8
  - 1.7.2 Quantum Chromodynamics . . . 9
  - 1.7.3 Weak Interactions and Electroweak unification . . . 10
  - 1.7.4 Spontaneous symmetry breaking and Higgs mechanism . . . 13
  - 1.7.5 Gauge and matter fields of the Standard Model . . . 14
- 1.8 Effective Field Theory . . . 15
  - 1.8.1 Counting dimensions and EFT renormalization . . . 16
  - 1.8.2 The Standard Model Effective Field Theory . . . 17
  - 1.8.3 SMEFT at dimension-six . . . 18
  - 1.8.4 Electroweak input parameters . . . 19
  - 1.8.5 Flavour assumptions at dimension-six . . . 22
- 1.9 Statistical Inference . . . 23
  - 1.9.1 Construction of the likelihood . . . 24
  - 1.9.2 Maximum likelihood fits and parameter estimation . . . 25
  - 1.9.3 Observed, Asimov, and data-Asimov datasets . . . 26
  - 1.9.4 Hypothesis testing, significance and exclusion . . . 26
  - 1.9.5 Treatment of uncertainties . . . 27
  - 1.9.6 Likelihood Scans and Systematic Breakdown . . . 28
  - 1.9.7 EFT statistical model in COMBINE . . . 29

2 The LHC and CMS **31**
- 2.1 The Large Hadron Collider . . . 31
- 2.2 The Compact Muon Solenoid . . . 35
  - 2.2.1 Tracking system . . . 36
  - 2.2.2 Electromagnetic Calorimeter . . . 37
  - 2.2.3 Hadron Calorimeter . . . 39
  - 2.2.4 Muon chambers . . . 40
  - 2.2.5 Trigger system . . . 42

3 Electroweak W Production **47**
3.1 Signal Topology . . . 48
3.2 Backgrounds . . . 49
3.2.1 QCD-induced W+jets . . . 49
3.2.2 Top-quark backgrounds . . . 49
3.2.3 Subdominant backgrounds . . . 51
3.3 Data and Simulation . . . 52
3.3.1 Overview of Run-II data taking . . . 52
3.3.2 Data samples and triggers . . . 53
3.3.3 Luminosity and pileup modeling . . . 54
3.3.4 Monte Carlo simulation . . . 54
3.4 Object definitions and event selection . . . 60
3.4.1 Electrons . . . 60
3.4.2 Muons . . . 60
3.4.3 Jets and b tagging . . . 62
3.4.4 Missing transverse momentum . . . 62
3.4.5 Event selection . . . 62
3.5 Background Modelling . . . 65
3.5.1 Top background . . . 66
3.5.2 W + Jets background . . . 67
3.5.3 Non-prompt background . . . 73
3.6 Signal Discrimination . . . 86
3.6.1 Training setup . . . 86
3.6.2 Architecture . . . 86
3.6.3 Input variable selection . . . 87
3.6.4 Training performance . . . 88
3.6.5 Performance comparison . . . 89
3.7 Systematic Uncertainties . . . 94
3.7.1 Uncertainties Specific to Individual Background Processes 94
3.7.2 Uncertainties affecting all the simulated processes . . . 95
3.7.3 Theoretical Uncertainties . . . 97
3.8 Inclusive Results . . . 110
3.8.1 Inclusive cross section measurement . . . 131

4 Fiducial Measurements **135**
4.1 Inclusive Cross Section . . . 135
4.2 Differential Cross Section . . . 142
4.2.1 Unfolding strategy . . . 142
4.2.2 Alternative multiclass approach . . . 143
4.2.3 Unfolded m_{jj} distribution . . . 150
4.2.4 Unfolded $p_T(jj)$ distribution . . . 156
4.2.5 Unfolded $\Delta\eta_{jj}$ distribution . . . 160
4.2.6 Unfolded $\Delta\phi_{jj}$ distribution . . . 163
4.2.7 Unfolded $p_T(j1)$ distribution . . . 166
4.2.8 Unfolded $p_T(j2)$ distribution . . . 169
4.2.9 Unfolded $p_T(l)$ distribution . . . 172
4.2.10 Unfolded $p_T(W)$ distribution . . . 175

5 SMEFT Interpretation **179**
5.1 Effective Field Theory Interpretation . . . 179
5.2 EFT Signal Modelling . . . 180
5.2.1 Implementation in COMBINE . . . 182
5.3 Single-channel EFT interpretation . . . 185
5.3.1 Strategy . . . 185
5.3.2 Results . . . 186
5.4 Multi-channel EFT interpretation . . . 190
5.4.1 Motivation for a global approach . . . 190
5.4.2 Inputs of the combination . . . 190
5.4.3 Contribution from the $W \to \ell\nu jj$ channel . . . 191
5.4.4 Results . . . 193

6 Conclusions **199**

Bibliography **200**

A W+Jets Correction **215**
A.1 Alternative rewieght . . . 215
A.2 Closure Test for The Two Methods . . . 217

B Fake and Prompt Rates **219**
B.1 FR PR - 2017 . . . 219
B.2 FR PR - 2016HIPM . . . 222
B.3 FR PR - 2016noHIPM . . . 224
B.4 FR PR - $\Delta\phi(\ell_1, \text{MET})$-binned . . . 226

C Input Variables to the binary DNN **229**

D Differential Measurement Impact Plots **233**
D.1 m_{jj} impacts . . . 234
D.2 $p_T(jj)$ impacts . . . 237
D.3 $\Delta\eta_{jj}$ impacts . . . 239
D.4 $\Delta\phi_{jj}$ impacts . . . 241
D.5 $p_T(j_1)$ impacts . . . 242
D.6 $p_T(j_2)$ impacts . . . 243
D.7 $p_T(\ell)$ impacts . . . 244
D.8 $p_T(W)$ impacts . . . 246

E EFT Impact plots **247**

Chapter 1

Mathematical foundations of the Standard Model

The Standard Model (SM) of particle physics, a quantum field theory based on local gauge symmetries, provides a remarkably accurate description of the fundamental interactions among elementary particles. Its formulation represents the culmination of decades of theoretical and experimental progress that took place throughout the 20th century.

From a mathematical standpoint, the SM is built upon two key principles that have shaped our understanding of high-energy phenomena: Noether's theorem and gauge invariance. These concepts, introduced respectively in Sections 1.5 and 1.6, establish the link between symmetries and conservation laws, and determine how fundamental fields interact with one another. Although these topics are purely mathematical in nature, they provide the essential framework from which all modern quantum field theories, including the SM itself, are derived.

The first part of this chapter is therefore devoted to introducing these mathematical foundations. The discussion then illustrates how local symmetries give rise to gauge fields and interactions, leading naturally to the formulation of the SM, comprising the gauge, Higgs, and Yukawa sectors (Section 1.7).

Building upon this structure, the framework of Effective Field Theories (Section 1.8) is introduced as a systematic way to extend the SM by incorporating possible higher-dimensional operators that encode effects of new physics at higher energy scales. Finally, the chapter concludes with a discussion of the statistical inference methods (Section 1.9) that allow these theoretical models to be quantitatively tested against experimental data.

1.1 Group Theory

The natural playground to deal with symmetries is group theory. Physicists are particularly interested in a special branch of group theory that studies continuous symmetries called Lie theory. Lie groups are differentiable manifolds that

provide the mathematical foundation for the concepts of spatial translations, Lorentz boosts or rotations in an arbitrary space. The interest of physics into continuously generated groups is based on the fact that the general element can be reached by the repeated action of infinitesimal elements. The process may be thought of as an infinitesimal symmetry motion. Any group element g, indinitesimaly close to the identity $\mathbb{1}$ can be written as $g(\alpha) = \mathbb{1} + i\alpha^a T^a$ where α^a is the infinitesimal group parameter and T^a are Hermitial operators called generators of the symmetry group. We can write a general transformation for a finite parameter θ as the repeated action of an infinitesimal transformation $h(\theta) = (\mathbb{1} + i\alpha^a T^a) \cdot (\mathbb{1} + i\alpha^a T^a) \cdot ... = (\mathbb{1} + i\alpha^a T^a)^k$. Expressing the infinitesimal transformation as a function of a large number N we get $h(\theta) = \lim_{N\to\infty}(\mathbb{1}+\theta^a T^a/N)^N = e^{\theta^a T^a}$. The object T^a generates the finite transformation h, which is why it's called the generator [1]. We learn from the Baker-Campbell-Hausdorff-Formula [2] that the natural product of the Lie algebra is the Lie bracket $[,]$. As the set of generators T^a must span the space of infinitesimal group transformations, the commutator of the generators must be a linear combination of them $[T^a, T^b] = T^a T^b - T^b T^a = i f^{abc} T^c$ where the numbers f^{abc} are called structure constants and summarize the entire group multiplication law. The Jacobi identity $[T^a, [T^b, T^c]] + [T^b, [T^c, T^a]] + [T^c, [T^a, T^b]] = 0$, which is satisfied by the Lie algebra generators, implies that the structure constants obey $f^{ade} f^{bcd} + f^{bde} f^{cad} + f^{cde} f^{abd} = 0$.

If one of the generators T^a commutes with all the others,

$$[T^a, T^b] = 0 \qquad \forall\, b,$$

it spans an independent continuous Abelian subgroup. Such a subgroup has the structure of the group of phase rotations

$$\psi \;\rightarrow\; e^{i\alpha}\psi,$$

which is denoted as $U(1)$ [3]. In contrast, when at least two generators do not commute, the corresponding Lie group is non-Abelian, namely its elements are not commutative under the group multiplication law.

Lie groups have been extensively classified by Killing and Cartan in the 19th century (for an historical review see [4]). Almost all compact and simple Lie groups can be identified as member of one of three families, known as classical groups namely unitary transformations of N-dimensional vectors ($SU(N)$), orthogonal transformations of N-dimensional vectors ($SO(N)$) and symplectic transformations of N-dimensional vectors ($S_p(N)$).

1.2 Poincaré group

The historical turning point for the development of modern QFTs is Einstein's theory of special relativity [5]. By postulating that phenomena are perceived the same in all inertial frames of reference and that the velocity of light c is invariant

under a change of inertial frame of reference, humanity discovered that space and time are deeply connected. The concept of symmetry is manifest when we look for invariants under a transformation of the frame of reference. As there is no preferred frame of reference, physics laws should be written equally for all observers. The group of transformations permitted by the special relativity constraints is called the Poincaré group [6] which is the semidirect product between the Lorentz group and the group of translations in spacetime, as the velocity of light should be invariant under translations.

If we want to preserve the postulates of special relativity the quadratic form $ds^2 = dx_\mu dx_\nu \eta^{\mu\nu} = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2$ (where η is the metric of the Minokowsky space [7] $\eta_{00} = 1$, $\eta_{11} = \eta_{22} = \eta_{33} = -1$) should be the same in all inertial frames of reference [8]:

$$ds'^2 = dx'_\mu dx'_\nu \eta^{\mu\nu} = ds^2 = dx_\mu dx_\nu \eta^{\mu\nu} \tag{1.1}$$

Therefore the transformed coordinates dx'_μ should be written as a function of an arbitrary constant a_μ and of a constant matrix Λ, such that the scalar product of the Minkowski spacetime is invariant [1]

$$x_\mu \to x'_\mu = \Lambda^\sigma_\mu dx_\sigma + a_\mu \quad ; \quad \eta_{\mu\nu} = \Lambda^\gamma_\mu \Lambda^\delta_\nu \eta_{\gamma\delta} \tag{1.2}$$

These transformations, forming the so-called inhomogeneous Lorentz group or Poincaré group, have the special property that the speed of light is the same in all inertial frames. If in one frame we measure $|d\bar{x}/dt| = c$ then $\eta_{\mu\nu} dx^\mu dx^\nu = dt^2 - d\bar{x}^2 = 0$ from which it follows that also in another inertial frame $\eta_{\mu\nu} dx'^\mu dx'^\nu = 0$ therefore $|d\bar{x}'/dt| = c$. From the right hand side of Equation 1.1 we can see that $\det(\Lambda) = \pm 1$ and if we look at the $\mu = \nu = 0$ component $\Lambda^\sigma_0 \eta_{\sigma\rho} \Lambda^\rho_0 = (\Lambda^0_0)^2 - \sum_{i>0} (\Lambda^i_0)^2 = \eta_{00} = 1$ from which two solutions arise $\Lambda^0_0 \geq 1$ and $\Lambda^0_0 \leq 1$. The four possible combinations of $\det(\Lambda)$ and value of Λ^0_0 form four disjoint groups not smoothly connected. The proper orthochronus Lorentz group, characterized by $\det(\Lambda) = +1$ and $\Lambda^0_0 \geq 1$, is of special interest as it is a Lie group, containing the identity and therefore all of its transformations can be built from infinitesimal transformations. Any Lorentz transformation other than the proper orthochronus ones can be built up as the product of a transformation from the latter group and one or both of two special discrete transformations known as time-reversal and space-inversion. For the inhomogeneous Lorentz group the identity is the transformation $\Lambda^\mu_\nu = \delta^\mu_\nu$, $a_\mu = 0$. If we want to study infinitesimal transformations around the identity we can write $\Lambda^\mu_\nu = \delta^\mu_\nu + \omega^\mu_\nu$, $a_\mu = \epsilon_\mu$ where ω^μ_ν is an antisymmetric second-rank tensor. An infinitesimal Lorentz transformation $U(\mathbb{1} + \omega, \epsilon)$ must then be equal to $\mathbb{1}$ plus terms linear in ω and ϵ [9]:

$$U(\mathbb{1} + \omega, \epsilon) = \mathbb{1} + \frac{1}{2} i \omega_{\mu\nu} J^{\mu\nu} - i \epsilon_\mu P^\mu + ... \tag{1.3}$$

Where P^μ and $J^{\mu\nu}$ are the generators of the inhomogeneous Lorentz group and must be Hermitian in order for $U(\mathbb{1} + \omega, \epsilon)$ to be unitary. In more familiar terms P^μ is the four-momentum operator while $J^{\mu\nu}$ is the relativistic angular

momentum tensor. In particular we see that P^1, P^2 and P^3 are the components of the three-momentum operator, P^0 is the energy operator and J^{23}, J^{31} and J^{12} are components of the angular momentum three-vector. The remaining generators J^{10}, J^{31} and J^{30} form a set of transformations that change the current reference frame into a coordinate system that moves with a different constant velocity, commonly referred to as a boost.

The operators, constructed from the generators, that commute with all other operators are Called Casimir operators [10]. It can be proved that for any simple Lie algebra the quadratic operator $T^2 = T^a T^a$ always commutes with the generators. The importance of these operators in QFT stems from the fact that they can serve as a label for any irreducible unitary representations of the Poincaré group. The latter has two Casimir operators:

$$\begin{aligned} &P_\mu P^\mu = m^2 \\ &W_\mu W^\mu = -m^2 s(s+1), \quad m > 0 \end{aligned} \tag{1.4}$$

Where P^μ is the four-momentum vector and $W_\mu = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} J^{\nu\rho} P^\sigma$ is called the Pauli-Lubanski four-vector which spacial components, in the rest frame, become the generators of the $SU(2)$ algebra [11]. The scalar quantity m is commonly referred to as the invariant mass of particles while s is referred to as spin.

The labels for the irreducible representations of the Poincaré group (m, s) are how elementary particles are classified and the origin of such quantities emerges in QFT in pure group-theoretical terms [12]. For example particles with spin 0, 1/2 and 1 are described respectively by the (0,0), $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ and the $(\frac{1}{2}, \frac{1}{2})$ representations of the Poincaré group.

Additional labels, called charges, arise spontaneously in the SM from internal symmetries and will be explained in the next chapters.

1.3 The Standard Model Formalism

The historical development of Quantum Field Theory (QFT) has been strongly linked to the canonical formalism [13–16]. In practice, however, it is often more convenient to work in the Lagrangian formulation, where Lorentz invariance can be implemented in a straightforward manner. A Lagrangian density that is manifestly covariant guarantees the construction of operators satisfying the commutation rules of the Poincaré algebra. This feature becomes crucial in non-Abelian gauge theories: attempting to define a consistent Hamiltonian without starting from a Lorentz-invariant Lagrangian would in fact be impractical [9].

Let us denote by x the spacetime coordinates, by $\phi(x)$ a generic field, and by $\partial_\mu \phi$ its derivatives. A Lagrangian density $\mathcal{L}(\phi, \partial_\mu \phi)$ then defines the action, a Lorentz scalar that encodes the dynamics of the system,

$$S = \int L \, dt = \int \mathcal{L}(\phi, \partial_\mu \phi) \, d^4 x. \tag{1.5}$$

From the Lagrangian formalism one can always recover the Hamiltonian formalism. Recalling the standard relation $H = \sum p\dot{q} - L$, the Hamiltonian density takes the form

$$H = \int d^3x \left[\frac{\partial \mathcal{L}}{\partial \dot{\phi}(\bar{x})} \dot{\phi}(\bar{x}) - \mathcal{L} \right], \tag{1.6}$$

where the conjugate momentum is defined in the usual way as $p(x) = \partial \mathcal{L} / \partial \dot{\phi}(x)$.

1.4 Euler-Lagrange Equation

The equation of motion for a system characterized by a Lagrangian density $\mathcal{L}$ are derived from the principle of least action. A system evolving from two configurations in a time Δt will do so along the path in configuration space that are stationary points of the action functional S namely $\delta S = 0$. The solution to the previous equation is known as the Euler-Lagrange equation of motion for a field

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \tag{1.7}$$

1.5 Noether's theorem

The concepts frequently employed by physicists to describe nature, such as conserved quantities, are directly associated with symmetries. The theorem summarizing this concept is known as Noether's theorem [17] and states that every continuous symmetry of the action of a physical system (S) with conservative forces has a corresponding conservation law. In classical field theory we can distinguish between spacetime symmetries acting only on spacetime coordinates such as translations $x_\mu \to x_\mu + a_\mu$ and internal symmetries that act on the fields themselves $\phi(x) \to \phi(x) + \alpha \Delta\phi(x)$. In the first case one can see a translation as a transformation of the field configuration $\phi(x) \to \phi(x) + a_\mu \partial^\mu \phi(x)$. Since the Lagrangian density is a scalar quantity, under an infinitesimal spacetime translation $x^\mu \to x^\mu + a^\mu$ it is expected to transform covariantly as $\Delta\mathcal{L} = a_\mu \partial^\mu \mathcal{L}$ Here $\Delta\mathcal{L}$ denotes the variation of the Lagrangian density due to the variation of the spacetime coordinates. On the other hand, the same translation can be viewed as a change in the fields entering $\mathcal{L}(\phi, \partial_\mu \phi)$, and therefore $\Delta\mathcal{L}$ can also be computed explicitly by varying ϕ and its derivatives. Comparing the expected geometric transformation with the explicit variation, one obtains $\Delta\mathcal{L}$ to the results obtained by varying fields

$$a_\mu \partial^\mu \mathcal{L} = a_\nu \partial^\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi \right) + a_\nu \left[\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) \right] \partial_\nu \phi = 0 \tag{1.8}$$

The term in square brackets vanishes by the Euler-Lagrange equation 1.7 and we are left with

$$a_\nu \partial^\mu (\delta^\nu_\mu \mathcal{L}) - a_\nu \partial^\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi \right) = 0; \quad \partial^\mu T^\nu_\mu = 0 \tag{1.9}$$

where T^{ν}_{μ} is the canonical energy–momentum tensor of the field ϕ.

We see from Equation 1.9 that we have four conserved currents (one for each ν). For time translations ($\nu = 0$) the conserved charge is the Hamiltonian or total energy of the system $\partial^0 T^0_0 = -\partial^i T^0_i$. For spacial translations ($\nu = i$, $i = 1, 2, 3$) the total momentum of the field configuration P_i is conserved. Also interestingly, the conserved charges associated with a particular symmetry (such as spatial translations, rotations as well as all internal symmetries) act as generators of the symmetry itself [9].

More generally, the conservation law for a quantity Q_a called charge, can be expressed with the help of the divergence theorem yielding

$$Q_a = \int d^{n-1}x j^0_a \tag{1.10}$$

where j^{μ}_a is a tensor (which rank depends on the transformation) called the conserved current and n is the dimension of the manifold.

1.6 Gauge transformations

From Maxwell equations for electrodynamics we know that fields are defined in terms of derivatives of the potentials so there is an infinite family of possible potentials that will lead to the same physical phenomena. For example electrostatic potential is defined up to an arbitrary additive constant. When dealing with QFT, the fundamental fields cannot be observed directly but we can observe some of their properties. In field theories, various arrangements of unobservable fields can lead to identical observable outcomes. The process of changing from one such field configuration to another is termed a gauge transformation. As the properties of the system are not changed by such transformation we usually refer to this as a gauge symmetry of the system and, therefore, we should also find conserved quantities. Suppose to have a theory that is invariant under the global symmetry $U(1)$ of phase rotations $\psi \to e^{i\alpha}\psi$. In QFT it is natural to consider local transformations, i.e. a point in spacetime is influenced only by its neighbouring points in order to avoid action at a distance. The local requirement for such transformations is translated into the space-time dependence of the α parameter: $\alpha \to \alpha(x)$ that implies $\psi(x) \to e^{i\alpha(x)}\psi(x) = C(x)\psi(x)$.

At this stage, the crucial difference with the global case is evident: the action of the ordinary derivative ∂_μ on the transformed field does not factorise anymore. For a global transformation, one simply has $\partial_\mu\psi(x) \to e^{i\alpha}\partial_\mu\psi(x)$, so the derivative transforms with the same phase as the field. For the local case, however,

$$\partial_\mu\psi(x) \;\longrightarrow\; e^{i\alpha(x)}\Big(\partial_\mu\psi(x) + i(\partial_\mu\alpha(x))\,\psi(x)\Big), \tag{1.11}$$

and an additional term proportional to $\partial_\mu\alpha(x)$ spoils the invariance of the kinetic term.

The same fact can be seen explicitly using the finite-difference definition of the derivative:

$$n^\mu \partial_\mu \psi(x) = \lim_{\epsilon \to 0} \frac{\psi(x^\mu + n^\mu \epsilon) - \psi(x)}{\epsilon}. \tag{1.12}$$

Under a local transformation the two fields transform differently,

$$\psi(x) \to C(x)\psi(x), \qquad \psi(x + n\epsilon) \to C(x + n\epsilon)\psi(x + n\epsilon),$$

so the subtraction of neighbouring points is not well defined without an additional structure. In order to compare field values at distinct spacetime points in a consistent way, we introduce the object $U(y, x)$, called Wilson line [18], that transforms under $U(1)$ as a pure phase, $U(y, x) \to e^{i\alpha(y)} U(y, x) e^{-i\alpha(x)}$, with $U(x, x) = \mathbb{1}$.

We can therefore define a covariant derivative that transforms in the same way as the fields:

$$\eta^\mu D_\mu \psi(x) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \Big[\psi(x + \epsilon\eta) - U(x + \epsilon\eta, x)\psi(x) \Big], \tag{1.13}$$

so that the undesired space-time dependence of $\alpha(x)$ is absorbed into $U(y, x)$. If $U(y, x)$ is continuous in its spacetime arguments, then as the $U(1)$ group is a Lie group we can Taylor expand around the identity for infinitesimal displacements:

$$U(x + \epsilon n, x) = \mathbb{1} - ie\epsilon n^\mu A_\mu(x) + \mathcal{O}(\epsilon^2). \tag{1.14}$$

Here e is a coupling constant, while the coefficient of the displacement is a new vector field $A_\mu(x)$, which in mathematical terms is called an Ehresmann connection [19]. Taking the infinitesimal limit, the covariant derivative reduces to the familiar form

$$D_\mu \psi(x) = \partial_\mu \psi(x) + ieA_\mu(x)\psi(x), \tag{1.15}$$

with the transformation law

$$A_\mu(x) \to A_\mu(x) - \frac{1}{e} \partial_\mu \alpha(x). \tag{1.16}$$

We are now able to construct all possible Lagrangians invariant under a local symmetry. Any combination of ψ and its covariant derivatives will also be locally invariant. To complete the construction one needs a kinetic term for the new vector field A_μ, which must be invariant by itself and depend only on A_μ and its derivatives. This term arises from the commutator of two covariant derivatives,

$$[D_\mu, D_\nu] = ie(\partial_\mu A_\nu - \partial_\nu A_\mu) = ieF_{\mu\nu}, \tag{1.17}$$

where $F_{\mu\nu}$ is the electromagnetic field strength tensor.

From this simple example — which provides the basis for quantum electrodynamics by interpreting ψ as the electron field, A_μ as the electromagnetic potential, and e as the electric charge — we understand the geometrical origin of gauge fields as Ehresmann connections [19], arising naturally from the requirement of local invariance.

1.7 The Standard Model of Particle Physics

This section is dedicated to the SM of particle physics, which provides the theoretical framework describing elementary particles and their interactions. Imposing local gauge invariance in the SM as a quantum field theory amounts to requiring that the SM Lagrangian remains invariant under local transformations belonging to specific Lie groups. The locality condition enforces that these transformations depend explicitly on space-time. The SM exhibits invariance under local gauge transformations of the symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_{Y_L} \tag{1.18}$$

which accounts for three of the four known fundamental interactions in nature: the strong, weak, and electromagnetic forces. The incorporation of the fourth interaction, gravity, within the same theoretical framework remains one of the most compelling open challenges in modern physics.

1.7.1 Quantum Electrodynamics

The first step toward the modern SM has been the development of Quantum Electro-Dynamics (QED). The Lagrangian density of the free electron field (Dirac's Lagrangian [20, 21]) ψ transforming under the $(1/2,0)\oplus(0,1/2)$ representation of the Lorentz group is $\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi$ where $\bar{\psi} = \psi^\dagger\gamma^0$ and m is the Dirac mass. The dimension of the γ^μ matrices in Minkowsky space is at least 4×4 and they obey the so-called Dirac algebra $\{\gamma^\mu,\gamma^\nu\} = 2\eta^{\mu\nu}\times\mathbb{1}_{n\times n}$. We can write a representation of the Lorentz algebra using γ matrices as $\mathcal{S}^{\mu\nu} = i/4\,[\gamma^\mu,\gamma^\nu]$. In particular $\mathcal{S}^{0i}$ are the generators of the three-dimensional rotations while $\mathcal{S}^{ij}$ are the generators of Lorentz boost. A four-component field ψ that transforms under boosts and rotations as $\mathcal{S}$ is called a Dirac spinor.

From diagonal properties of the γ matrices in the chiral basis, one can show that the Dirac representation of the Lorentz group is reducible and we can form two 2-dimensional representation by considering diagonal blocks separately $\psi = (\psi_L,\psi_R)$ where ψ_L and ψ_R are eigenstates of the chirality γ_5

$$\psi_L = \frac{1-\gamma_5}{2}\psi; \quad \psi_R = \frac{1+\gamma_5}{2}\psi \tag{1.19}$$

It can be seen that $\mathcal{L}$ is manifestly invariant under a global $U(1)$ symmetry. It was already showed in chapter 1.6 that if one promotes the symmetry by requiring locality, Lorentz invariance is guaranteed by replacing normal derivatives ∂_μ to covariant ones D_μ transforming the same way as the fields ψ namely $\mathcal{L} = i\bar{\psi}(x)(\gamma^\mu\partial_\mu - m)\psi(x) - e\bar{\psi}(x)\gamma^\mu A_\mu\psi(x)$ where we notice a new term summarising the interaction between the fermion field and gauge field. The requirement of the free Lagrangian to be invariant under the local gauge transformation has led to a new term describing the dynamics of the fields. The only Lorentz invariant term, summarising the kinetic behaviour of the new gauge field A_μ, is

the tensor $F^{\mu\nu}$ described in chapter 1.6. In order to recover classical electrodynamics, we can write a physical spin 1 Proca Lagrangian [22, 23] for the gauge field with the relative coupling to the fermionic field via covariant derivative. A mass term for A_μ would explicitly break local gauge invariance, hence it is forbidden. As a result, the gauge field remains massless, as expected for the photon.

$$\mathcal{L}_{QED} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.20}$$

The coupling constant e is usually identified with the electric charge and its conservation follows from Noether theorem 1.5. When the Euler-Lagrange procedure is applied, we recover the Dirac equation in the electromagnetic field $i\gamma^\mu\partial_\mu - m\psi = e\gamma^\mu A_\mu\psi$

1.7.2 Quantum Chromodynamics

The other sectors of the SM are built on the same principles of the earlier QED but with more complex symmetry structures. Yang-Mills theories [24] generalize QED in a deep way by expanding local invariance to non Abelian symmetry groups while retaining renomalizability [25, 26].

Quantum Chromodynamics (QCD) is the SM sector devoted to the strong force. Its symmetry group is the non Abelian $SU(3)_C$ with 8 independent generators i.e. 8 massless spin 1 fields as force carriers. The introduction of a non vanishing commutator between the generators of such group implies that the gauge charges, that are conserved thanks to Noether's theorem, are gauge-dependent (i.e. they are not invariant under gauge transformations) and the gauge fields themselves carry gauge charges. The fundamental charge associated with the symmetry is called color and comes in N = 3 types.

The only matter field interacting via strong force is the quark $q_f(x)$ which is a spin 1/2 fermion with fractional electric charge and color charge thus obeying Dirac's equation. Quarks come in different flavours f. The coviariant derivative associated with a local $SU(3)$ symmetry reads as $D_\mu = \partial_\mu - ig_S G^a_\mu \lambda^a/2$ where the G^a_μ are the eight gluon gauge fields. The kinetic term of such gauge fields presents the peculiar self-interaction term $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_S f^{abc} G^b_\mu G^c_\nu$ coming from the Lie algebra of $SU(3)$. As for QED, the gauge-invariant kinetic term for the gauge fields reads as $-G^a_{\mu\nu}G^{a\mu\nu}/4$. The Lagrangian density for QCD [3] is build analogously to what was done in QED namely a kinetic term for the gauge fields and the Dirac's term for spin 1/2 fermions

$$\mathcal{L}_{QCD} = \sum_f \bar{q}_f(x)(i\gamma^\mu D_\mu - m_f)q_f(x) - \frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} \tag{1.21}$$

Color-charged particles cannot be isolated. For sufficiently strong couplings, QCD exhibits color confinment [18], meaning that the only finite-energy asymptotic states are singlets of color $SU(3)_C$ [27].

The QCD Lagrangian density has additional $U(1)$ and $SU(N_f)$ global symmetries. The former rotates by the same angle R and L components and the

associated conserved quantity is the baryon number B i.e. the net number of quarks (right-handed plus left-handed) is conserved in QCD interactions. The $SU(N_f)$ is decomposed in the so-called isospin symmetry (with charges I, I_3), which is a rotation in flavour space of the N_f quark fields as QCD interactions are flavour-blind, and a chiral symmetry where L and R quark field components are separately invariant as QCD interactions preserve parity. However the $SU(N_f)$ symmetry is spontaneously broken [28] meaning that the QCD vacuum state is not invariant under the same symmetries as $\mathcal{L}$. QCD has a non vanishing vacuum expectation value for a quark pair condensate $\left\langle 0|\bar{Q}Q|0\right\rangle = \left\langle 0|\bar{Q}_L Q_R + \bar{Q}_R Q_L|0\right\rangle \neq 0$ and this explains why only 1% of the mass of the proton comes from its constituents. $SU(N_f)$ is almost a symmetry for $N_f = 2$ that rotates up and down type quarks due to the fact that $m_u \simeq m_d$. $N_f = 3$ implies that rotating u, d and s quark fields leads to invariance under $SU(3)$ but this is only approximate as the strange quark is more than an order of magnitude heavier than the up and down quarks.

1.7.3 Weak Interactions and Electroweak unification

The observations of a third force began in the early 1900 while studying the radioactive properties of elements. The experimental evidences puzzled physics for half a century. The first proposed interpretation of the neutron β-decay came from Fermi, describing it as a four-point interaction [29]. Following the experiments of C. S. Wu et al. [30] it was understood that left and right eigenstates of the chirality defined in Eq. 1.20 behave differently under the weak force thus violating spatial parity $P : x \to -x$. Christenson et. al. [31] also proved that weak interactions violate the product of charge conjugation and parity by examining the decay of the long-lived neutral kaon K^0_L. The conserved Noether's current (mediated by electrically charged carriers due to conservation of the electromagnetic charge) for a theory describing weak interactions should maximally violate parity, thus presenting a vector and an axial component $V - A$. For leptons l such a current reads as $J_{\mu,l} = \bar{\psi}_{\nu_l}\gamma_\mu(1-\gamma_5)\psi_l = J^V_\mu - J^A_\mu = 2\bar{\psi}^L_{\nu_l}\gamma_\mu\psi^L_l$ where it is explicit that right-handed fermions do not interact.

A theory of charged currents with SM formalism would require a gauge symmetry group with two generators. Such a group does not exist, the closest would be SU(2) with three generators. Such a theory would generate two charged gauge fields $W^\pm$ behaving correctly under the weak force hypothesis of left-handedness and a neutral weak gauge boson Z which would couple to both left and right handed states. However the weak force acts feebly at a short distance thus it should be mediated by massive spin-1 vector bosons (to recover Fermi theory at low energy). On the contrary, QED is mediated by the massless photon and acts at an infinite range. Moreover, the construction of fermion mass terms like $-m_l(\bar{l}_L l_R + \bar{l}_R l_L)$ is forbidden by gauge invariance as l_L and l_R belong to different $SU(2)$ representations thus will have different $U(1)$ charges.

The solution to all these problems came with the electroweak unification from Glashow [32], Salam [33] and Weinberg [34] exploiting the mechanism of spon-

taneous symmetry breaking, precisely the Higgs [35], Englert, Brout [36] (BEH) mechanism [37]. The EW theory is a chiral theory, meaning that ψ_L and ψ_R behave differently under the gauge group: ψ_L transforms as a doubled under $SU(2)_L$ while ψ_R as a singlet. We can define left-handed quark and lepton $SU(2)_L$ doublets and will denote as ψ_L a general left-handed fermion field in this representation

$$q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \quad l_L = \begin{pmatrix} v_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} v_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} v_{\tau L} \\ \tau_L \end{pmatrix} \tag{1.22}$$

We assume for the electroweak Yang-Mills theory a gauge group $SU(2)_L \times U(1)_{Y_L}$ which generators are called respectively weak isospin T^a and weak hypercharge Y_L. The gauge fields associated with such generators are called W^a_μ for $SU(2)_L$ and B_μ for $U(1)_{Y_L}$. Before electroweak symmetry breaking the Lagrangian density is composed of four pieces $\mathcal{L}_{EW} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_h + \mathcal{L}_y$ denoting respectively the gauge bosons kinetic term, kinetic term of SM fermions, the Higgs field Lagrangian and lastly Yukawa fermion interactions. Focusing on the first two terms, the Lagrangian density reads as

$$\mathcal{L}_{g+f} = -\frac{1}{4} W^{\mu\nu}_a W^a_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} + \bar{\psi}_L i\gamma^\mu D_\mu \psi_L + \bar{\psi}_R i\gamma^\mu D_\mu \psi_R, \tag{1.23}$$

where $W^a_{\mu\nu}$ and $B_{\mu\nu}$ are the non-abelian and abelian field-strength tensors, defined as in Chapters 1.7.1 and 1.7.2. For a generic fermion multiplet the covariant derivative is

$$D_\mu = \partial_\mu - ig\, T^a W^a_\mu - ig'\, Y_L B_\mu. \tag{1.24}$$

The generator of the unbroken electromagnetic group is $Q = T^3 + Y_L$.

To express the gauge fields in a basis where the action of Q is diagonal, it is convenient to form linear combinations of the original $SU(2)_L$ and $U(1)_{Y_L}$ fields. This procedure makes explicit which combinations correspond to physical states of definite electric charge.

The two charged weak bosons are defined as

$$W^\pm_\mu = \frac{1}{\sqrt{2}} \left(W^1_\mu \mp i W^2_\mu \right), \tag{1.25}$$

which transform with electric charge ± 1 and therefore mediate charged weak interactions.

The remaining neutral fields W^3_μ and B_μ mix through an orthogonal transformation parametrised by the Weinberg angle θ_ω:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_\omega & -\sin\theta_\omega \\ \sin\theta_\omega & \cos\theta_\omega \end{pmatrix} \begin{pmatrix} W^3_\mu \\ B_\mu \end{pmatrix}. \tag{1.26}$$

The angle is fixed by requiring that the unbroken $U(1)_{\text{em}}$ couples with strength e, leading to the relation $e = g \sin\theta_\omega = g' \cos\theta_\omega$, which guarantees that the photon field A_μ remains massless and couples universally to the electromagnetic current.

In this basis the gauge fields correspond to the physical electroweak bosons: the charged mediators $W^{\pm}$, the neutral massive boson Z_μ, and the massless photon A_μ. The covariant derivative in terms of the mass eigenstates takes the form

$$D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}\left(W_\mu^+ T^+ + W_\mu^- T^-\right) - \frac{ig}{\cos\theta_\omega} Z_\mu \left(T^3 - \sin\theta_\omega{}^2 Q\right) - ieA_\mu Q \tag{1.27}$$

Charged current couplings are then defined as [38]

$$\mathcal{L}_{CC} = \frac{g}{\sqrt{2}}\left[\bar{\psi}_L T^+ W_\mu^+ \psi_L + \bar{\psi}_L T^- W_\mu^- \psi_L\right] \tag{1.28}$$

While with A_μ one recovers the traditional QED neutral currents, we have an additional neutral vector current mediated by a gauge field Z_μ

$$\mathcal{L}_{NC} = \frac{g}{2cos\theta_\omega}\left[\bar{\psi}_L T^3 Z\mu - 2\bar{\psi}Qsin^2\theta_\omega \psi Z^\mu\right] \tag{1.29}$$

We can note some things about weak interactions. Charged weak interactions are currently the sole recognized interactions capable of altering particle flavor. Weak neutral currents can couple both left and right-handed fermion fields however the flavour change is suppressed.

Gauge self-interactions

The self-interactions of the electroweak gauge bosons are encoded in the gauge-kinetic term

$$\mathcal{L}_g = -\frac{1}{4} W^a_{\mu\nu} W^{a\,\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.30}$$

with

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\,\epsilon^{abc} W^b_\mu W^c_\nu, \qquad B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \tag{1.31}$$

The non-abelian term proportional to $g\,\epsilon^{abc}$ generates self-interactions among the weak gauge fields. After rotating to the mass eigenstate basis $W^\pm_\mu, Z_\mu, A_\mu$ defined above, the gauge Lagrangian can be rewritten as

$$\mathcal{L}_g = -\frac{1}{2} W^+_{\mu\nu} W^{-\,\mu\nu} - \frac{1}{4} Z_{\mu\nu} Z^{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \mathcal{L}_{WWV} + \mathcal{L}_{WWVV}, \tag{1.32}$$

where $F_{\mu\nu}$ and $Z_{\mu\nu}$ are the photon and Z-boson field-strength tensors, and $\mathcal{L}_{WWV}$ and $\mathcal{L}_{WWVV}$ denote, respectively, the triple and quartic gauge couplings. The vertices in $\mathcal{L}_{WWV}$ correspond to the $WW\gamma$ and WWZ interactions, while $\mathcal{L}_{WWVV}$ contains the quartic couplings $WW\gamma\gamma$, $WWZZ$, and $WWZ\gamma$. These non-abelian self-interactions are a characteristic prediction of the $SU(2)_L$ gauge structure and play a central role in processes such as vector-boson fusion and vector-boson scattering.

1.7.4 Spontaneous symmetry breaking and Higgs mechanism

The solution to the problem of mass generation of electroweak vector bosons and fermions while retaining gauge invariance is known as BEH mechanism and is described in $\mathcal{L}_h$. Consider a system of complex scalar fields $\phi \equiv (\phi_u, \phi_d)$ transforming as a doublet under $SU(2)$. If one applies an $SU(2)$ infinitesimal transformation on this system, the difference of the transformed and the original fields depends on the commutator of the $SU(2)$ generators $\tau_k/2$, $k = 1, 2, 3$ (τ_k referred to as Pauli matrices) as $\delta S(k) \equiv \phi' - \phi \simeq i\epsilon_k(x)[\tau_k/2, \phi]$ where $\epsilon_k(x)$ is the infinitesimal parameter. This equation is valid even when we take the expectation value of the fields on vacuum (v.e.v) $\langle 0|\phi|0\rangle = \langle\phi\rangle_0$. If one postulates a non-vanishing v.e.v. for at least one component of ϕ, then none of the generators can annihilate the vacuum $\tau_k|0\rangle \neq 0$ and all the commutators $[\tau_k/2, S]$ are not zero. Thus if we want to impose $SU(2)$ symmetry on this system we find that it is possible to write an invariant Lagrangian but the vacuum will not be invariant. This behaviour is called spontaneous symmetry breaking and a $SU(2)$ invariant Lagrangian with such a property can be written by allowing S to obey to a Coleman-Weinberg potential $V(\phi)$ [39]

$$\mathcal{L}_h = (D_\mu\phi)^\dagger(D^\mu\phi) + \underbrace{\mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2}_{-V(\phi^\dagger\phi)} \tag{1.33}$$

where $\lambda > 0$ ensures the stability of the potential and $\mu^2 > 0$ makes its minimum occur at a nonzero field value. The potential thus acquires the characteristic "Mexican-hat" shape, featuring a continuous set of degenerate minima lying on a circle of radius $|\phi| = \sqrt{\mu^2/2\lambda}$.

Assuming, without loss of generality, that the vacuum expectation value points along the real direction, we set $\langle\phi_d\rangle_0 = v/\sqrt{2} = \sqrt{\mu^2/2\lambda}$ and $\langle\phi_u\rangle_0 = 0$. The field can then be expanded around this new minimum as

$$\phi(x) = \frac{e^{i\theta_k\frac{\tau_k}{2}}}{\sqrt{2}}\begin{pmatrix}0\\ v+h(x)\end{pmatrix} \approx \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_1(x)+i\eta_2(x)\\ v+h(x)+i\eta_3(x)\end{pmatrix} \xrightarrow{SU(2)} \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v+h(x)\end{pmatrix} \tag{1.34}$$

where $\eta_i(x)$ are massless spin-0 Nambu-Goldstone bosons [28, 40], one for each spontaneosuly broken $SU(2)$ generator. No such boson has been observed in nature so they should mathematically disappear. This is true only because the symmetry $SU(2)$ was promoted to a local symmetry thus we can gauge away the Nambu-Goldstone bosons with a suitable gauge choice getting rid of nonphysical degrees of freedom. This is not true if the symmetry is global. For this task it is convenient to work in the unitary gauge which comprises a polar parametrization of the doublet field and a rotation in the $SU(2)$ space to get rid of the exponential factor and is highlighted by $\xrightarrow{SU(2)}$ in equation 1.34. The only physical field h is referred to as the Higgs field and in unitary gauge the potential $V(\phi^\dagger\phi)$ takes the form $V(\phi^\dagger\phi) = -\mu^2h^2 - \lambda vh^3 - \lambda h^4/4$ showing that the quantum of the Higgs field is a scalar boson with mass $m_h = \sqrt{2}\mu = \sqrt{2\lambda}v$.

The origin of the electroweak gauge boson masses comes from the kinetic term of ϕ $(D_\mu\phi)(D^\mu\phi)$ under the $SU(2)_L \otimes U(1)_{Y_L}$ symmetry group. If we take equation 1.24 for the covariant derivative and evaluate it on the v.e.v. retaining only interesting terms

$$\begin{aligned}\langle (D_\mu\phi)(D^\mu\phi)\rangle_0 &= \frac{1}{2}\begin{pmatrix}0 & v\end{pmatrix}\left(gW_\mu^a T^a + g'Y_L B_\mu\right)^2\begin{pmatrix}0\\ v\end{pmatrix} \\ &= \frac{1}{2}\frac{v^2}{4}\left[g^2(A_\mu^1)^2 + g^2(A_\mu^2)^2 + (-gA_\mu^3 + g'B_\mu)^2\right]\end{aligned} \tag{1.35}$$

Where we can see that there are three massive gauge bosons with mass proportional to the v.e.v. and a massless one.

$$\begin{aligned}W_\mu^\pm &= \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2); \quad m_W = g\frac{v}{2} \\ Z_\mu &= \frac{1}{\sqrt{g^2+g'^2}}(gA_\mu^3 - g'B_\mu); \quad m_Z = \sqrt{g^2+g'^2}\frac{v}{2} \\ A_\mu &= \frac{1}{\sqrt{g^2+g'^2}}(g'A_\mu^3 + gB_\mu); \quad m_A = 0\end{aligned} \tag{1.36}$$

The combinations $W^\pm$, Z_μ, and A_μ, previously introduced (Eqs. 1.25–1.26) as the fields diagonalising the electromagnetic charge operator, now acquire a direct physical meaning: after electroweak symmetry breaking, they become the actual mass eigenstates of the theory.

If one allows the Higgs field to couple with fermion fields through a Yukawa term, then it can provide the fermion with mass as well. The scalar field allows for gauge invariant terms which couple the left-handed doublet and right-handed singlet states of the form $\bar{\psi}_L\Gamma\psi_R\phi$ where Γ include the coupling constants and make the Yukawa couplings invariant under the Lorentz and gauge groups. After the symmetry is broken, the Yukawa Lagrangian can be written as

$$\mathcal{L}_y = -\frac{v+h}{\sqrt{2}}\sum_f^{fermions} y_f\bar{\psi}_L^f\psi_R^f \tag{1.37}$$

Where y_i are the Yukawa couplings or proportionality constant between the mass of a fermion and the Higgs v.e.v. The fermion mass can be written as $m_f = vy_f/\sqrt{2}$.

1.7.5 Gauge and matter fields of the Standard Model

The content of the SM is commonly categorized into matter and gauge fields. The former encompasses all spin 1/2 fermionic matter fields obeying Fermi-Dirac statistics, comprising leptons and quarks. On the other hand, the latter encompasses the spin 1 gauge fields following Bose-Einstein statistics and responsible for mediating forces. An extra scalar field, unrelated to any fundamental force, is present – the Higgs field. The matter fields within the SM are divided into three nearly identical generations, possessing identical quantum numbers but distinct masses. Table 1.1 summarises spin 1/2 fields while Table 1.2 gauge bosons of the $SU(3)_C \times SU(2)_L \times U(1)_{Y_L}$ gauge symmetry.

	ψ	Mass	Q	Color	$\mathbf{T_3}$	$\mathbf{Y_L}$
$\mathbf{1}^{st}$	$\boldsymbol{\nu_e}$	< 225 eV	0	–	L: $+\frac{1}{2}$; R: 0	L: -1; R: 0
	$\boldsymbol{e}$	510.99895000(15) keV	-1	–	L: $-\frac{1}{2}$; R: 0	L: -1; R: -2
	$\boldsymbol{u}$	$2.16^{+0.49}_{-0.26}$ MeV	$+\frac{2}{3}$	r,g,b	L: $+\frac{1}{2}$; R: 0	L: $+\frac{1}{3}$; R: $+\frac{4}{3}$
	$\boldsymbol{d}$	$4.67^{+0.48}_{-0.17}$ MeV	$-\frac{1}{3}$	r,g,b	L: $-\frac{1}{2}$; R: 0	L: $+\frac{1}{3}$; R: $-\frac{2}{3}$
$\mathbf{2}^{nd}$	$\boldsymbol{\nu_\mu}$	< 0.19 MeV	0	–	L: $+\frac{1}{2}$; R: 0	L: -1; R: 0
	$\boldsymbol{\mu}$	105.6583755(23) MeV	-1	–	L: $-\frac{1}{2}$; R: 0	L: -1; R: -2
	$\boldsymbol{c}$	1.27 ± 0.02 GeV	$+\frac{2}{3}$	r,g,b	L: $+\frac{1}{2}$; R: 0	L: $+\frac{1}{3}$; R: $+\frac{4}{3}$
	$\boldsymbol{s}$	$93.4^{+8.6}_{-3.4}$ MeV	$-\frac{1}{3}$	r,g,b	L: $-\frac{1}{2}$; R: 0	L: $+\frac{1}{3}$; R: $-\frac{2}{3}$
$\mathbf{3}^{rd}$	$\boldsymbol{\nu_\tau}$	< 18.2 MeV	0	–	L: $+\frac{1}{2}$; R: 0	L: -1; R: 0
	$\boldsymbol{\tau}$	1.77693(9) GeV	-1	–	L: $-\frac{1}{2}$; R: 0	L: -1; R: -2
	$\boldsymbol{t}$	172.69 ± 0.30 GeV	$+\frac{2}{3}$	r,g,b	L: $+\frac{1}{2}$; R: 0	L: $+\frac{1}{3}$; R: $+\frac{4}{3}$
	$\boldsymbol{b}$	$4.18^{+0.03}_{-0.02}$ GeV	$-\frac{1}{3}$	r,g,b	L: $-\frac{1}{2}$; R: 0	L: $+\frac{1}{3}$; R: $-\frac{2}{3}$

Table 1.1: Lepton (e, μ, τ) and quark (u, d, c, s, t, b) masses and charges, treated as spin-$\frac{1}{2}$ fermions. Upper limits, where reported, correspond to a 95% confidence level [41].

	Spin	Mass	Q	Color	$\mathbf{T_3}$	$\mathbf{Y_L}$
Photon ($\boldsymbol{\gamma}$)	1	$< 10^{-18}$ eV	0	–	0	0
Gluon ($\boldsymbol{g}$)	1	~ 0	0	r,g,b	0	0
$\boldsymbol{Z^0}$	1	91.1876 ± 0.0021 GeV	0	–	0	0
$\boldsymbol{W^\pm}$	1	80.377 ± 0.012 GeV	± 1	–	± 1	0
$\boldsymbol{h}$	0	125.25 ± 0.17 GeV	0	–	0	0

Table 1.2: Gauge bosons: spin, masses [41], and charges.

1.8 Effective Field Theory

In the absence of the observation of a new particle, indirect searches for new physics can be fundamental in the discovery of a new theoretical framework. Often, indirect searches are ambiguous as they can be an indication of several possible models. This is due to the decoupling theorem [42] which essentially states that if the energy scale of new physics Λ is far greater than the probed energy E then Feynman amplitudes involving heavy masses as internal lines may be safely neglected. However this also means that we can study the low energy approximation of many different UV-complete and decoupled theories in a model-independent fashion. This is the concept underlying effective field theories (EFT) which, simply stated, are just general quantum field theories

based on the principles of relativity and quantum mechanics but without the restriction of renormalizability and valid up to a certain energy scale Λ. In many cases, EFTs are the low-energy limit of a more complete theory, just as the Fermi 4-point interaction is the low-energy approximation of the electroweak theory.

Effective Field Theories enable the calculation of an experimentally measurable quantity with a finite error. Formally, this is done by expanding the existing theory with new higher-dimensional terms as a function of a small expansion parameter called the power counting. Such EFT expansion is systematic in the sense that there is a well-defined procedure to compute higher order corrections in the power counting, thus reducing the theoretical error of the predictions.

1.8.1 Counting dimensions and EFT renormalization

The fact that the SM Lagrangian has mass dimension 4 can be understood by requiring that the action be dimensionless in natural units ($\hbar = c = 1$). Since the action is given by

$$S = \int d^4x \, \mathcal{L}(x), \tag{1.38}$$

and the space-time measure has dimension $[d^4x] = -4$, it follows that

$$[\mathcal{L}] = 4. \tag{1.39}$$

This fixes the overall dimension of the Lagrangian density.

For the covariant derivative we note that it is defined as

$$D_\mu = \partial_\mu - igA_\mu. \tag{1.40}$$

The ordinary derivative has dimension $[\partial_\mu] = 1$, since coordinates scale as $[x] = -1$ and derivatives invert this. Consistency then requires the gauge field to have $[A_\mu] = 1$, so that the product gA_μ has the same dimension as ∂_μ. Therefore

$$[D_\mu] = 1. \tag{1.41}$$

With these assignments all the kinetic and interaction terms of the SM — such as $D_\mu\phi^\dagger D^\mu\phi$ for scalars, $\bar{\psi}D_\mu\psi$ for fermions, and $X_{\mu\nu}X^{\mu\nu}$ for gauge bosons — indeed have total dimension 4, as required. From this condition we can directly extract the canonical mass dimensions of the fields and couplings. For instance, the scalar kinetic term fixes the dimension of the scalar field, the gauge kinetic term fixes that of the gauge field, and the fermionic kinetic term fixes the fermion field dimension. Finally, Yukawa interactions determine the dimension of the Yukawa couplings. Explicitly:

$$\begin{aligned}
&[D_\mu\phi^\dagger D^\mu\phi] = 4 \to [\phi] = 1 \\
&[X_{\mu\nu}X^{\mu\nu}] = 4 \to [X_\mu] = 1 \\
&[\bar{\psi}D_\mu\psi] = 4 \to [\psi] = \frac{3}{2} \\
&[\bar{\psi}y_L\phi\psi] = 4 \to [y_L] = 0
\end{aligned} \tag{1.42}$$

Going into higher dimensions we introduce effective operators $\mathcal{L}+ = c\mathcal{O}^d$ capturing the new physics effects where the coupling constant c is called the Wilson coefficient and gauges the action of the operator $\mathcal{O}$ while d is the mass dimension of the operator. However, it is known that the SM Lagrangian at $d = 4$ is renormalizable while $d > 4$ theories are not. It can be shown that EFTs are renormalizable order by order i.e. cutting the expansion at a mass dimension d. Any EFT prediction at a certain order in the power counting is, within its uncertainty, a valid description of nature.

Let's assume a Feynman diagram that contains the loop integral $\int F(p, ...)d^4p$. We define the superficial degree of divergence [3] D by the UV behavior of the intergral such that it is given by $\int dp\, p^{D-1}$. This implies that if $D < 0$ the integral is convergent, $D = 0$ the integral is log-divergent and if $D > 0$ it is divergent. We can derive D from the topology of the diagram considering the number of external lines E_f where f is the particle specie, the spin of the propagator s_f, the number of involved verices N_k and the dimension of the operator involved in the interaction d_k as

$$D = 4 - \sum_f E_f(1 + s_f) + \sum_k N_k d_k \tag{1.43}$$

If one assumes to approximate the loop with an EFT vertex, in the limit of infinite mass for the propagator in the loop, we see that the insertion of one ($N_k = 1$) or two ($N_k = 2$) operators of dimension $d_k = 6$ make the integral divergent with $D = 6$ and $D = 8$. If we want the theory to be renormalizable we need counter terms of this dimension to balance the divergencies. The term $-\sum_k N_k(d_k - 4)$ is exactly the power of the of mass dimension that regularises the theory. For example using the canonical mass dimension Λ as the power counting parameter we obtain $c\mathcal{O}^d \to c\mathcal{O}^d/\Lambda^{d-4}$.

1.8.2 The Standard Model Effective Field Theory

The Standard Model Effective Field Theory (SMEFT) [43] is a consistent EFT generalization of the SM constructed out of a series of $SU(3)_C \times SU(2)_L \times U(1)_{Y_L}$ invariant higher dimensional operators built from SM fields and including an h field embedded in an $SU(2)$ doublet (see [44] for a recent review). The SMEFT Lagrangian is a systematic expansion of the SM Lagrangian with higher order terms sorted by canonical dimension

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + \mathcal{L}^{(7)} + ...; \quad \mathcal{L}^{(d)} = \sum_i^{n_d} \frac{c_i^{(d)}}{\Lambda^{d-4}} \mathcal{O}_i^{(d)} \tag{1.44}$$

Here, Λ represents the scale of new physics, and it is assumed to be significantly greater than the masses of SM particles and the kinematic scales of the processes under investigation. This is indicated by $(E/\Lambda) << 1$, suggesting a state of near decoupling between the new physics sector and the observed processes.

At dimension five, $\mathcal{L}^{(5)}$ contains a single lepton-number-violating $\Delta L = 2$ operator and its Hermitian conjugate $\Delta L = -2$ operator. This operator, called

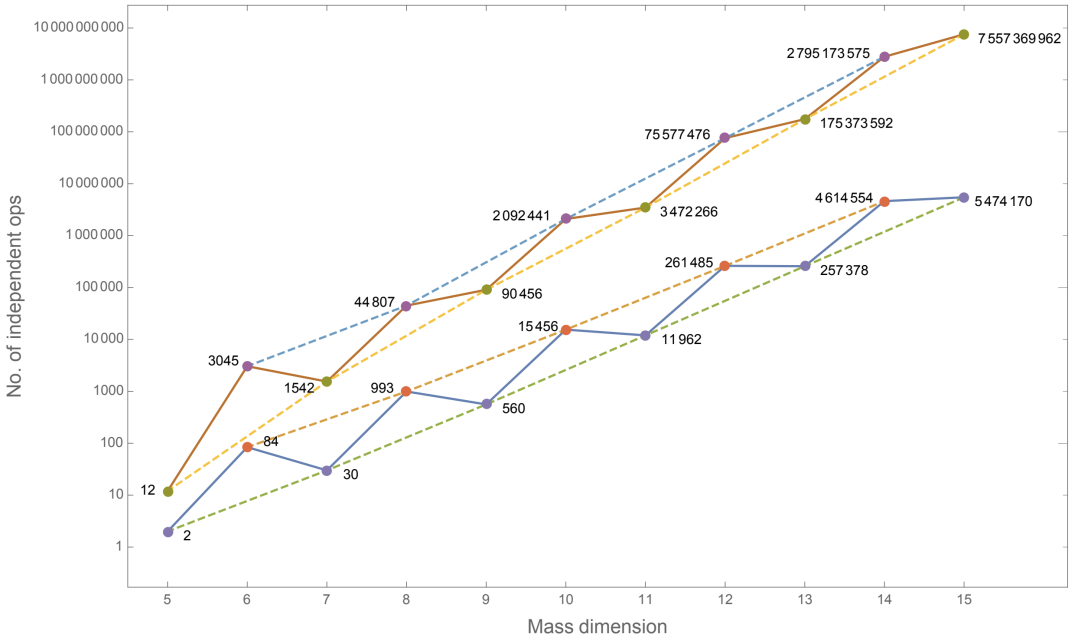


Figure 1.1: Growth of the number of independent operators in the SMEFT from [52]. At each mass dimension, lower points indicate the number of operators assuming one fermion generation while upper ones assume three generations. Dashed lines connect odd and even dimension mass term.

Weinberg operator [45] $\mathcal{L}_5 = c_{ll'}/\Lambda \left[\phi \cdot \bar{l}_L^c\right] [l'_L \cdot \phi]$ + h.c., give rise to dimension-three Majorana mass terms $m_{ll'} = c_{ll'} \cdot v^2/\Lambda$ for the left-handed neutrinos l_L in the spontaneously broken theory [46].

At dimension six, for three generations of fermions, there are 2499 independent dimension-six operators [47] that do not violate baryon number and lepton number $\Delta B = \Delta L = 0$. These operators give the leading SMEFT contributions under the assumption that the new physics energy scale is well below the lepton-number violation scale $\Lambda \ll \Lambda_{\not L}$. Additional 273 dimension-six terms can be constructed with $\Delta B = \Delta L = 1$ allowing for proton decays in SMEFT. Higher order SMEFT operators with $d = 7, 8, ...$ tamed by larger inverse powers of Λ have been extensively studied [48–52] and the number of operators at each mass dimension has been determined and summarised in Figure 1.1.

1.8.3 SMEFT at dimension-six

The leading baryon and lepton number conserving SMEFT contributions come from dimension-six operators. To identify the minimal set of independent operators, one should first enumerate all possible structures built from SM fields and invariant under the SM gauge group $SU(3)_C \times SU(2)_L \times U(1)_{Y_L}$. The naive counting gives rise to thousands of operators, precisely 2499, but many of them are not independent. Redundancies can be systematically removed by means of integration by parts, equations of motion, Fierz identities and symme-

try relations, all of which ensure that physically equivalent structures are not double counted. Incomplete sets of operators are often used in the literature as basis, such as the so-called HISZ basis [53], SILH basis [54] or the Buchmüller and Wyler result [43]. However, the first non-redundant and complete operator basis for $\mathcal{L}^{(6)}$ has come to be known as the "Warsaw basis" [47], comprising 59 independent operator classes (before accounting for flavour indices). Further developments led to a complete SILH basis in recent years [55].

	Fermions					**Scalars**
Field	$l^j_{L,p}$	$e_{R,p}$	$q^{\alpha,j}_{L,p}$	$u^\alpha_{R,p}$	$d^\alpha_{R,p}$	ϕ^j
$\boldsymbol{Y_L}$	$-\frac{1}{2}$	-1	$\frac{1}{6}$	$\frac{2}{3}$	$-\frac{1}{3}$	$\frac{1}{2}$

Table 1.3: SM matter content in $SU(2)$ representation where $j = 1, 2$ denotes the weak isospin, $\alpha = r, g, b$ colour and $p = 1, 2, 3$ is the fermion generation index .

The Warsaw basis is defined in terms of SM fields under in the $SU(2)$ representation and summarised in table 1.3, where $j = 1, 2$ denotes the weak isospin, $\alpha = r, g, b$ colour, $p = 1, 2, 3$ is the fermion generation index and ϕ is the Higgs $SU(2)$ doublet. The conjugate of the Higgs field ($\phi^\dagger$ or $\tilde{\phi}$) is defined though the two-index Levi-Civita tensor ϵ as $\phi^{\dagger,j} = \epsilon_{jk}(\phi^k)^*$. The Hermitian derivative acting on ϕ is defined as $\phi^\dagger i \overleftrightarrow{D}_\mu \phi = \phi^\dagger i D_\mu \phi - i D_\mu \phi^\dagger \phi$ and $\phi^\dagger i \overleftrightarrow{D}^i_\mu \phi = \phi^\dagger i \sigma^i D_\mu \phi - i D_\mu \phi^\dagger \sigma^i \phi$ where σ^i are Pauli matrices. Gauge field strength tensors (generically referred to as $X_{\mu\nu}$) are defined as in the SM for $SU(3)$, namely $G_{\mu\nu}$, $W^a_{\mu\nu}$ for $SU(2)$ and $B_{\mu\nu}$ for $U(1)$, as described in section 1.7.3.

According to the dimension counting described in section 1.8.1, we have various ways to construct dimension-six operators starting from SM terms and we can categorise them according to their field content. The bosonic classes are labelled X^3, ϕ^6, $\phi^4 D^2$, $X^2\phi^2$, $\psi^2 X\phi$, $\psi^2\phi^2 D$. The so-called four-fermion operators modelling point interactions are divided by fermions chirality as $(\bar{L}L)(\bar{L}L)$, $(\bar{R}R)(\bar{R}R)$, $(\bar{L}L)(\bar{R}R)$ and the mixed $(\bar{L}R)(\bar{R}L) + (\bar{L}R)(\bar{L}R)$. The full list of dimension-six operators in the Warsaw basis is given in Table 1.4.

1.8.4 Electroweak input parameters

The SM can make meaningful predictions only if some parameters of the theory, as measured by experiments, are provided as inputs. If one assumes massless neutrinos we need to provide the SM with 19 measured constants. If we allow neutrinos to be massive, 7 more parameters appear leading to a total of 26 as summarised in the following table, where θ_{ij} are the elements of the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix [56, 57] for flavour mixing, δ the CP-violating phase of the CKM matrix, and by U_{ij} the elements

X^3		ϕ^6 and $\phi^4 D^2$		$\psi^2\phi^3$	
Q_G	$f^{abc}G^{a\nu}_{\mu}G^{b\rho}_{\nu}G^{c\mu}_{\rho}$	Q_ϕ	$(\phi^\dagger\phi)^3$	$Q_{e\phi}$	$(\phi^\dagger\phi)(\bar{l}_p e_r \phi)$
$Q_{\tilde{G}}$	$f^{abc}\tilde{G}^{a\nu}_{\mu}G^{b\rho}_{\nu}G^{c\mu}_{\rho}$	$Q_{\phi\Box}$	$(\phi^\dagger\phi)\Box(\phi^\dagger\phi)$	$Q_{u\phi}$	$(\phi^\dagger\phi)(\bar{q}_p u_r \tilde{\phi})$
Q_W	$\epsilon^{ijk}W^{i\nu}_{\mu}W^{j\rho}_{\nu}W^{k\mu}_{\rho}$	$Q_{\phi D}$	$(D^\mu\phi)^\dagger(D_\mu\phi)$	$Q_{d\phi}$	$(\phi^\dagger\phi)(\bar{q}_p d_r \phi)$
$Q_{\tilde{W}}$	$\epsilon^{ijk}\tilde{W}^{i\nu}_{\mu}W^{j\rho}_{\nu}W^{k\mu}_{\rho}$				

$X^2\phi^2$		$\psi^2 X\phi$		$\psi^2\phi^2 D$	
$Q_{\phi G}$	$\phi^\dagger\phi\, G^a_{\mu\nu}G^{a\mu\nu}$	Q_{eW}	$(\bar{l}_p\sigma^{\mu\nu}e_r)\sigma^i\phi\, W^i_{\mu\nu}$	$Q^{(1)}_{\phi l}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{l}_p\gamma^\mu l_r)$
$Q_{\phi\tilde{G}}$	$\phi^\dagger\phi\, \tilde{G}^a_{\mu\nu}G^{a\mu\nu}$	Q_{eB}	$(\bar{l}_p\sigma^{\mu\nu}e_r)\phi\, B_{\mu\nu}$	$Q^{(3)}_{\phi l}$	$(\phi^\dagger i\overleftrightarrow{D}^i_\mu\phi)(\bar{l}_p\sigma^i\gamma^\mu l_r)$
$Q_{\phi W}$	$\phi^\dagger\phi\, W^i_{\mu\nu}W^{i\mu\nu}$	Q_{uG}	$(\bar{q}_p\sigma^{\mu\nu}T^a u_r)\tilde{\phi}\, G^a_{\mu\nu}$	$Q_{\phi e}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{e}_p\gamma^\mu e_r)$
$Q_{\phi\tilde{W}}$	$\phi^\dagger\phi\, \tilde{W}^i_{\mu\nu}W^{i\mu\nu}$	Q_{uW}	$(\bar{q}_p\sigma^{\mu\nu}u_r)\sigma^i\tilde{\phi}\, W^i_{\mu\nu}$	$Q^{(1)}_{\phi q}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{q}_p\gamma^\mu q_r)$
$Q_{\phi B}$	$\phi^\dagger\phi\, B_{\mu\nu}B^{\mu\nu}$	Q_{uB}	$(\bar{q}_p\sigma^{\mu\nu}u_r)\tilde{\phi}\, B_{\mu\nu}$	$Q^{(3)}_{\phi q}$	$(\phi^\dagger i\overleftrightarrow{D}^i_\mu\phi)(\bar{q}_p\sigma^i\gamma^\mu q_r)$
$Q_{\phi\tilde{B}}$	$\phi^\dagger\phi\, \tilde{B}_{\mu\nu}B^{\mu\nu}$	Q_{dG}	$(\bar{q}_p\sigma^{\mu\nu}T^a d_r)\phi\, G^a_{\mu\nu}$	$Q_{\phi u}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{u}_p\gamma^\mu u_r)$
$Q_{\phi WB}$	$\phi^\dagger\sigma^i\phi\, W^i_{\mu\nu}B^{\mu\nu}$	Q_{dW}	$(\bar{q}_p\sigma^{\mu\nu}d_r)\sigma^i\phi\, W^i_{\mu\nu}$	$Q_{\phi d}$	$(\phi^\dagger i\overleftrightarrow{D}_\mu\phi)(\bar{d}_p\gamma^\mu d_r)$
$Q_{\phi\tilde{W}B}$	$\phi^\dagger\sigma^i\phi\, \tilde{W}^i_{\mu\nu}B^{\mu\nu}$	Q_{dB}	$(\bar{q}_p\sigma^{\mu\nu}d_r)\phi\, B_{\mu\nu}$	$Q_{\phi ud}$	$i(\tilde{\phi}^\dagger D_\mu\phi)(\bar{u}_p\gamma^\mu d_r)$

$(\bar{L}L)(\bar{L}L)$		$(\bar{R}R)(\bar{R}R)$		$(\bar{L}R)(\bar{L}R)+(\bar{L}R)(\bar{R}L)$	
Q_{ll}	$(\bar{l}_p\gamma_\mu l_r)(\bar{l}_s\gamma^\mu l_t)$	Q_{ee}	$(\bar{e}_p\gamma_\mu e_r)(\bar{e}_s\gamma^\mu e_t)$	Q_{le}	$(\bar{l}_p\gamma_\mu l_r)(\bar{e}_s\gamma^\mu e_t)$
$Q^{(1)}_{qq}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{uu}	$(\bar{u}_p\gamma_\mu u_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{lu}	$(\bar{l}_p\gamma_\mu l_r)(\bar{u}_s\gamma^\mu u_t)$
$Q^{(3)}_{qq}$	$(\bar{q}_p\gamma_\mu\sigma^i q_r)(\bar{q}_s\gamma^\mu\sigma^i q_t)$	Q_{dd}	$(\bar{d}_p\gamma_\mu d_r)(\bar{d}_s\gamma^\mu d_t)$	Q_{ld}	$(\bar{l}_p\gamma_\mu l_r)(\bar{d}_s\gamma^\mu d_t)$
$Q^{(1)}_{lq}$	$(\bar{l}_p\gamma_\mu l_r)(\bar{q}_s\gamma^\mu q_t)$	Q_{eu}	$(\bar{e}_p\gamma_\mu e_r)(\bar{u}_s\gamma^\mu u_t)$	Q_{qe}	$(\bar{q}_p\gamma_\mu q_r)(\bar{e}_s\gamma^\mu e_t)$
$Q^{(3)}_{lq}$	$(\bar{l}_p\gamma_\mu\sigma^i l_r)(\bar{q}_s\gamma^\mu\sigma^i q_t)$	Q_{ed}	$(\bar{e}_p\gamma_\mu e_r)(\bar{d}_s\gamma^\mu d_t)$	$Q^{(1)}_{qu}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{u}_s\gamma^\mu u_t)$
		$Q^{(1)}_{ud}$	$(\bar{u}_p\gamma_\mu u_r)(\bar{d}_s\gamma^\mu d_t)$	$Q^{(8)}_{qu}$	$(\bar{q}_p\gamma_\mu T^a q_r)(\bar{u}_s\gamma^\mu T^a u_t)$
		$Q^{(8)}_{ud}$	$(\bar{u}_p\gamma_\mu T^a u_r)(\bar{d}_s\gamma^\mu T^a d_t)$	$Q^{(1)}_{qd}$	$(\bar{q}_p\gamma_\mu q_r)(\bar{d}_s\gamma^\mu d_t)$
				$Q^{(8)}_{qd}$	$(\bar{q}_p\gamma_\mu T^a q_r)(\bar{d}_s\gamma^\mu T^a d_t)$

$(\bar{L}R)(\bar{L}R)+(\bar{L}R)(\bar{R}R)$		Baryon number violating	
Q_{leqd}	$(\bar{l}^j_p e_r)(\bar{d}_s\, q^j_t)$	Q_{duq}	$\epsilon^{\alpha\beta\gamma}\epsilon_{jk}\left[(d^\alpha_p)^T C\, u^\beta_r\right]\left[(q^{\gamma j}_s)^T C\, l^k_t\right]$
$Q^{(1)}_{quqd}$	$(\bar{q}^j_p u_r)\epsilon_{jk}(\bar{q}^k_s d_t)$	Q_{qqu}	$\epsilon^{\alpha\beta\gamma}\epsilon_{jk}\left[(q^{\alpha j}_p)^T C\, q^{\beta k}_r\right]\left[(u^\gamma_s)^T C\, e_t\right]$
$Q^{(8)}_{quqd}$	$(\bar{q}^j_p T^a u_r)\epsilon_{jk}(\bar{q}^k_s T^a d_t)$	$Q^{(1)}_{qqq}$	$\epsilon^{\alpha\beta\gamma}\epsilon_{jk}\epsilon_{mn}\left[(q^{\alpha j}_p)^T C\, q^{\beta k}_r\right]\left[(q^{\gamma m}_s)^T C\, l^n_t\right]$
$Q^{(1)}_{lequ}$	$(\bar{l}^j_p e_r)\epsilon_{jk}(\bar{q}^k_s u_t)$	$Q^{(3)}_{qqq}$	$\epsilon^{\alpha\beta\gamma}(\sigma^i\epsilon)_{jk}(\sigma^i\epsilon)_{mn}\left[(q^{\alpha j}_p)^T C\, q^{\beta k}_r\right]\left[(q^{\gamma m}_s)^T C\, l^n_t\right]$
$Q^{(3)}_{lequ}$	$(\bar{l}^j_p\sigma_{\mu\nu}e_r)\,\epsilon_{jk}(\bar{q}^k_s\sigma^{\mu\nu}u_t)$	Q_{duu}	$\epsilon^{\alpha\beta\gamma}\left[(d^\alpha_p)^T C\, u^\beta_r\right]\left[(u^\gamma_s)^T C\, e_t\right]$

Table 1.4: Dimension-six operators in the Warsaw basis [47].

of the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [58, 59] with associated CP-violating phase δ' that accounts for neutrino flavour mixing. The Lagrangian parameters $g = \{g_1, ...g_N\}$ are fixed imposing a set of defining conditions that relate them to (pseudo-)observables $\mathcal{O} = \{\mathcal{O}_1, ...\mathcal{O}_M\}$ for $M \geq N$ as $\mathcal{O}_n = F^{(0)}_n(g)$ for $n = 1, ..., M$ where $F^{(0)}_n$ is a function of the parameters g at a chosen perturbative order [60]. If $M = N$ the system is invertible and we can recover the measured values of g from functions K of the observables $\mathcal{O}$ as $g_i = K^{(0)}_i(\mathcal{O})$ for $i = 1, ..., N$.

$\boldsymbol{m_\nu = 0}$	
α_S, θ_{QCD}	QCD
$y_e, y_\mu, y_\tau, y_u, y_d, y_c, y_s, y_t, y_b$	Yukawas
g, g', v, λ	EW+Higgs
$\theta_{12}, \theta_{13}, \theta_{23}, \delta$	CKM
$\boldsymbol{m_\nu \neq 0}$	
$y_{\nu_e}, y_{\nu_\mu}, y_{\nu_\tau}$	Yukawas
$U_{12}, U_{13}, U_{23}, \delta'$	PMNS

Going from SM to SMEFT several additional parameters are introduced, namely the cutoff scale Λ and the Wilson coefficients $c_i^{(d)}$. The problem is that the observables $\mathcal{O}$ used to fix the value of the SM parameters g will receive contributions from the SMEFT parameters

$$\begin{aligned} \mathcal{O}_n &= F_n^{(0)}(g) + \frac{1}{\Lambda^2} F_n^{(2)}(g, c^{(6)}) + \frac{1}{\Lambda^4} F_n^{(4)}(g, c^{(8)}) + ... \\ g_i &= K_i^{(0)}(\mathcal{O}) + \frac{1}{\Lambda^2} K_i^{(2)}(\mathcal{O}, c^{(6)}) + \frac{1}{\Lambda^4} K_i^{(4)}(\mathcal{O}, c^{(8)}) + ... \end{aligned} \tag{1.45}$$

A generic predicted observable $\mathcal{P} = P^{(0)}(g) + \sum_{d=2,4,8,...} P^{(d)}(g, c^{4+d})/\Lambda^d$ will receive direct SMEFT contributions from $P^{d\geq 2}$ and indirect SMEFT contributions from g, namely $F_n^{d\geq 2}$. The net effect is that all the EFT corrections to input measurements are recast into corrections to predicted quantities. It can be shown that if $\mathcal{P}$ is an input observable $\mathcal{O}_p$ used to measure g then all EFT corrections cancel order by order. If $\mathcal{P}$ is not an input observable, then its prediction will depend on the chosen set of observables $\mathcal{O}$ used to fix the SM values g. The difference in the predicted quantity $\mathcal{P}$ due to a different set of input observables is

$$\mathcal{P}(\mathcal{O}) - \mathcal{P}(\mathcal{O}') = \frac{1}{\Lambda^2} \left[-\frac{\partial P^{(0)}}{\partial F_n^{(0)}} F_n^{(2)} + \frac{\partial P^{(0)}}{\partial F'^{(0)}_n} F'^{(2)}_n \right] + \frac{1}{\Lambda^4} ... \tag{1.46}$$

It is essential to choose carefully the input observables set $\mathcal{O}$ as results for $\mathcal{P}$ depend upon the input parameter set chosen [61].

For the Higgs and electroweak sector the SM contains 4 independent quantities $g = \{g, g', v, \lambda\}$ that can be fixed by measuring observables in the set $\mathcal{O} = \{\alpha_{em}, G_F, m_W, m_Z, m_h\}$. The Higgs boson mass m_h is always needed to fix λ, but the choice of the remaining three is arbitrary. Three alternative schemes for electroweak input observables are available. The $\{\alpha_{em}, G_F, m_Z\}$ set was historically used by LEP and features the most precisely known parameters from the set. The downside is that m_W now receives SMEFT corrections that will enter in many other observables and in the propagator definition of the W boson, leading to a highly non-linear dependence of predictions upon Wilson

coefficients. The second set solves the previous problem by incorporating m_W in the input set, thus freezing its value to the measured one $\{G_F, m_Z, m_W\}$. Here "freezing" means that the experimental number is kept fixed, but one must still account for SMEFT corrections to the theoretical relation that links G_F to the vacuum expectation value v, so that the effects of higher dimensional operators reappear indirectly in other predictions. The last set presents $\{\alpha_{em}, m_Z, m_W\}$ and has the advantage of avoiding shifts in electroweak couplings that arise from processes depending on G_F, such as the muon decay. However it can lead to slightly worse electroweak convergences both in the SM and in SMEFT [62]. At the time being, the developers of SMEFT tools tend to favour the $\{G_F, m_Z, m_W\}$ set [63].

1.8.5 Flavour assumptions at dimension-six

While the number of independent electroweak structures amounts to only 59 independent dimension-six operators [47], a proliferation of Wilson coefficients occurs when accounting for all the possible flavour structures. If no flavour symmetry is imposed, the SMEFT Lagrangian of equation 1.44 will depend on 1350 CP-even plus 1149 CP-odd independent coefficients for the dimension-six operators [64]. To date, it is hopeless to attempt to constrain such a large parameter-space, therefore a dimensionality reduction strategy is needed. This is practically done by imposing a flavour symmetry (and symmetry-breaking sector) at the expense of introducing some model dependence. These symmetries are not necessarily fundamental symmetries of the underlying ultraviolet (UV) theory but could be accidental if e.g. the theory discriminates between fermion generations [65]. In the following non-exhaustive list, the most popular flavour symmetries will be explained.

Maximal $U(3)^5$ symmetry

The largest group of flavour-symmetry transformations compatible with the gauge symmetries of the SM Lagrangian is $U(3)^5$

$$U(3)^5 = U(3)_l \times U(3)_q \times U(3)_e \times U(3)_u \times U(3)_d = SU(3)^5 \times U(1)^5 \quad (1.47)$$

$U(3)^5$ essentially states that fermions with the same electroweak quantum numbers are not distinguished by gauge interactions [66]. The most restrictive assumption that can be made is that $U(3)^5$ is an exact symmetry of the beyond-the-SM sector. This is only valid if one neglects Yukawa couplings.

Linear Minimal Flavor Violation

In the SM Yukawa couplings $Y_{e,u,d}$ and the CKM phase δ explicitly break the $U(3)^5$ flavour symmetry. Flavour violation is realised by elevating Yukawa couplings to spurions of the flavour symmetry [67]. This involves enforcing a $U(3)^5$

symmetry on the fermion fields, which permits unrestricted $U(3)^5$-invariant insertions of these spurious parameters in the currents. These insertions are responsible for generating flavour-violating effects [60]. Spurion insertions are organized in a power expansion of the Yukawa couplings, e.g. for a single insertion $\mathcal{O}(Y_{u,d,e})$ or double insertions $\mathcal{O}(Y_u^2)$, $\mathcal{O}(Y_u Y_d)$, $\mathcal{O}(Y_u^2 Y_d)$ and so on [65].

$U(2)^3$ symmetry in the quark sector

The CKM description of flavour mixing and CP-violation in the quark sector is incredibly precise however the underlying mechanism generating quark masses and mixing is still unknown. The Minimal Flavour Violation anastz [68] with a broken $U(3)^3$ does not provide any clue about the origin of the hierarchal nature of the Yukawa couplings. In order to recover the level of flavour violation suppression observed in experiments, a minimally broken $U(2)^3$ flavor symmetry acting on the first two generations of quarks can arguably be more interesting [69, 70]. In this way, quarks of the first two generation transform differently under this symmetry. Denoting by (q_p, u_p, d_p) left and right handed fields for the first and second generations (with generation index $p = 1, 2$), by (Q, t, b) left and right handed fields for the third generation and with Ω a generic $U(2)$ transformation, the symmetry under which only light quarks transform can be written as

$$\begin{aligned} &U(2)^3 = U(2)_q \times U(2)_u \times U(2)_d \\ &q \to \Omega_q q, \quad u \to \Omega_u, \quad d \to \Omega d, \quad Q \to Q, \quad t \to t, \quad b \to b \end{aligned} \tag{1.48}$$

The lepton sector decouples from the quark sector and there is freedom to relax or tighten the lepton flavour restrictions. Several options are at use. The most restrictive case is the $U(3)^2 = U(3)_l \times U(3)_e$ symmetry, as done in the maximal assumption and Minimal Flavour Violation scenarios. One can increasingly relax this requirement imposing for example $U(2)^2 = U(2)_l \times U(2)_e$ (when combining with quark sector leads to $U(2)^5$ [69]) or even assume flavour diagonality $U(1)^3_{l+e} = U(1)_e \times U(1)_\mu \times U(1)_\tau$ [60, 71].

The number of parameters for each flavour assumption is given in Table 1.5.

1.9 Statistical inference

In high-energy physics, the extraction of physical parameters from observed data relies on the construction of a likelihood function that encodes all statistical and systematic information about the measurement. The COMBINE framework, developed within the CMS Collaboration, provides a unified implementation of this approach, allowing a wide range of statistical procedures, including hypothesis testing, confidence interval estimation, and signal strength measurements, to be performed within a common and reproducible environment [72].

	General		$U(3)^5$		**MFV**		$U(2)^3 \times U(1)^3$		$U(2)^3 \times U(3)^2$	
	All	CP	All	CP	All	CP	All	CP	All	CP
X^3	4	2	4	2	2	-	4	2	4	2
ϕ^6 and $\phi^4 D^2$	3	-	3	-	3	-	3	-	3	-
$\psi^2 X \phi$	8	4	8	4	4	-	8	4	8	4
$\psi^2 \phi^3$	54	27	6	3	7	-	14	7	10	5
$\psi^2 X \phi$	144	72	16	8	20	-	36	18	28	14
$\psi^2 \phi^2 D$	81	30	9	1	14	-	21	2	15	2
$(\bar{L}L)(\bar{L}L)$	297	126	8	-	10	-	31	-	16	-
$(\bar{R}R)(\bar{R}R)$	450	195	9	-	19	-	40	2	27	2
$(\bar{L}L)(\bar{R}R)$	648	288	8	-	28	-	54	4	31	4
$(\bar{L}R)(\bar{L}R) + (\bar{L}R)(\bar{R}L)$	810	405	14	7	13	-	64	32	40	20
total	2499	1149	85	25	120	-	275	71	182	53

Table 1.5: Number of independent parameters for each class of dimension-six operators as defined in the Warsaw basis under different flavour assumptions. Table from [60]

1.9.1 Construction of the likelihood

In statistical inference, the choice of the likelihood as the central object of analysis stems from the Neyman–Pearson lemma [73], which states that the ratio of two likelihood functions provides the most powerful test statistic for discriminating between competing hypotheses. This principle motivates the use of the likelihood ratio as the foundation of both parameter estimation and hypothesis testing in high-energy physics.

The fundamental object of any statistical analysis is the likelihood function,

$$L(\vec{\Phi}) = p(\text{data}; \vec{\Phi}), \tag{1.49}$$

which represents the probability density for the observed data given a set of model parameters $\vec{\Phi}$. For practical purposes, the parameter space is partitioned into *parameters of interest* (POIs), denoted by $\vec{\mu}$, and nuisance parameters, $\vec{\nu}$, which account for experimental and theoretical uncertainties. The likelihood used in COMBINE is factorized into primary and auxiliary components as:

$$p(\vec{x}, \vec{y}; \vec{\mu}, \vec{\nu}) = p(\vec{x}; \vec{\mu}, \vec{\nu}) \prod_k p_k(y_k; \nu_k), \tag{1.50}$$

where $\vec{x}$ represents the primary observables (e.g. binned event yields or distributions used in the fit) and $\vec{y}$ are auxiliary measurements constraining the nuisance parameters. Each nuisance parameter ν_k has an associated auxiliary observable y_k and probability density $p_k(y_k; \nu_k)$, which encodes the prior knowledge of that uncertainty. This structure ensures a consistent treatment of both statistical and systematic sources of uncertainty, and allows their propagation to the parameters of interest through profiling or marginalization.

1.9.2 Maximum likelihood fits and parameter estimation

Parameter estimation in COMBINE is based on the profile likelihood method. For a given choice of the parameters of interest (POIs) $\vec{\mu}$, the nuisance parameters $\vec{\nu}$ are profiled, that is, they are adjusted to the values that maximise the likelihood for each $\vec{\mu}$.

$$\hat{\hat{\vec{\nu}}}(\vec{\mu}) = \arg\max_{\vec{\nu}} L(\vec{\mu}, \vec{\nu}). \tag{1.51}$$

The global maximum of the likelihood defines the maximum-likelihood estimates (MLEs) $\hat{\vec{\mu}}$ and $\hat{\vec{\nu}}$.

The relative compatibility of different hypotheses is quantified through the profile likelihood ratio,

$$\lambda(\vec{\mu}) = \frac{L(\vec{\mu}, \hat{\hat{\vec{\nu}}}(\vec{\mu}))}{L(\hat{\vec{\mu}}, \hat{\vec{\nu}})}. \tag{1.52}$$

In COMBINE, this ratio is expressed through the test statistic

$$q(\vec{\mu}) = -2 \ln \lambda(\vec{\mu}), \tag{1.53}$$

whose distribution, under asymptotic conditions, can be approximated by a χ^2 function with one degree of freedom [74, 75]. The denominator of $\lambda(\vec{\mu})$ corresponds to the overall best-fit configuration, while the numerator encodes the constrained likelihood for a fixed value of $\vec{\mu}$.

Likelihood formulation for binned counting experiments. In most high-energy physics analyses, the observables entering the likelihood are binned event yields n_i corresponding to distinct regions of phase space or categories. Assuming independent Poisson statistics, the likelihood function is written as

$$L(\vec{n} \,|\, \vec{\mu}, \vec{\nu}) = \prod_i \frac{\left[\mu\, s_i(\vec{\nu}) + b_i(\vec{\nu})\right]^{n_i}}{n_i!}\, e^{-[\mu\, s_i(\vec{\nu}) + b_i(\vec{\nu})]}, \tag{1.54}$$

where $s_i(\vec{\nu})$ and $b_i(\vec{\nu})$ denote the expected signal and background contributions in bin i, and $\vec{\nu}$ represents the set of nuisance parameters describing systematic effects.

Results obtained before and after the fit optimization are referred to as pre-fit and post-fit, respectively, the latter incorporating the full constraints from data.

In analyses involving multiple correlated signal contributions, COMBINE can perform simultaneous fits of several POIs. This configuration, often referred to as a multi-signal model, is employed whenever different components or model parameters are correlated and must be measured together. Typical examples include global fits of models beyond the Standard Model (BSM), where several parameters contribute coherently to the prediction, or unfolding analyses in which multiple bins of a differential observable are fitted simultaneously, allowing correlations and shared systematic uncertainties to be properly propagated across all bins.

1.9.3 Observed, Asimov, and data-Asimov datasets

In addition to the observed dataset, corresponding to the real data collected by the experiment, two deterministic configurations of the likelihood model are widely used in the statistical inference procedures: the *Asimov* and the *data-Asimov* datasets.

Asimov dataset. Following the terminology introduced by Franchise and formalised by Cowan et al. [74], the Asimov dataset is a deterministic, fluctuation-free realization of the model in which each bin content is set equal to its expected value for a given signal strength μ_{inj} and nominal nuisance parameters $\vec{\nu}_0$:

$$n_i^A = \mu_{\text{inj}}\, s_i(\vec{\nu}_0) + b_i(\vec{\nu}_0). \tag{1.55}$$

In the asymptotic regime, the use of this idealised sample yields the same median behaviour of the test statistic as would be obtained from an ensemble of pseudo-experiments. It therefore allows expected sensitivities and confidence intervals to be evaluated without explicitly generating large numbers of toy datasets.

Data-Asimov dataset. A second configuration, referred to as the *data-Asimov* dataset, is constructed using the post-fit nuisance parameter values $\hat{\vec{\nu}}_{\text{data}}$ obtained from the fit to the observed data, while keeping the injected signal strength fixed to its nominal value $\mu_{\text{inj}} = 1$:

$$n_i^{A_{\text{data}}} = s_i(\hat{\vec{\nu}}_{\text{data}}) + b_i(\hat{\vec{\nu}}_{\text{data}}). \tag{1.56}$$

When this dataset is subsequently refitted, both the signal strength and the nuisance parameters are floated, but the likelihood is maximised exactly at $\hat{\mu} = 1$ and $\hat{\vec{\nu}} = \hat{\vec{\nu}}_{\text{data}}$. This construction reproduces the pattern of pulls and correlations observed in data while removing statistical fluctuations, and provides a smooth reference model for diagnostic studies such as impact plots and post-fit shapes.

1.9.4 Hypothesis testing, significance and exclusion

Once the likelihood model is defined, it can be used to perform statistical hypothesis testing. Two competing hypotheses are generally considered: the background-only hypothesis (H_b), corresponding to the SM expectation without a signal contribution ($\mu = 0$), and the signal-plus-background hypothesis (H_s), which includes both the SM background and a possible signal component scaled by one or more signal strength parameters $\vec{\mu}$.

The test statistic used to discriminate between the two hypotheses is the same profile-likelihood ratio introduced above, where large values of q_μ, defined as in Eq. 1.53, indicate increasing incompatibility between the data and the tested hypothesis H_μ.

If an excess consistent with a nonzero signal is observed, the probability of obtaining such a deviation under the background-only hypothesis is quantified using a one-sided p-value:

$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|H_b)\, dq_0, \tag{1.57}$$

where $f(q_0|H_b)$ denotes the probability density of the test statistic under H_b. The corresponding local significance, expressed in standard deviations, is

$$Z = \Phi^{-1}(1 - p_0), \tag{1.58}$$

where Φ^{-1} denotes the inverse cumulative distribution function of the standard normal distribution $\mathcal{N}(0, 1)$.

For exclusion limits, COMBINE employs the modified frequentist approach known as the CL_s method [76]. In this framework, the likelihood ratio test statistic q_μ is used to construct the probability distributions of the data under both the signal-plus-background (H_{s+b}) and background-only (H_b) hypotheses. From these distributions, one defines the corresponding p-values

$$p_{s+b} = \int_{q_{\text{obs}}}^{\infty} f(q|H_{s+b})\, dq, \qquad p_b = \int_{q_{\text{obs}}}^{\infty} f(q|H_b)\, dq, \tag{1.59}$$

which quantify the probability of observing a test statistic value at least as extreme as the one obtained in data, assuming each hypothesis to be true.

The modified frequentist confidence level is then defined as

$$CL_s = \frac{p_{s+b}}{1 - p_b}. \tag{1.60}$$

A signal hypothesis is excluded at 95% confidence level if $CL_s < 0.05$.

In practice, the distributions $f(q_\mu)$ are often approximated using the asymptotic results from Wilks' theorem [75] and Cowan et al. [74], which link the test statistic q_μ to a χ^2 distribution in the large-sample limit. This approach provides accurate and computationally efficient estimates of significances and confidence intervals without requiring extensive ensembles of pseudo-experiments.

1.9.5 Treatment of uncertainties

Systematic uncertainties enter the likelihood through dedicated nuisance parameters, each constrained by a corresponding probability term $p(y; \nu)$. The functional form of these constraints depends on the nature of the uncertainty and is specified in the COMBINE datacard. The most common types include:

Log-normal uncertainties are typically used for multiplicative effects such as luminosity or cross section variations. Gamma distributions are used for Poisson-limited control samples, while `rateParam` terms allow the normalization of specific processes to float freely or to be linked across channels according to analytical expressions, enabling data-driven background estimations such as ABCD methods.

Type	Directive	Mult. factor	Constraint term
Log-normal	`lnN`	κ^{ν}	Gaussian in ln space
Asymmetric log-normal	`lnN` (asym.)	$\kappa^{\nu}_{\text{up/down}}$	Gaussian in ln space
Log-uniform	`lnU`	κ^{ν}	Uniform in ln space
Gamma (Poisson-limited)	`gmN`	ν/N	Poisson term $P(y;\nu)$
Flat	`flatParam`	–	Uniform prior $U(a,b)$
Rate parameter	`rateParam`	user-defined	Uniform or analytic prior

Table 1.6: Summary of the most common uncertainty types supported in COMBINE.

Nuisance parameter impacts

An essential diagnostic of the fitting procedure is provided by the impact plot, which summarizes the effect of each nuisance parameter on the parameter of interest (POI), typically the signal strength μ. The computation of impacts proceeds through a series of profile-likelihood fits. For each nuisance parameter ν_k, the fit is repeated while varying ν_k by $\pm 1\sigma$ around its post-fit value, re-profiling all other parameters. The corresponding change in the best-fit POI is defined as the impact:

$$\Delta\mu_k^{\pm} = \hat{\mu}(\hat{\nu}_k \pm 1\sigma) - \hat{\mu}, \tag{1.61}$$

and the total impact is expressed as

$$\Delta\mu_k = \frac{1}{2}\left|\Delta\mu_k^{+} - \Delta\mu_k^{-}\right|. \tag{1.62}$$

The ranking of nuisance parameters is obtained by sorting them in descending order of their absolute impact $|\Delta\mu_k|$. Parameters with the largest impacts contribute the most to the overall uncertainty in the parameter of interest (POI), highlighting which experimental or theoretical sources dominate the measurement precision. Impact plots provide an intuitive visualization of how systematic uncertainties propagate to the result and their relative importance. Each row corresponds to one nuisance: the left panel shows its post-fit value (pull) and uncertainty, while the right panel shows the corresponding variation of the POI when the nuisance is shifted by $\pm 1\sigma$.

1.9.6 Likelihood Scans and Systematic Breakdown

Likelihood scans. To evaluate the compatibility of the data with a given signal hypothesis, and to extract the best-fit value of a POI, a likelihood scan is performed using the profile likelihood ratio defined in Eq. 1.53. In this procedure, the likelihood is repeatedly maximized for fixed values of the POI,profiling all nuisance parameters at each step.The resulting curve $-2\Delta \ln L(\mu)$, commonly referred to as the likelihood scan, encodes the statistical uncertainty in the parameter through its curvature around the minimum. Under the asymptotic

approximation, the interval corresponding to $\Delta(-2\ln L) = 1$ defines the 1σ (68% CL) confidence interval, while $\Delta(-2\ln L) = 3.84$ corresponds to 95% CL. This method is also used to compute the likelihood contours in two dimensions when fitting pairs of correlated POIs, for instance in EFT interpretations where two Wilson coefficients are allowed to vary simultaneously.

Breakdown of systematic uncertainties. To quantify the relative contribution of each group of systematic uncertainties, the fit is repeated multiple times while sequentially freezing subsets of nuisance parameters to their prefit (nominal) values. This effectively removes the corresponding sources of uncertainty from the likelihood model. The change in the fitted uncertainty in the parameter of interest, before and after freezing a given group, is interpreted as that group's fractional contribution to the total uncertainty.

1.9.7 EFT statistical model in Combine

The COMBINE framework, originally developed for Higgs boson measurements and searches, has been extended to accommodate effective field theory interpretations. In this approach, the expected event yield in each phase-space region can be expressed as a quadratic expansion in the Wilson coefficients:

$$N = \mathrm{SM} + \sum_i \left(c_i \cdot \mathrm{Lin}_i + c_i^2 \cdot \mathrm{Quad}_i\right) + \sum_{i<j} c_i c_j \cdot \mathrm{Mix}_{ij}, \tag{1.63}$$

where SM denotes the SM contribution, Lin_i the interference term between the SM and the operator Q_i, Quad_i the pure EFT quadratic term, and Mix_{ij} the mixed interference between operators Q_i and Q_j. The coefficients c_i represent the Wilson coefficients of the EFT operators.

A direct implementation of Eq. 1.63 in the statistical model may lead to negative bin contents, since interference terms can produce deficits with respect to the SM expectation. As the Poisson-based likelihoods used in ROOFIT and ROOSTAT [77, 78] require non-negative predictions, the model is reformulated to ensure that all templates entering the likelihood are positive-defined. This algebraic redefinition, originally introduced in the ANALYTICANOMALOUSCOUPLING framework developed by the Milano CMS group and publicly available at [79], allows EFT fits to be implemented in COMBINE using only positive templates.

For a single operator, Eq. 1.63 can be rewritten as

$$N = (1-c)\cdot \mathrm{SM} + c\cdot(\mathrm{SM} + \mathrm{Lin} + \mathrm{Quad}) + (c^2 - c)\cdot \mathrm{Quad}, \tag{1.64}$$

which is mathematically equivalent to using the three non-negative quantities $|\mathcal{A}_{\mathrm{SM}}|^2$, $|\mathcal{A}_{\mathrm{SM}} + \mathcal{A}_{Q_i}|^2$, and $|\mathcal{A}_{Q_i}|^2$. For multiple operators, the model generalizes

to include mixed terms while maintaining positive-definiteness:

$$
\begin{aligned}
N = \mathrm{SM} \cdot (1 - \sum_i c_i + \sum_{i<j} c_i c_j) \\
+ \sum_i (c_i - \sum_{j \neq i} c_i c_j)\,(\mathrm{SM} + \mathrm{Lin}_i + \mathrm{Quad}_i) \\
+ \sum_i (c_i^2 - c_i)\,\mathrm{Quad}_i + \sum_{i<j} c_i c_j\,(\mathrm{SM} + \mathrm{Lin}_i + \mathrm{Lin}_j + \mathrm{Quad}_i + \mathrm{Quad}_j + 2\,\mathrm{Mix}_{ij}).
\end{aligned}
\tag{1.65}
$$

This formulation, equivalent to expressing $|\mathcal{A}_{\mathrm{SM}} + \sum_i c_i\,\mathcal{A}_{Q_i}|^2$ in terms of measurable quantities, guarantees that all input templates are positive-defined and therefore compatible with the likelihood construction in COMBINE.

Chapter 2

The Large Hadron Collider and the Compact Muon Solenoid

The remarkable precision achieved by the SM and its extensions can only be tested through experiments capable of reaching the highest energies ever produced in a laboratory. This chapter presents the experimental facilities where the analyses are carried out: the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector at the European Organization for Nuclear Research (CERN).

The first section (Section 2.1) describes the LHC, the most powerful particle accelerator ever built, outlining its operating principles, design, and performance. The subsequent sections (Section 2.2) focus on the CMS detector, a general-purpose apparatus designed to explore a broad range of physics phenomena, from precision measurements to searches for new particles.

A detailed overview of its subdetectors is then presented, covering the tracking system, electromagnetic and hadronic calorimeters, muon chambers, and the two-level trigger system responsible for selecting the most relevant events for further analysis. Together, these systems form an integrated instrument that allows the reconstruction and identification of particles with high efficiency and precision, enabling the wide physics program pursued by the CMS Collaboration.

2.1 The Large Hadron Collider

The LHC is a superconducting hadron accelerator and collider based on a double-ring structure, located at CERN in the 27 km long tunnel that previously hosted the Large Electron-Positron collider (LEP) [80]. Its main design motivation is the exploration of the scalar sector of the SM, aiming to shed light on the origin of electroweak symmetry breaking, for which the Higgs mechanism provides the theoretical framework. The LHC physics programme extends beyond the discovery of the Higgs boson, offering a unique opportunity to probe

the validity of the SM at the TeV scale. In addition, it enables precision studies of QCD, electroweak, and flavour physics, as well as direct searches for new particles and interactions predicted in various extensions of the SM.

The unprecedented performance of the LHC, with a centre-of-mass energy of 13 TeV and an instantaneous luminosity of order $\mathcal{O}(10^{34})$ $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, constitutes an extraordinary technological achievement. The stringent requirements on beam intensity and collimation rule out the use of a particle–antiparticle collider, which would otherwise benefit from a shared vacuum and magnetic system. Consequently, the LHC was conceived as a proton–proton collider, where separate magnetic fields and vacuum chambers for the two beams are implemented in a twin-bore magnet design, imposed by the limited internal tunnel diameter of 3.7 meters. Among the eight potential interaction regions, only four host detectors, a choice made to reduce beam crossings and limit unnecessary perturbations to the circulating beams. Achieving a beam energy of 7 TeV within the LEP tunnel demands a dipole field of 8.33 T, attainable exclusively with superconducting technology. To this end, the LHC relies on 1232 superconducting dipole magnets and 858 superconducting quadrupole magnets, built from a Niobium–Titanium alloy and cooled with superfluid helium to an operational temperature of 1.9 K.

Protons in the LHC are accelerated by means of 16 superconducting Radio Frequency (RF) cavities, eight per beam, operating at a frequency of 400 MHz. Their function is to decelerate faster protons and accelerate slower ones, thereby preserving the longitudinal compactness of the proton bunches in both coordinate and momentum space. This mechanism relies on the principle of phase stability. To ensure stable operation, the cavities are housed in cylindrical cryomodules and maintained at a temperature of 4.5 K. In the standard configuration of the LHC, a proton bunch is injected every ten RF cycles. Given the 27 km ring circumference, this would allow for about 3600 bunches; however, additional spacing must be reserved for beam injection and extraction, reducing the effective capacity to 2808 bunches per beam. This scheme corresponds to a mean bunch spacing of 25 ns, leading to a nominal collision rate of approximately 40 MHz.

To gradually accelerate protons up to the design energy, the LHC relies on a dedicated injection chain. Protons are first accelerated in Linac4 to 160 MeV, then transferred to the Proton Synchrotron Booster (PSB) before being injected into the Proton Synchrotron (PS), where their energy is raised to 26 GeV. From the PS, they are transferred to the Super Proton Synchrotron (SPS), which accelerates them further to 450 GeV, providing the final injection energy for transfer into the LHC.

Four large experiments sit at the four crossing points of the LHC beams. Among these, two are general-purpose detectors positioned at the high-luminosity cross locations with a peak luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Specifically, there is the Toroidal LHC ApparatuS (ATLAS) detector situated at Point 1, and the Compact Muon Solenoid (CMS) detector positioned at Point 5. Conversely, the remaining two detectors, the Large Ion Collider Experiment (ALICE) and the

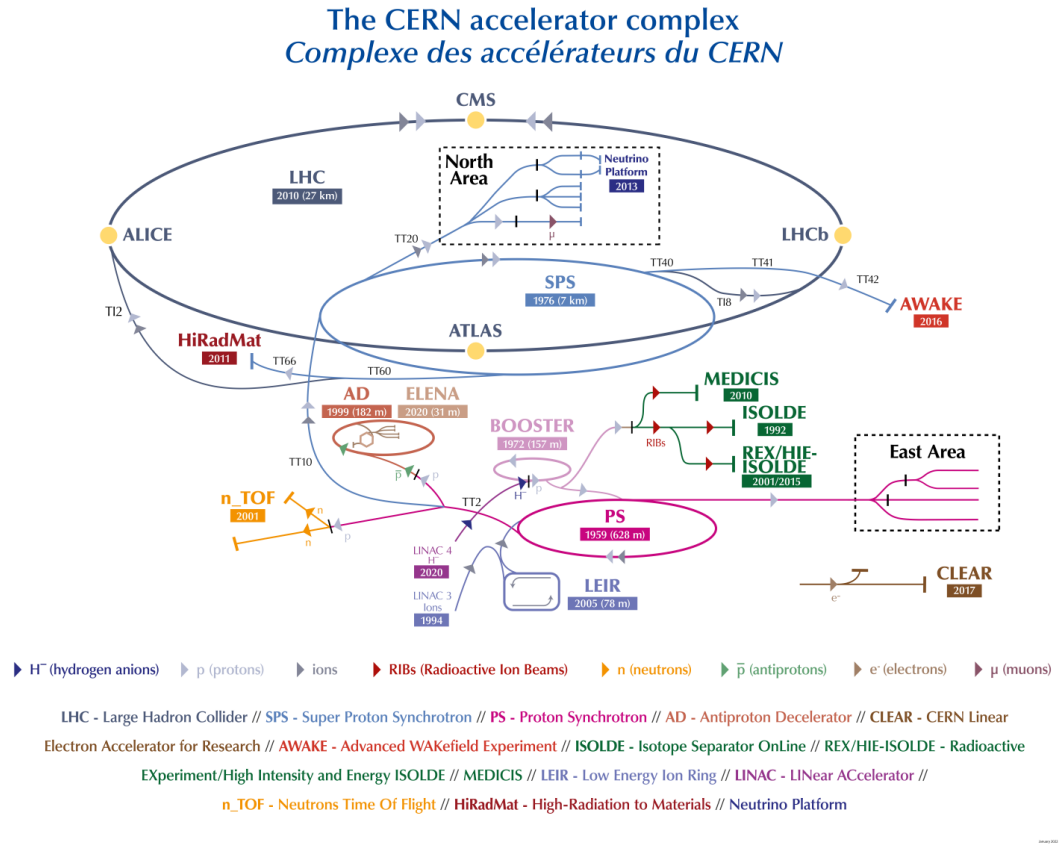


Figure 2.1: CERN accelerator complex, experiments and injection chain

Large Hadron Collider beauty (LHCb) detectors, have more specialized purposes. These detectors are stationed at cross points with lower luminosity levels of 2×10^{30} cm^{-2}s^{-1} and 5×10^{32} cm^{-2}s^{-1} respectively. The layout of the LHC is illustrated in Fig. 2.1.

The probabilistic character of the physical processes being investigated at the LHC makes searches for Beyond the Standard Model (BSM) processes and accurate measurements of previously understood physics processes only possible when a statistically significant number of these events can take place at the LHC. The production cross section of a process, σ, the machine luminosity L, the detector efficiency ϵ and the detector geometrical acceptance A are used to calculate the expected number of events N

$$N = L\sigma \cdot \epsilon \cdot A \tag{2.1}$$

The machine luminosity depends on the beam parameters and, for a Gaussian beam distribution, can be written as

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{2.2}$$

Where N_b and n_b are respectively the number of particles per bunch and the

number of bunches in the beam, f_{rev} the revolution frequency, γ_r a relativistic factor, ϵ_n the normalized emittance of the beams and β^* is the machine betatron function related to the size of the beam at the interaction point. F is the so-called geometric luminosity reduction factor and is related to the beam crossing angle, transverse RMS beam size and RMS bunch length at the interaction point. The integrated machine luminosity $\int_0^t Ldt$ measures the amount of data delivered to the experiments. The LHC began its operational period in 2009

Symbol	Description	2016	2017	2018
$\sqrt{s}$	centre-of-mass energy (TeV)	13	13	13
Δt	Bunch spacing (ns)	25	25	25
L	Instantaneous luminosity peak (10^{34} cm^{-2}s^{-1})	1.4	2.1	2.1
μ	Average pileup	25	38	37
n_b	Bunches per beam	2220	2556/1868	2556
N_b	Protons per bunch (stable beam) (10^{11})	1.0-1.25	1.0-1.25	1.0-1.25
ϵ_n	Transverse emittance (μm)	2.2	2.2	1.9
f_{rev}	Revolution frequency (Hz)	11245	11245	11245
β^*	Beta function (cm)	40	40-30	30-25

Table 2.1: Main operational parameters of the LHC during Run 2 (2016–2018). The centre-of-mass energy remained constant at 13 TeV, while progressive improvements in the accelerator optics and beam intensity led to higher instantaneous luminosities and average pileup conditions over the years.

with the first delivered proton-proton collision at a centre-of-mass energy of 0.9 TeV. During 2010 the centre-of-mass energy was increased to 7 TeV, collecing 45 pb^{-1}. In 2011 the peak luminosity was recorder at 4×10^{33} cm^{-2}s^{-1}, collecting 6 fb^{-1} of data. The first stable run (Run-I) came in 2012 with a centre-of-mass energy of 8 TeV and a luminosity of 6×10^{33} cm^{-2}s^{-1}, stable for almost the entire year. After approximately two years of shutdown in order to consolidate the high current splices between the superconducting magnets, the machine entered in a long-term phase of operations at 13 TeV known as the Run-II. In this period from 2015 to 2018 the instantaneous luminosity exceeded the design specifications with peaks of 2.1×10^{34} cm^{-2}s^{-1} with consequently higher number of simultaneous interactions, or pileup (PU), with a mean of 40 and peaks of 70 in 2018. The LHC Run-II delivered an integrated luminosity of approximately 150 fb^{-1} to each experiment. A second long shutdown (2019–2021) prepared the accelerator complex for Run-III, which started in 2022 at a centre-of-mass energy of 13.6 TeV, close to the design value. During this phase, the LHC is expected to deliver an integrated luminosity of about 350 fb^{-1}, with an instantaneous luminosity roughly twice the nominal one. A further long shutdown is foreseen from mid-2026 to around 2030, during which the High-Luminosity LHC (HL-LHC) upgrade will be implemented. In this configuration, the machine is expected to reach an instantaneous luminosity between five and seven times the nominal value, corresponding to 5–7 $\times 10^{34}$, cm^{-2}s^{-1}.

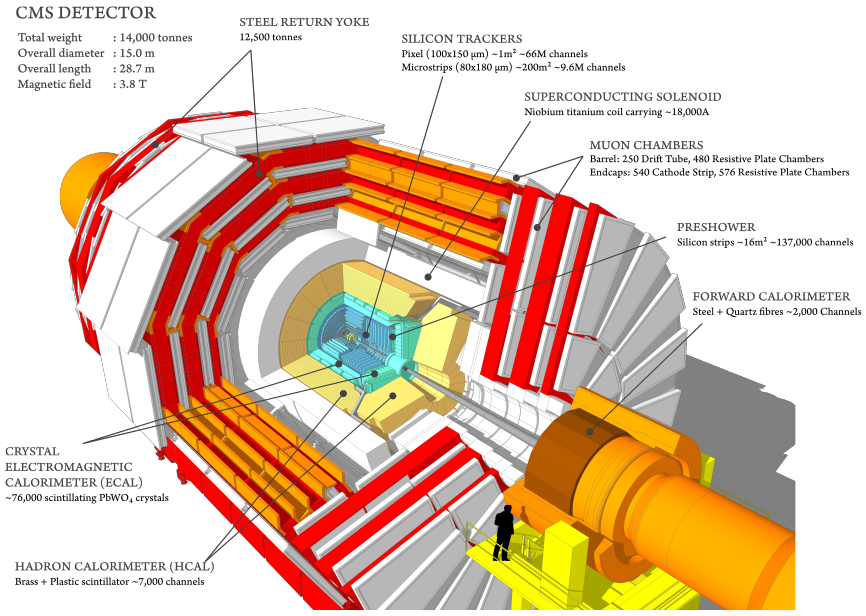


Figure 2.2: CMS detector layout

2.2 The Compact Muon Solenoid

The CMS detector, one of the two general-purpose experiments at the LHC, was conceived to explore high-momentum interactions and provides excellent performance in the reconstruction of electrons, photons, and muons. At its core, CMS hosts a niobium–titanium superconducting solenoid capable of producing a 3.8 T magnetic field across its 6-meter diameter, making it one of the largest magnets ever built, with a total weight of about 2000 tons. Its compact geometry allows the entire detector to fit within the experimental cavern located roughly 100 meters underground at LHC Point 5. The broad range of physics goals guided the detector design, leading to outstanding performance of the electromagnetic calorimeter, the muon system, and the inner tracker, all optimized for precise particle identification and momentum measurements. Since many BSM signatures are expected to interact only weakly with the detector material, new physics could manifest as missing transverse energy relative to the energy of the colliding beams. To ensure sensitivity to such signatures, both the electromagnetic and the hadron calorimeters were designed to be nearly hermetic, providing extensive geometric coverage that approaches the full solid angle around the interaction point and thereby maximising the detection efficiency of collision products. The innermost tracking detectors, positioned as close as 3 cm from the beams, are also located inside the solenoid, exploiting its magnetic field to measure the curvature and momentum of charged particles. Outside the solenoid, within the return yoke where the magnetic field is about 2 T, lie the muon chambers, the outermost components of CMS, which combine three different detector technologies [81]. Owing to the varying particle flux, each CMS subdetector is divided into a central barrel region and two forward endcaps, often employing distinct technological solutions. A schematic view of

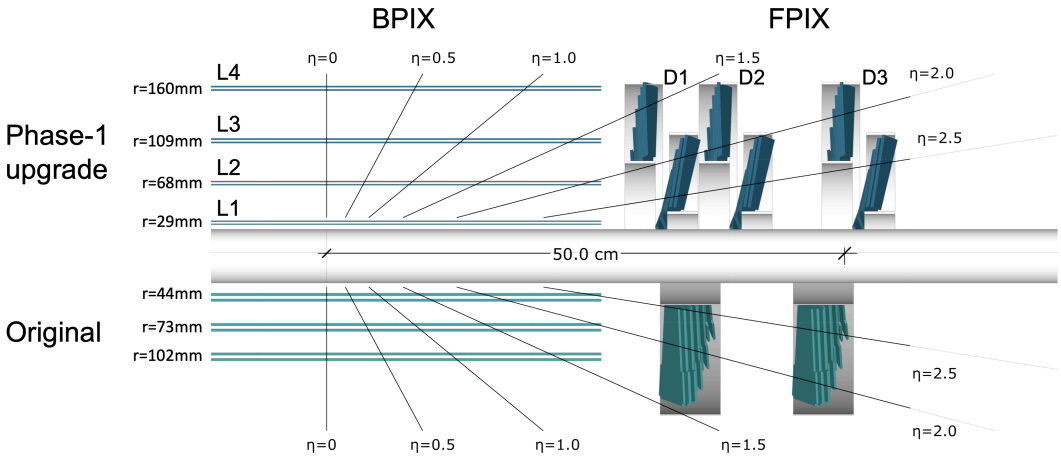


Figure 2.3: Schematic view of the CMS tracking system in the $r - z$ plane before (bottom half) and after (top half) the upgrades introduced during the 2016/2017 year-end technical stop [82]

the CMS detector is shown in Figure 2.2.

A right-handed Cartesian coordinate system is used to describe the CMS detector and its collision products. Its origin coincide with the nominal interaction point, the x-axis points toward the centre of the detector, the y-axis points up (perpendicular to the LHC plane) and the z-axis along the anticlockwise beam direction. A polar coordinate system is also used where θ indicates the polar angle with respect to the z-axis and ϕ represents the azimuthal angle in the $x - y$ plane. The pseudorapidity $\eta = -\log\tan(\theta/2)$ is often used instead of θ to describe the angle of emission for a particle with respect to the beam axis because differences in rapidity proves to be Lorentz-invariant and the production of particles at hadron colliders is approximately constant as a function of η. Spatial separation is often defined in terms of the angular euclidean distance in the $\eta - \phi$ plane as $(\Delta R)^2 = (\Delta\eta)^2 + (\Delta\phi)^2$. Transverse quantities with respect to the beam axis are often used as they are invariant under Lorentz boosts resulting from the initial momentum of the interacting partons, e.g. the projected momentum of a particle onto the transverse plane is defined as $(p_T)^2 = (p_x)^2 + (p_y)^2$

2.2.1 Tracking system

The CMS tracking system is the innermost subdetector, positioned closest to the interaction point [83]. It extends 5.6 m along the beam axis and up to 2.4 m radially, covering an active area of nearly 200 m^2 with about 80 million readout channels. Owing to the harsh radiation environment in this region, the tracker was designed with a two-tiered layout optimized to minimize material usage. It is entirely based on silicon technology, providing large surfaces of thin, finely segmented sensors. Within the uniform magnetic field of the solenoid, the spatial measurements of charged particles (hits) are combined into reconstructed

tracks, enabling momentum and charge determination. Correlating multiple tracks allows the identification of the primary interaction vertex and its separation from pileup vertices originating from other simultaneous pp collisions. The tracker also enables the reconstruction of displaced secondary vertices from in-flight decays of B and C hadrons or τ leptons.

Due to the steeply decreasing particle flux with radial distance r, the innermost region ($r < 20$ cm) is instrumented with high-granularity pixel detectors. Between $20 < r < 250$ cm, silicon strip detectors with a pitch increasing with r are used to adapt to the lower particle density. The pixel detector covers up to $|\eta| < 2.5$ and consists of three barrel layers and two endcap disks. Following the 2017 upgrade, an additional barrel and endcap layer were installed, placing the barrel layers at radii of 2.9, 6.8, 10.2, and 16.0 cm. Each pixel cell has a size of $100 \times 150, \mu\text{m}^2$, yielding a total of 66 million channels (almost 90% of the tracker) over an active area of about 1 m^2, with spatial resolutions of 10 μm in (r, ϕ) and 20 μm along z.

The outer tracker uses larger silicon strip modules with pitches ranging from 80 to 180 μm, a compromise between resolution and the cost of covering 200 m^2 with fine segmentation. The barrel strip system is split into the Tracker Inner Barrel (TIB) with four layers up to $|z| < 65$ cm, and the Tracker Outer Barrel (TOB) with six layers up to $|z| < 110$ cm. The gap between TIB and TOB is filled by four Tracker Inner Disks (TID). In the forward region, coverage up to $|\eta| < 2.5$ is provided by two Tracker Endcaps (TEC), each composed of nine disks spanning $120 < |z| < 280$ cm. Depending on r, the single-hit resolution ranges from 20–50 μm transversely to 200–500 μm longitudinally.

To mitigate radiation damage and dissipate heat from the on-detector electronics, the pixel and strip detectors are cooled to operating temperatures of about -20°C and -15°C, respectively. A schematic overview of the CMS tracker system is shown in Figure 2.3.

2.2.2 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) [84] is a homogeneous calorimeter made of 75,848 lead tungstate ($PbWO_4$) scintillating crystals, designed to measure the energy of mainly photons and electrons. An overview of the geometry and layout is shown in Fig. 2.4. The crystals serve simultaneously as absorbers and active media, providing excellent energy resolution.

$PbWO_4$ was chosen for its density (8.28 g/cm^3), short radiation length (0.89 cm), and small Molière radius (2.19 cm), which ensure compact containment of electromagnetic showers. The material is also radiation tolerant and has a fast response: 80% of the scintillation light is emitted within 25 ns, matching the 40 MHz bunch crossing rate of the LHC. Its main drawback is the low light yield, about 30 photons per MeV of deposited energy (temperature dependent), which requires photodetectors with internal amplification capable of operating in the 3.8 T axial magnetic field of CMS. In the barrel, avalanche photodi-

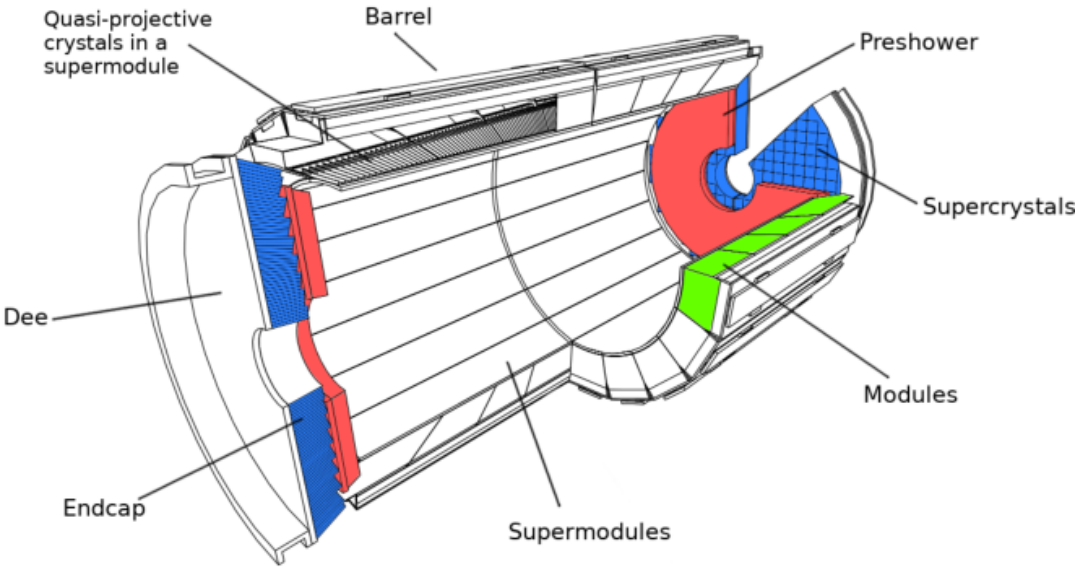


Figure 2.4: ECAL layout and geometry

odes are used, while in the endcaps vacuum phototriodes are preferred for their higher radiation tolerance. The front-end electronics amplify the photodetector signals and digitize them with a 12-bit ADC at 40 MHz.

The barrel (EB) consists of 61,200 crystals arranged in 36 supermodules, 18 per half barrel, each spanning 20° in ϕ. Crystals have a truncated-pyramid shape with a transverse section of 22×22 mm^2 and a length of 23 cm, covering the pseudorapidity range $|\eta| < 1.479$. The endcaps (EE) extend coverage to $|\eta| = 3$ with two disks, each made of 7324 crystals with a front face of 2.86×2.86 cm^2, placed about 314 cm from the interaction point. Endcap crystals are grouped into super-crystals (SCs) of 5×5 units. Each endcap is divided into two halves ("Dees"), with 3662 crystals forming 138 standard SCs and 18 partial SCs. In the forward region, a Si–Pb sampling detector (the preshower) helps discriminate between single photons and pairs from π^0 decays.

In both barrel and endcaps, the crystals are mounted in a quasi-projective geometry, their axes tilted by about 3° with respect to the interaction point in both η and ϕ [84], minimizing energy losses in inter-crystal gaps.

The crystal and photodetector light yields strongly depend on temperature (–2%/°C for $PbWO_4$ and –2.3%/°C for photodetectors). A dedicated cooling system maintains stable conditions: ±0.05°C in the barrel and ±0.1°C in the endcaps, with a nominal working temperature of 18°C. This choice also mitigates the wavelength-dependent transmission loss caused by radiation damage. To monitor and correct for transparency variations, a laser system injects 440 nm light into each crystal, providing time-dependent calibration constants.

The calorimeter energy resolution can be parametrized as

$$\left(\frac{\sigma}{E}\right) = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C, \tag{2.3}$$

where E is expressed in GeV, and S, N, and C represent the stochastic, noise,

and constant terms, respectively. The stochastic term reflects fluctuations in scintillation photon statistics, scaling as $\sqrt{n}$ with E. The noise term is energy-independent, while the constant term is dimensionless and accounts for inhomogeneities such as longitudinal non-uniformities, leakage, channel response dispersion, and thermal stability.

Beam tests with electrons [85], in conditions without a magnetic field or upstream material, yielded for a 3×3 crystal matrix: $S = 2.8\%$, $N = 12\%$, $C = 0.3\%$. Inter-channel calibration is essential due to variations in crystal light yield ($\sim 15\%$) and dispersion in endcap phototriode gains ($\sim 25\%$). Initial laboratory and cosmic-ray calibrations are refined in situ with collision data. These procedures yield an energy resolution of about 2% for 45 GeV electrons in the barrel, and 2–5% elsewhere. In the central barrel, where bremsstrahlung losses are minimal, the resolution improves to $\sim 1.5\%$. A comprehensive review of ECAL calibration and performance in pp collisions at $\sqrt{s} = 7$ TeV is given in Ref. [86].

2.2.3 Hadron Calorimeter

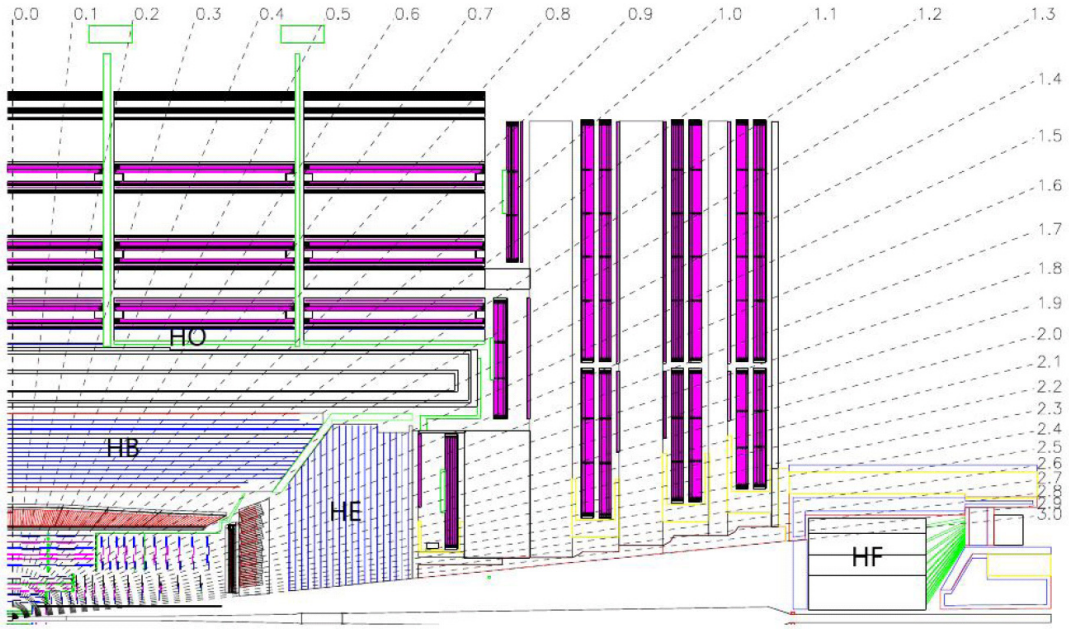


Figure 2.5: HCAL layout and geometry. The longitudinal view shows the barrel (HB) in between ECAL and the solenoid, the endcaps (HE), the outer barrel (HO) outside the solenoid, and the forward calorimeter (HF)

The CMS hadron calorimeter (HCAL), composed of alternating layers of brass and scintillator, is designed to measure the energies of hadrons that typically traverse the ECAL without being fully absorbed. Unlike electromagnetic showers, hadronic showers are intrinsically more difficult to measure. Their development is affected by nuclear interactions and non-Poissonian fluctuations, which result in the production of numerous undetectable particles. In addition, hadronic

showers contain an electromagnetic component from $\pi^0 \to \gamma\gamma$ decays, to which the detector responds differently compared to the purely hadronic part. These effects degrade the intrinsic energy resolution, which can only be partially recovered with advanced offline techniques, such as particle flow reconstruction. Despite these limitations, the HCAL plays a crucial role in reconstructing final states containing hadronic jets and in determining the missing transverse energy from non-interacting particles such as neutrinos.

The HCAL design is strongly constrained by the parameters of the CMS solenoid. To achieve nearly hermetic coverage up to $|\eta| < 5$, essential for accurate measurement of the missing transverse energy imbalance, a compact sampling calorimeter was adopted. The system uses alternating layers of non-magnetic brass absorber and plastic scintillator. Brass was chosen as the absorber material due to its short nuclear interaction length ($\lambda_I \approx 16.2$ cm) and mechanical robustness. As incoming hadrons traverse the absorber, they initiate hadronic showers through interactions with brass nuclei. The secondary particles deposit energy in the scintillator layers, which emit light collected by wavelength-shifting fibers and detected by hybrid photodiodes (HPDs).

The HCAL comprises approximately 70,000 scintillating tiles and is segmented into several subdetectors covering different pseudorapidity regions. In the central region, the HCAL barrel (HB) extends up to $|\eta| < 1.3$ and is located between the ECAL and the solenoid. Since the HB volume does not provide full shower containment, an additional HCAL outer detector (HO) is placed outside the solenoid, covering $|\eta| < 1.26$. This extends the total calorimeter depth to about 11 nuclear interaction lengths, thereby improving the energy resolution. The HCAL endcap (HE) covers $1.3 < |\eta| < 3.0$, while the HCAL forward (HF), located 11.2 m away from the interaction point, extends the coverage up to $|\eta| = 5.2$ (see Fig. 2.5).

Unlike HB, HO, and HE, which employ brass absorbers and plastic scintillators, the HF must withstand extremely high radiation levels. It is therefore built with 1.65 m thick steel absorbers and 0.6 mm radiation-hard quartz fibres as the active medium, which produce Cherenkov light detected by photomultiplier tubes (PMTs). To disentangle the electromagnetic and hadronic components of the showers, two sets of fibres of different lengths are deployed.

The transverse segmentation of the calorimeter is defined by towers of scintillator tiles: $\Delta\eta \times \Delta\phi \approx 0.087 \times 0.087$ in the HB and about 0.17×0.17 in the HE. The overall HCAL energy resolution is dominated by the imperfect containment of hadronic showers. Test beam measurements with charged pions in the 4–300 GeV range yielded a stochastic term of approximately $S = 110\%$ and a constant term of $C = 9\%$ [87].

2.2.4 Muon chambers

Located outside the solenoid, the outermost part of CMS accommodates the muon detector, which covers a vast area of about 25,000 m^2 [88]. Muons, being

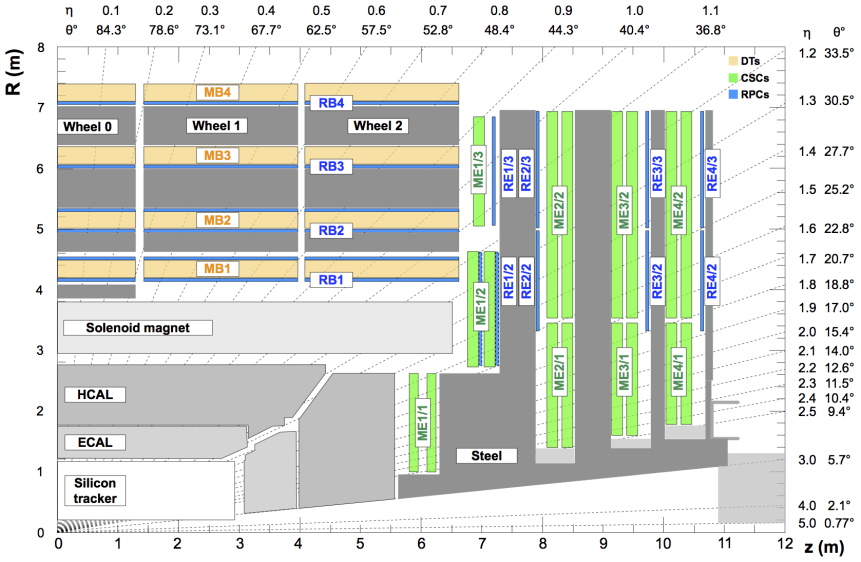


Figure 2.6: $r-z$ cross section of a quarter of CMS in its longitudinal direction. Orange, green, and blue regions denote respectively the DTs, CSCs, and RPCs subsystems while the dark grey areas represent the steel support disks.

more massive than electrons, undergo fewer radiative losses within the tracker material and are capable of penetrating the ECAL, HCAL, and the solenoid. The muon chambers therefore play a dual role: they provide identification of muons and also supply an additional measurement of their momentum, exploiting the 2 T return magnetic field inside the iron flux-return yoke in which they are embedded. Since isolated muons are clear signatures of interesting events, they are measured twice in CMS: first in the inner tracker, and subsequently in the outer muon system. The CMS muon system has a cylindrical layout, consisting of a barrel section and two endcap sections. It employs three types of gaseous detectors, chosen according to the expected background rates and magnetic field conditions: drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). This technological choice ensures large-area coverage in a cost-effective manner while remaining robust against different radiation environments. An overview of the geometry is shown in Fig. 2.6.

In the barrel region, $|\eta| < 1.2$, where muon and neutron background rates are low and the magnetic field is uniform ($B < 0.4$ T), CMS employs 250 DTs distributed in five barrel sections along the z axis (the so-called *wheels*: W0 at $z = 0$, W$\pm$1, and W$\pm$2). Each wheel is divided into 12 sectors, each covering 30° in ϕ. Along the radial direction, DT chambers are grouped in four stations at radii of approximately 4.0, 4.9, 5.9, and 7.0 m from the beam axis [89], interspersed with the steel plates of the return yoke. The basic element is the drift cell, with transverse size 42×13 mm^2 and a central anode wire of 50 μm diameter, filled with an Ar/CO_2 (85%/15%) gas mixture. This provides good quenching properties, a saturated drift velocity of about 55 μm/ns, and a maximum drift time of 400 ns. The field configuration is defined by four electrodes: two cathode strips on the sides of the cell and two ground electrodes

on the planes above and below the wires. A DT chamber is built from three *superlayers* (SLs), each consisting of four staggered layers of parallel cells. Two SLs measure the r–ϕ coordinate (wires parallel to the beam axis), while the third SL measures the r–z coordinate (wires orthogonal). Each DT cell achieves a spatial resolution of ~ 200 μm, and when combined across a chamber this yields 80–120 μm in position and about 1 mrad in angle. In addition, DTs provide a trigger signal with a time resolution of $\sim$ 5 ns, correlated with the muon transverse momentum.

In the endcap regions, $0.9 < |\eta| < 2.4$, background rates are higher and the magnetic field is both strong and non-uniform. Here, CMS uses CSCs, which tolerate non-uniform magnetic fields, provide fine segmentation, and have a fast response. A CSC is a multi-wire proportional chamber with cathode-strip readout [90]. Each trapezoidal chamber consists of six layers of anode wires interleaved with seven layers of orthogonal cathode strips. Four stations per endcap are arranged perpendicular to the beam axis, with cathode strips oriented radially to measure the z coordinate, while wires measure the bending r–ϕ plane. The gas mixture (50% CO_2, 40% Ar, 10% CF_4) is ionized by passing muons, and hits are reconstructed at the intersection of fired strips and wires. The timing is calibrated so that hits from promptly produced muons are centered at zero in the bunch-crossing time distribution. A single CSC hit provides ~ 100 μm spatial resolution and ~ 10 mrad angular resolution. Together, DTs and CSCs cover the full pseudorapidity interval $|\eta| < 2.4$ without gaps, ensuring excellent muon identification.

For precise bunch-crossing identification and redundancy in triggering, CMS employs RPCs, installed in both barrel and endcaps up to $|\eta| < 1.6$. RPCs consist of two parallel bakelite resistive plates (2 mm thick) separated by a thin gas gap filled with a mixture of 95.2% $C_2H_2F_4$ (Freon), 4.5% i-C_4H_{10} (isobutane), and 0.3% SF_6 [91], with added water vapor to maintain 40–50% humidity. Operated in avalanche mode, RPCs provide strong signals induced on readout strips aligned along the η direction, with strip pitch up to 2 cm. The spatial resolution ranges from 0.8 to 1.2 cm, reconstructed as the centroid of clusters of fired strips. RPCs deliver excellent timing resolution of about 1 ns, enabling unambiguous identification of the correct bunch crossing, particularly critical at the highest LHC luminosities.

2.2.5 Trigger system

At the LHC design luminosity of 10^{34} cm^{-2}s^{-1}, the proton-proton collision rate exceeds 1 GHz. The majority of these collisions are inelastic, producing low energy multi-jet events with a cross section $\sigma_{\text{inel}} \sim 70$ mb several orders of magnitude greater than the cross section of interesting processes for the CMS physics programme ($\sigma < 10^5$ pb). These collisions happen every 25 ns that, combined with the highly granular information recorded by the detector, generate approximately 70 Tb/s of data. At the time being, it is unfeasible to read out, store and analyse such large volumes of data. The trigger system of CMS

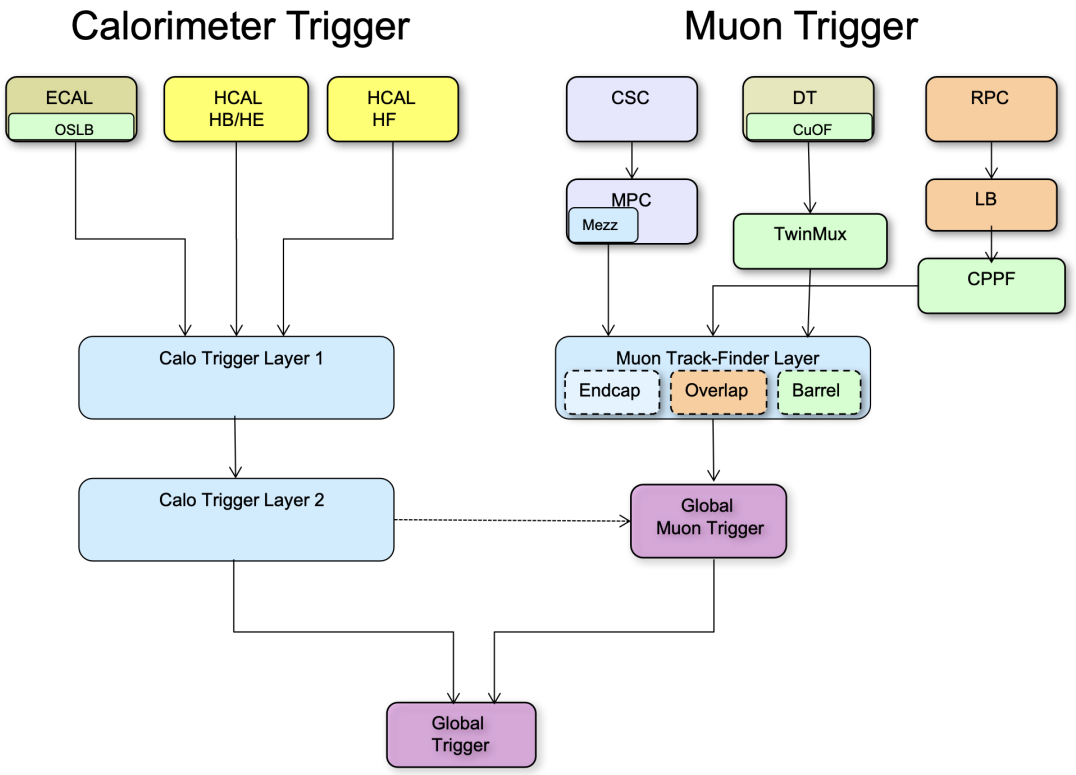


Figure 2.7: Schematic representation of the CMS L1 trigger system from [92]

is designed to discriminate collisions of interest from the sea of proton proton inelastic collisions, as to reduce the rate of events to a sustainable amount of about 10^5 events/s for detector readout, and about 10^3 events/s for offline storage. To achieve this goal CMS employs a two-tiered solution, with components are referred to as Level-1 Trigger (L1) and High Level Trigger (HLT).

Level-1 Trigger

The Level-1 (L1) trigger is implemented in custom hardware and can access calorimetric and muon information only with reduced granularity compared to the full detector capabilities. It operates with a fixed latency of 3.8 μs, during which the critical decision to accept a collision must be made. This reduces the event rate from the 40 MHz bunch-crossing frequency to approximately 100 kHz. Tracker information is not available at L1, since the system cannot sustain its readout at 40 MHz. Dedicated boards are being designed for the HL-LHC upgrade to enable the readout of the outer tracker at the full bunch-crossing rate [93].

The absence of tracking information at L1 prevents discrimination between objects originating from different collisions within the same bunch crossing. Moreover, electrons and photons cannot be distinguished and produce similar signatures. The L1 selection process is organized into algorithms referred to as *seeds*, which define conditions on one or more candidate objects such as muons,

electrons, photons, jets, τ leptons, or missing transverse energy E_T^{miss}. These candidates are reconstructed from low-level inputs called *trigger primitives*, provided by each CMS subdetector. Most seeds require N candidate objects passing minimum E_T and pseudorapidity thresholds. Single- and double-object seeds account for nearly 75% of the available bandwidth. As of 2018, the L1 trigger system supports up to 512 algorithms.

A schematic representation of the L1 trigger architecture is shown in Fig. 2.7, taken from Ref. [92]. Two parallel paths process calorimetric information (from ECAL and HCAL subsystems HB, HE, HF) and muon information (from DT, RPC, CSC). Their outputs are sent to the micro global trigger (μGT), which combines them and issues the final accept/reject decision.

Calorimeter trigger Experience from Run 1 led to a major upgrade of the calorimeter trigger [94], replacing the earlier 4×4 trigger-tower regional granularity with full-tower readout. The new system achieves a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ over most of the calorimeter acceptance (slightly coarser at high $|\eta|$). This requires transmitting the full trigger tower information for each bunch crossing to a single board, which is computationally demanding. To address this, a time-multiplexed architecture was adopted, eliminating regional boundaries and enabling full-granularity energy sums.

The calorimeter primitives are processed in parallel by 18 Layer-1 boards, dedicated to preprocessing operations such as calibration and synchronization. These boards are advanced mezzanine cards equipped with Xilinx Virtex-7 FPGAs (CTP7) [95]. Their outputs are sent to nine Layer-2 boards, where reconstruction algorithms identify jets, τ leptons, photons, and electrons, using dedicated shape vetoes. Layer-2 nodes employ master processor (MP7) cards, also based on Virtex-7 FPGAs, designed for flexible algorithm development [92]. Finally, a demultiplexer (DeMux) board reorders and formats the events for μGT processing.

Muon trigger The muon L1 trigger combines redundant information from the different subsystems to reconstruct muon tracks. For Run 2, three dedicated track-finder systems were implemented in distinct pseudorapidity regions: the barrel muon track finder (BMTF), the overlap muon track finder (OMTF), and the endcap muon track finder (EMTF).

The BMTF covers $|\eta| < 0.83$ and combines primitives from DT and RPC chambers using the TwinMUX system to form *super-primitives*. The redundancy of DT and RPC measurements improves position accuracy, and the bending angle information from DTs is used in road-search algorithms to estimate the muon p_T.

The OMTF covers $0.83 < |\eta| < 1.24$, using inputs from all three subsystems (DT, CSC, RPC) in the overlap region. For the forward region, $|\eta| > 1.24$, the EMTF uses CSC and RPC information to reconstruct endcap muons. These

systems achieve latencies of ~ 750 ns, relying on advanced pattern-recognition algorithms to convert hit patterns into p_T assignments and quality criteria. Each track finder can transmit up to 36 muon candidates to the global muon trigger, which resolves duplicates and forwards up to eight ranked muons to the μGT.

The three muon trigger systems employ different hardware solutions: BMTF uses MP7 cards to provide high computing power, while OMTF and EMTF rely on modular track-finder cards with large memory optimized for pattern recognition [96].

Global trigger The final L1 decision is taken by the μGT, based on the combined calorimeter and muon information. The μGT computes multi-object correlations and global event quantities such as invariant masses, angular separations, or momentum sums. It runs all seed algorithms in parallel on MP7 boards, ensuring large processing capacity. During Run 2, the μGT configuration included approximately 350–400 active algorithms.

High-Level Trigger

Events accepted by the L1 trigger are further processed by the High-Level Trigger (HLT). While the primary goal of L1 is to reduce the input rate to a level sustainable for detector readout, the HLT aims to select events of physics interest at a rate compatible with long-term offline storage. The main limitations of the HLT come from timing constraints and from the requirement that trigger efficiencies must be measurable directly from data.

Unlike the hardware-based L1 system, the HLT is software-based and runs on a processor farm composed of commercial CPUs, known as the Event Filter Farm (EVF). The HLT latency is not fixed, but on average it is about 260 ms per event, with processing times scaling with event complexity and occasionally reaching $\mathcal{O}(1\text{ s})$. The EVF, running Scientific Linux, consists of builder and filter units. The builder units retrieve event fragments corresponding to a specific L1 accept and assemble them into complete events, while the filter units unpack raw data, perform partial reconstruction, and apply the trigger algorithms. In practice, builder and filter functionalities are implemented within a single multi-core machine. In total, the EVF hosts about 36,000 CPU cores, capable of processing the full 100 kHz L1 input rate and reducing it to a final output of a few kHz in Run 2, with design rates reaching up to about 10 kHz in Run 3. Accepted events are handled by the storage manager, temporarily stored on disk, and subsequently transferred to the CMS Tier-0 facility, which performs prompt offline reconstruction and permanent storage [97]. Roughly half of these events are promptly reconstructed (within 48 h), while the other half are "parked" for deferred processing.

The HLT is organized around the concept of paths, defined as sequences of algorithmic steps executed in a fixed order. Each HLT path is seeded by a specific

L1 decision. The algorithms within a path become progressively more complex, first applying coarse selections and later refined reconstructions. Early stages use calorimeter and muon information to reduce the event rate before invoking the more computationally demanding tracking reconstruction. A typical coarse reconstruction of jets, τ leptons, electrons, photons, and muons at the HLT has a time budget of ~ 50 ms. Tracking reconstruction is more expensive, but enables distinguishing electrons from photons, identifying heavy-flavour jets and hadronically decaying τ leptons, evaluating lepton isolation, and measuring the lepton momentum with high precision. To reduce CPU load, tracking is performed only on a subset of events and within limited pseudorapidity regions (*regional tracking*). Finally, the particle-flow (PF) [98] sequence provides the most accurate object reconstruction at the HLT. The PF sequence is the most time-consuming step and represents a streamlined version of the offline PF algorithm, optimized for fast execution.

Chapter 3

Measurement of Electroweak W Production via Vector-Boson Fusion in Run II

The theoretical and experimental ingredients introduced so far come together in the measurement presented in this chapter. It focuses on the analysis strategy developed to measure the electroweak (EW) production of a W boson in association with two jets. Among the various EW topologies, particular attention is devoted to the vector boson fusion configuration, which is especially sensitive to the non-Abelian self-interactions of the EW gauge bosons. This study represents a key test of the EW sector of the SM and provides a benchmark for searches of deviations that could hint at new physics.

This measurement constitutes one of the cleanest probes of the EW sector accessible at the LHC, and provides sensitivity to anomalous trilinear and quartic gauge couplings. The analysis makes use of the full Run II dataset collected by CMS at $\sqrt{s} = 13$ TeV, and is based on the reconstruction of events containing a leptonic W decay and two energetic jets with high invariant mass and large pseudorapidity separation.

The chapter begins by introducing the topology of the EW Wjj signal and the main background processes that mimic its signature (Section 3.1–3.2). It then presents the datasets, trigger selections, and simulated samples used to model the signal and the various background components (Section 3.3). The reconstruction and selection of physics objects are detailed in Section 3.4, together with the event-level requirements that define the regions of the analysis.

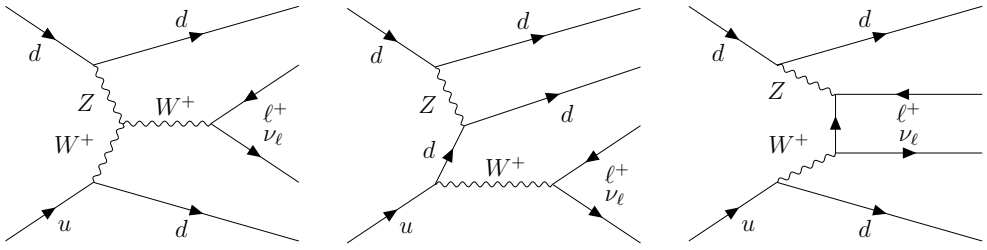
Particular attention is devoted to the validation of the simulation and to the application of data-driven corrections and background estimates, which ensure a reliable modeling of the most relevant processes. The strategy adopted to separate the EW Wjj signal from the overwhelming QCD-induced background is described in Section 3.6, where a deep neural network classifier is employed to exploit the full kinematic information of the event. The chapter concludes with a discussion of the systematic uncertainties affecting the measurement and

the extraction of the inclusive signal cross section (Section 3.8).

3.1 Signal Topology

In proton–proton collisions at the LHC, *Vector Boson Fusion* (VBF) occurs when a valence quark from each incoming proton radiates a weak boson, which subsequently interact or "fuse." In these interactions, the quarks exchange a t-channel four-momentum with Q^2 typically of the order of m_W^2 or m_Z^2, resulting in the emission of two energetic forward jets within the detector acceptance. The experimental hallmark of VBF processes is thus the presence of two hadronic "tagging" jets in opposite hemispheres, characterized by a large pseudorapidity separation ($\Delta\eta_{jj}$) and a large dijet invariant mass (m_{jj}).

VBF production plays a central role at the LHC, both as a process of intrinsic EW interest and as a key benchmark for precision Higgs measurements. The study of Z and W bosons production via EW VBF-like topologies provides an important test of the SM, probing the same color-singlet exchange and triple-gauge-boson couplings involved in Higgs boson production through VBF. In contrast to Higgs VBF, however, the production of Z or W bosons in association with two jets involves several classes of EW diagrams at the same order in α_W^4, including vector boson fusion, W-strahlung, and multiperipheral topologies. These amplitudes interfere destructively, a consequence of the non-abelian gauge structure of the SM that ensures unitarity at high energies [99]. Such cancellations, while reducing the overall rate, make the process an especially sensitive probe of the EW sector.



(a) Vector Boson Fusion (b) W-strahlung (c) Multiperipheral

Figure 3.1: Feynman diagrams for $pp \to l\nu jj$ via purely electroweak interactions (α_W^4).

The EW production of a W boson in association with two jets ($pp \to Wjj$) thus provides a unique laboratory to study vector boson scattering and the self-interactions of the weak gauge bosons. This process is also directly sensitive to potential deviations from the SM, such as anomalous triple-gauge couplings or higher-dimensional operators in an EFT framework. In this context, precise measurements of the Wjj cross section, both inclusive and differential, represent a crucial step toward constraining the structure of the EW Lagrangian.

Experimentally, the process is rare compared to QCD-induced W+jets produc-

tion, but can be isolated by exploiting the distinctive kinematic features of VBF topologies: large m_{jj}, wide $\Delta\eta_{jj}$ and suppressed hadronic activity in the central region.

The EW Wjj process was first observed by the CMS Collaboration at $\sqrt{s} =$ 8 TeV, with preliminary results released in 2015 [100] and results published in 2016 [101]. The ATLAS Collaboration reported an independent observation at $\sqrt{s} = 7$ and 8 TeV [102], achieving a signal significance above five standard deviations. More recently, CMS presented the first measurement at $\sqrt{s} = 13$ TeV using the 2016 dataset [103], extending the exploration of EW Wjj production to Run II energies. The analysis described in this chapter builds upon these results, providing a full Run-II measurement with improved statistical precision and extended differential sensitivity.

3.2 Backgrounds

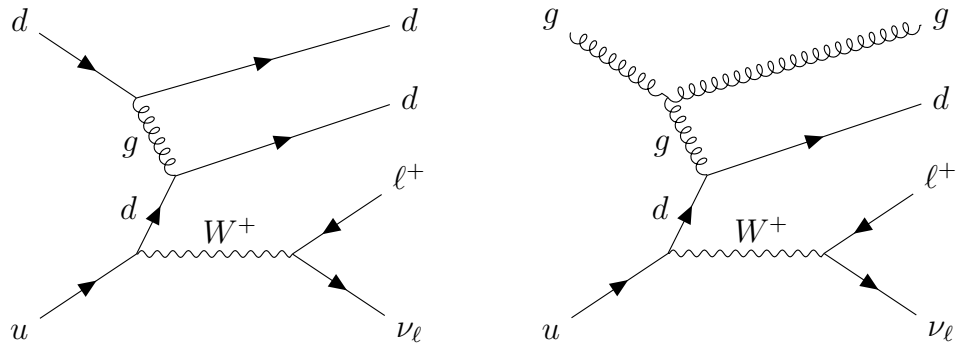
From a theoretical standpoint, several SM processes can mimic the $l\nu jj$ final state targeted in the EW Wjj (VBF-like) measurement. The most relevant backgrounds can be grouped into three broad classes: (i) QCD-induced production of W bosons in association with jets; (ii) top-quark processes (single-top and double-top production $t\bar{t}$); (iii) subdominant EW and instrumental contributions. The following discussion summarizes their leading-order topologies and main characteristics.

3.2.1 QCD-induced W+jets

The dominant irreducible background is the mixed EW–QCD production of a W boson with two jets, schematically of order $\mathcal{O}(\alpha_W^2\alpha_s^2)$. Representative partonic channels include $qg \to q'W$ with an additional hard jet from initial/final-state radiation, and $qq' \to qq'W$ via gluon exchange (Fig.3.2). Compared to the EW color-singlet exchange, QCD color flow favours more central hadronic activity and typically yields smaller m_{jj} and $\Delta\eta_{jj}$; nevertheless, hard fluctuations can populate the VBF-like phase space. The QCD-induced W+jets background dominates the inclusive $l\nu jj$ production rate, with a total cross section approximately two orders of magnitude larger than the purely EW process.

3.2.2 Top-quark backgrounds

Top-quark production represents one of the most significant backgrounds to the EW Wjj process. Both single-top and $t\bar{t}$ production can yield the same $l\nu jj$ final state targeted in this analysis, where a leptonically decaying W boson is accompanied by two hadronic jets. Although these processes are not purely EW, their final-state topology can closely resemble that of the signal, especially when b jets are misidentified or fall outside the detector acceptance.

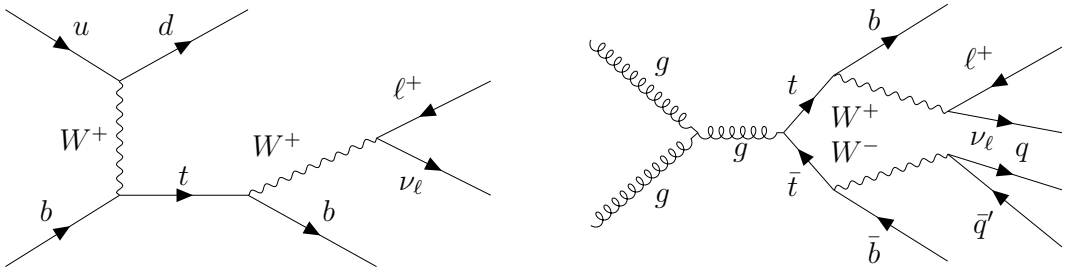


(a) $qq \to qq\,W$ via gluon exchange. (b) $qg \to q'g\,W$ channel.

Figure 3.2: Representative QCD-induced W+jets diagrams ($\alpha_W^2\alpha_S^2$) leading to $l\nu jj$ final states. The gluon exchange introduces a color connection across the event, yielding more central hadronic activity than the purely EW (color-singlet) production.

At $\sqrt{s} = 13$ TeV, both single-top and $t\bar{t}$ production occur at rates roughly one to two orders of magnitude higher than the EW Wjj signal. The single-top t-channel process is particularly relevant, since it naturally yields one isolated lepton, missing transverse momentum from the neutrino, and two jets, one of which is typically forward, closely reproducing the topology of VBF-like events. Semileptonic $t\bar{t}$ events also contribute when one W decays leptonically and the other hadronically. Although the nominal topology contains two b jets and several additional jets from the hadronic W decay, acceptance effects, imperfect b-tagging, and hard jets from initial- or final-state radiation can produce configurations entering the signal region. In principle, the larger jet multiplicity could be used to discriminate against $t\bar{t}$, but in realistic LHC conditions most events, signal and background alike, contain extra jets from radiation or pileup. As a result, jet counting offers limited discrimination power, whereas the presence of a well-identified isolated lepton from the W decay remains a much cleaner signature.

A powerful feature enabling the rejection of top quark backgrounds in CMS is the presence of long-lived b hadrons from top quark decays. These b hadrons have a proper lifetime of about 1.5 ps ($c\tau \simeq 450\,\mu$m) and typically travel several millimetres before decaying, with decay lengths that can extend to the centimetre scale depending on their boost. Their displaced decay vertices are reconstructed within the silicon tracker with a spatial precision of about 10 μm in the transverse plane and 20 μm along z, thanks to the fine granularity of the pixel detector. Such displaced secondary vertices (SVs), together with the identification of high-p_T leptons from semileptonic b decays, are exploited by modern b-tagging algorithms to efficiently identify jets originating from b quarks, thereby improving the rejection of single-top and $t\bar{t}$ backgrounds.



(a) t-channel single-top production with $t \to bW(\to \ell\nu)$.

(b) Semileptonic $t\bar{t}$ production via gluon fusion in the s channel.

Figure 3.3: Representative diagrams for top-quark production processes contributing to the $l\nu jj$ final state.

3.2.3 Subdominant backgrounds

Diboson (WW, WZ). Diboson production via mixed EW–QCD interactions ($pp \to VVjj$, $\mathcal{O}(\alpha_W^2\alpha_s^2)$) constitutes a subdominant but non-negligible background. At the inclusive level, its production rate is roughly an order of magnitude larger than that of the purely EW Wjj signal, with typical cross sections of a few tens of picobarns at $\sqrt{s} = 13$ TeV. The color-octet exchange in the t channel generates enhanced central hadronic activity and a dijet system characterized by lower m_{jj} and smaller $\Delta\eta_{jj}$ compared to the VBF topology.

Drell–Yan + jets. Production of Z/γ^* bosons in association with jets ($\mathcal{O}(\alpha_W^2\alpha_s^n)$) can contribute to the background when one of the leptons from $Z \to \ell^+\ell^-$ is not reconstructed or fails identification. In such cases, the event is misinterpreted as containing a single lepton, while apparent missing transverse energy arises from detector resolution effects or from neutrinos in heavy-flavor decays.

A second, genuinely irreducible contribution arises from $Z \to \tau^+\tau^-$ decays. In particular, the semileptonic channel, where one τ decays leptonically ($\tau \to \ell\nu_\ell\nu_\tau$) and the other hadronically, closely reproduces the Wjj topology, featuring one isolated lepton, genuine missing energy from multiple neutrinos, and two jets originating from the hadronic τ decay. This mode is therefore the most relevant Drell–Yan (DY) contamination. Nevertheless, the hadronic τ jets are typically narrower, less energetic, and more central than the quark-initiated jets from EW Wjj production, allowing their suppression through dedicated τ vetoes and VBF-like kinematic selections.

Nonprompt leptons. A further source of background arises from events in which the reconstructed lepton does not originate from the prompt decay of a W boson. Such nonprompt leptons can originate from semileptonic decays of heavy-flavor hadrons ($b \to \ell X$, $c \to \ell X$), from the decay-in-flight of light hadrons (mainly π or K mesons), or from hadronic jets or photon conversions

misidentified as leptons. In all cases, the apparent isolated lepton is embedded in a QCD-dominated hadronic environment, with significant additional jet activity and poorly balanced missing transverse momentum. These processes form a reducible background to the EW Wjj signal: although tight lepton identification and isolation criteria greatly suppress their contribution, the residual contamination remains non-negligible.

3.3 Datasets, triggers, and simulated samples

The analysis presented in this thesis is based on the full Run II dataset of the CMS experiment at the CERN LHC, corresponding to proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. A detailed understanding of the recorded datasets, the corresponding trigger selections, and the simulated samples used to model the signal and background processes is essential for achieving an accurate measurement of the cross section and for estimating systematic uncertainties. This section provides an overview of the data-taking conditions and reconstruction campaigns during Run II, describes the datasets and triggers used in the analysis, and summarizes the Monte Carlo (MC) simulations employed for signal and background modeling.

3.3.1 Overview of Run-II data taking

During Run II of the LHC (2016–2018), the CMS experiment collected a total integrated luminosity of nearly 138 fb^{-1} of proton–proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. Each data-taking year corresponds to a distinct detector configuration, trigger menu, and pileup environment, requiring dedicated reconstruction and simulation campaigns. The 2016 data are further subdivided into two distinct reconstruction eras, referred to as *pre-VFP* (also known as 2016APV or 2016HIPM) and *post-VFP* (or 2016 noHIPM). The split accounts for a change in the readout conditions of the silicon strip detector, following the mitigation of the APV25 preamplifier saturation issue during the year-end technical stop. Although the terms HIPM and noHIPM originated from an incorrect assumption related to highly ionizing particles (HIP), they have since become standard nomenclature for distinguishing these eras in all Run-II UltraLegacy (UL) reprocessings.

Table 3.1 summarizes the main features of the Run-II data-taking periods, including their integrated luminosities and typical pileup conditions. All datasets used in this analysis correspond to the centrally produced UL reconstruction, processed into NanoAODv9 format.

Year	Period / Era	Label	Integrated Luminosity [fb^{-1}]	$\langle\text{PU}\rangle$
2016	pre/post VFP	`2016HIPM`, `2016noHIPM`	36.3	∼27
2017	—	`2017`	41.5	∼38
2018	—	`2018`	59.7	∼37

Table 3.1: Summary of Run-II CMS data-taking periods used in this analysis, including era labels, total integrated luminosities, and typical average pileup conditions.

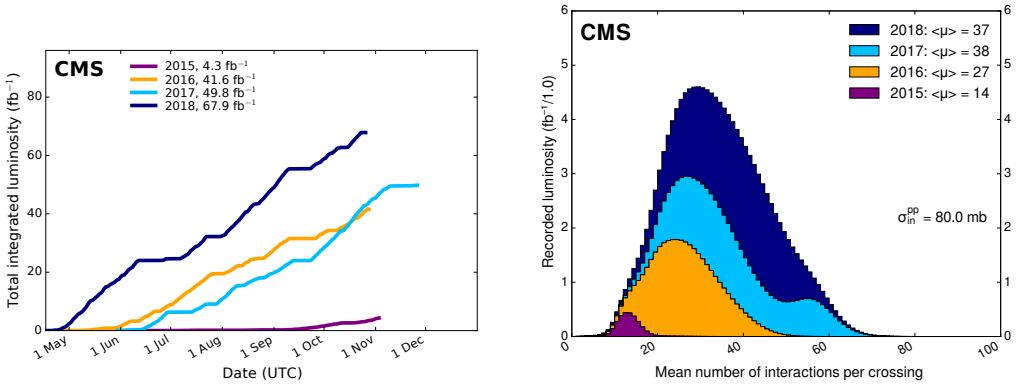


Figure 3.4: Left: integrated luminosity collected by CMS during each Run-II year. Right: average number of reconstructed primary interactions per bunch crossing (*pileup*) as a function of the data-taking period. The pileup (PU) refers to multiple proton–proton interactions occurring in the same bunch crossing, which can overlap with the primary hard-scattering event and affect the reconstruction of physics objects.

3.3.2 Data samples and triggers

Events are selected from the SingleMuon and SingleElectron (or EGamma, for 2018) primary datasets. Only events passing at least one of the high-level trigger (HLT) paths listed in Table 3.2 are retained. All trigger efficiencies are measured in data using the tag-and-probe method in $Z \to \ell\ell$ events, and scale factors correcting the simulation to data are derived as functions of the lepton transverse momentum and pseudorapidity [104].

Year	Primary Datasets	Single Muon HLT	Single Electron / EGamma HLT
2016	HIPM: SingleMuon, SingleElectron	`HLT_IsoTkMu24` ∨ `HLT_IsoMu24`	`HLT_Ele25_eta2p1_WPTight_Gsf`
	noHIPM: SingleMuon, SingleElectron	`HLT_IsoTkMu24` ∨ `HLT_IsoMu24`	`HLT_Ele27_WPTight_Gsf`
2017	SingleMuon, SingleElectron	`HLT_IsoMu27`	`HLT_Ele35_WPTight_Gsf`
2018	SingleMuon, EGamma	`HLT_IsoMu24`	`HLT_Ele32_WPTight_Gsf`

Table 3.2: Primary datasets and HLT paths used in the analysis for each Run-II data-taking year. The 2016 data are divided into pre-VFP (HIPM) and post-VFP (noHIPM) periods due to different readout conditions of the strip tracker.

The trigger selections for 2016 and 2018 employ similar p_T thresholds of approximately 24–27 GeV, while 2017 required slightly tighter thresholds due to

increased instantaneous luminosity and pileup.

3.3.3 Luminosity and pileup modeling

The total amount of data analyzed corresponds to an integrated luminosity of

$$\mathcal{L}_{\text{tot}} = (36.3 + 41.5 + 59.7)\ \text{fb}^{-1} \simeq 138\ \text{fb}^{-1},$$

recorded by CMS during the Run-II period. The instantaneous luminosity of the LHC increased steadily from 2016 to 2018, reaching peak values above $2 \times 10^{34}\ \text{cm}^{-2}\text{s}^{-1}$. As a consequence, the average number of simultaneous proton–proton interactions per bunch crossing rose from approximately $\langle \text{PU} \rangle = 27$ in 2016 to about 37 in 2018. This PU environment significantly affects object reconstruction, isolation, and missing transverse momentum, making its accurate modeling essential for precision measurements.

Pileup in simulation. In MC samples, pileup interactions are simulated by overlaying additional inelastic *pp* collisions, so-called *minimum-bias* events, on top of the hard-scattering process. These additional interactions are generated with Pythia8 [105] using the CP5 tune [106], and the number of overlaid vertices per event is drawn from a Poisson distribution matching the expected LHC conditions for each data-taking year.

To reproduce the pileup conditions observed in data, a dedicated reweighting procedure is applied event-by-event. The weight assigned to each simulated event depends on the true number of interactions n_{PU} generated in the event:

$$w(n_{\text{PU}}) = \frac{P_{\text{data}}(n_{\text{PU}})}{P_{\text{MC}}(n_{\text{PU}})}, \tag{3.1}$$

where P_{data} and P_{MC} denote the normalized distributions of the number of interactions in data and simulation, respectively. These distributions are derived from the official CMS luminosity information and pileup profiles corresponding to each year:

- 2016: `2016_25ns_UltraLegacy_PoissonOOTPU`,
- 2017: `2017_25ns_UltraLegacy_PoissonOOTPU`,
- 2018: `2018_25ns_UltraLegacy_PoissonOOTPU`.

3.3.4 Monte Carlo simulation

All simulated samples used in this analysis are based on the UL MC productions corresponding to the Run-II data-taking years. Events are processed into the NanoAODv9 format, which provides a compact representation optimized for large-scale analysis. The production campaigns and their corresponding tags are listed below:

- `Summer20UL16_106x_nAODv9_HIPM_Full2016v9` (2016 pre-VFP),
- `Summer20UL16_106x_nAODv9_noHIPM_Full2016v9` (2016 post-VFP),
- `Summer20UL17_106x_nAODv9_Full2017v9` (2017),
- `Summer20UL18_106x_nAODv9_Full2018v9` (2018).

The underlying event is modeled using the CP5 tune of Pythia8, and the parton distribution functions (PDFs) are described by the NNPDF3.1 NNLO [107] set.

Detector simulation

All events are processed through the Geant4-based [108] CMS detector simulation, which accurately reproduces the detector geometry, material distribution, and electronic response. The same reconstruction software used for data is applied to the simulated events, ensuring consistent definitions of physics objects (leptons, jets, and missing transverse momentum).

Signal modeling

The EW Wjj signal, corresponding to the pure $\mathcal{O}(\alpha_W^4)$ component, is generated at leading order (LO) using Madgraph5_aMC@NLO [109]. Events are required at generator level to contain exactly one lepton–neutrino pair and two quarks in the final state. Separate samples are produced for positively and negatively charged W bosons:

$$pp \to W^+(\to \ell^+\nu_\ell)jj, \qquad pp \to W^-(\to \ell^-\bar{\nu}_\ell)jj.$$

The parton shower and hadronization are simulated with Pythia8 using the dipole recoil scheme (`SpaceShower:dipoleRecoil = on`), which provides a more accurate description of additional hadronic radiation in color-singlet processes such as VBF and VBS [110]. At the generator level, minimal technical cuts of $p_T(j) > 10$ GeV and $m_{jj} > 10$ GeV are applied to ensure a stable and efficient integration of the matrix element over the phase space. These loose requirements have negligible impact on the analysis phase space, where a dedicated selection of $m_{jj}^{\mathrm{LHE}} > 200$ GeV is later applied to isolate the EW Wjj topology and to suppress the contribution from diboson (WV) production in the semileptonic final state, which is formally included in the signal definition.

The EW production of a W boson in association with two jets represents a rare process compared to the QCD-induced Wjj production, which dominates the total yield by roughly two orders of magnitude. Interference effects between the EW and mixed QCD amplitudes have been evaluated at leading order and found to contribute less than 0.2% of the total yield in the signal phase space, and are therefore neglected in the nominal signal model.

Process	Dataset name	σ [pb]
Wp_LNuJJ_EWK	`EWKWPlus2Jets_WToLNu_M-50`	39.04
Wm_LNuJJ_EWK	`EWKWMinus2Jets_WToLNu_M-50`	32.17

Table 3.3: Summary of the simulated signal samples used in the analysis. Each dataset corresponds to the EW production of a W boson in association with two jets, generated at LO with MADGRAPH5_aMC@NLO and showered with PYTHIA8.

Background modeling

All major SM background processes are simulated using centrally produced samples, generated and maintained within the CMS Collaboration. These samples are produced by the dedicated CMS Physics Object Groups and generator-level working groups, ensuring a consistent and validated configuration of the MC event generation, parton shower (PS), and detector simulation. The complete list of samples, including dataset names, cross sections, and generator settings, is provided in Tables 3.4 and 3.5.

- **W+jets and Drell–Yan:** generated at leading order with MADGRAPH5_aMC@NLO using the MLM [111] jet–parton matching scheme and interfaced to PYTHIA8 for parton showering and hadronization. To increase statistical precision in the 2-jet–enriched phase space relevant for this analysis, H_T-binned samples are used and stitched to the inclusive samples following Ref. [112]. Here, H_T denotes the scalar sum of the transverse momenta of all partons entering the matrix-element calculation.

- **Top-quark processes ($t\bar{t}$ and single-top):** simulated at next-to-leading order in QCD using POWHEG [113–116] interfaced with PYTHIA8. When explicit spin correlations are required, top quark decays are handled with MADSPIN [117]. The single-top s-channel is generated with MADGRAPH5_aMC@NLO

- **Diboson (WW, WZ, ZZ) and VBS:** generated at next-to-leading order with MADGRAPH5_aMC@NLO and interfaced to PYTHIA8 for parton showering and hadronization. For purely EW samples, the dipole-recoil configuration (`SpaceShower:dipoleRecoil = on`) is enabled to improve the description of radiation in color-singlet topologies. All samples include up to two additional partons in the matrix element.

- **Triboson (WWW, WWZ, WZZ, ZZZ):** produced with MADGRAPH5_aMC@NLO at NLO accuracy and showered with PYTHIA8 using the CP5 tune.

- **Vector-boson+photon ($W\gamma$, $Z\gamma$):** generated either at LO with MADGRAPH5_aMC@NLO using MLM matching or at NLO with MADGRAPH5_aMC@NLO using FXFX [118] merging, both interfaced to PYTHIA8.

- **VBF Z+jets:** the EW Zjj process is simulated with MADGRAPH5_aMC@NLO and PYTHIA8, adopting the dipole-recoil scheme for a consistent treatment of color-singlet radiation across the VBF processes.

- **Higgs processes:** gluon-fusion, vector-boson fusion, and associated (VH) production are generated with POWHEG and PYTHIA8; decays to WW are modeled with JHUGEN [119, 120] to preserve spin correlations.

All samples are normalised to the most precise theoretical cross sections available (typically NNLO in QCD for inclusive modes) and scaled to the integrated luminosity of each data-taking year. Tables 3.4 and 3.5 summarise all the MC samples used in this analysis, including their generation setup, nominal cross sections, and reference links to the corresponding CMS or LHC documentation.

Simulation–data corrections

To ensure that simulated events reproduce the detector and reconstruction performance observed in data, each event is assigned a weight that combines a number of corrections. These account for differences in detector conditions, trigger and reconstruction efficiencies, and physics-object identification across data-taking years.

Corrections related to the detector and beam conditions include the *pileup reweighting*, which adjusts the distribution of additional proton–proton interactions per bunch crossing to match that observed in data, the *Level-1 prefiring correction* that compensates for trigger inefficiencies in the ECAL endcaps during the early Run-II period [121], and the correction to the efficiency of the *pileup jet identification* algorithm, ensuring consistent modeling of spurious jets in the forward region.

Physics-object–related corrections are derived from control samples in data and include the scale factors for *lepton reconstruction, identification, and isolation*, measured as functions of the lepton transverse momentum and pseudorapidity using tag-and-probe methods in $Z \to \ell\ell$ events [104]. The *trigger efficiency* is similarly corrected to account for differences between online and offline selection. Jets receive a *b-tagging efficiency scale factor*, derived from samples enriched in heavy-flavor jets, to ensure a consistent identification of b hadrons between data and simulation.

All these factors are applied multiplicatively to each simulated event so that the resulting samples reproduce, as closely as possible, the detector response and object efficiencies observed in data. The associated uncertainties are treated as nuisance parameters in the final statistical interpretation.

Sample Name	nanoAOD Path	σ(pb)	Ref.
Wjets_HT	WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	61526.7	E
	WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8	1264.0	W
	WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	1256.00	W
	WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	335.500	W
	WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	45.2500	W
	WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	10.9700	W
	WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	4.93300	W
	WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	1.16000	W
	WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.02678	I
top	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	87.310	E
	ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8	3.34368	D
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	26.2278	D
	ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8	44.07048	D
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85	D
	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	35.85	D
	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	364.35	E
DY_M-50	DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6077.22	E
	DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	146.5	W
	DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	160.7	W
	DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	48.63	W
	DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	6.993	W
	DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	1.761	W
	DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.8021	W
	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.1937	W
	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8	0.003514	W
DY_else	DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	18610.0	E
	DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	54.39	W
	DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	5.697	W

Table 3.4: Table containing all the samples included in the analysis. References are clickable links.

Sample Name	nanoAOD Path	σ(pb)	Ref.
VV	WminusToLNuWminusTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.03319	I
	WminusToLNuZTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	1.144	I
	WminusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.3393	I
	WplusTo2JWminusToLNuJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	4.86	I
	WplusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.5615	I
	WplusToLNuWplusTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.0757	I
	WplusToLNuWminusTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	4.853	I
	WplusToLNuZTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	1.884	I
	ZTo2LZTo2JJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.3345	I
VBS	WminusToLNuZTo2JJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.1028	I
	WminusTo2JZTo2LJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.03055	I
	WplusTo2JZTo2LJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.05631	I
	WplusToLNuZTo2JJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.1888	I
	WplusToLNuWplusTo2JJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.09107	I
	WminusToLNuWminusTo2JJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.03373	I
	WplusToLNuWminusTo2JJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.9442	I
	WplusTo2JWminusToLNuJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.9445	I
	ZTo2LZTo2JJJ_dipoleRecoil_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.01646	I
VVV	ZZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.01476	W
	WZZ_TuneCP5_13TeV-amcatnlo-pythia8	0.05709	W
	WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.1707	W
	WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8	0.2158	W
Vg	WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	412.7	W
	ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8	51.1	W
VBF-Z	EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8	6.215	W
Higgs	GluGluHToWWToLNuQQ_M-125_TuneCP5_13TeV_powheg_jhugen751_pythia8	28.88	I
	HWminusJ_HToWWTo2L2Nu_WTo2L_M-125_TuneCP5_13TeV-powheg-pythia8	0.001348	F
	HWplusJ_HToWWTo2L2Nu_WTo2L_M-125_TuneCP5_13TeV-powheg-pythia8	0.002124	F
	VBFHToWWToLNuQQ_M-125_TuneCP5_withDipoleRecoil_13TeV_powheg_jhugen751_pythia8	3.892	I

Table 3.5: Table containing all the samples included in the analysis. References are clickable links.

3.4 Object definitions and event selection

This section describes the reconstruction and selection of the physics objects entering the analysis, namely leptons, jets, and the missing transverse momentum (p_T^{miss}). All definitions follow the recommendations of the relevant CMS Physics Object Groups (POGs) for the UL Run II datasets, with additional requirements designed to suppress backgrounds and enhance the efficiency for EW Wjj events in a VBF-like topology.

3.4.1 Electrons

Electrons are reconstructed using the Particle-Flow (PF) algorithm and identified through the multivariate (MVA) electron identification developed by the EGamma POG. The `Fall17V2Iso` MVA discriminator is employed [122], with the 90% efficiency working point (WP90), optimized to separate prompt electrons from nonprompt or misidentified hadrons. Two identification levels are used in the analysis: a *Loose* ID, designed to be HLT-safe and employed for the definition of fakeable objects in the estimation of nonprompt backgrounds (see Sec. 3.5.3), and a *Tight* ID, used for the selection of signal leptons.

Tight electrons are further required to satisfy the following kinematic and isolation criteria: $|\eta| < 2.5$, a relative isolation smaller than 0.06 in a cone of $\Delta R = 0.3$, and small impact parameters with respect to the primary vertex ($|d_{xy}| < 0.05$ cm, $|d_z| < 0.1$ cm). Candidates consistent with photon conversions are vetoed, and an additional requirement of `ttHMVA` > 0.7 is applied to further suppress nonprompt contamination. The `ttHMVA` discriminator combines several variables sensitive to the origin of the lepton—such as impact parameters, isolation, and properties of the associated jet—into an MVA classifier optimised to distinguish prompt electrons from those arising in heavy-flavour decays.

Data-to-simulation scale factors are applied to correct the reconstruction, identification, and isolation efficiencies, derived via the tag-and-probe technique in $Z \to ee$ events. A summary of all selection criteria for electrons is provided in Table 3.6.

3.4.2 Muons

Muon reconstruction follows the recommendations of the CMS Muon POG. Candidates must satisfy $|\eta| < 2.4$ and a relative isolation smaller than 0.15 in a cone of $\Delta R = 0.4$. Impact parameter requirements with respect to the primary vertex are applied: $|d_{xy}| < 0.01$ cm (or < 0.02 cm for $p_T > 20$ GeV) and $|d_z| < 0.1$ cm. Two identification levels are defined: a *Loose* ID, used to define fakeable leptons for the estimation of nonprompt backgrounds, and a *Tight* ID, applied to select signal muons. The Tight ID includes a further requirement on the multivariate `ttHMVA` discriminator, `ttHMVA`> 0.8, to enhance the rejection

ID	Variable	Cut (EB)	Cut (EE)
Loose	$\|\eta\| <$	2.5	2.5
	$\sigma_{\eta\eta} <$	0.0106	0.03
	$\|\Delta\eta_{in\ seed}\| <$	0.0032	0.00632
	$\|\Delta\phi_{in\ seed}\| <$	0.0547	0.0394
	H/E $<$	$0.046 + \frac{1.16}{E_{\mathrm{SC}}} + \frac{0.0324\rho}{E_{\mathrm{SC}}}$	$0.0275 + \frac{2.52}{E_{\mathrm{SC}}} + \frac{0.183\rho}{E_{\mathrm{SC}}}$
	relIsoWithEA $<$	$0.0478 + \frac{0.506}{p_T}$	$0.0658 + \frac{0.963}{p_T}$
	$\left\|\frac{1}{E} - \frac{1}{p}\right| <$	0.184	0.014
	Missing inner hits $\leq$	1	1
	Conversion veto	pass	pass
	$\|d_{xy}\| <$	0.05	0.1
	$\|d_z\| <$	0.1	0.2
Tight	mvaFall17V2Iso_WP90	pass	pass
	Conversion veto	pass	pass
	pfRelIso03 all $<$	0.06	0.06
	$\|d_{xy}\| <$	0.05	0.1
	$\|d_z\| <$	0.1	0.2
	ttHMVA $>$	0.7	0.7

Table 3.6: Selection criteria for electron Tight and Loose identification in the barrel (EB) and endcap (EE) regions.

of nonprompt muons.

Table 3.7 summarizes the full muon selection.

ID	Variable	Cut (EB)	Cut (EE)
Loose	$\|\eta\| <$	2.4	2.4
	(POG) tightId	pass	pass
	$\|d_z\| <$	0.1	0.1
	pfRelIso04 all $<$	0.4	0.4
	$\|d_{xy}\| <$	0.01 (for $p_T \leq 20$ GeV) 0.02 (for $p_T > 20$ GeV)	0.01 (for $p_T \leq 20$ GeV) 0.02 (for $p_T > 20$ GeV)
Tight	Loose	pass	pass
	pfRelIso04 all $<$	0.15	0.15
	ttHMVA $>$	0.8	0.8

Table 3.7: Selection criteria for Loose and Tight muon IDs in the barrel (EB) and endcap (EE) regions. The Loose ID corresponds to the HLT-safe working point.

3.4.3 Jets and b tagging

Jets are reconstructed from PF candidates using the anti-k_T algorithm [123] with a distance parameter $R = 0.4$ (AK4 jets). Charged hadrons not associated with the primary vertex are excluded to reduce pileup contamination. Jets overlapping with identified leptons within $\Delta R < 0.3$ are removed to avoid double-counting of energy deposits. All jets are required to pass the `Tight` Jet ID [104], and those with $p_T < 50$ GeV must also satisfy the `Loose` PU ID to further reject pileup jets. Jet energy corrections and resolutions follow the official JME POG prescriptions.

b jets are identified using the DeepCSV algorithm [124, 125]. Jets are required to have $p_T > 30$ GeV and $|\eta| < 2.5$, and the `Medium` working point is adopted, corresponding to an average misidentification probability of about 1% for light-flavor jets. The discriminator thresholds are 0.6001, 0.5847, 0.4506, and 0.4168 for the 2016 (HIPM), 2016 (noHIPM), 2017, and 2018 datasets, respectively.

The b-tagging efficiency is defined as the probability for a jet of a given flavour (b, c, or light) to satisfy the chosen working point. Since this efficiency differs between data and simulation, due to differences in detector response, tracking performance, and heavy-flavour modelling, corrections are applied to simulated events in the form of scale factors. These scale factors, provided by the BTV POG [126], are applied on an event-by-event basis and ensure that the tagging efficiencies in simulation reproduce those measured in data.

3.4.4 Missing transverse momentum

The missing transverse momentum vector, p_T^{miss}, is defined as the negative vectorial sum of all PF candidates in the event and is reconstructed using the PUPPI algorithm [127], which mitigates the impact of pileup interactions by assigning per-particle weights based on local event properties. Dedicated event filters recommended by the JME POG [128] are applied to remove events affected by detector noise or reconstruction failures.

3.4.5 Event selection

The event selection is designed to identify EW Wjj production in a topology consistent with the VBF one, while efficiently rejecting backgrounds from QCD-induced W+jets and top-quark processes. The selection proceeds in two steps: a common preselection applied to all regions, followed by additional requirements defining signal and control regions used to constrain the background contributions in the statistical analysis.

Preselection

Candidate events are required to contain exactly one isolated lepton (electron or muon) passing the "tight" identification criteria described in Secs.3.6 and 3.7, and at least two reconstructed jets. Events with an additional loose lepton with $p_T > 15$ GeV are vetoed to suppress dileptonic backgrounds from Drell–Yan and top quark decays.

The two leading jets in transverse momentum are used to define the dijet system. They are required to satisfy $p_T(j_1) > 40$ GeV and $p_T(j_2) > 30$ GeV, and to be widely separated in pseudorapidity ($|\Delta\eta_{jj}| > 2$) with a large invariant mass $m_{jj} > 400$ GeV. Jets in the transition region $2.5 < |\eta_j| < 3.0$ are excluded due to known modeling limitations in this detector region.

The selected lepton must have $|\eta_\ell| < 2.0$ and transverse momentum above the trigger efficiency plateau for the corresponding single-lepton paths. For each data-taking year, this corresponds to $p_T > 30$GeV for electrons (25GeV for muons) in 2016, $p_T > 38$GeV (30GeV) in 2017, and $p_T > 35$GeV (25GeV) in 2018.

Table 3.8 summarizes the preselection criteria applied uniformly to all data-taking periods and simulated samples.

Object	Requirements
AK4 Jet	At least 2 jets (j) with: $p_T(j_1) > 40$ GeV, $p_T(j_2) > 30$ GeV $m_{jj} > 400$ GeV, $\|\Delta\eta_{jj}\| > 2$ $\|\eta_j\| < 2.5$ or $\|\eta_j\| > 3$
Lepton	Exactly 1 tight lepton (l) with: $p_T(l) > 25$ GeV, $\|\eta_l\| < 2$

Table 3.8: Summary of preselection cuts applied to all MC and data samples. The region defined by $2.5 < |\eta_j| < 3.0$ is excluded since it is not well modeled by the MC simulations. Events with additional leptons with $p_T(l) > 15$ GeV are also excluded.

Control and signal regions

To enhance sensitivity and constrain the main backgrounds, events are categorized into orthogonal regions according to the jet, lepton, and b-tag content, as well as the transverse mass of the reconstructed W boson:

$$M_T(W) = \sqrt{2p_T^\ell p_T^{miss}(1 - \cos\Delta\phi(\ell, p_T^{miss}))}. \tag{3.2}$$

Top control region (Top CR). Events enriched in top-quark backgrounds are selected by requiring the presence of at least two b-tagged jets, identified

with the medium working point of the DeepCSV algorithm. Using a single b tag would still define a region orthogonal to the SR, but it would lead to a noticeably lower purity because of the larger contamination from W+jets. Requiring two b-tagged jets instead yields a much cleaner top-enriched control region, which is used to validate the modeling of $t\bar{t}$ and single-top processes and to constrain their normalization in the final fit.

Nonprompt lepton control region (Nonprompt CR). The nonprompt control region targets QCD multijet events where jets are misidentified as leptons. It is defined by requiring a transverse mass $20 < M_T(W) < 50$ GeV, where the genuine W contribution is suppressed while nonprompt leptons dominate. A veto on b-tagged jets is applied to reduce contamination from top-quark processes, and events are further required to satisfy $|\Delta\eta_{jj}| \leq 5$ and $p_{T,j_2} \leq 70$ GeV in order to avoid overlap with the VBF- and W+jets-enriched regions. The method used to estimate this background is detailed in Sec. 3.5.3.

W+jets control region (W+jets CR). The first control region targets QCD-induced W+jets events in which both leading reconstructed jets originate from the hard scattering. Events are required to pass $M_T(W) > 50$ GeV, to contain no b-tagged jets, and to satisfy $|\Delta\eta_{jj}| < 4$ and $m_{jj} \leq 750$ GeV. This selection enriches the sample in prompt-jet W+jets production while maintaining negligible overlap with the signal region. At generator level, both leading reconstructed jets are matched to particle-level jets within a cone of $\Delta R < 0.3$.

W+1 PU jet control region (W+1 PU jet CR). A second control region isolates W+jets events where one of the two leading reconstructed jets originates from pileup activity rather than from the hard interaction. Such events are characterized by a low transverse momentum of the subleading jet and a large pseudorapidity separation. They are selected by requiring a b-jet veto, $M_T(W) > 50$ GeV, PuppiMET > 70 GeV, $p_{T,j_2} < 70$ GeV, and $|\Delta\eta_{jj}| > 5$. At generator level, these events correspond to configurations in which only one of the two leading reconstructed jets can be matched to a generated jet within $\Delta R < 0.3$. Because the prompt-jet and pileup-jet components have distinct matrix-element origins and exhibit different kinematic shapes, they are modeled separately using the two dedicated control regions.

Signal region (SR). The signal region targets events consistent with EW Wjj production. To maximize purity, events must satisfy $M_T(W) > 50$ GeV, $|\Delta\eta_{jj}| > 4$, and the subleading jet is required to have $p_T > 70$ GeV, which implicitly ensures a high-p_T leading jet as well. A veto on b-tagged jets is applied to suppress top-quark backgrounds.

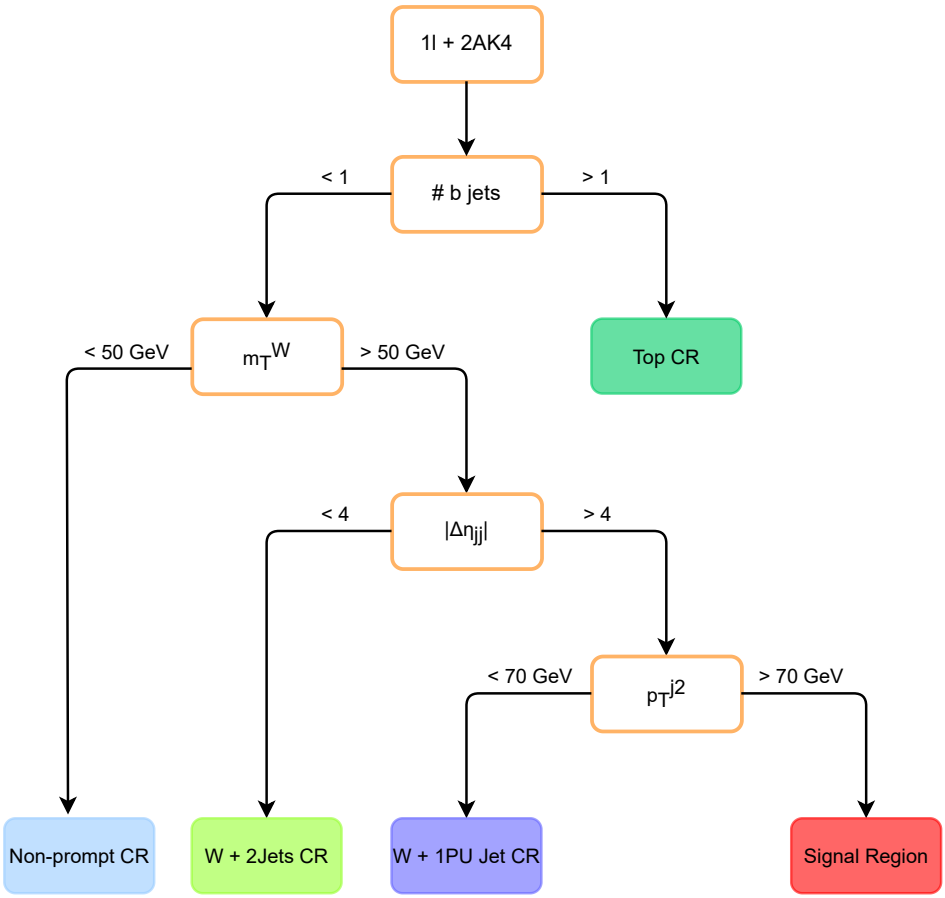


Figure 3.5: Schematic representation of the event categorization workflow used in the analysis. Only the key selection criteria that define orthogonal regions are shown, illustrating how events are partitioned into the signal and control regions. Additional selection requirements, applied within each region, are used to enhance the purity in the targeted process (signal or specific background). A complete summary of the selection criteria defining all regions is provided in Table 3.9.

3.5 Experimental Modelling of Backgrounds

The main SM backgrounds to the EW Wjj signal were discussed in Section 3.2 from a theoretical perspective. This section focuses instead on their experimental modelling and on the strategies adopted to ensure a precise data-to-simulation agreement. A combination of data-driven methods and MC corrections is employed to estimate and constrain the different background contributions. Where simulations are used, MC events are reweighted or normalized to match distributions observed in data, while for purely instrumental or non-prompt sources, data-driven techniques are applied.

Region	Requirements
Top CR	≥ 2 b-tagged jets
Non-prompt CR	b-jet veto $p_{T,j_2} \leq 70$ GeV $\lvert\Delta\eta_{jj}\rvert \leq 5$ $20 < M_T(W) < 50$ GeV
W+Jets CR	b-jet veto $\lvert\Delta\eta_{jj}\rvert \leq 4$ $M_T(W) > 50$ GeV $m_{jj} \leq 750$ GeV
W+1 PU Jet CR	b-jet veto $p_{T,j_2} \leq 70$ GeV $\lvert\Delta\eta_{jj}\rvert > 5$ $M_T(W) > 50$ GeV PuppiMET > 70 GeV
Signal Region	b-jet veto $p_{T,j_2} > 70$ GeV $\lvert\Delta\eta_{jj}\rvert > 4$ $M_T(W) > 50$ GeV

Table 3.9: Summary of the selection cuts defining all regions used in the analysis. These selections are applied on top of the preselection cuts listed in Table 3.8.

3.5.1 Top background

Top quark production, which includes $t\bar{t}$ pair production, associated production of a top quark and a W boson, and single top quark production, is a significant background in this analysis. This category is collectively referred to as the top quark background. To suppress this background, events with b-tagged jets are rejected based on b-jet identification. The DeepCVS b-tagging algorithm is used to reject events containing at least two jets originating from B hadron decays, thereby substantially reducing the top quark background in the signal region.

Additionally, the transverse momentum of the top quark in $t\bar{t}$ data events is observed to be significantly softer than predicted by simulations, which are based on either leading-order (LO) or next-to-leading-order (NLO) matrix elements interfaced with parton showers. This residual discrepancy is likely due to higher-order QCD and/or EW corrections. To address this, a correction is applied to $t\bar{t}$ events in the form of a parametrized weight as a function of the generator-level top and antitop quarks p_T. This weight is derived from the ratio of the p_T distribution in NNLO+EWK and NLO samples:

$$SF(p_T) = 0.103 \cdot e^{-0.0118 \cdot p_T} - 0.000134 \cdot p_T + 0.973$$

The normalisation of the top quark background will be left floating in the final fit, and the top control region will be used in the combined fit to measure its normalisation.

3.5.2 W + Jets background

The mixed QCD production of a W boson in association with jets represents the primary background in this analysis, contributing up to 65% of the total events in the signal region. The nominal W+jets sample used in the analysis corresponds to the *hard-scattering component*, in which both leading jets originate from the hard matrix element. It is simulated at LO with separate H_T energy bins. Discrepancies between data and simulation are observed in several kinematic distributions, including the m_{jj} and $|\Delta\eta_{jj}|$ of the dijet system and the p_T of the individual jets and of the leptonic W boson. These discrepancies, consistent across data-taking years and event categories, are attributed to limitations in the LO matrix-element description, parton showering, and hadronisation modelling in the W+jets simulation.

On the left of Figures 3.7 and 3.8, several jet-related observables are shown for the electron and muon channels, illustrating these discrepancies.

To address this issue, two data-driven techniques have been developed to improve the modeling of the W+Jets background using control regions. The main method, based on a functional reweighting of the W+Jets simulation, is described in detail in Section 3.5.2 and constitutes the nominal approach used in the analysis. An alternative strategy, relying on the introduction of rate parameters to correct the W+Jets yield and shape directly within the statistical model, is presented in Appendix A.1. Although conceptually similar, this second method is not used in the nominal fit and serves as a validation of the robustness of the adopted correction.

Alternative MC samples with higher-order matrix element calculations and lower parton multiplicities were also tested (see Section 3.5.2). However, these samples suffer from limited statistical precision in the phase space relevant for this analysis, particularly in events with two or more jets and large m_{jj}, resulting in sizable fluctuations and no significant improvement in the modelling of the kinematic distributions.

The normalization of the W+Jets background is left floating in the final fit, and a dedicated control region is included in the combined likelihood to constrain it. Additional details on the systematic uncertainties associated with this process are provided in Section 3.7.

Reweight through a fitted function

The first strategy adopts a data-driven correction procedure developed to account for potential mismodeling in the simulation of W+Jets events. Within the W+Jets CR, all background processes estimated from simulation, including the non-prompt component, are subtracted from the observed data, assuming their normalisations are well constrained. The remaining yield, dominated by genuine W+Jets events, is then compared to the simulated prediction on a bin-by-bin basis. The correction is derived as a function of the transverse momentum of the leading jet, identified as the observable with the largest modelling discrepancies. The resulting ratio is fitted with a falling exponential function to suppress statistical fluctuations at high p_T, and the nominal fit function is subsequently used as a smooth correction weight applied consistently across all regions of the analysis.

Systematic uncertainties arising from the fitting procedure are evaluated by employing, in addition to the nominal model, a second-order polynomial function. This allows potential biases due to the choice of analytical form to be assessed. The resulting systematic uncertainty is propagated to the final statistical fit. The nominal falling exponential function used in the reweighting is defined with four free parameters. To evaluate the statistical uncertainty associated with the fit, all exponential functions consistent with the same overall normalization are considered, varying each of the four parameters up and down by its fitted uncertainty. Figure 3.6 shows the ratio distribution for the W+Jets sample in the corresponding control region, together with the nominal fit and the up/down statistical variations from the reweighting procedure, shown separately for electrons (left) and muons (right).

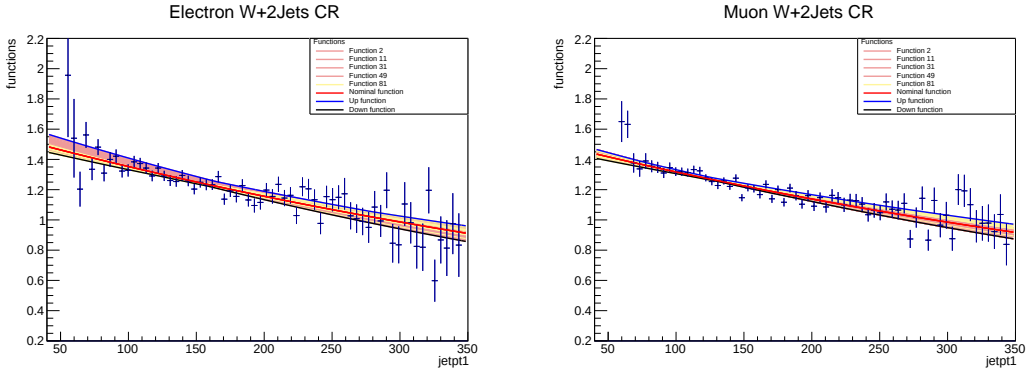


Figure 3.6: Reweight function derived in W+Jets CR for the W+Jets background. Red: nominal reweighting function. Blue and black: up/down statistical variations due to the statistical fluctuation of the sample. Overflow and underflow bin content is considered when necessary. Uncertainties are propagated in the final fit.

In figures 3.7 and 3.8, some data/MC distributions in the W+Jets CR before (left) and after (right) the reweighting procedure is applied, for electron and muon final states, respectively.

Comparison of LO HT-binned and NLO Jet-binned samples

Two different NLO (QCD) samples are available from central production. The comparison was performed using the 2018 dataset as a benchmark. Several key kinematic distributions were examined, showing that the NLO sample differs from the LO one by an overall factor of 0.76. After normalising both samples to emphasise only shape differences, statistical fluctuations in the NLO sample became visible in the signal region (SR). The results are presented in Figure 3.9.

Dataset Name
/WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-pythia8
/WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8
/WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8
/WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8

Table 3.10: Available centrally produced NLO (QCD) samples for W+Jets. The first entry is an inclusive NLO sample, while the following entries are n-jet binned samples. The inclusive sample covers the high-multiplicity jet phase space.

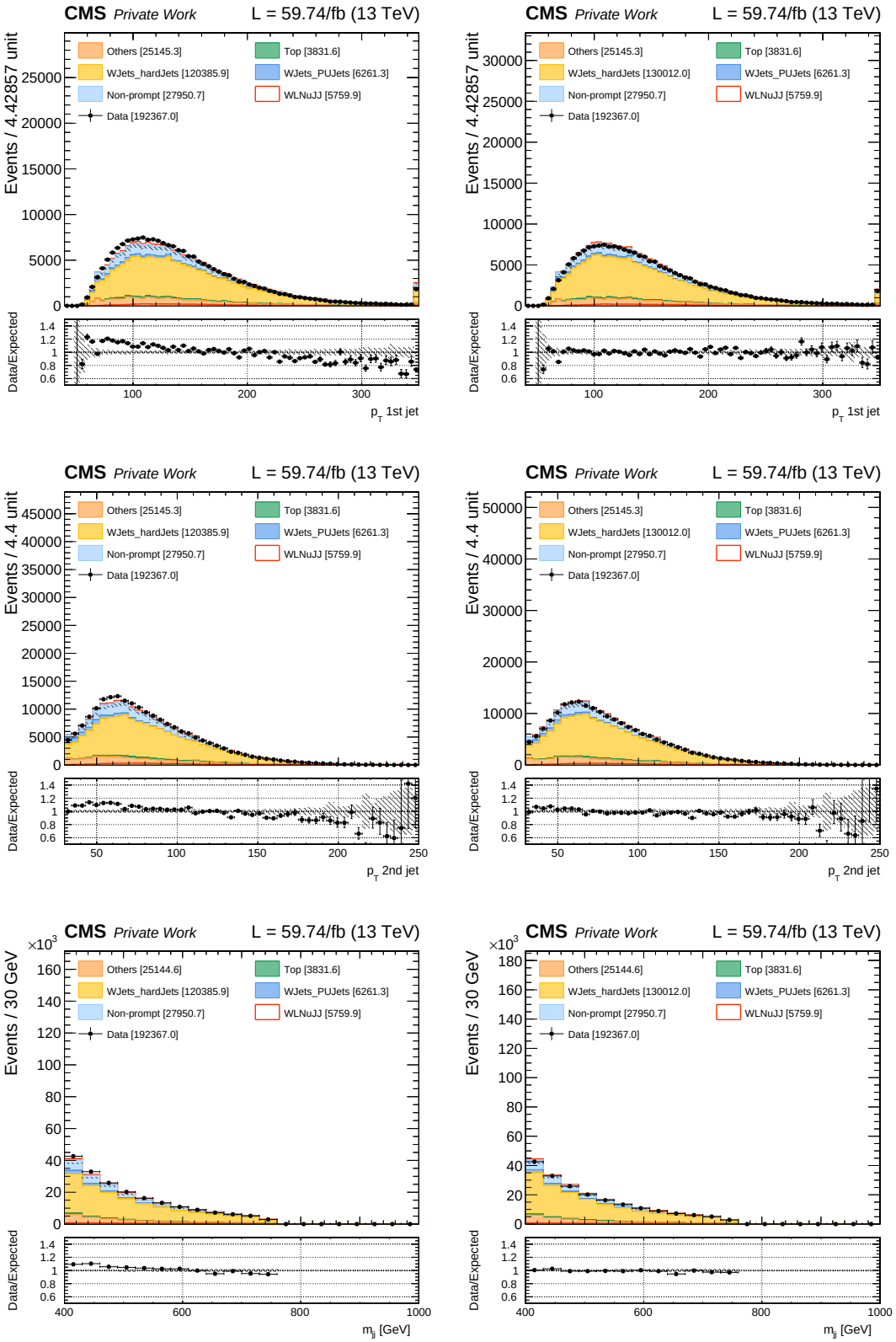


Figure 3.7: Data/MC agreement in the W+Jets CR for the electron channel. From top to bottom: distributions of the p_T of the leading and sub-leading jet in the event and the m_{jj} of the di-jet system. From left to right: pre- and post-reweighting the sample with the procedure in Sec. 3.5.2. Statistical uncertainty is shown in shaded grey in the ratio plot.

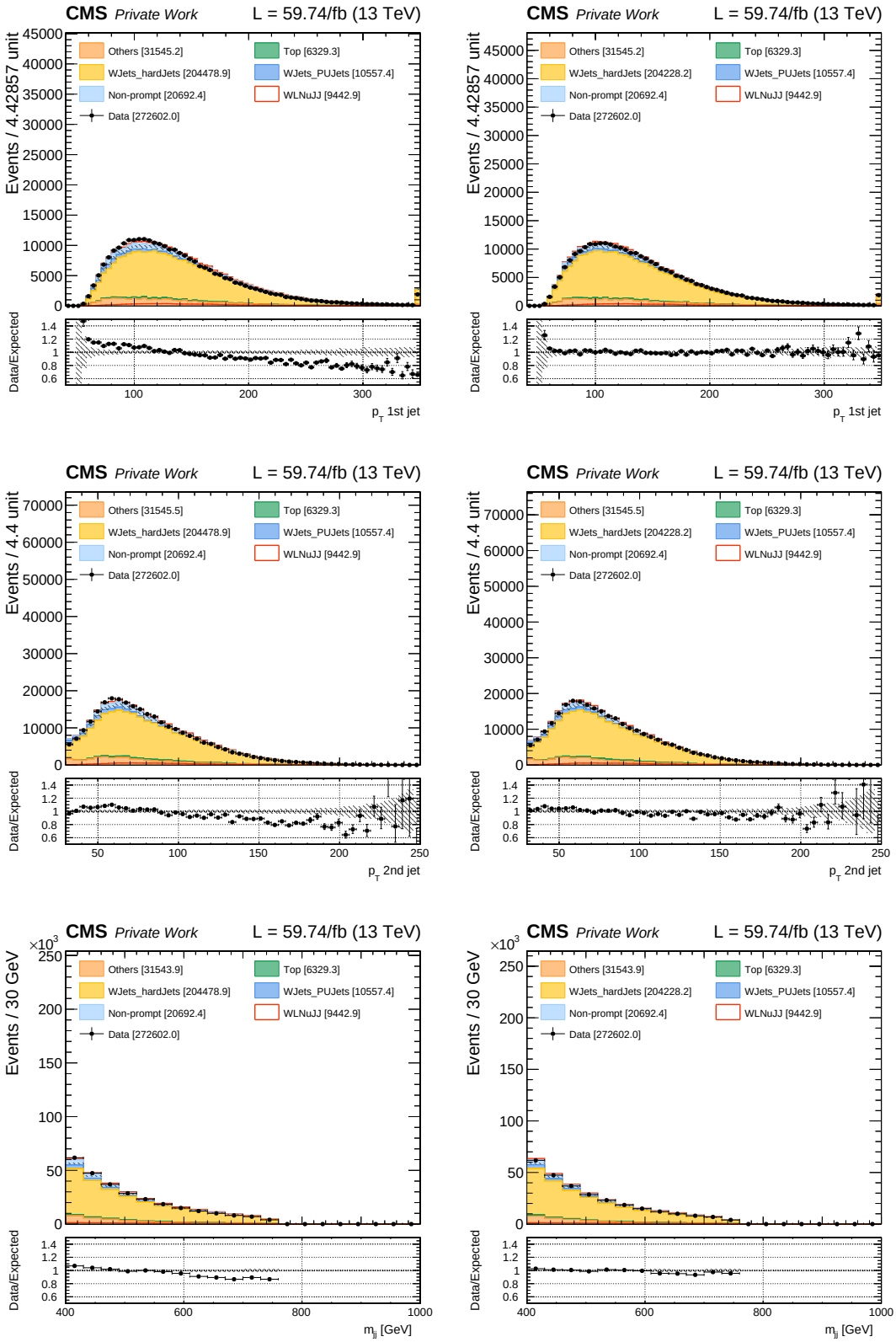


Figure 3.8: Data/MC agreement in the W+jets control region for the muon channel. From top to bottom: distributions of the p_T of the leading and subleading jets and of the di-jet invariant mass m_{jj}. Left: before reweighting. Right: after applying the functional reweighting described in Sec. 3.5.2. The grey band represents the statistical uncertainty in the ratio plot.

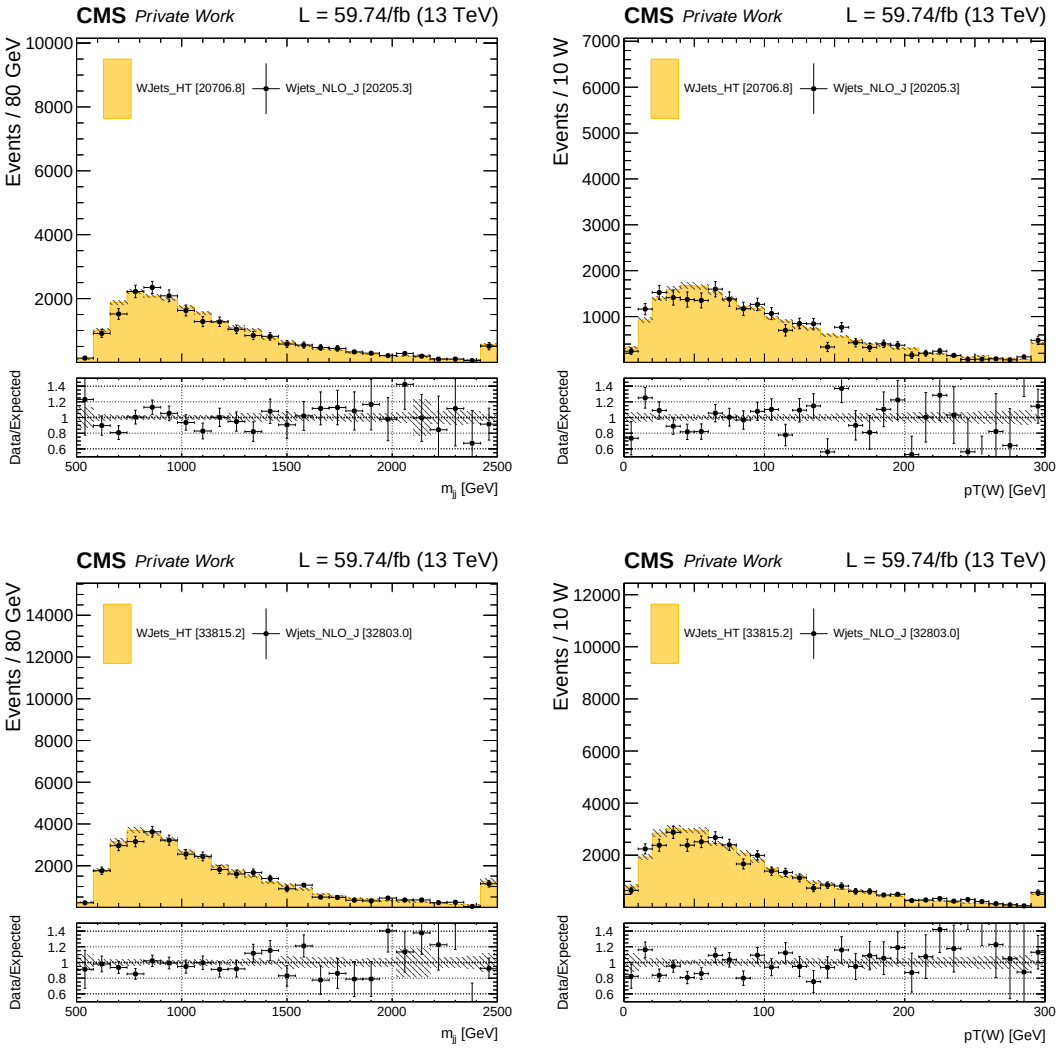


Figure 3.9: NLO/LO comparison in the SR after normalizing the NLO to the LO (reweighted) sample. The grey band represents the statistical uncertainty of the LO sample, with no systematic uncertainty shown. Left: m_{jj} of the di-jet system. Right: p_T of the lepton–neutrino system. Top: electron channel. Bottom: muon channel.

3.5.3 Non-prompt background

Among the selected events, a small fraction of leptons does not originate from the prompt decay of a W boson but from hadron decays or jet misreconstruction. These are referred to as *non-prompt* or *fake* leptons. Their contribution is particularly relevant in the electron channel, where it accounts for about 10% of the total yield in the signal region, while in the muon channel it is typically below 4%.

This difference can be traced back to both detector effects and the physical mechanisms that generate fake leptons. The main sources of non-prompt leptons are: (i) jets misidentified as leptons, for instance when a charged hadron accompanied by a nearby neutral pion produces an electromagnetic shower resembling that of an electron; (ii) semileptonic decays of hadrons containing heavy quarks, which affect both lepton flavours similarly; and (iii) photon conversions or pair production in the detector material, which predominantly lead to electrons because the e^+e^- production cross section is orders of magnitude larger than that for $\mu^+\mu^-$. As a result, the non-prompt background is expected to be substantially larger in the electron channel, while muons, thanks to their distinct penetration pattern and dedicated tracking in the muon system, are much less affected.

As described in Section 3.4, two lepton working points are defined: a *loose* selection, which relaxes identification and isolation requirements and therefore retains both prompt and non-prompt leptons, and a *tight* selection, specifically optimised to suppress leptons that are only partially isolated or originate from secondary processes such as hadron decays or photon conversions. The tight working point is used in the final analysis regions, where the residual non-prompt contribution becomes small but non-negligible.

To estimate this residual contamination, a data-driven method is employed. Its goal is to determine the probability that a lepton originating from a non-prompt source satisfies the tight selection. This probability, known as the *fake rate*, is measured in a control region enriched in QCD multijet events, characterised by low transverse mass $M_T(W)$ and small missing transverse momentum p_T^{miss}, where non-prompt leptons dominate the sample. The measured fake rate is then used to predict the non-prompt contamination in all analysis regions directly from data.

The method, commonly referred to as the *fakeable object method*, relies on events containing loosely selected leptons and quantifies the fraction that also pass the tight selection. The control regions employed in the method are illustrated schematically in Figure 3.10. The QCD-enriched region, dominated by multijet events, is used to measure the fake rate $f(p_T, \eta)$, while the DY and top-enriched regions are employed to determine the prompt rate $p(p_T, \eta)$ and to extract the scale factors applied to simulated samples. Since these control regions are defined using leptons that satisfy at least the loose selection, no dedicated scale factors exist to correct the simulation-to-data efficiency in this domain. The DY- and top-enriched regions therefore play a crucial role in validating that the

MC description remains accurate also for loosely selected leptons, ensuring that the subtraction of residual prompt contamination in the QCD-enriched region is reliable.

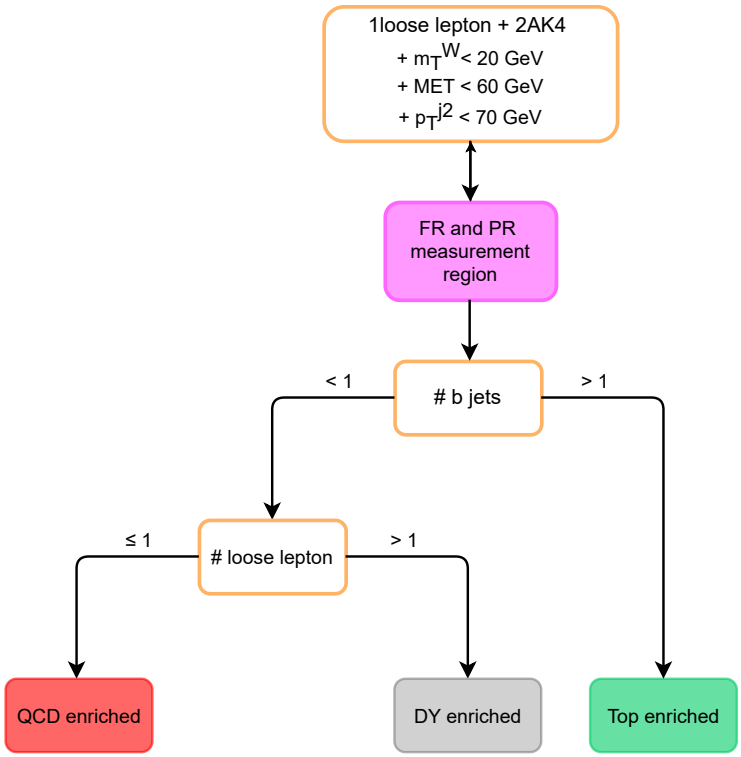


Figure 3.10: Schematic representation of the control regions employed in the measurement of the fake and prompt rates. All regions are defined in the low-$M_T(W)$ and low-p_T^{miss} phase space, outside the kinematic domain of the main analysis. The QCD-enriched region, dominated by multijet events, is used to extract the fake rate $f(p_T, \eta)$, while the DY- and top-enriched regions, defined with similar kinematics, are used to determine the prompt rate and to validate the Monte Carlo description of loosely selected leptons.

A detailed summary of the event selections defining the three control regions is reported in Table 3.11. The common preselection criteria ensure similar kinematic conditions across all regions, while the specific requirements on b-jets, lepton multiplicity, and invariant mass $m_{\ell\ell}$ select the desired QCD-, DY-, and top-enriched samples respectively.

Mathematical formulation

The underlying concept of the method can be formalised as follows. Assuming that the total number of leptons, N_l, passing the loose selection consists of a prompt component (N_p) and a non-prompt component (N_f), one can write:

$$N_l = N_p + N_f = N_{t0} + N_{t1}, \tag{3.3}$$

Region	Requirements
Common selections	At least 2 jets (j) with: $p_T(j_1) > 40$ GeV, $30 < p_T(j_2) < 70$ GeV $\|\eta_j\| < 2.5$ or $\|\eta_j\| > 3$ At least one loose lepton with: $p_T > 25$ GeV (or according to HLT), $\|\eta_l\| < 2$ $M_T(W) < 20$ GeV, $p_T^{\text{miss}} > 60$ GeV,
QCD-enriched	b-veto and exactly one loose lepton
DY-enriched	b-veto and at least two loose leptons with $60 < m_{\ell\ell} < 120$ GeV
Top-enriched	At least two b-jets

Table 3.11: Summary of the selection cuts defining the three control regions used for the measurement of the fake and prompt rates. Common preselection criteria ensure comparable kinematics, while specific requirements on b-jets, lepton multiplicity, and invariant mass $m_{\ell\ell}$ isolate the QCD-, DY-, and top-enriched samples respectively.

where N_{t0} and N_{t1} represent the number of events with zero and one tight lepton, respectively.

Although N_p and N_f are unknown, they can be expressed in terms of the fake rate (f) and the prompt rate (p), which represent the probabilities for a non-prompt and a prompt lepton, respectively, to satisfy the tight selection criteria:

$$N_{t0} = (1-p)N_p + (1-f)N_f \; N_{t1} \quad = pN_p + fN_f \tag{3.4}$$

By measuring p and f, which depend on the lepton (p_T, η), the values of N_f can be inferred from N_{t0} and N_{t1} as:

$$N_f = \frac{1}{p-f}, [pN_{t0} - (1-p)N_{t1}], . \tag{3.5}$$

Finally, the number of non-prompt leptons passing the tight selection is given by $N_f^{\text{pass}} = f \times N_f$.

The fake rate $f(p_T, \eta)$ is measured from a QCD-enriched region, requiring exactly one lepton with the same p_T threshold as in the analysis, and two jets with p_T greater than 40 and 30 GeV, respectively. Additional selection criteria include PuppiMET < 60 GeV, subleading jet $p_T < 70$ GeV, $M_T(W) < 20$ GeV and b-veto. In this phase space, contributions from top quark processes are controlled by requiring two tightly b-tagged AK4 jets with $p_T > 30$ GeV and $\eta < 2.5$. Similarly, the DY contribution is assessed by selecting events with an additional lepton and $60 < m_{ll} < 120$ GeV. The prompt rate $p(p_T, \eta)$ is measured in the DY-enriched mentioned region.

Table 3.11 summarizes the selection criteria and the definitions of the different regions used to measure the fake and prompt rates. These regions are defined

before applying the analysis tight lepton selection, as outlined in Tables 3.8 and 3.9.

Measurement

The fake and prompt rates are computed as the ratio of leptons passing the *tight* selection to those passing the *loose* selection, evaluated in different control regions. In each (p_T, η) bin, the quantities $tight_{\text{DATA}}$ and $loose_{\text{DATA}}$ denote the number of events in data containing, respectively, a tight or a loose lepton. Similarly, $tight_{\text{MC}}$ and $loose_{\text{MC}}$ correspond to the same counts in MC simulation. Since the QCD-enriched region used for the fake rate measurement still contains a small fraction of prompt leptons, this contribution is subtracted using Monte Carlo prediction, corrected by a scale factor (SF) that accounts for differences between data and simulation.

The scale factors are derived as the integral ratio of data to MC yields in dedicated control regions enriched in the relevant prompt processes. Specifically, the DY- and top-enriched regions introduced in Section 3.5.3 are used to determine the SF for DY and top-quark processes, respectively, by computing the inclusive ratio $SF = N_{\text{DATA}}/N_{\text{MC}}$ after common preselection. For W+jets events, no pure control region can be defined in data. Therefore, the same scale factor derived from the DY-enriched region is used as a proxy, as the lepton reconstruction and isolation efficiencies are expected to be similar between W+jets and DY processes. The subtraction of the residual prompt contamination is then performed using $SF \cdot tight_{\text{MC}}$ and $SF \cdot loose_{\text{MC}}$ in the numerator and denominator of Eq. 3.6, respectively.

The resulting expressions for the fake and prompt rates are:

$$\begin{aligned} f(p_T, \eta) &= \frac{tight_{\text{DATA}} - SF \cdot tight_{\text{MC}}}{loose_{\text{DATA}} - SF \cdot loose_{\text{MC}}} \quad \text{(QCD-enriched region)} \\ p(p_T, \eta) &= \frac{tight_{\text{DATA}}}{loose_{\text{DATA}}} \quad \text{(DY-enriched region)} \end{aligned} \tag{3.6}$$

The fake rate $f(p_T, \eta)$ thus represents the probability that a non-prompt lepton passing the loose selection also satisfies the tight requirements, after correcting for prompt contamination. Both f and p are measured in bins of lepton transverse momentum and pseudorapidity to capture the dependence of the efficiencies on kinematics. The binning schemes used for all years are:

1. η binning: [-2.0, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5, 2.0]

2. p_T binning (2016, 2018): [25.0, 30.0, 35.0, 40.0, 45.0, 60.0, 75.0, 120.0, 200.0]

3. p_T binning (2017): [30.0, 34.0, 38.0, 41.0, 45.0, 60.0, 75.0, 120.0, 200.0]

At high transverse momentum, statistical uncertainties become dominant due to the limited number of events available in the QCD-enriched region. In particular, the last two p_T bins ($p_T > 75$ GeV) suffer from large fluctuations and are therefore not used in the analysis. For leptons above 75 GeV, the fake rate is instead taken to be constant and equal to the value measured in the last statistically reliable bin.

Figures 3.11 and 3.12 show the measured fake rates as a function of p_T and η for electrons and muons, respectively. Analogous distributions for the prompt rates are shown in Figures 3.15 and 3.16. Further details, including alternative parameterisations and studies of the p_T–$\Delta\phi_{jj}$ dependence, are provided in Appendix B.4.

Although the electron fake rate tends to plateau at a roughly constant value, the behaviour observed for muons is markedly different: the muon fake rate increases with p_T and approaches unity at high transverse momenta, independently of the detector region. While this trend may appear counterintuitive, since higher-p_T muons are generally expected to be easier to distinguish from nonprompt objects, it is a direct consequence of how the fake rate is defined.

The fake rate is the ratio between the number of candidates passing the tight selection and the number of candidates passing a loose preselection. In the electron case, the loose selection is genuinely loose, and the denominator retains many objects that fail the tight identification, leading to a stable ratio. For muons, however, the loose selection is effectively constrained by the requirements imposed by the single-muon trigger. `HLT_IsoMu24` enforces L1 $p_T > 22$ GeV, L3 $p_T > 24$ GeV with stringent quality criteria, and an online isolation requirement ($I_{\text{trk}}/p_T < 0.07$). These conditions remove most non-isolated or low-quality muons before they ever enter the fake-rate denominator. At high p_T, nearly all muons that satisfy the trigger also satisfy the tight offline identification, causing the ratio to approach one.

Using a looser trigger would mitigate this effect, but such triggers are prescaled and would lead to fake-rate measurements dominated by statistical fluctuations. The observed increase of the muon fake rate with p_T is therefore an intrinsic consequence of the available trigger strategy rather than a feature of the muon identification itself.

A closure test of the full method is performed in the non-prompt control region defined in Section 3.4.5, using a selection of $20 < M_T(W) < 50$ GeV. The agreement between data and the background prediction obtained with the fake-rate method is shown in Figures 3.17 and 3.18.

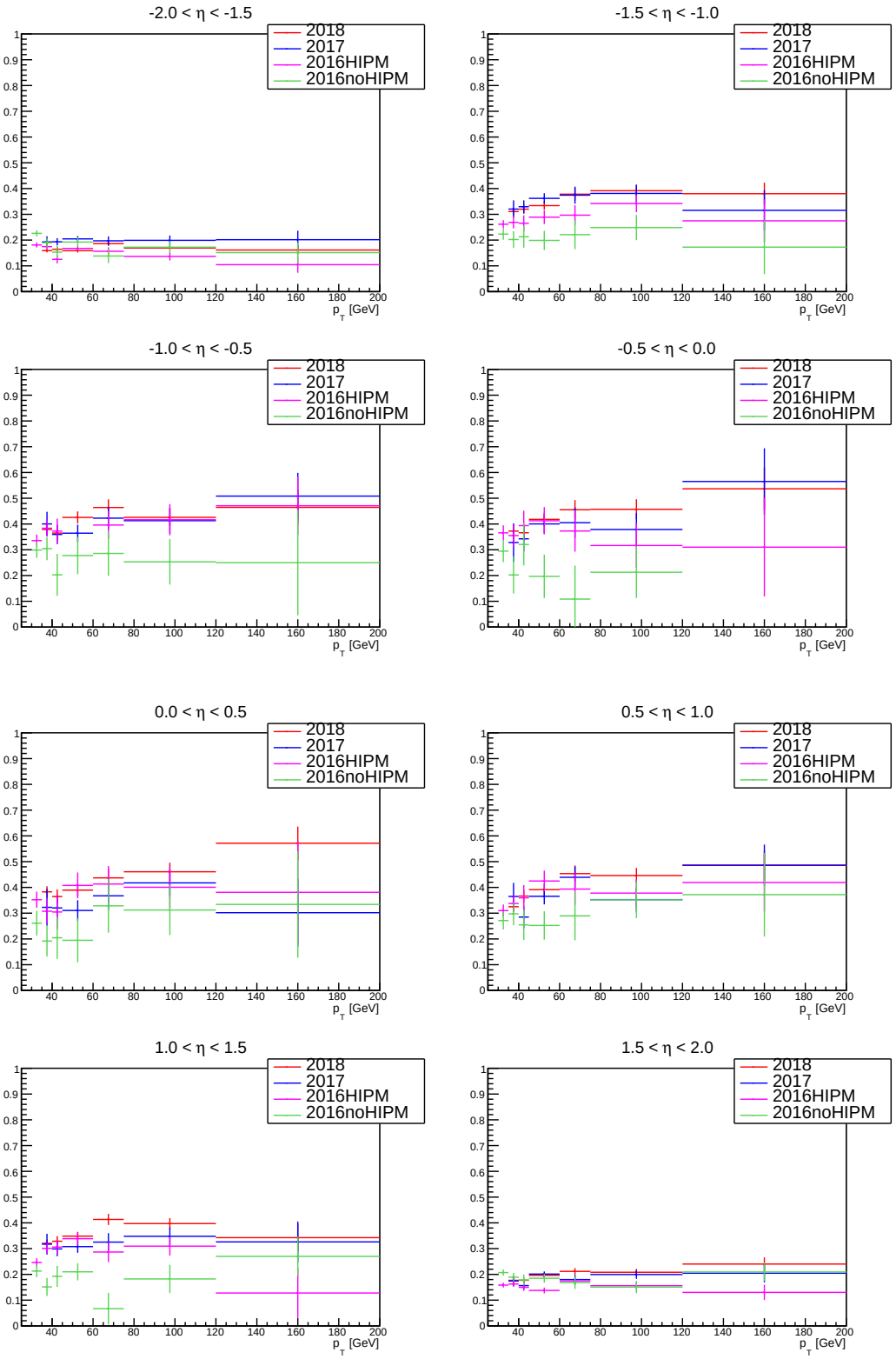


Figure 3.11: Fake Rate (FR) measurement as a function of p_T and η of the electron. All the years are shown here.

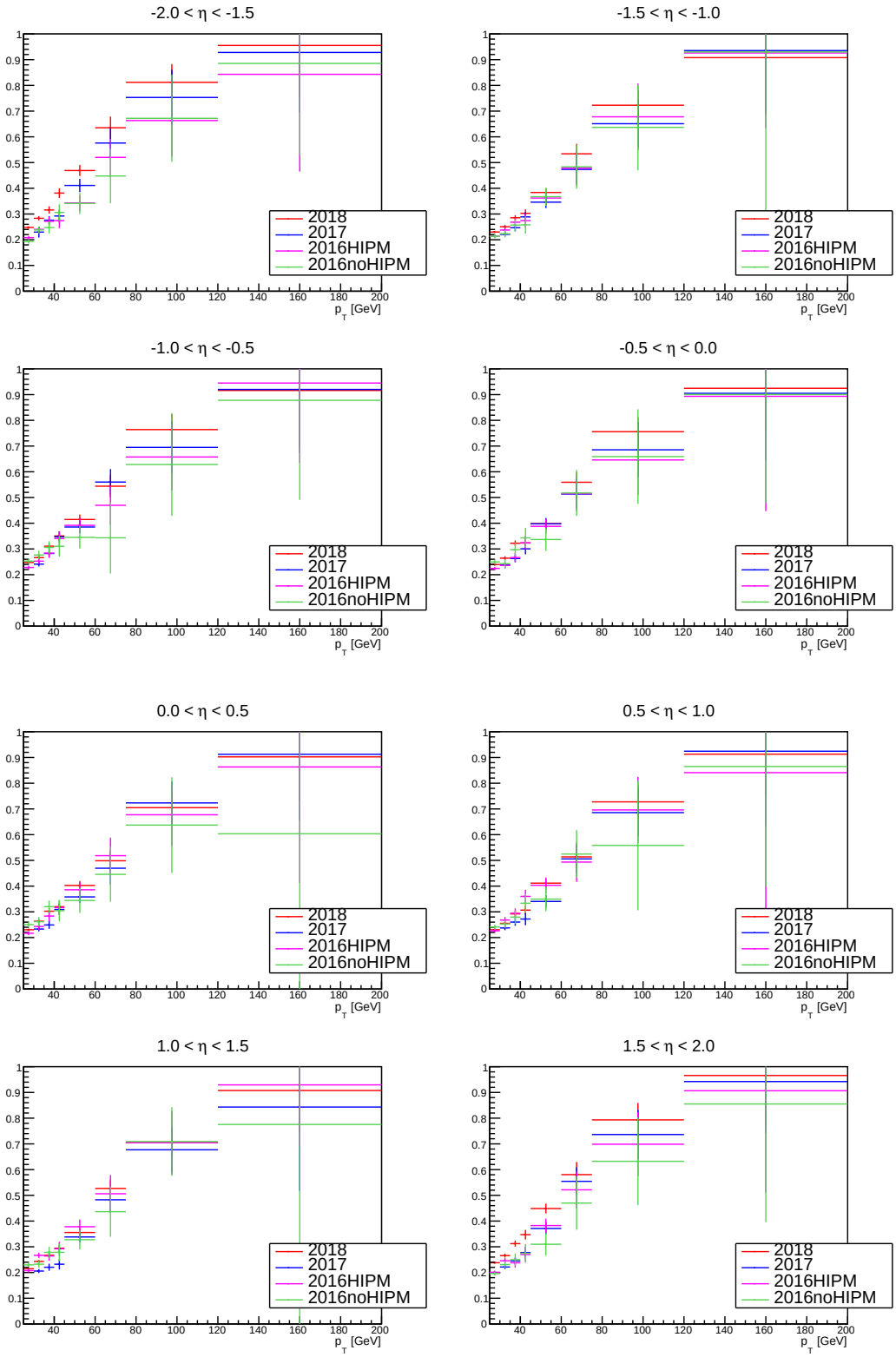


Figure 3.12: Fake Rate (FR) measurement as a function of p_T and η of the muon. All the years are shown here.

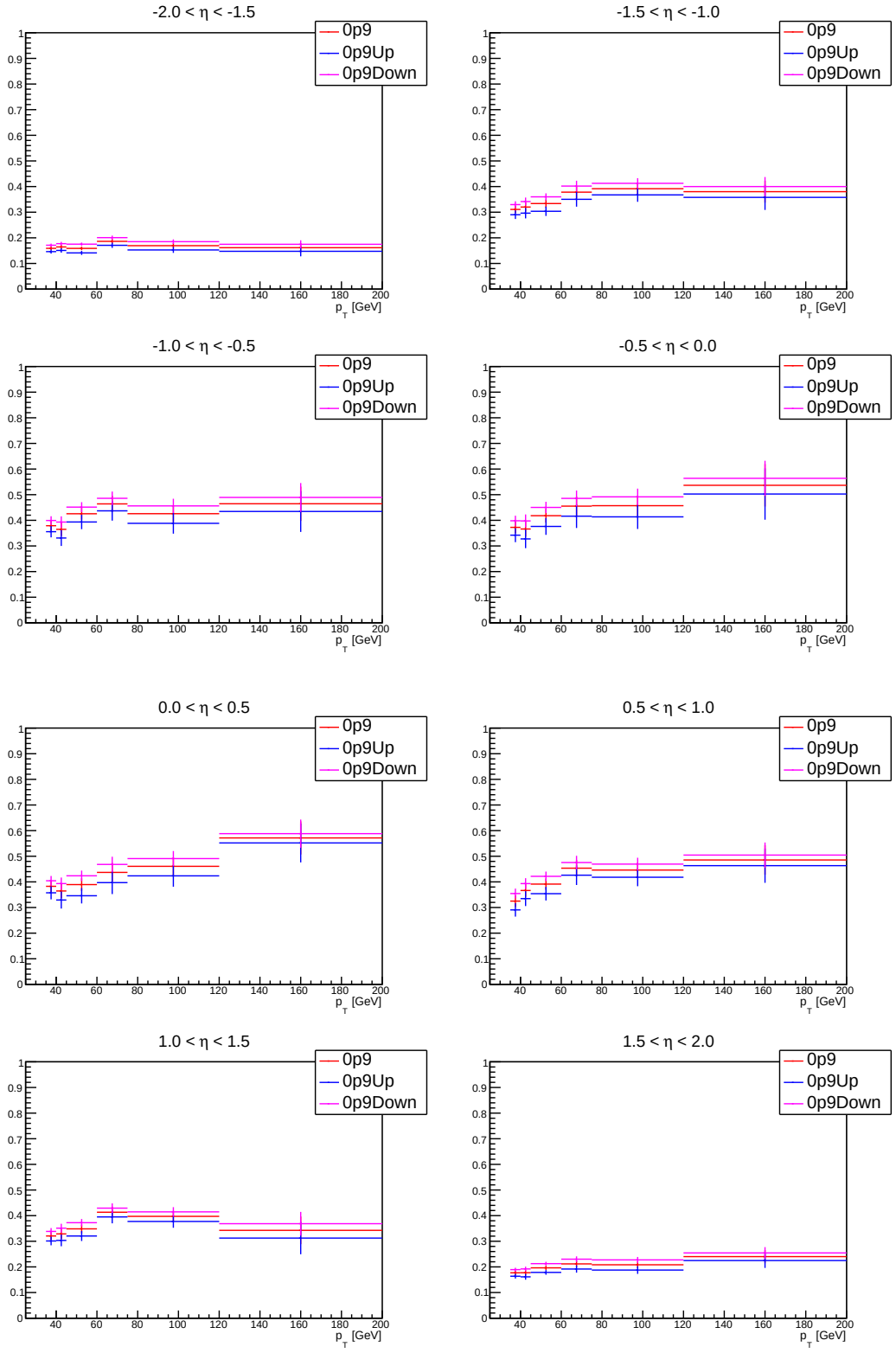


Figure 3.13: Electron FR ewk scale factor (SF) 10% variation up/down. This variation is due to a unique SF used to subtract all the residual prompt contamination from all the different processes in the QCD measurement region (Eq. 3.6). Nominal is considered to be the SF measured for the DY process. Only the 2018 year is depicted here, see B for the remaining years.

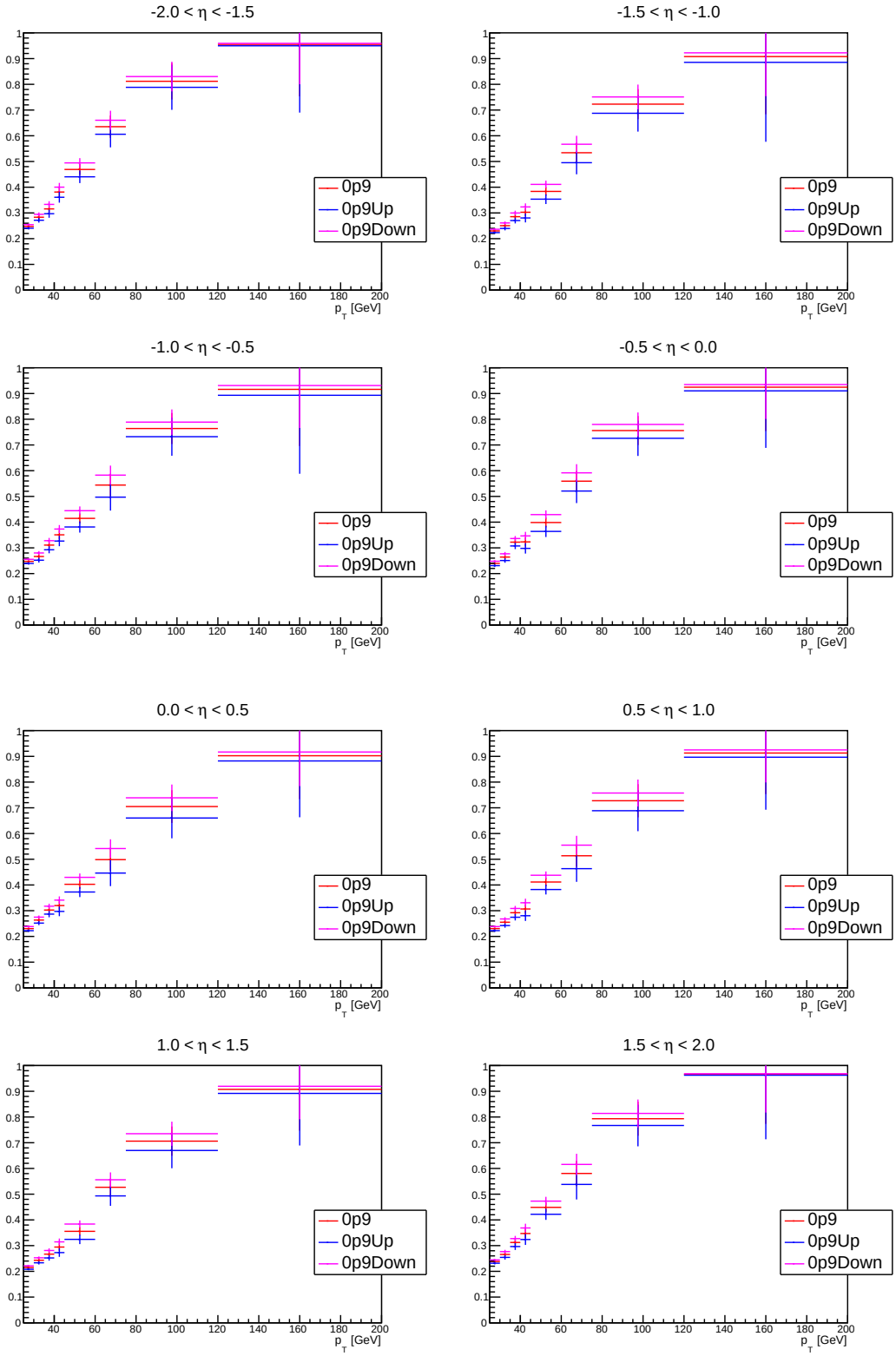


Figure 3.14: Muon FR ewk scale factor (SF) 10% variation up/down. This variation is due to a unique SF used to subtract all the residual prompt contamination from all the different processes in the QCD measurement region (Eq. 3.6). Nominal is considered to be the SF measured for the DY process. Only the 2018 year is depicted here, see B for the remaining years.

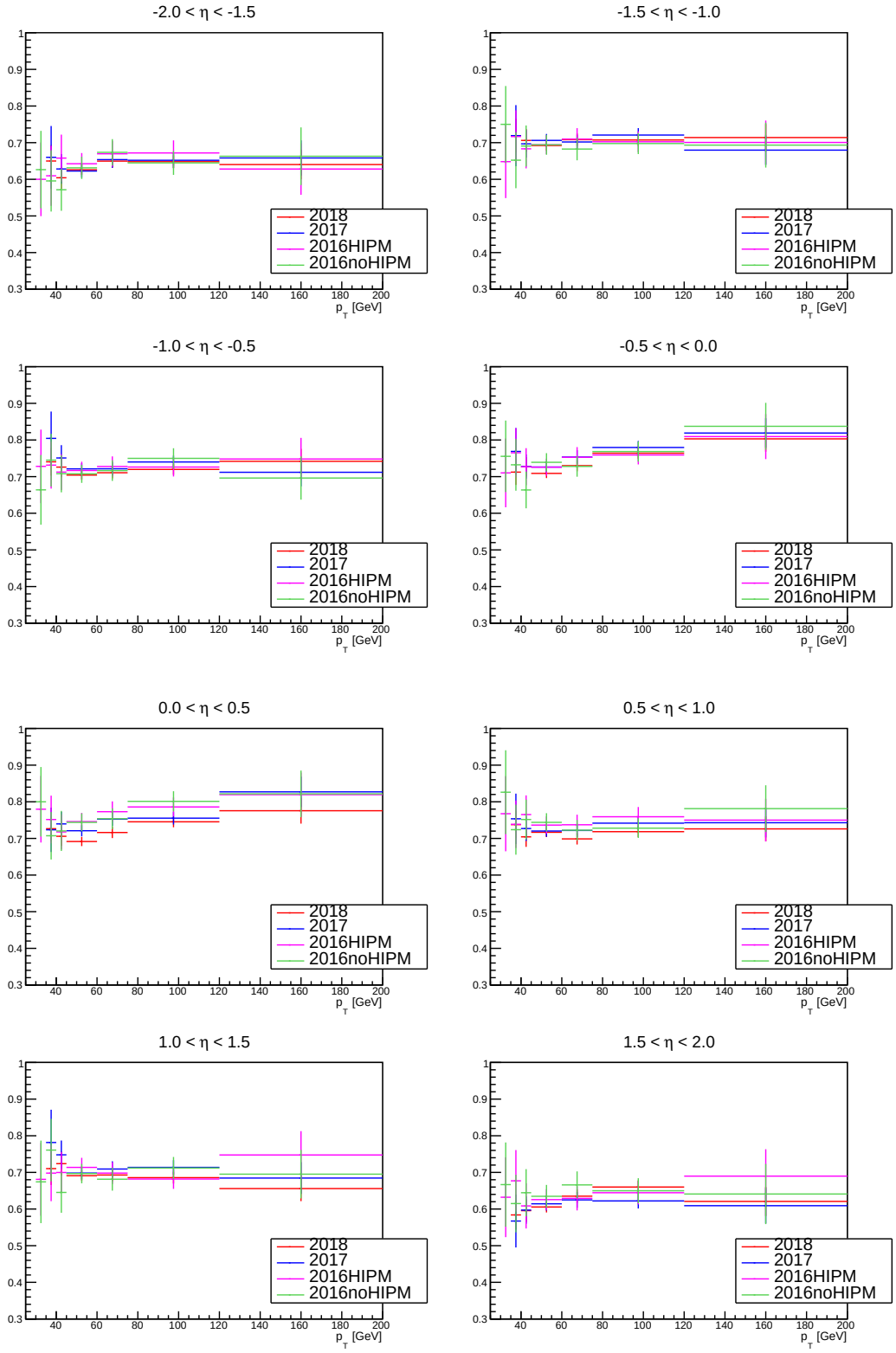


Figure 3.15: Prompt Rate (PR) measurement as a function of p_T and η of the electron. All the years are shown here. Note the very different Y-axis scale with respect to the FR plots.

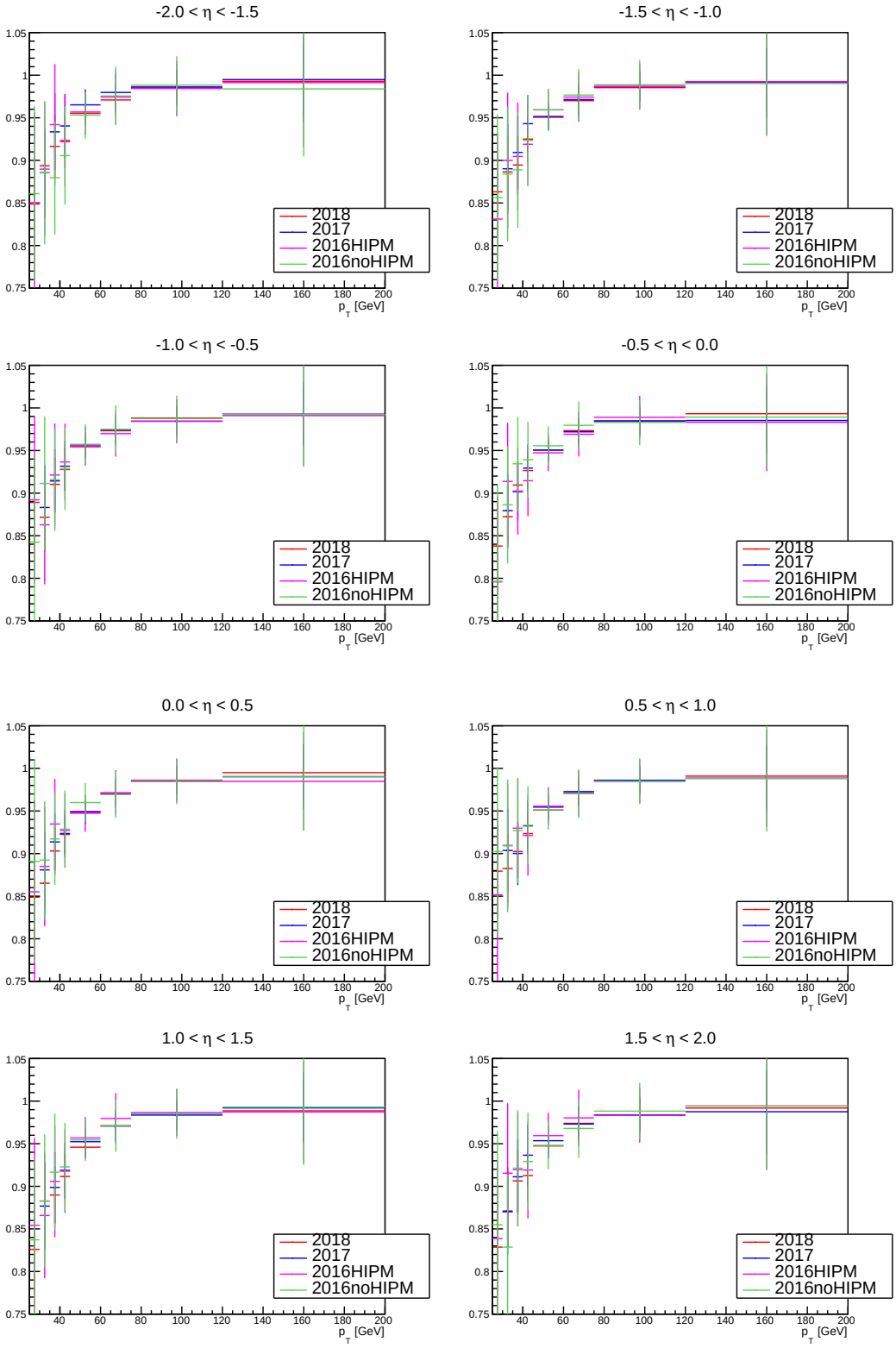


Figure 3.16: Prompt Rate (PR) measurement as a function of p_T and η of the muon. All the years are shown here. Note the very different Y-axis scale with respect to the FR plots.

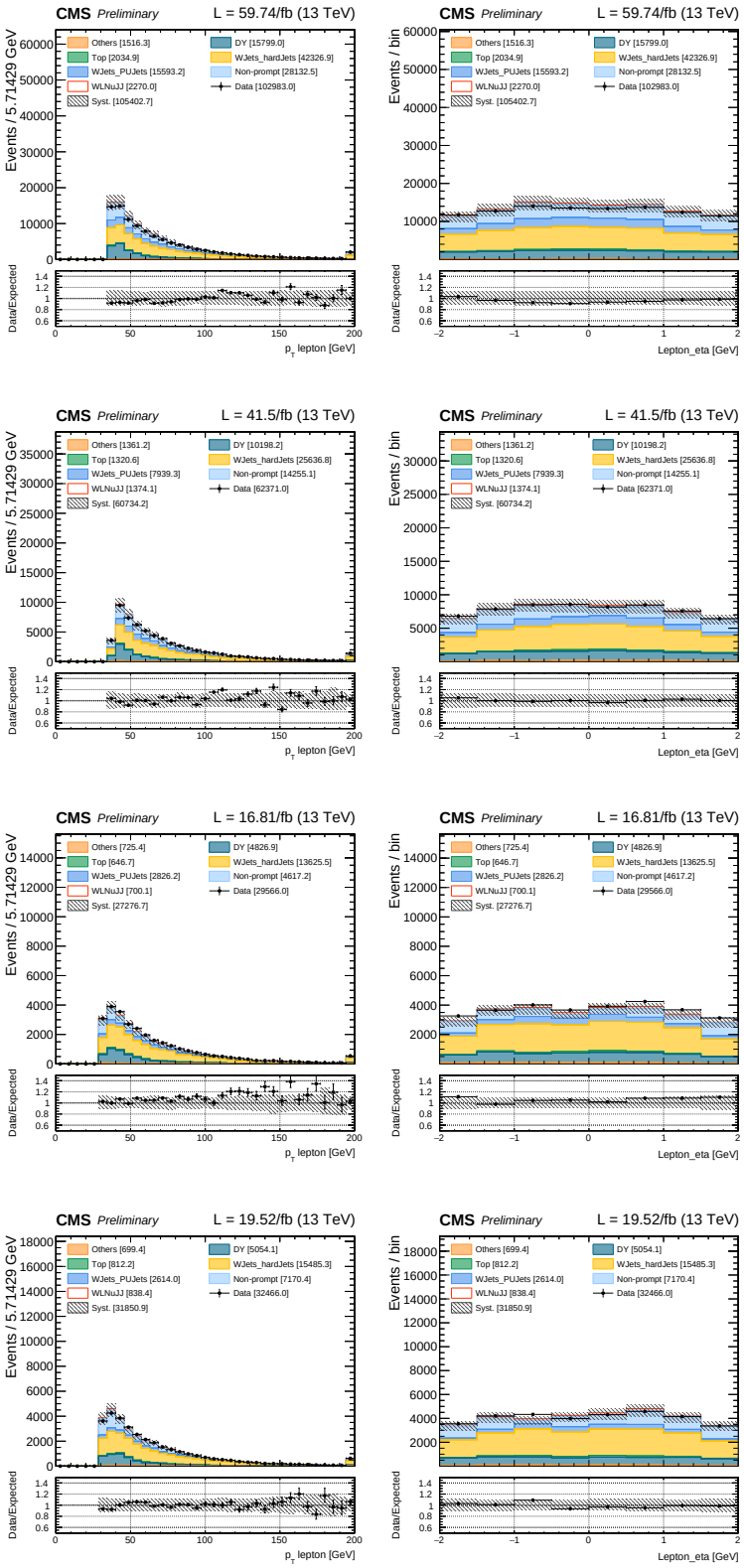


Figure 3.17: Agreement in the non-prompt CR for the electron p_T (left) and η (right). From top to bottom: 2018, 2017, 2016 noHIPM, and 2016 HIPM. The grey shaded areas represent the systematic uncertainties related to the non-prompt estimation. The dominant uncertainty is the sample statistics.

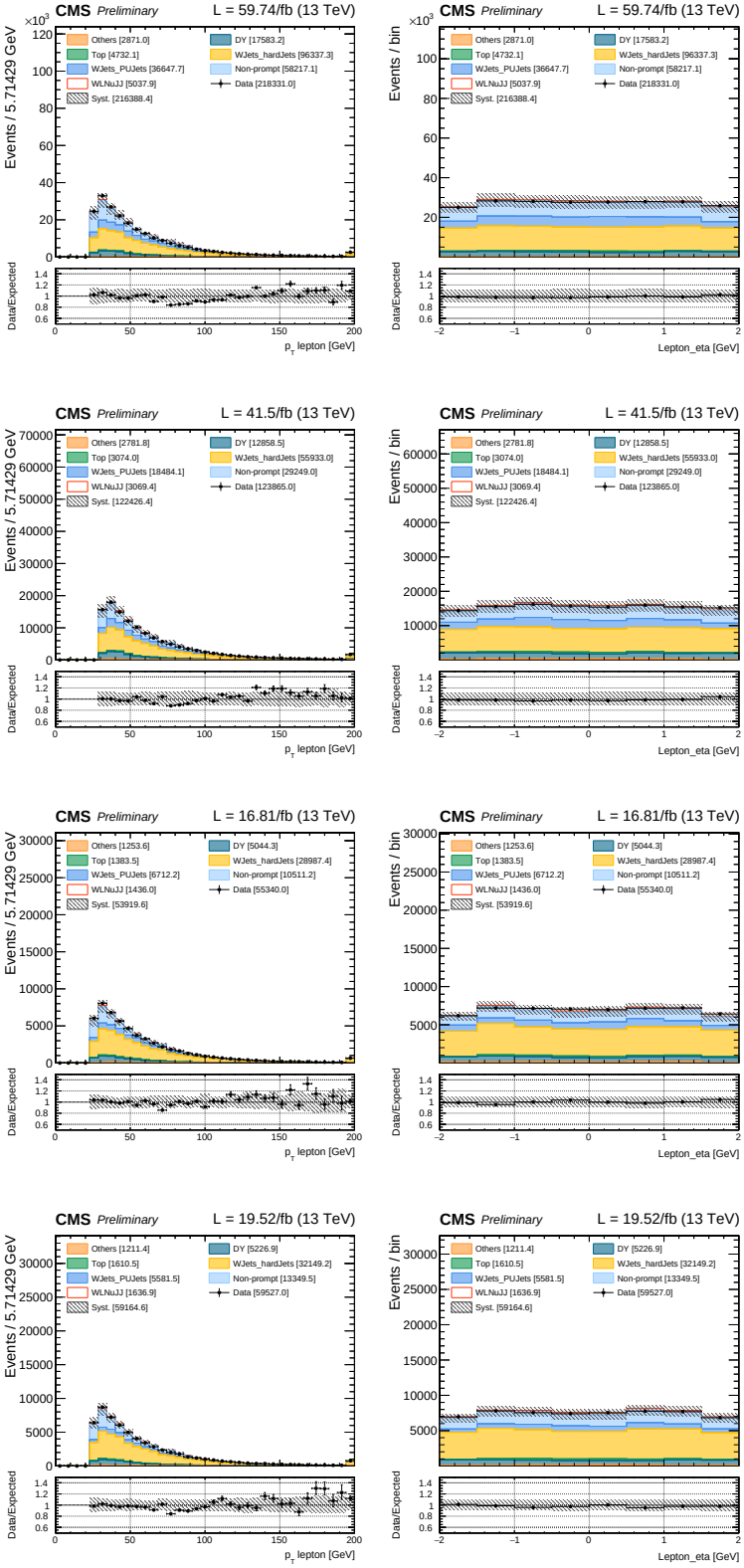


Figure 3.18: Agreement in the non-prompt CR for the muon p_T (left) and η (right). From top to bottom: 2018, 2017, 2016 noHIPM, and 2016 HIPM. The grey shaded areas represent the systematic uncertainties related to the non-prompt estimation. The dominant uncertainty is the sample statistics.

3.6 Signal Discrimination

A feed-forward Deep Neural Network (DNN) is employed to maximise the separation between the pure EW Wjj signal and the dominant QCD-induced W+jets background. The output of this DNN, referred to as the DNN score, is later used as an input observable in the maximum likelihood fit. As the invariant mass of the dijet system is one of the most powerful single kinematic variables to distinguish signal from background, it is used as a benchmark to quantify the DNN performance. Comparisons are also made against other relevant observables, such as the Zeppenfeld variable (Z_l) and the two-dimensional combination (m_{jj}, Z_l). For consistency in the comparison, the Zeppenfeld variable is transformed as $-|Z_l|$, such that larger values correspond to signal-like topologies. The same convention is applied in the two-dimensional (m_{jj}, Z_l) representation, ensuring that both observables vary monotonically with the signal probability.

3.6.1 Training setup

The DNN is trained using simulated events from the SR, as defined in Section 3.4.5. The training is designed to separate the EW Wjj signal from the QCD-induced W+jets background, which shares identical final states but differs in colour flow and jet kinematics. To ensure a balanced representation, the two samples are rescaled such that the signal and background contributions amount to 50% each in the training. This choice allows the model to focus purely on the kinematic differences between the two processes, rather than on their relative production rates. Each event is further weighted by the corresponding dataset luminosity, ensuring that all data-taking years (2016–2018) are proportionally represented in the overall training sample and that the classifier output remains unbiased with respect to year-dependent effects. Additional metadata, such as the data-taking year or lepton flavour, were also tested as input features, but no improvement in discrimination power was observed.

To prevent statistical correlations between training and evaluation, two identical models are trained independently using disjoint subsets of events: one trained on even-numbered events and applied to odd-numbered ones, and vice versa. This "even/odd" scheme ensures a perfect phase-space match while preserving orthogonality between training and application. In what follows, this two-model system is referred to as a single DNN.

3.6.2 Architecture

The model is implemented using TensorFlow and the Keras API. The chosen architecture consists of three fully connected hidden layers with 64 neurons each. Each layer is followed by a dropout of 0.3 and batch normalisation to stabilize training and mitigate overfitting. The ReLU activation function is

used throughout, except for the final output neuron, which employs a sigmoid activation to represent a binary probability between signal and background:

$$f_{\text{DNN}} : \mathbb{R}^{N_{\text{features}}} \to [0, 1].$$

The network is optimized using the Adam algorithm with a learning rate of 10^{-3}, and trained to minimize the binary cross-entropy loss:

$$\mathcal{L} = -\sum_i w_i \Big[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \Big],$$

where w_i are the event weights. Early stopping is employed to prevent overfitting and unnecessary training once the model has converged. At each epoch, the validation loss is monitored, and training stops if no significant improvement is observed for 30 consecutive epochs. A minimum improvement threshold of $\Delta\text{loss} = 0.002$ is required for the validation loss to be considered better than in the previous best epoch. If this condition is not met for the specified patience period, the training is terminated and the model weights from the best-performing epoch are restored. This strategy ensures efficient convergence while avoiding fluctuations due to statistical noise in the validation sample.

During training, two control metrics are monitored: the binary cross-entropy loss and the classification accuracy. While the loss quantifies the agreement between predicted and true class probabilities, the accuracy measures the fraction of correctly classified events, using a threshold of 0.5 on the DNN output to distinguish signal-like from background-like events:

$$\text{Accuracy} = \frac{1}{N} \sum_i \mathbb{1} \Big[(\hat{y}_i > 0.5) = y_i \Big],$$

where $\hat{y}_i$ is the DNN output for event i, y_i is the true label (1 for signal, 0 for background), and $\mathbb{1}(\cdot)$ is the indicator function that equals 1 when its argument is true and 0 otherwise. Both quantities are tracked epoch-by-epoch on the training and validation samples to monitor the learning behaviour, while the test sample is used only after training to provide an independent performance estimate.

3.6.3 Input variable selection

An initial pool of 26 physics-motivated observables was considered, covering all relevant aspects of the event topology: the VBF dijet system, the reconstructed W boson, and the leptonic final state. Correlation matrices were evaluated separately for signal and background samples, and the difference matrix was used to identify variables that most effectively distinguish the two processes. From this preselection, the SHapley Additive exPlanations (SHAP) algorithm [129] was employed to quantify each variable's contribution to the DNN decision function. Variables found to be highly correlated or with negligible SHAP importance were discarded, resulting in the final set of eight inputs listed in

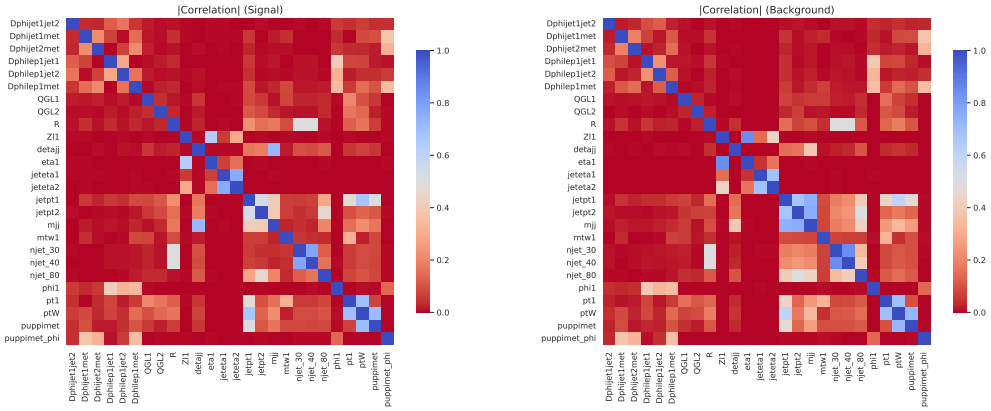


Figure 3.19: Correlation matrices for the initial set of 26 input variables considered in the DNN training, shown separately for the EW signal (left) and the mixed QCD W+jets background (right). Strong positive correlations are indicated by blue colours. Only the absolute value of the correlation is shown here.

Variable	Description
Z_l	$\eta_l - \frac{1}{2}(\eta_{j_1} + \eta_{j_2})$
$\lvert\Delta\eta_{jj}\rvert$	Pseudorapidity separation between the two tagging jets
$p_T(j_1)$	Transverse momentum of the leading jet
$p_T(W)$	Transverse momentum of the lepton + p_T^{miss} system
R	$\frac{\lvert\vec{p_T}(jj)+\vec{p_T}(W)\rvert}{\lvert\vec{p_T}(j_1)\rvert+\lvert\vec{p_T}(j_2)\rvert+\lvert\vec{p_T}(W)\rvert}$
m_{jj}	Invariant mass of the two tagging jets
η_{j_1}	Pseudorapidity of the leading jet
$p_T(l)$	Transverse momentum of the lepton

Table 3.12: Final input variables used for DNN training.

Table 3.12. Normalised signal and background distributions for all input variables, illustrating their discriminating power and motivation for inclusion in the network, are provided in Appendix C.

The SHAP summary plot in Figure 3.20 confirms that m_{jj}, $|\Delta\eta_{jj}|$, and Z_l dominate the model's discrimination power, while the remaining variables contribute complementary event-level information.

3.6.4 Training performance

The model is trained on a dataset of approximately 3 million (unweighted) events, with 1.3 million events reserved for validation and testing. Each set is enriched with 15% signal events. The evolution of the training and validation loss and accuracy is shown in Figure 3.21, demonstrating stable convergence and no sign of overtraining.

The resulting DNN output distribution is shown in Figure 3.22. Signal events

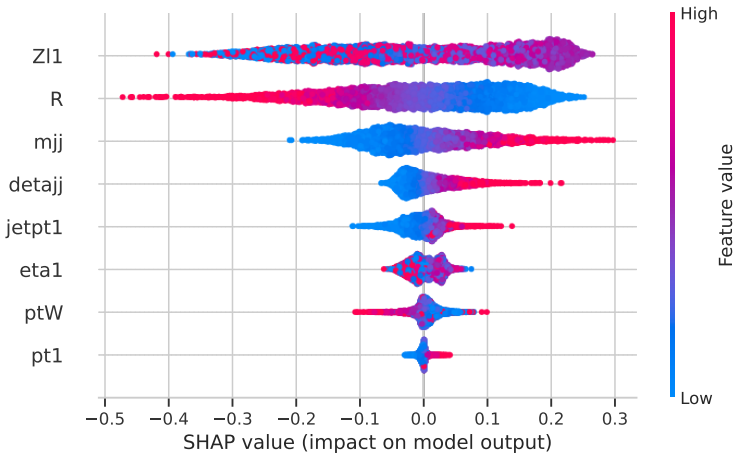


Figure 3.20: SHAP values for the input variables. Variables are ordered by their average contribution to the DNN output (from most to least important). Each point represents a single event: the horizontal position indicates the SHAP value, i.e. the strength and direction of the variable's contribution to the final DNN score, while the colour encodes the actual value of the variable (low: blue, high: red). For instance, high values of m_{jj} (red points) tend to produce positive SHAP values, pushing the DNN output toward the signal-like region, as expected from the physics of the process.

populate the high-score region, while QCD W+jets backgrounds are concentrated at low DNN values.

3.6.5 Performance comparison

The discriminating power of the DNN is quantified using several complementary metrics and compared with those obtained from single kinematic observables, such as m_{jj} and Z_l, as well as from their two-dimensional combination (m_{jj}, Z_l). As illustrated in Figure 3.23, the DNN achieves the best overall performance, significantly outperforming both individual and combined observables, particularly in the high-purity region dominated by the EW signal.

Receiver Operating Characteristic. The Receiver Operating Characteristic (ROC) curve is constructed by varying the threshold on the DNN output and, for each threshold value, computing the signal efficiency

$$\epsilon_{\text{sig}} = \frac{N_{\text{sig}}^{\text{passed}}}{N_{\text{sig}}^{\text{total}}},$$

and the background efficiency

$$\epsilon_{\text{bkg}} = \frac{N_{\text{bkg}}^{\text{passed}}}{N_{\text{bkg}}^{\text{total}}}.$$

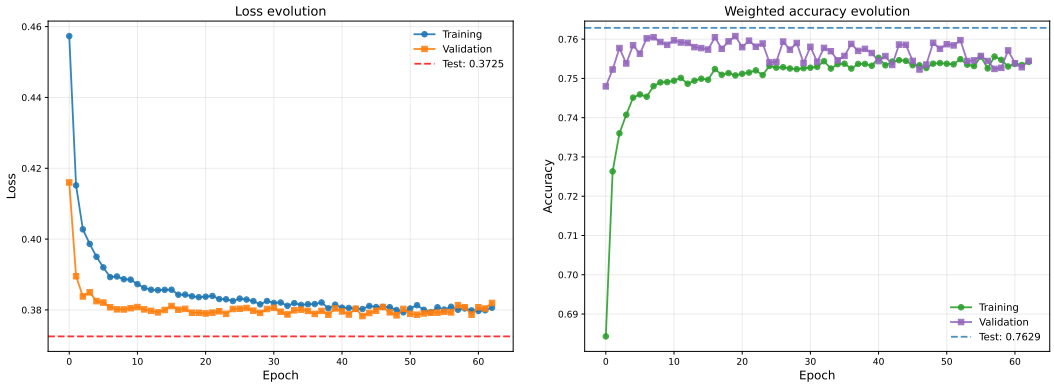


Figure 3.21: Training evolution of the binary cross-entropy loss (left) and accuracy (right). On the left, the training loss (blue solid) and validation loss (orange solid) converge smoothly to similar values, while the test loss (orange dashed) shows a slightly better performance, consistent with the absence of dropout during inference. On the right, the accuracy for the training (green solid), validation (blue solid), and test (blue dashed) samples exhibits a similar trend. The close agreement between training and validation curves indicates that the model generalises well without signs of overfitting. The marginally higher performance observed in the test set is consistent with the expected behavior when dropout layers are inactive during inference (test).

The ROC curve therefore represents ϵ_{sig} as a function of ϵ_{bkg}, with an ideal classifier lying in the top-left corner of the plane (high ϵ_{sig}, low ϵ_{bkg}). The global separation power is quantified by the area under the curve (AUC), while particular attention is given to the region $\epsilon_{\text{bkg}} < 0.04$, corresponding to a background contamination below 4%, which is most relevant for the EW signal extraction.

Purity and significance. In addition to the ROC, more physically motivated quantities are evaluated as a function of the selection threshold on the discriminant.

The signal purity is defined as

$$P = \frac{N_{\text{sig}}^{\text{passed}}}{N_{\text{sig}}^{\text{passed}} + N_{\text{bkg}}^{\text{passed}}},$$

and represents the fraction of signal events in the selected sample after applying a cut on the discriminant. The purity is shown as a function of the signal efficiency ϵ_{sig} in Figure 3.24. A higher curve indicates a better separation between signal and background, corresponding to a cleaner sample for a given efficiency.

To ensure a consistent and model-independent comparison, all discriminants are first normalised and flattened into signal quintiles, such that the cumulative distribution of signal events is uniform in the transformed variable. This monotonic mapping, based on the empirical cumulative distribution function (CDF), preserves the discriminating power while aligning the scale of different observables. In this way, purity and significance curves reflect only the

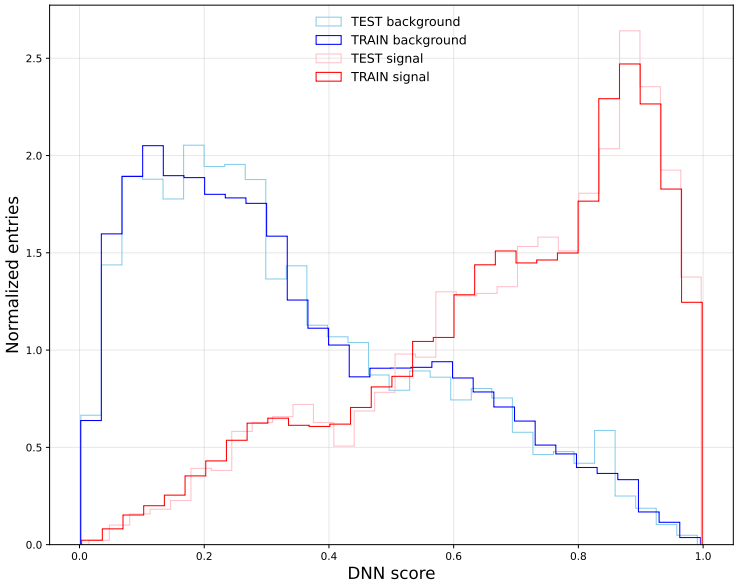


Figure 3.22: DNN output distributions for signal (red) and background (blue) in the training (dark) and test (light) samples. The close agreement between the two indicates no overtraining, while the clear separation between classes shows the discriminative power of the DNN.

intrinsic separation capability of each discriminant, not their original kinematic distributions.

The expected significance is also evaluated as a function of the selection threshold on the discriminant. It is estimated under the assumption of large background counts as

$$Z = \frac{N_{\text{sig}}^{\text{passed}}}{\sqrt{N_{\text{sig}}^{\text{passed}} + N_{\text{bkg}}^{\text{passed}}}},$$

and provides an approximate measure of the statistical sensitivity of a simple cut-based analysis using the given discriminant. The resulting significance curves are shown in Figure 3.25. The DNN yields the highest expected significance across most of the threshold range, confirming its superior separation power compared to m_{jj}, Z_l, and their combination.

In particular, the DNN significance curve exhibits a broad maximum around a threshold value of DNN $\simeq 0.5$, corresponding to an optimal balance between signal efficiency and background rejection. This working point, which nominally maximises the $S/\sqrt{S+B}$ ratio, is therefore adopted as the reference threshold for the discriminant in the subsequent statistical fits.

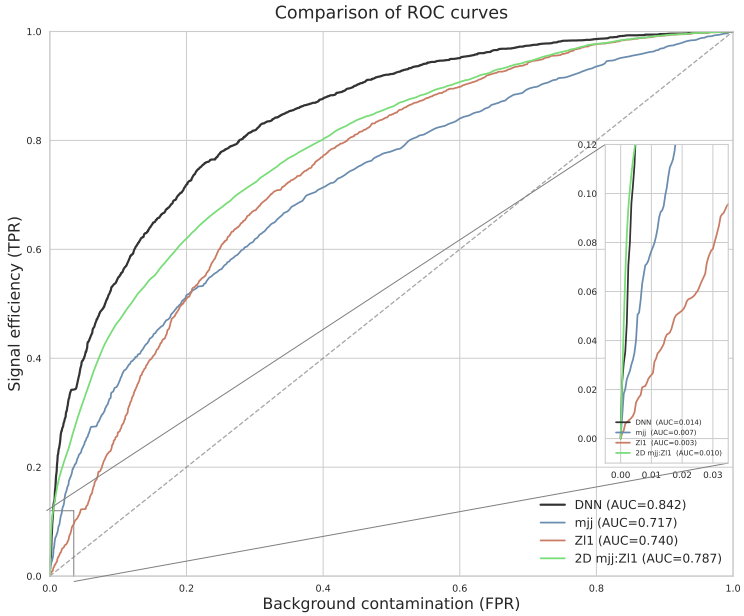


Figure 3.23: Comparison of ROC curves obtained using m_{jj} (blue), the DNN output (black), Z_l (red), and the two-dimensional (m_{jj}, Z_l) combination (green). The inset highlights the high-purity region, corresponding to a background contamination below 3.5%. The DNN provides the best overall discrimination across the full range, while its performance becomes comparable to that of the two-dimensional (m_{jj}, Z_l) observable in the extreme low-background regime.

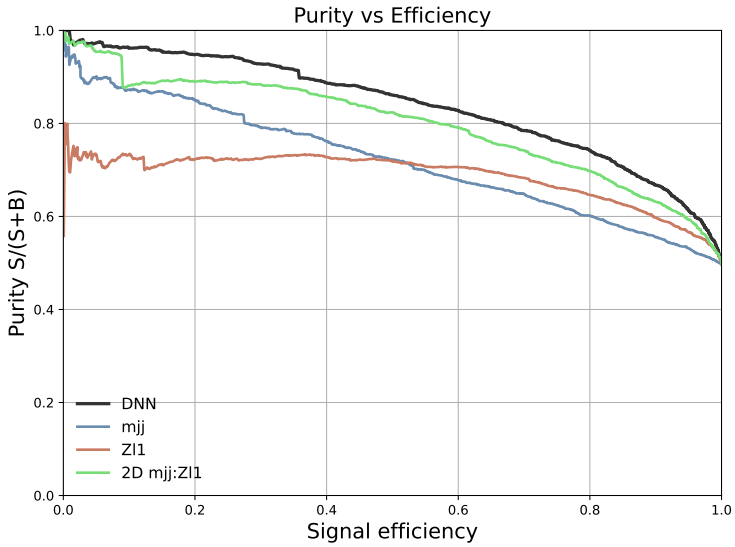


Figure 3.24: Signal purity as a function of signal efficiency for the m_{jj} (blue), DNN output (black), Z_l (red), and (m_{jj}, Z_l) (green) discriminants. All discriminants have been normalized and flattened into signal quintiles to ensure a uniform comparison. The DNN achieves higher purity at any given signal efficiency, reflecting its superior background rejection capability.

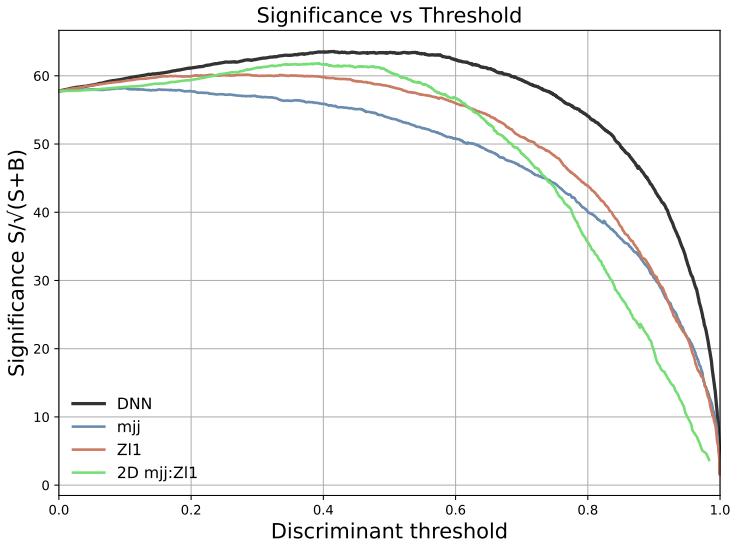


Figure 3.25: Expected significance Z as a function of the selection threshold on each discriminant: m_{jj} (blue), DNN output (black), Z_l (red), and (m_{jj}, Z_l) (green). All discriminants have been normalized and flattened into signal quintiles, ensuring that the comparison reflects only their discriminating power. The DNN provides the highest expected sensitivity across almost the entire threshold range, confirming its optimal performance for signal extraction.

3.7 Systematic Uncertainties

This section details the systematic uncertainties affecting the measurement and their influence on the final results. Both normalization uncertainties (e.g., luminosity uncertainty) and shape uncertainties (e.g., jet energy scale uncertainty) are considered. Shape uncertainties impact not only the overall normalization but also the distribution of key kinematic variables. To account for these, template distributions are produced for the nominal value of each relevant parameter and for its corresponding $\pm 1\sigma$ variations, and the associated nuisance parameters are then profiled in the statistical fit. Systematic uncertainties that are correlated across different years are modelled as a single nuisance parameter in the fit, following the guidelines provided by the relevant POGs. See Tables 3.13 and 3.14 for the specific correlation scheme used for each systematic uncertainty.

The implementation of these uncertainties in the likelihood model follows the formalism described in Section 1.9, where each source of systematic uncertainty enters the fit as a dedicated nuisance parameter, constrained by a Gaussian or log-normal prior depending on its nature. The corresponding shape and normalisation variations are thus profiled during the statistical inference to properly propagate their impact to the measured signal strength.

3.7.1 Uncertainties Specific to Individual Background Processes

- **Non-prompt background:** The non-prompt The background is estimated by applying weights to a subset of data events where leptons satisfy the loose selection criteria, as described in Sec. 3.5.3. The dominant systematic uncertainty for this background originates from the limited size of the control samples used to measure the fake rates, leading to a statistical uncertainty of up to 20%. Although the statistical uncertainty of the fake-rate measurement is not particularly large on its own, its effect on the final yield can be significantly amplified. This happens because the event weight depends on the quantity $(p - f)$ in the denominator of Eq. 3.6: when the fake rate f is shifted by its statistical uncertainty, it can become close to the prompt rate p, causing the denominator to approach zero. This purely mathematical amplification results in a large variation of the predicted non-prompt yield, even if the underlying statistical uncertainty on f is moderate. Additional uncertainties stem from the scale factor (SF) used to subtract residual prompt contamination. Indeed, the nominal value measured in the DY-enriched region is applied to all the MC residual contributions, except the top, which has its own SF measured. Separate measurements were performed for the loose and tight selections with a $\pm 10\%$ shift in the nominal SF value, generating two distinct (up and down) estimations for the fake rate, which are used

as systematic uncertainties. Finally, an overall systematic uncertainty of 30% is assigned to account for residual effects.

- **Top background:** The full weight of top quark p_T reweighting is considered as a systematic uncertainty in the top quark background cross section. The $t\bar{t}$ and single top quark production processes are treated as a single background. The kinematic shapes are modelled using simulations, with normalization set to theoretical predictions. The overall normalization of the top quark background is left floating in the fit independently for each year and category, controlled via the dedicated top control regions.

- **W+Jets background:** A data-driven approach for correcting the W+Jets background is detailed in 3.5.2. The main sources of systematic uncertainty in this correction stem from the statistical uncertainty of the samples used to derive the reweighting function, as well as uncertainty in the analytical form of the function itself.

3.7.2 Uncertainties affecting all the simulated processes

This category includes uncertainties in the MC modelling of the physical processes. Unless otherwise stated, these uncertainties are treated as uncorrelated across the three years while being correlated across the processes.

- **Luminosity:** All signal and background process yields are allowed to vary within the uncertainty in the total integrated luminosity, as determined by the CMS luminosity monitoring. The total luminosity uncertainty for a given year is composed of two parts: a component that is uncorrelated across data-taking years (dominated by year-specific calibration and stability effects), and a component that is correlated across years, reflecting the common systematic uncertainties in the luminosity scale determination. For the uncorrelated component of the 2016 (pre- and post-VFP) [130], 2017 [131], and 2018 [132] datasets, the uncertainties are 1.0%, 2.0%, and 1.5%, respectively. The correlation model between the three data-taking years follows the official prescription provided by the CMS LUM POG [133].

- **Trigger Scale Factor:** Data-to-simulation scale factors are applied to MC samples as a function of the lepton transverse momentum p_T and pseudorapidity η. These corrections are derived using the standard CMS tag-and-probe method, as documented in Ref. [104]. The associated uncertainties are estimated by varying the tag selection criteria and the Z boson mass window in the tag-and-probe fit, and by adding in quadrature the resulting variations with the statistical uncertainty. These uncertainties are treated as uncorrelated across the different data-taking years.

- **L1 Prefiring:** During the 2016 and 2017 data-taking periods, a fraction of events was lost due to a L1 trigger "prefiring" effect, in which

the trigger occasionally fired one bunch crossing earlier than the correct one [121]. Due to the design of the CMS data acquisition (DAQ) system, consecutive L1 triggers cannot be accepted; as a result, the genuine events corresponding to the subsequent bunch crossing are discarded whenever a prefire occurs. This effect, originating mainly from the ECAL trigger logic in the forward region ($2.0 < |\eta| < 3.0$), led to an efficiency loss that depends on the transverse energy and pseudorapidity of reconstructed photons and jets. A smaller contribution also arises from the muon L1 trigger, particularly in the barrel–endcap transition region, although its overall impact is subdominant. To correct for this effect, per-event scale factors ("prefiring weights") are applied to all simulated samples. These weights, provided by the CMS L1 POG, account for the probability that an event is affected by prefiring. The corresponding systematic uncertainty is evaluated by varying these weights within their uncertainties and propagating the effect to all observables. This uncertainty is considered independently for each year and treated as uncorrelated among different data-taking periods.

- **Lepton Reconstruction and Identification Efficiency:** The reconstruction and identification efficiencies for electrons and muons are measured using the standard CMS tag-and-probe technique applied to $Z \to \ell^+\ell^-$ events, as documented in Ref. [104]. To correct for residual differences between data and simulation, data-to-simulation scale factors are applied to all simulated samples as functions of the lepton transverse momentum p_T and pseudorapidity η. The corresponding uncertainties are estimated from variations of the tag selection and fit model used in the tag-and-probe method, and amount to approximately 1–2% for both electrons and muons, depending on the data-taking year. These uncertainties are treated as uncorrelated among different years.

- **Muon and Electron Momentum Scale:** The momentum scale of reconstructed muons and electrons is corrected using calibration constants and scale factors provided by the CMS MUO and EGM POGs. These corrections account for detector alignment, magnetic field, and calibration effects that can lead to small biases in the measured lepton momentum. Residual systematic uncertainties are evaluated by applying $\pm 1\sigma$ variations to the lepton momentum scale, as functions of the lepton transverse momentum p_T and pseudorapidity η, and propagating the resulting changes to all analysis observables. The uncertainties are treated as independent among different data-taking years.

- **Jet Pileup Identification (PU ID) Scale Factor:** An uncertainty is assigned to the scale factors correcting for the efficiency of the jet pileup identification algorithm, which distinguishes genuine jets from those originating from additional proton–proton interactions. This uncertainty affects jets with $30 < p_T < 50$ GeV and is provided centrally by the CMS JME POG [134]. It is treated as uncorrelated among the different data-taking years.

- **Jet Energy Scale (JES):** The jet energy scale uncertainty accounts for imperfect knowledge of the calibration applied to reconstructed jets. The uncertainty is evaluated by varying the jet energy by $\pm 1\sigma$, corresponding to a rescaling of the jet four-momentum according to $P^\mu \rightarrow P^\mu \cdot \left(1 + \frac{\delta^{\text{JES}}_{p_T}}{p_T}\right)$, where $\delta^{\text{JES}}_{p_T}$ represents the jet energy scale uncertainty as a function of the jet transverse momentum p_T and pseudorapidity η. Eleven independent sources of JES uncertainty are considered, following the recommendations of the JME POG: Absolute, AbsoluteY*, BBEC1, BBEC1Y*, EC2, EC2Y*, FlavorQCD, HF, HFY*, RelativeBal, and RelativeSampleY*, where Y* corresponds to 2016, 2017, or 2018. These components are implemented as shape variations; six of them are correlated across the three years, while the remaining ones are independent [135–140].

- **Jet Energy Resolution (JER):** The jet energy resolution uncertainty reflects the limited precision in modelling the detector response to jets. It is evaluated by varying the resolution by $\pm 1\sigma$ according to the recommendations of the Jet Energy Resolution and Corrections (JERC) group. The effect of this variation is propagated to all observables and treated as uncorrelated between data-taking years [135].

- **Missing Transverse Energy (MET):** The uncertainty in the MET originates from imperfect modelling of jets, leptons, and unclustered energy contributions. It is estimated by varying by $\pm 1\sigma$ the transverse momentum of each particle-flow candidate according to its resolution, and recomputing the event-level MET. This uncertainty is implemented as a shape variation and is treated as uncorrelated across years.

- ***b*-tagging Scale Factor:** Uncertainties on the b-tagging efficiency scale factors are derived by shifting the per-jet weights within the ranges recommended by the BTV POG [126, 141, 142]. Nine independent variations are considered: four (hfstats1, hfstats2, lfstats1, lfstats2) account for the statistical uncertainties in the calibration samples and are uncorrelated among years, while five (jes, lf, hf, cferr1, cferr2) represent uncertainties related to the flavour composition and are treated as correlated across years.

- **Pileup Reweighting:** The uncertainty associated with pileup reweighting is estimated by varying the minimum-bias cross section used to model the pileup distribution in simulation by $\pm 1\sigma$. Simulated events are then reweighted according to these alternative pileup profiles. This uncertainty is treated as independent across data-taking years.

3.7.3 Theoretical Uncertainties

All nuisance parameters associated with theoretical modelling uncertainties are treated as correlated across the three data-taking years, unless otherwise spec-

ified.

- **Parton Shower Modelling:** Uncertainties in PS modelling primarily affect the jet multiplicity and the overall event kinematics. They are evaluated by reweighting simulated events with dedicated variations of the initial-state radiation (ISR) and final-state radiation (FSR) scales in the Pythia 8 parton shower. For each of the two components, up and down variations are applied corresponding to scale factors of two and one half, respectively, and are treated as independent shape uncertainties. For processes simulated with the dipole recoil scheme (e.g. VBS and VBF samples), a dedicated uncertainty is assigned to account for alternative recoil treatments. In the case of the EW Wjj signal, this uncertainty is separated into two components: one affecting the overall production rate and one affecting the acceptance of events passing the analysis selections. Since the signal strength parameter already accounts for variations in the total production cross section, the corresponding normalization effect is not included as a separate uncertainty. Only the acceptance component, which alters the fraction of selected events within the fiducial phase space, is evaluated explicitly through event reweighting. This uncertainty is correlated for all the processes except the signal.

- **Renormalization and Factorization Scales (QCD Scales):** Missing higher-order corrections in the perturbative QCD expansion are accounted for by varying the renormalization (μ_R) and factorization (μ_F) scales up and down by factors of two and one half in the simulation. The uncertainty is estimated from six scale variations: $(0.5, 0.5)$, $(0.5, 1)$, $(1, 0.5)$, $(1, 2)$, $(2, 1)$, and $(2, 2)$, excluding the nominal configuration $(1, 1)$ and the opposite correlated cases $(0.5, 2)$ and $(2, 0.5)$. The envelope of the remaining variations defines a one-standard-deviation uncertainty. For the EW Wjj signal process, the uncertainty is separated into two components: one affecting the total cross section and one affecting the event acceptance. Since variations in the overall production rate are already accounted for by the floating signal strength parameter in the fit, only the acceptance component, affecting the fraction of events passing the analysis selection, is explicitly evaluated through event reweighting. The uncertainty is uncorrelated among different processes.

- **Parton Distribution Functions (PDFs):** The choice of parton distribution functions (PDFs) affects both the normalization and the shape of simulated kinematic distributions. The corresponding uncertainty is evaluated by reweighting simulated events using the alternative PDF eigenvectors recommended by the PDF4LHC working group [143]. The total PDF uncertainty is obtained by adding in quadrature the variations across all eigenvectors. An additional contribution associated with the value of the strong coupling constant, α_S, is evaluated independently through dedicated up and down reweights. This uncertainty is applied to the EW Wjj signal and to the dominant simulated background processes, namely the

mixed-QCD W+jets and top quark backgrounds. For the EW Wjj signal process, the PDF uncertainty is further separated into two components: one affecting the total production cross section and one affecting the acceptance of the selected events. Since variations in the total production rate are already absorbed by the signal strength parameter in the statistical model, only the acceptance effect is explicitly evaluated through event reweighting. This uncertainty is uncorrelated among processes.

- **Underlying Event (UE) Modelling:** The uncertainty associated with the underlying event (UE) modelling is evaluated using dedicated Pythia 8 samples with varied CP5 tune parameters. These variations probe the impact of the UE description on kinematic observables and event yields. A conservative log-normal (LnN) uncertainty of 1.5% is assigned to all simulated processes to cover residual effects. This uncertainty is not applied to samples whose normalisation or cross section is directly constrained or measured from data, such as the main background processes.

Nuisance Name	Processes Affected	Years Affected				Type	Years Correlated				Samples Correlated
		HIPM	noHIPM	2017	2018		HIPM	noHIPM	2017	2018	
lumi_13TeV	All MC processes	✓	✓	✓	✓	lnN	✓	✓			All MC processes
lumi_13TeV_correlated	All MC processes	✓	✓	✓	✓	lnN	✓	✓	✓	✓	All MC processes
lumi_13TeV_1718	All MC processes			✓	✓	lnN			✓	✓	All MC processes
CMS_fake_syst_ele	Non-prompt	✓	✓	✓	✓	lnN	✓	✓	✓	✓	Non-prompt
CMS_fake_syst_mu	Non-prompt	✓	✓	✓	✓	lnN	✓	✓	✓	✓	Non-prompt
CMS_fake_stat_e	Non-prompt	✓	✓	✓	✓	Shape	✓	✓	✓	✓	Non-prompt
CMS_fake_stat_m	Non-prompt	✓	✓	✓	✓	Shape	✓	✓	✓	✓	Non-prompt
VBFW_fake_e_EWKsub	Non-prompt	✓	✓	✓	✓	Shape					Non-prompt
VBFW_fake_m_EWKsub	Non-prompt	✓	✓	✓	✓	Shape					Non-prompt
CMS_eff_prefiring	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_btag_jes	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_lf	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_hf	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_hfstats1	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_hfstats2	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_lfstats1	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_lfstats2	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_cferr1	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_btag_cferr2	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_eff_trigger_eff	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_eff_e	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_eff_m	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_scale_e	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_scale_m	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_jetpuid	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_scale_JESAbsolute	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_scale_JESAbsolute_Year*	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_scale_JESBBEC1	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes

Table 3.13: Recap table of systematic uncertainty considered in the analysis. In the table also the processes they affect, the years they affect, how the systematic uncertainty is treated in the fit and the correlation scheme used. ***Year** = 2016HIPM, 2016noHIPM, 2017, 2018

Nuisance Name	Processes Affected	Years Affected				Type	Years Correlated				Samples Correlated
		HIPM	noHIPM	2017	2018		HIPM	noHIPM	2017	2018	
CMS_scale_JESBBEC1_Year	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_scale_JESEC2	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_scale_JESEC2_Year*	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_scale_JESFlavorQCD	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_scale_JESHF	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_scale_JESHF_Year*	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_scale_ESRelativeBal	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_scale_JESRelativeSample_Year*	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
CMS_res_j	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
CMS_scale_met	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All MC processes
singleTopToTTbar	Top	✓	✓	✓	✓	Shape	✓	✓	✓	✓	-
CMS_top_pT_reweighting	Top	✓	✓	✓	✓	Shape	✓	✓	✓	✓	-
CMS_PS_ISR	All but VBS, VBF processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All but VBS, VBF
CMS_PS_FSR	All but VBS, VBF processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	All but VBS, VBF
CMS_PS_ISR_dipole	VBS, VBF processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	VBS, VBF processes
CMS_PS_FSR_dipole	VBS, VBF processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	VBS, VBF processes
CMS_PU	All MC processes	✓	✓	✓	✓	Shape	✓	✓			All MC processes
QCDscale	All MC processes	✓	✓	✓	✓	Shape	✓	✓	✓	✓	-
QCDscale_WLNuJJ_accept	WLNuJJ	✓	✓	✓	✓	Shape	✓	✓	✓	✓	-
pdf_weight	WJets_HT, top	✓	✓	✓	✓	Shape	✓	✓	✓	✓	-
pdf_weight_WLNuJJ_accept	WLNuJJ	✓	✓	✓	✓	Shape	✓	✓	✓	✓	-
WJets_reweight_stat_Year*	WJets_hardJets	✓	✓	✓	✓	Shape					-
WJets_reweight_functional_Year*	WJets_hardJets	✓	✓	✓	✓	Shape					-
Top_norm_ele_Year*	Top	✓	✓	✓	✓	rateParameter					-
Top_norm_mu_Year*	Top	✓	✓	✓	✓	rateParameter					-
WPUJet_norm_ele_Year*	WJets_PUJets	✓	✓	✓	✓	rateParameter					-
WPUJet_norm_mu_Year*	WJets_PUJets	✓	✓	✓	✓	rateParameter					-
WhardJet_norm_ele_Year*	WJets_hardJets	✓	✓	✓	✓	rateParameter					-
WhardJet_norm_mu_Year*	WJets_hardJets	✓	✓	✓	✓	rateParameter					-

Table 3.14: Recap table of systematic uncertainty considered in the analysis. In the table also the processes they affect, the years they affect, how the systematic uncertainty is treated in the fit and the correlation scheme used. ***Year** = 2016HIPM, 2016noHIPM, 2017, 2018

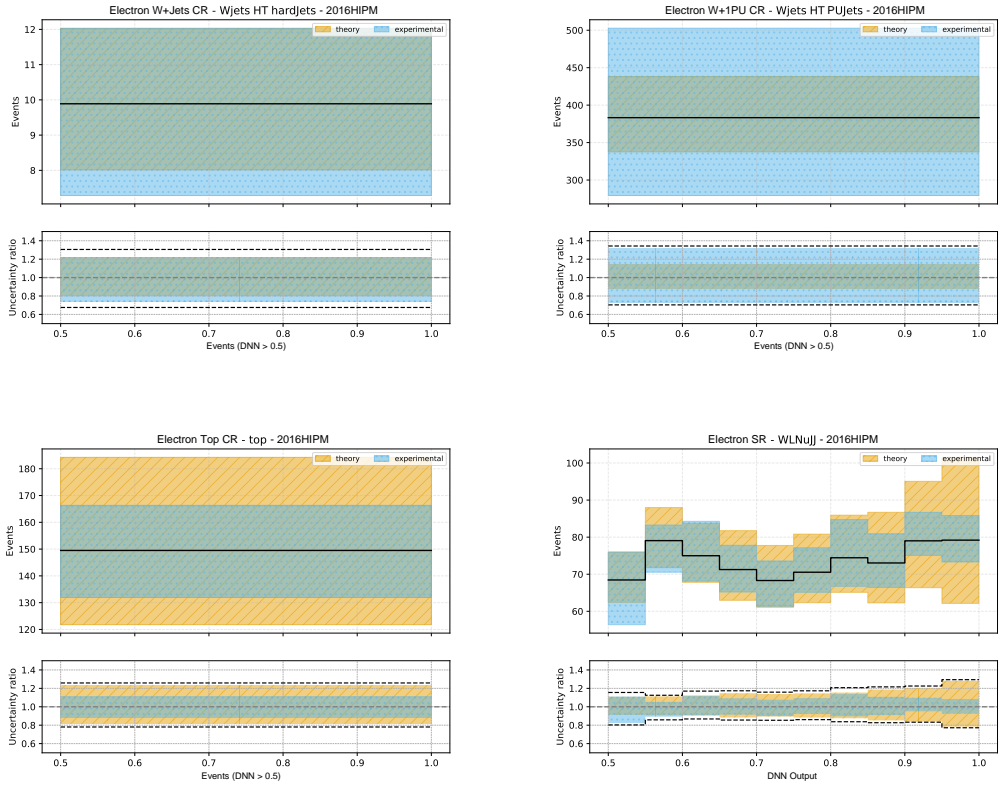


Figure 3.26: CR and SR inputs used to validate the modelling and associated uncertainties in the 2016 HIPM dataset electron final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

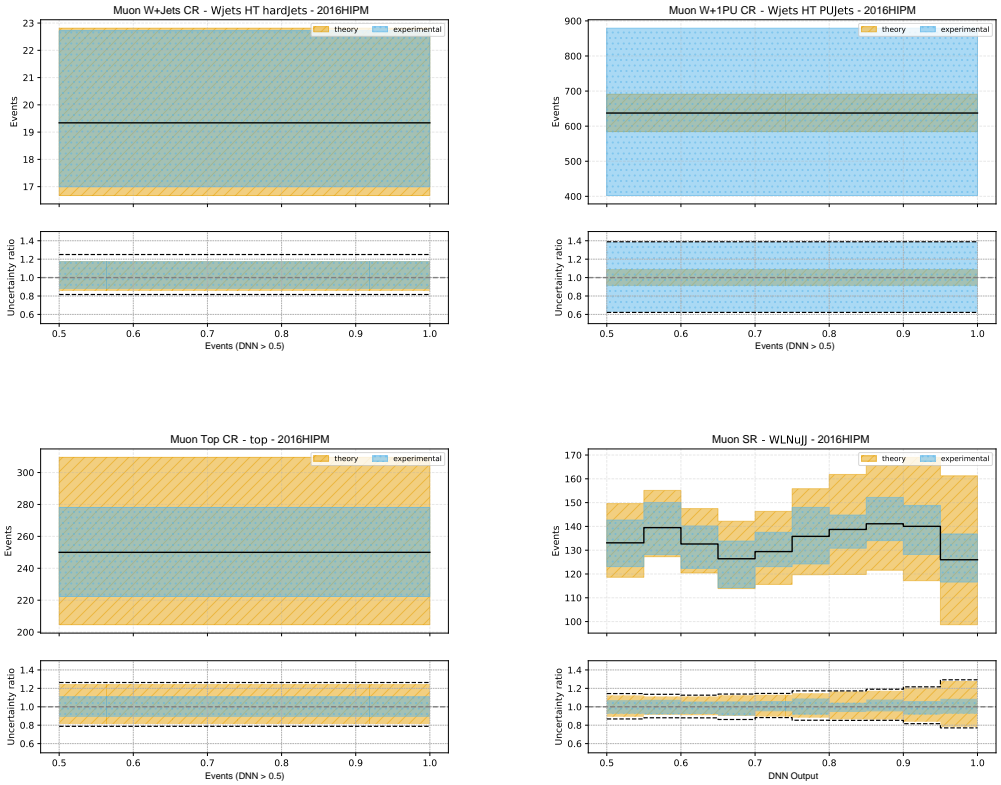


Figure 3.27: CR and SR inputs used to validate the modelling and associated uncertainties in the 2016 HIPM dataset muon final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

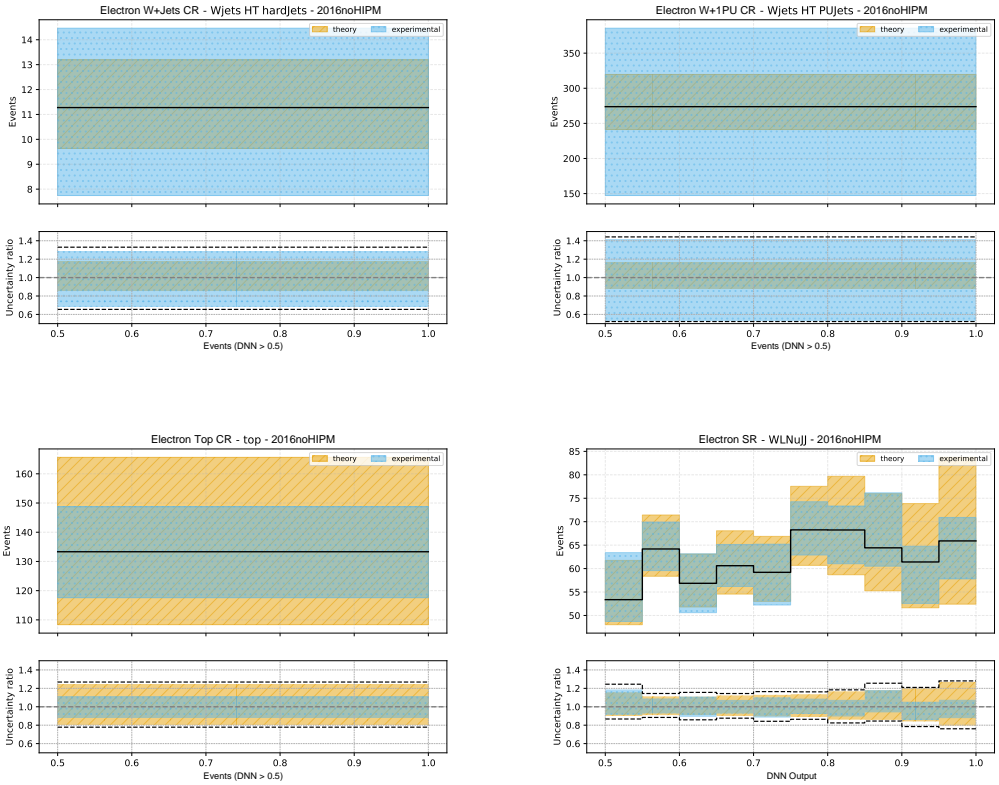


Figure 3.28: CR and SR inputs used to validate the modelling and associated uncertainties in the 2016 noHIPM dataset electron final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

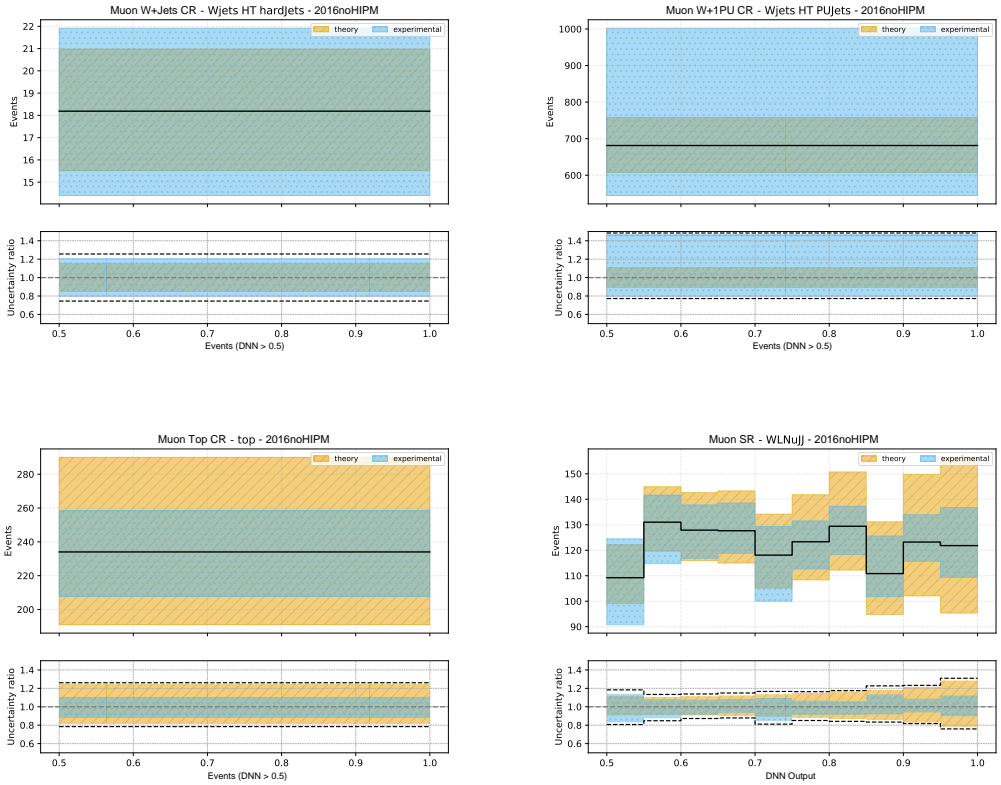


Figure 3.29: CR and SR inputs used to validate the modelling and associated uncertainties in the 2016 noHIPM dataset muon final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

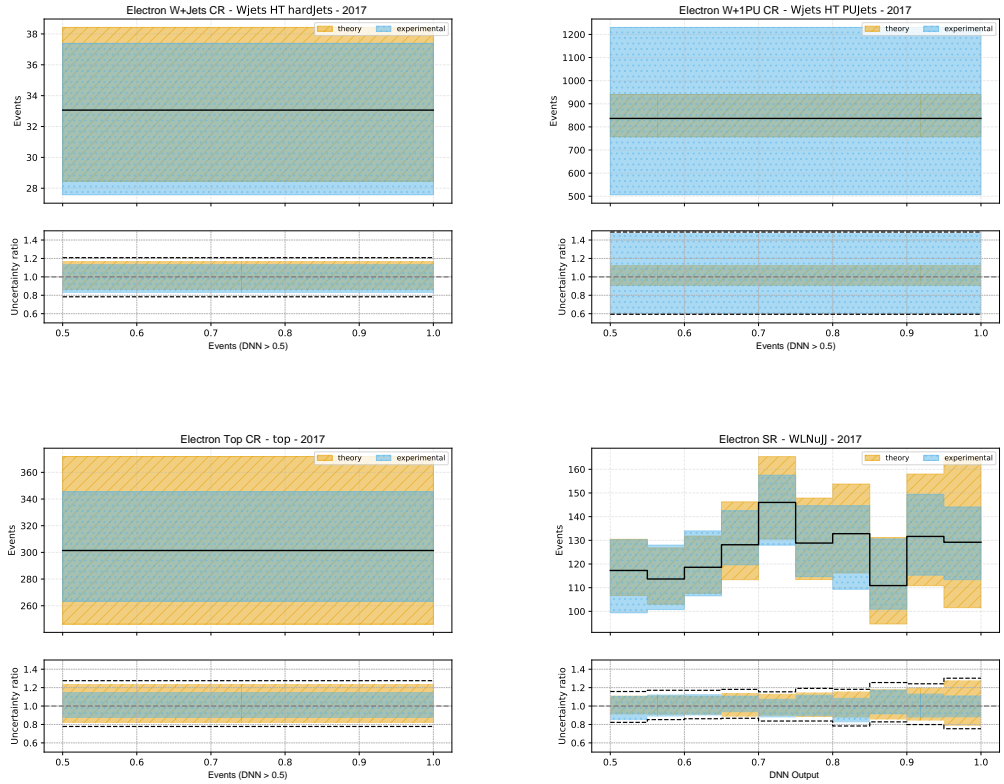


Figure 3.30: CR and SR inputs used to validate the modelling and associated uncertainties in the 2017 dataset electron final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

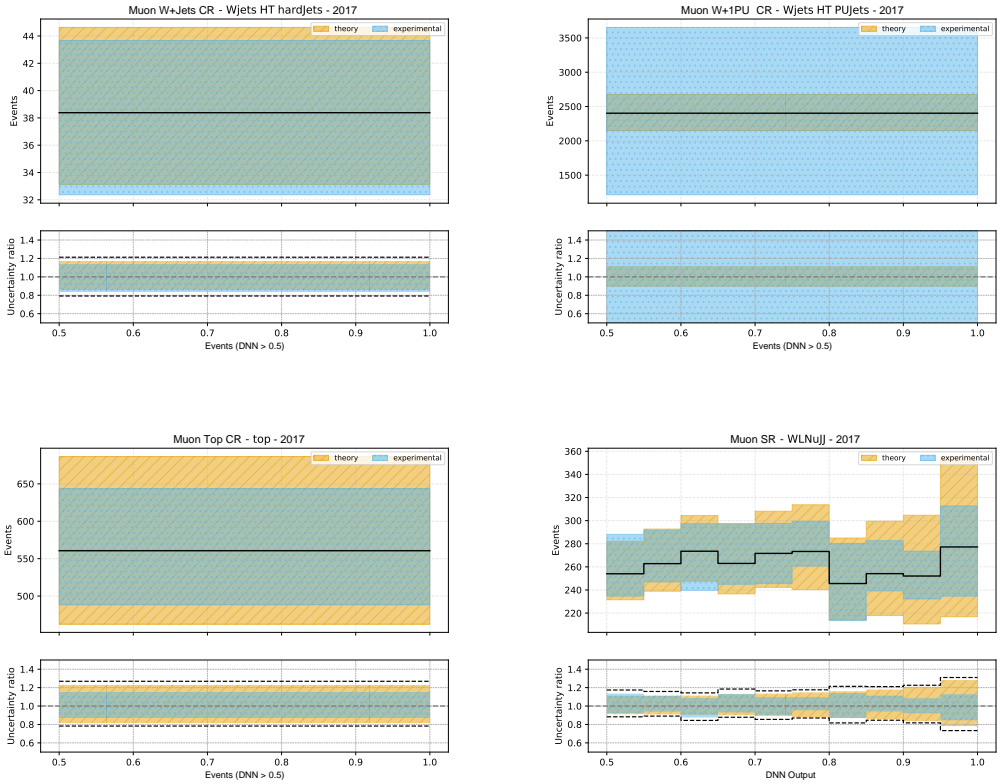


Figure 3.31: CR and SR inputs used to validate the modelling and associated uncertainties in the 2017 dataset muon final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

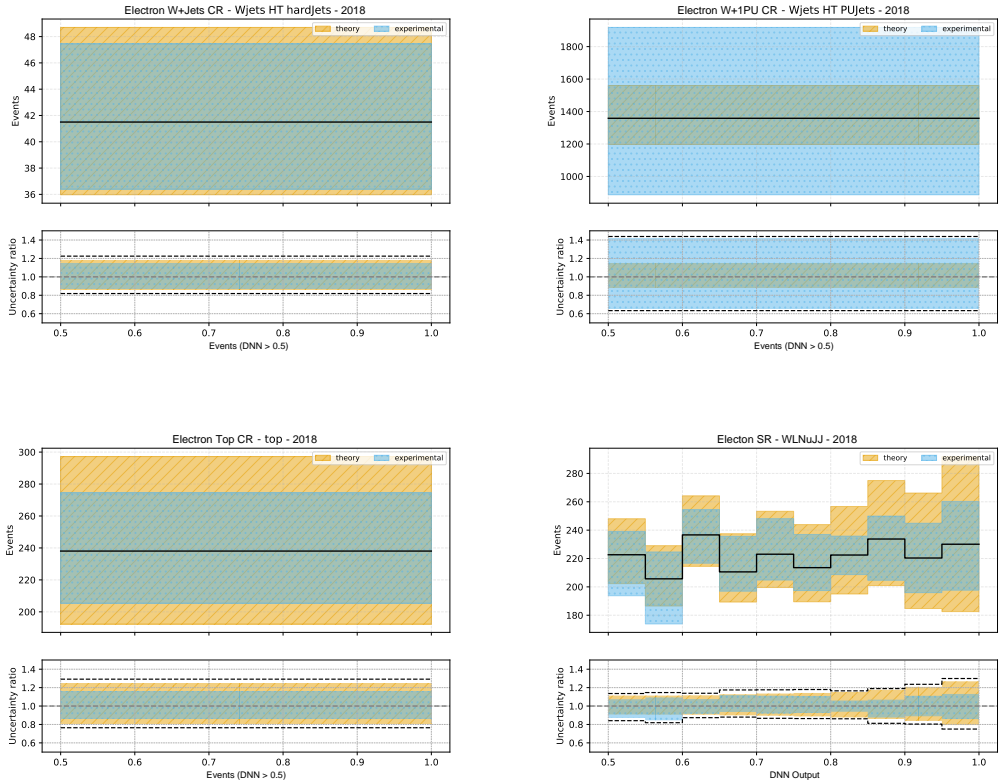


Figure 3.32: CR and SR inputs used to validate the modelling and associated uncertainties in the 2018 dataset electron final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

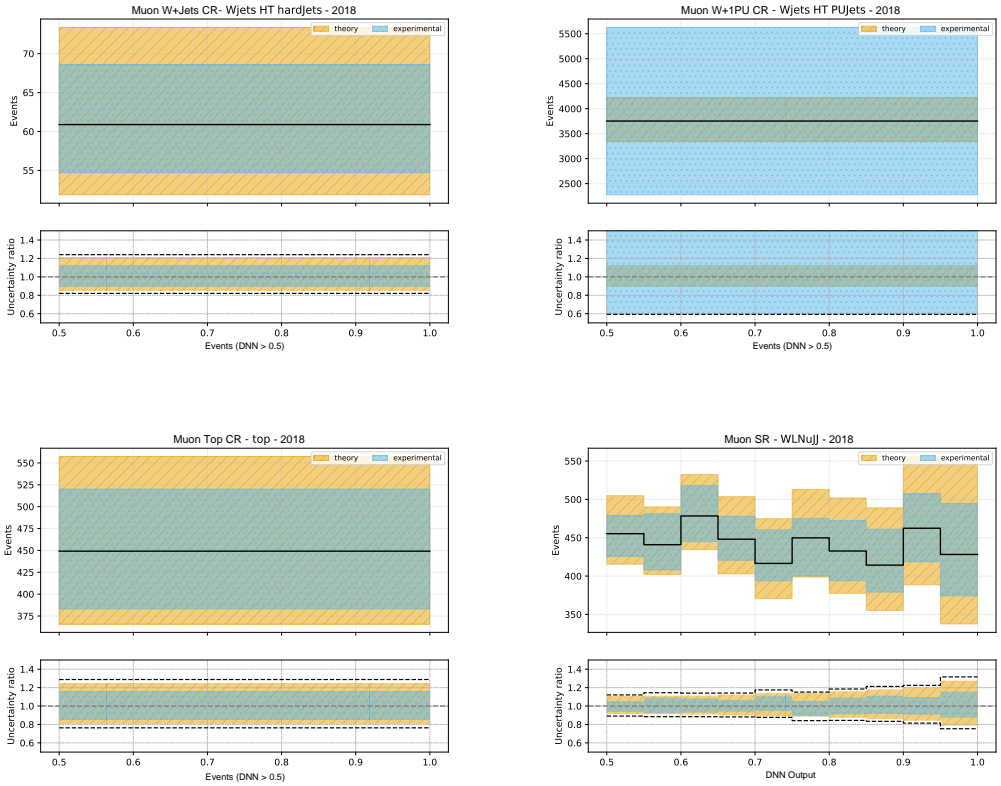


Figure 3.33: CR and SR inputs used to validate the modelling and associated uncertainties in the 2018 dataset muon final state. In the W+Jets CRs (top), a single-bin event count in the high-DNN-score region is shown, together with its uncertainty. Same for Top CR (left, bottom). In the signal region (right, bottom row), the high-score output of the DNN is displayed using 10 bins. A breakdown of the uncertainty contributions is also provided: in the W+jets control regions, the experimental uncertainties are dominated by the jet energy scale and resolution (JES/JER), while in all regions, the dominant theoretical uncertainty arises from the QCD renormalisation and factorisation scale variations. PDF and parton-shower uncertainties are comparatively much smaller.

3.8 Inclusive Results

In this section, the inclusive cross section for the EW production of a W boson in association with two jets is presented, combining the electron and muon final states. The measurement is obtained through a simultaneous binned profile maximum-likelihood fit, as described in Section 1.9, which makes use of the statistical framework implemented in COMBINE [144]. This procedure allows for the extraction of the signal strength and the evaluation of the pulls and constraints on the nuisance parameters encoding the various systematic uncertainties.

The fit is performed on the output of the DNN classifier in the signal regions, while the normalisations of the dominant background processes, namely W+jets and top quark production, are constrained using dedicated control regions that enter the likelihood simultaneously. This strategy ensures that background contributions are accurately modelled, leading to a reliable extraction of the signal yield. The same procedure is applied independently to each of the four data-taking years, and the individual results are then combined to exploit the full statistical power of the Run II dataset.

Figures from 3.34 to 3.37 show the agreement for the variables used to train the DNN in the non-prompt CR for the full Run II dataset. Similarly, figures from 3.38 to 3.41 show the agreement in the W+jets CR, figures from 3.42 to 3.45 show the agreement in the W+1PU jets CR, and finally figures 3.46 and 3.49 show the agreement in the top CR. From figure 3.50 to 3.53, a partial unblinded comparison is shown in the SR, restricted to bins with $\frac{S}{\sqrt{B}} < 5$.

In all ratio panels, the shaded band represents the total uncertainty, including both statistical and systematic components. In addition, the expected post-fit uncertainties associated with the nuisance parameters that control the normalisation of the major background processes (implemented as `rateParam` terms, see Sec. 1.9.5) are propagated to the ratio and included in the uncertainty band.

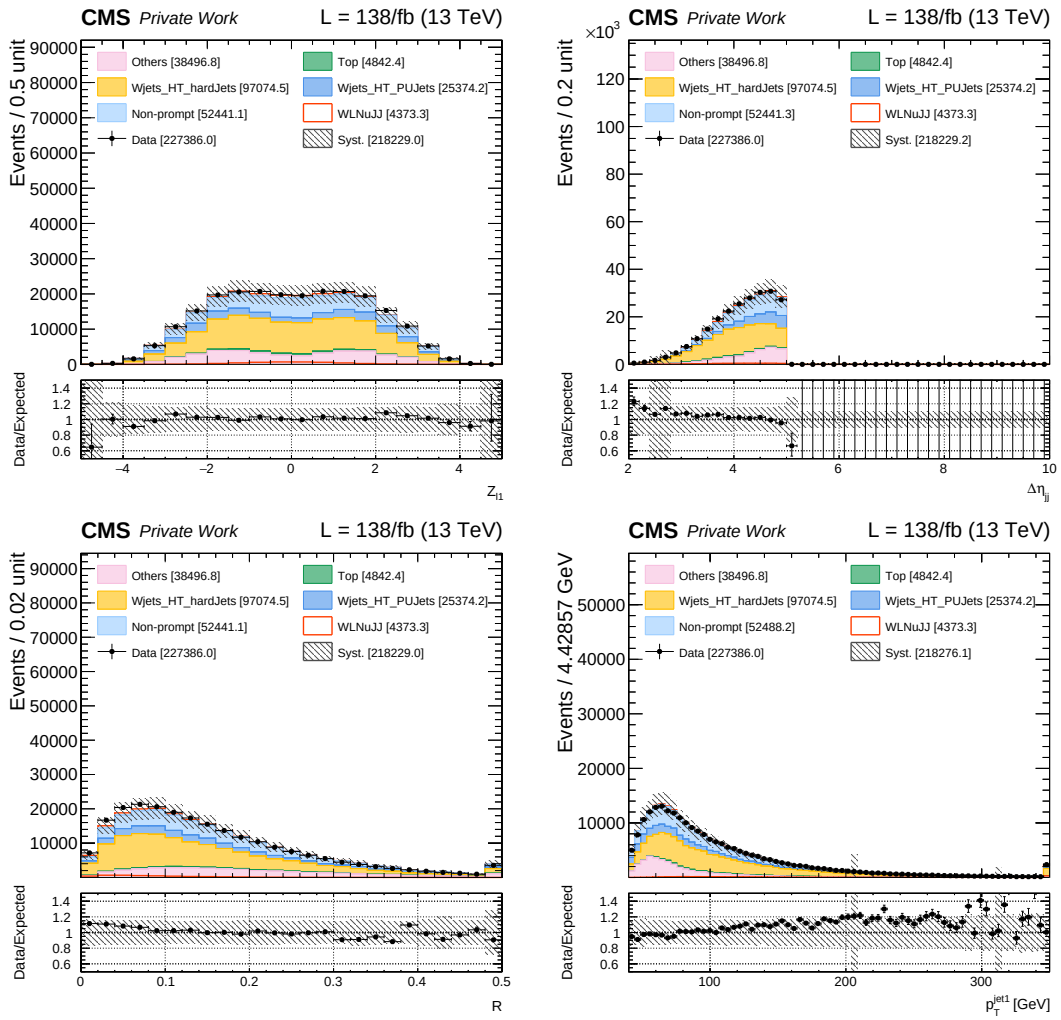


Figure 3.34: Data/MC agreement in the non-prompt CR for the electron final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: R total transverse momentum projection and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

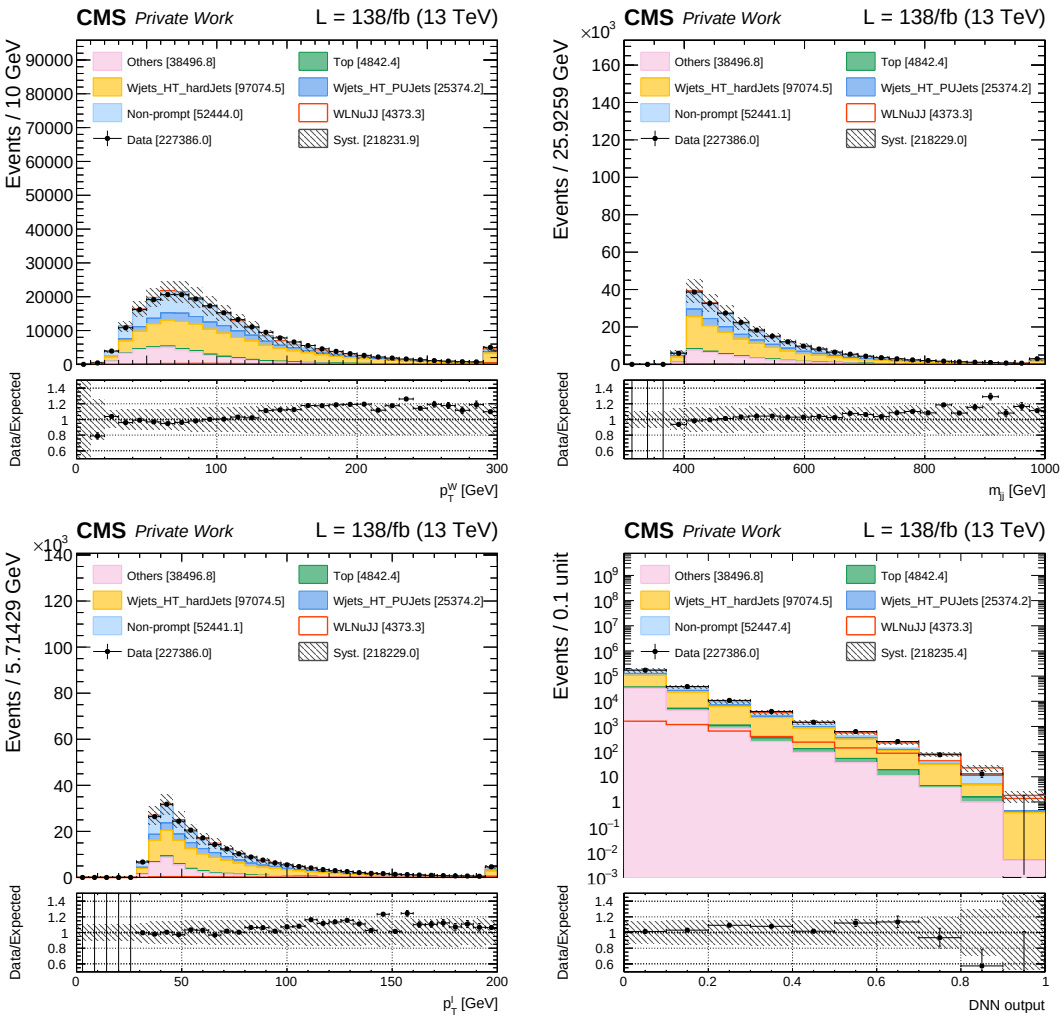


Figure 3.35: Data/MC agreement in the non-prompt CR for the electron final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

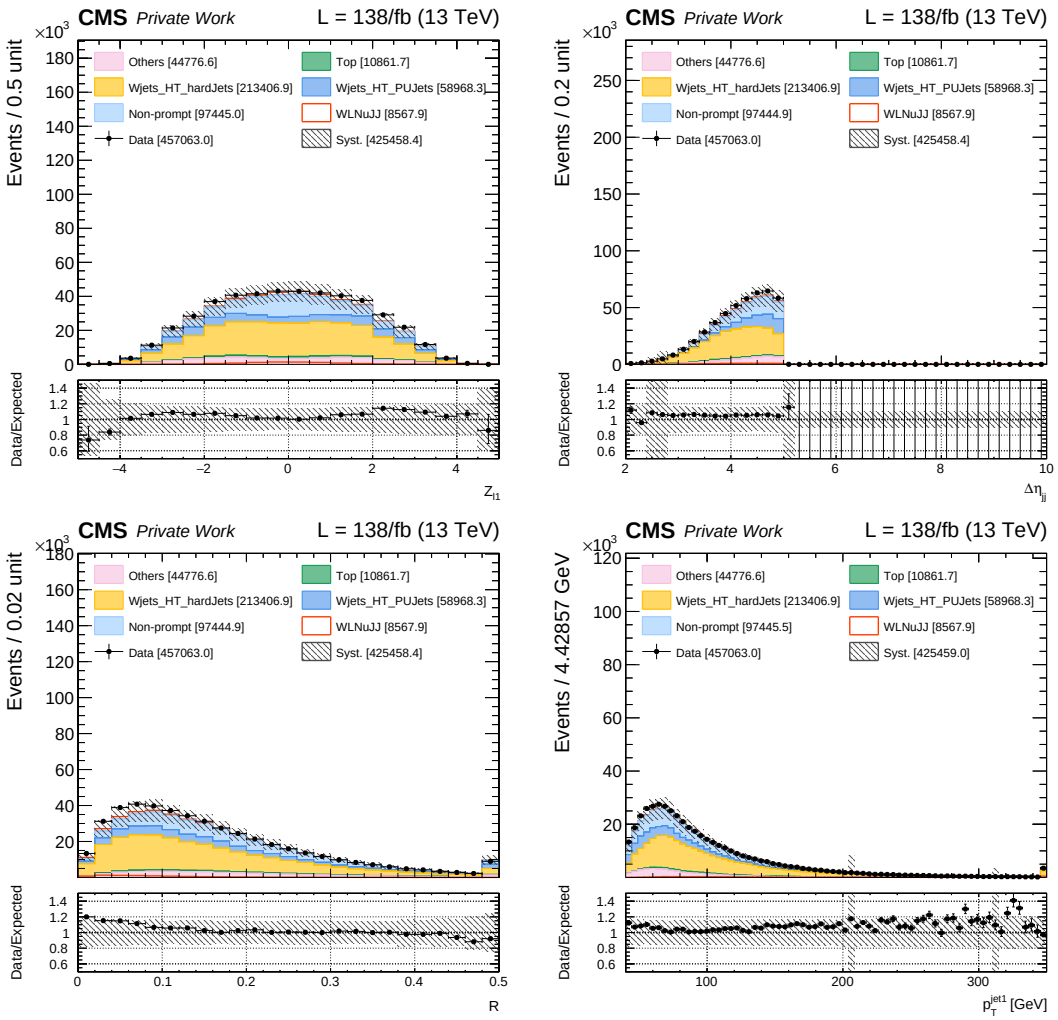


Figure 3.36: Data/MC agreement in the non-prompt CR for the muon final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

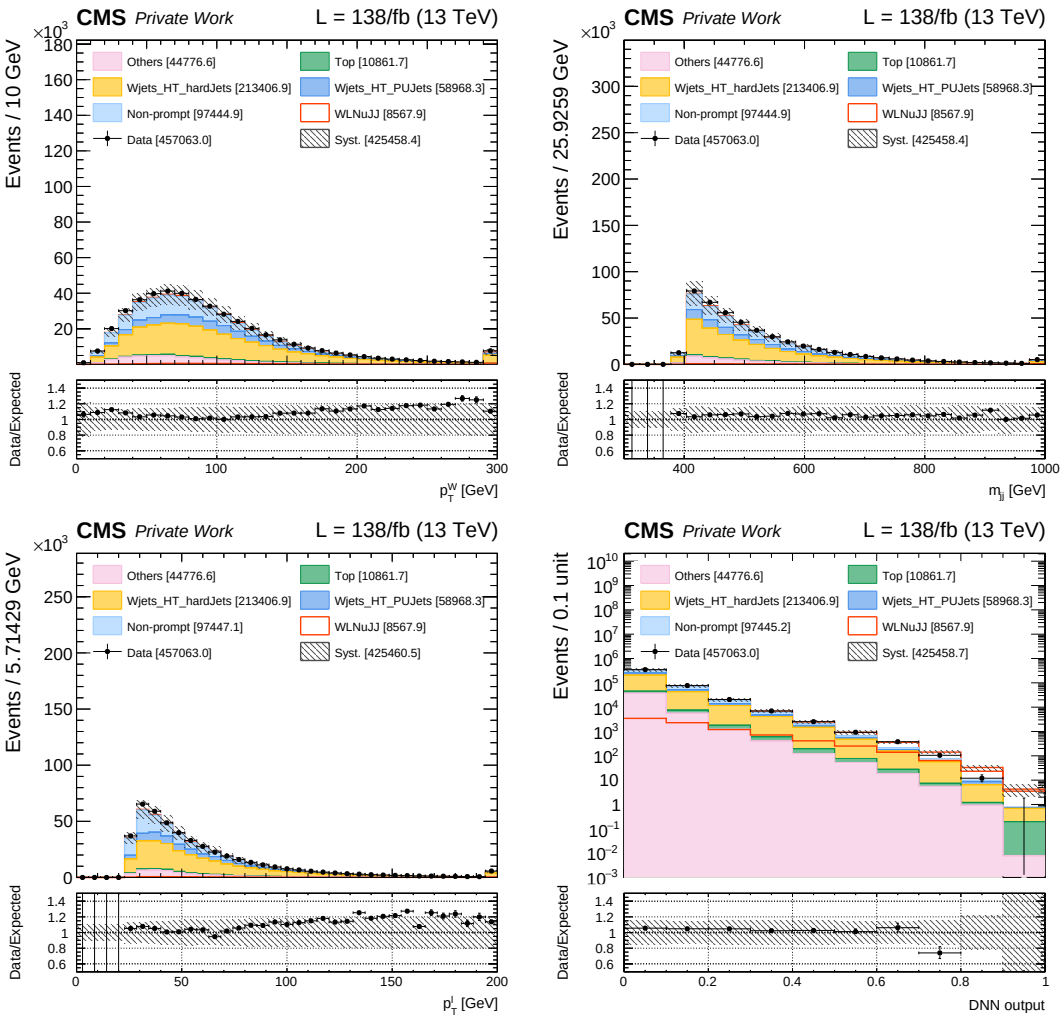


Figure 3.37: Data/MC agreement in the non-prompt CR for the muon final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

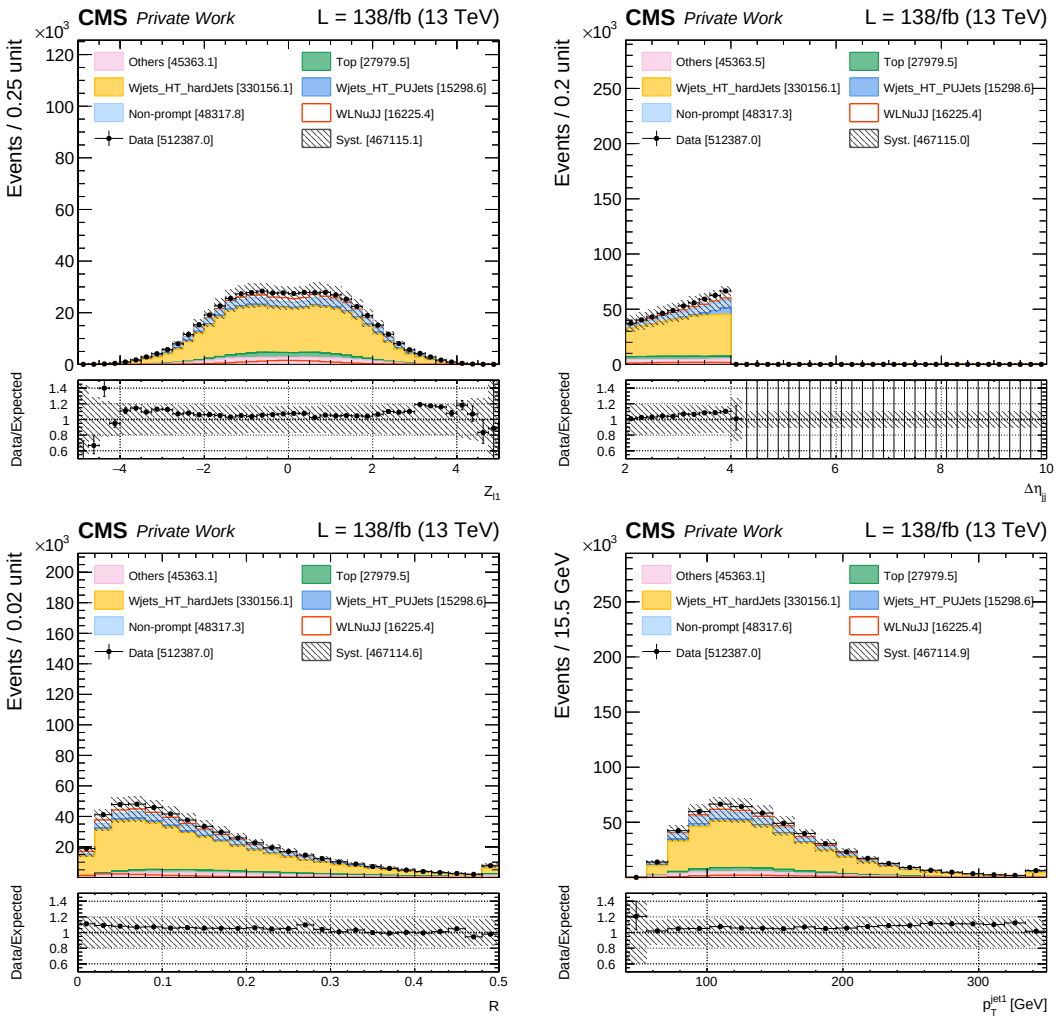


Figure 3.38: Data/MC agreement in the W+Jets CR for the electron final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

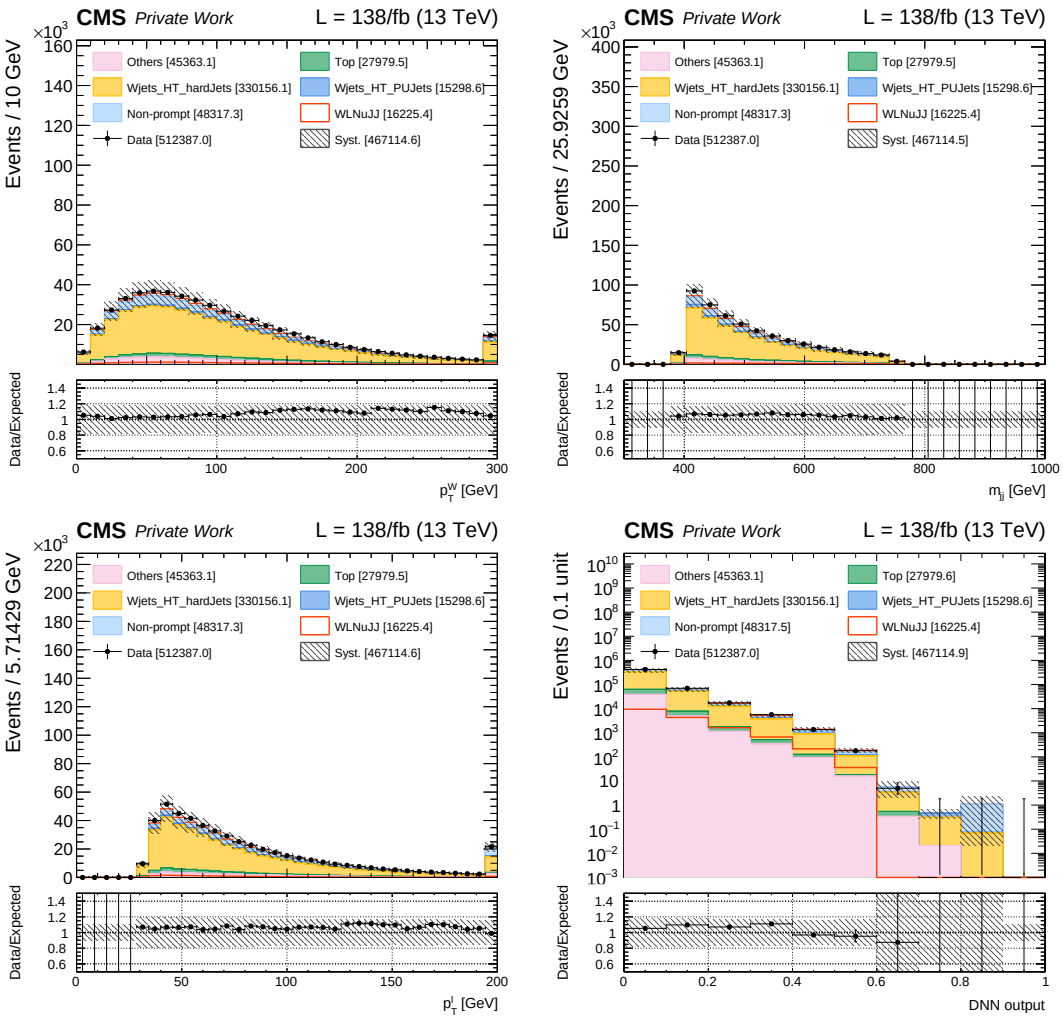


Figure 3.39: Data/MC agreement in the W+Jets CR for the electron final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

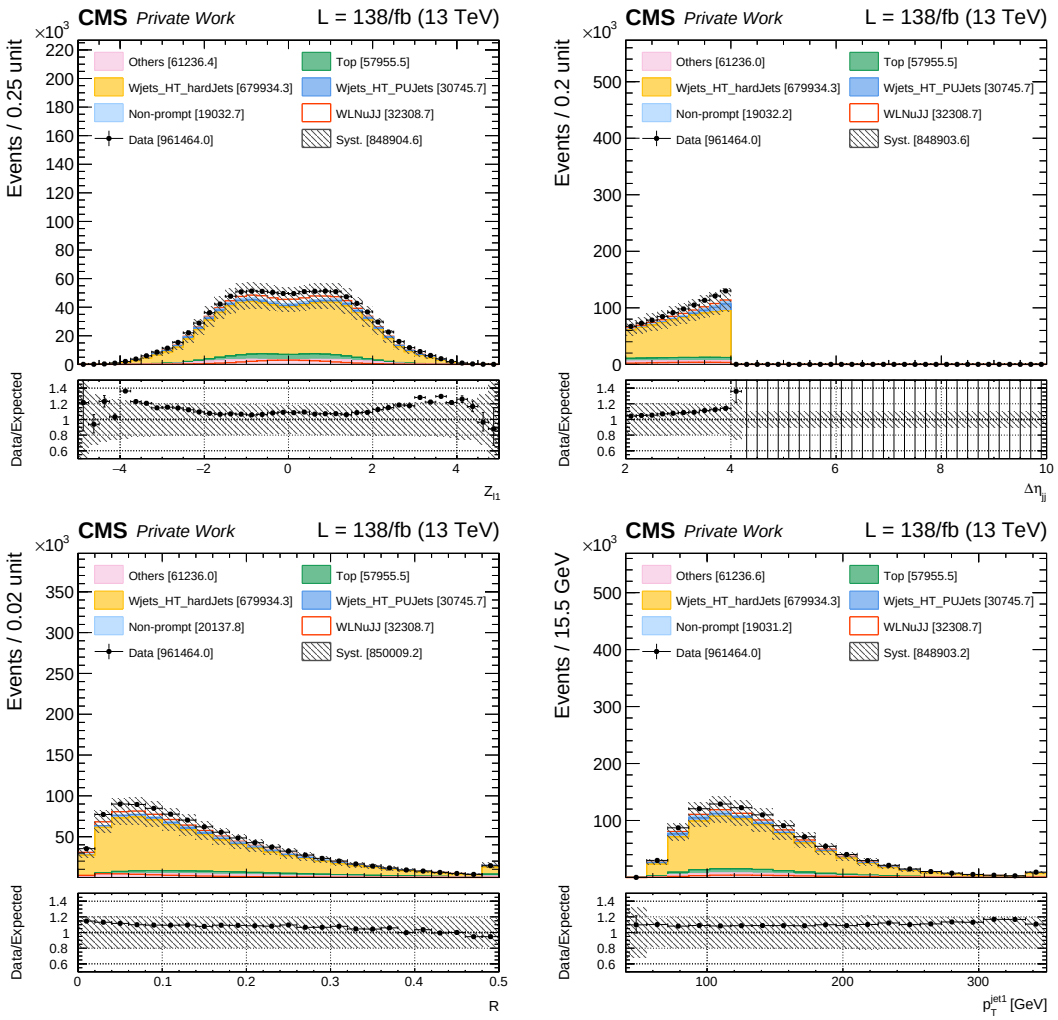


Figure 3.40: Data/MC agreement in the W+Jets CR for the muon final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

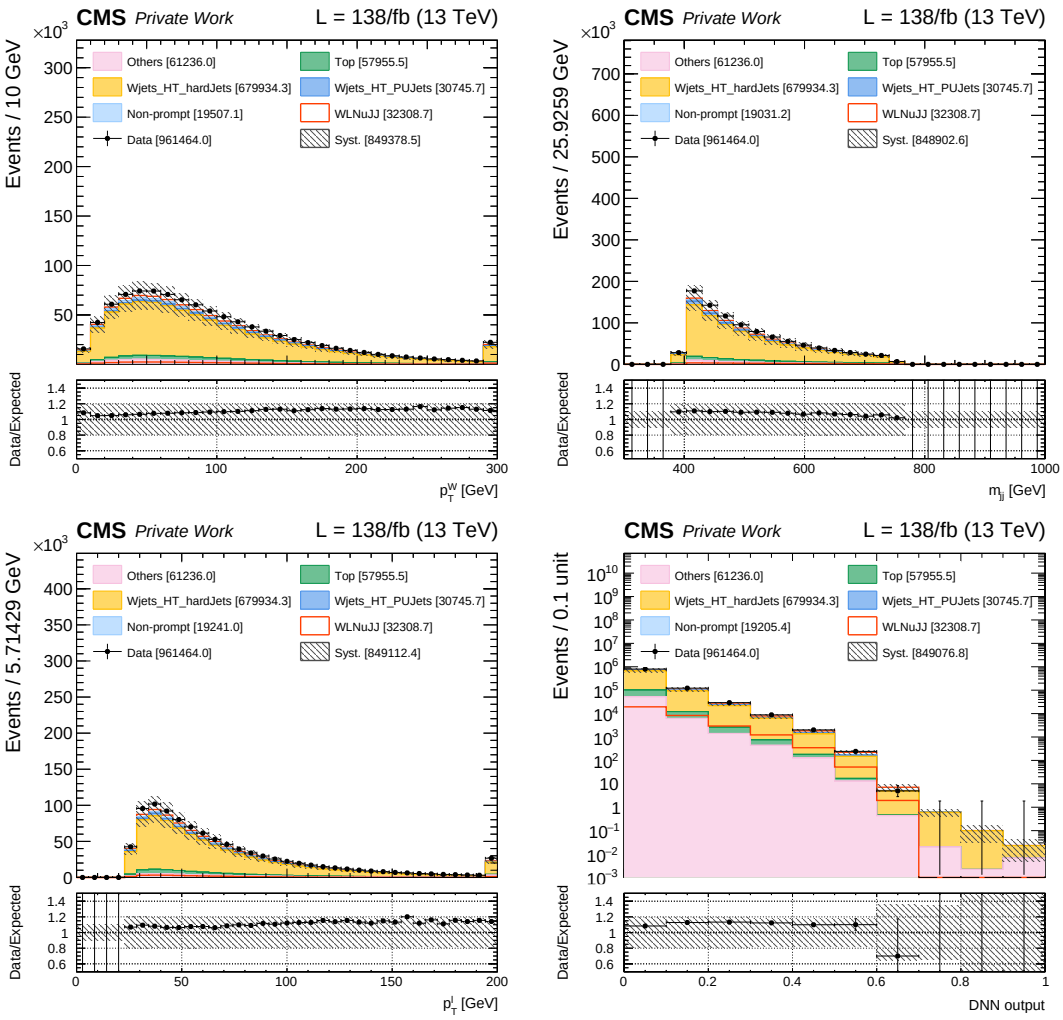


Figure 3.41: Data/MC agreement in the W+Jets CR for the muon final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

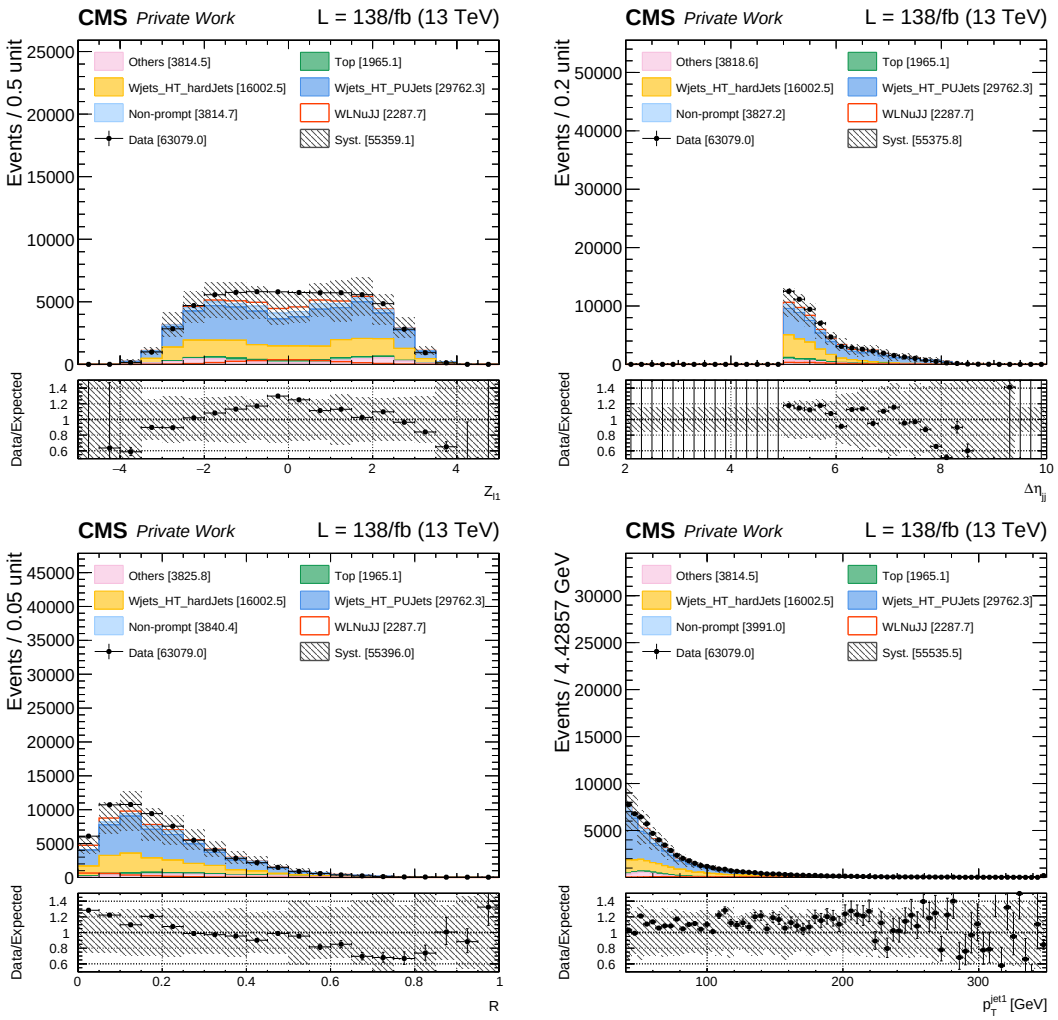


Figure 3.42: Data/MC agreement in the W+1PU Jets CR for the electron final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

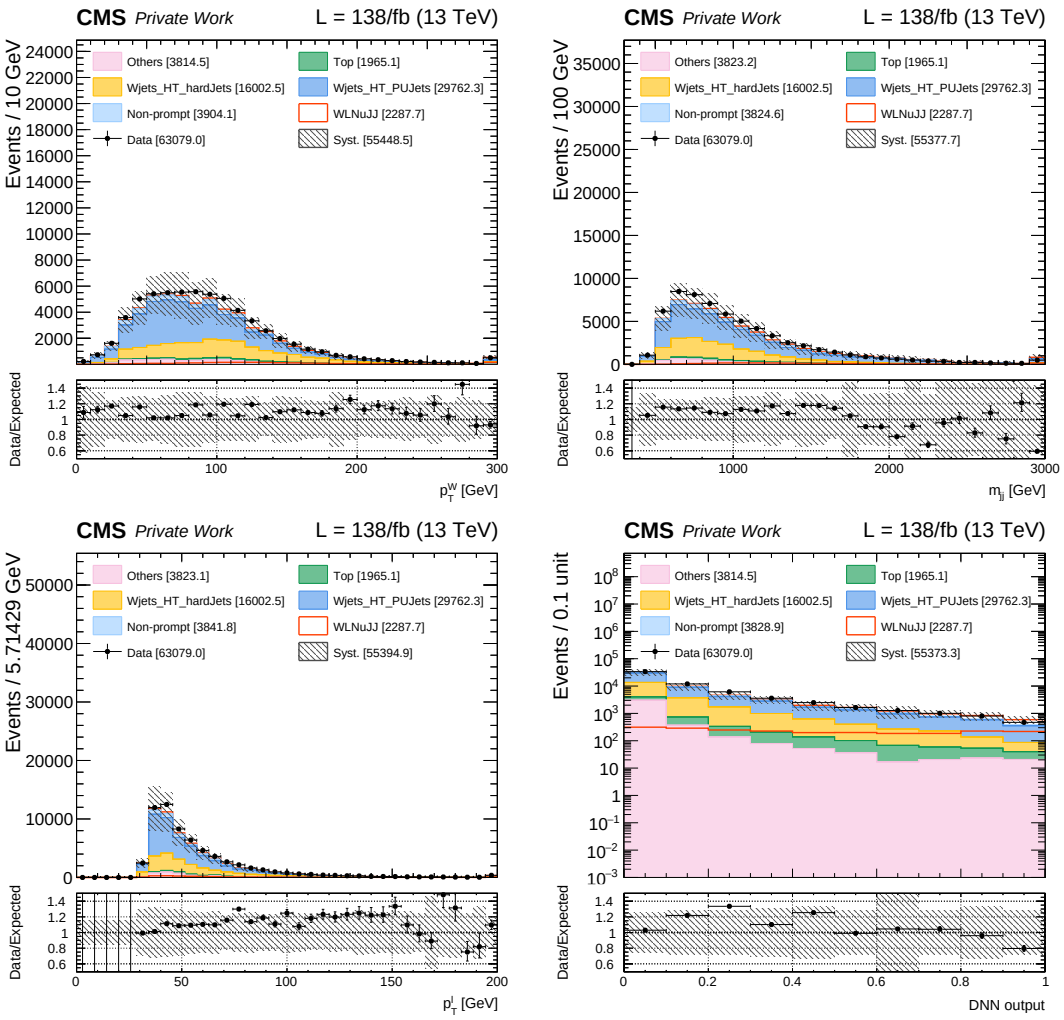


Figure 3.43: Data/MC agreement in the W+1PU Jets CR for the electron final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

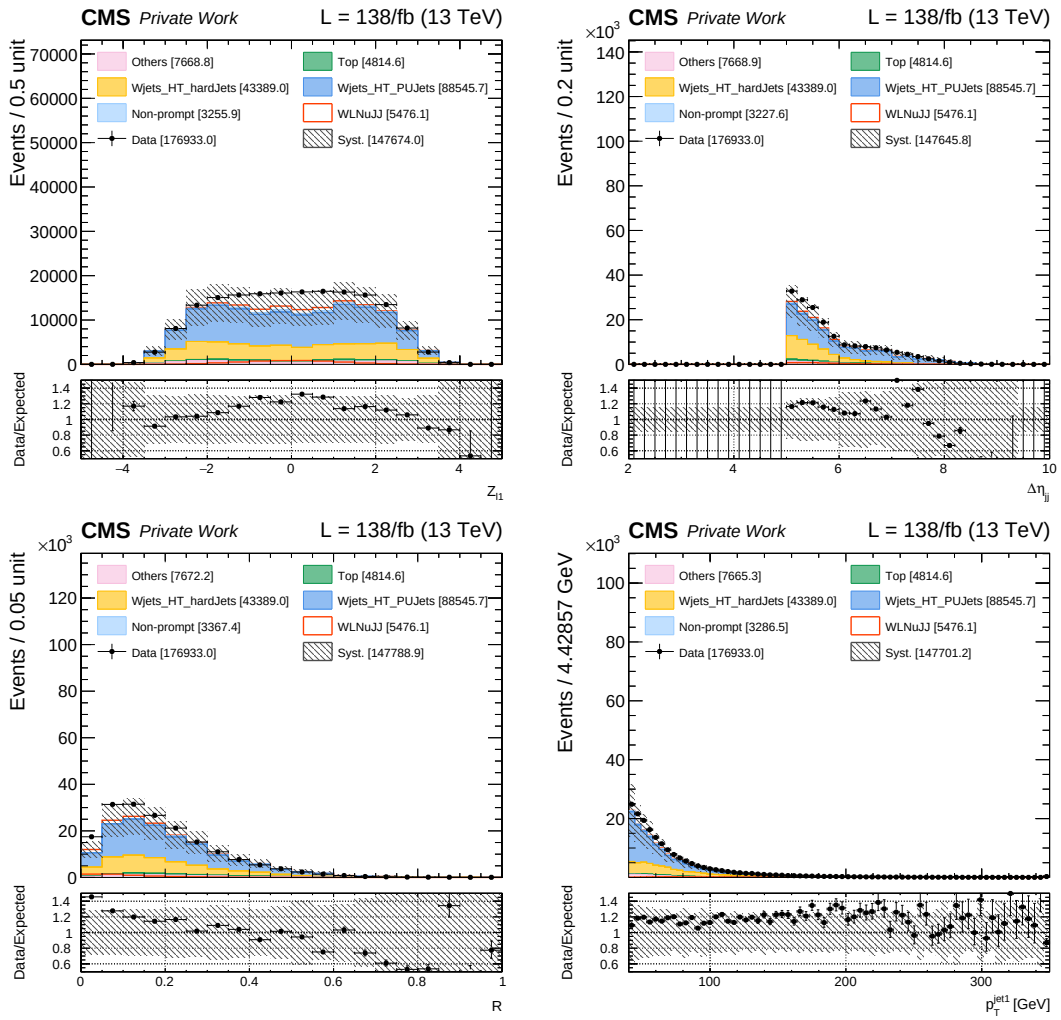


Figure 3.44: Data/MC agreement in the W+1PU Jets CR for the muon final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

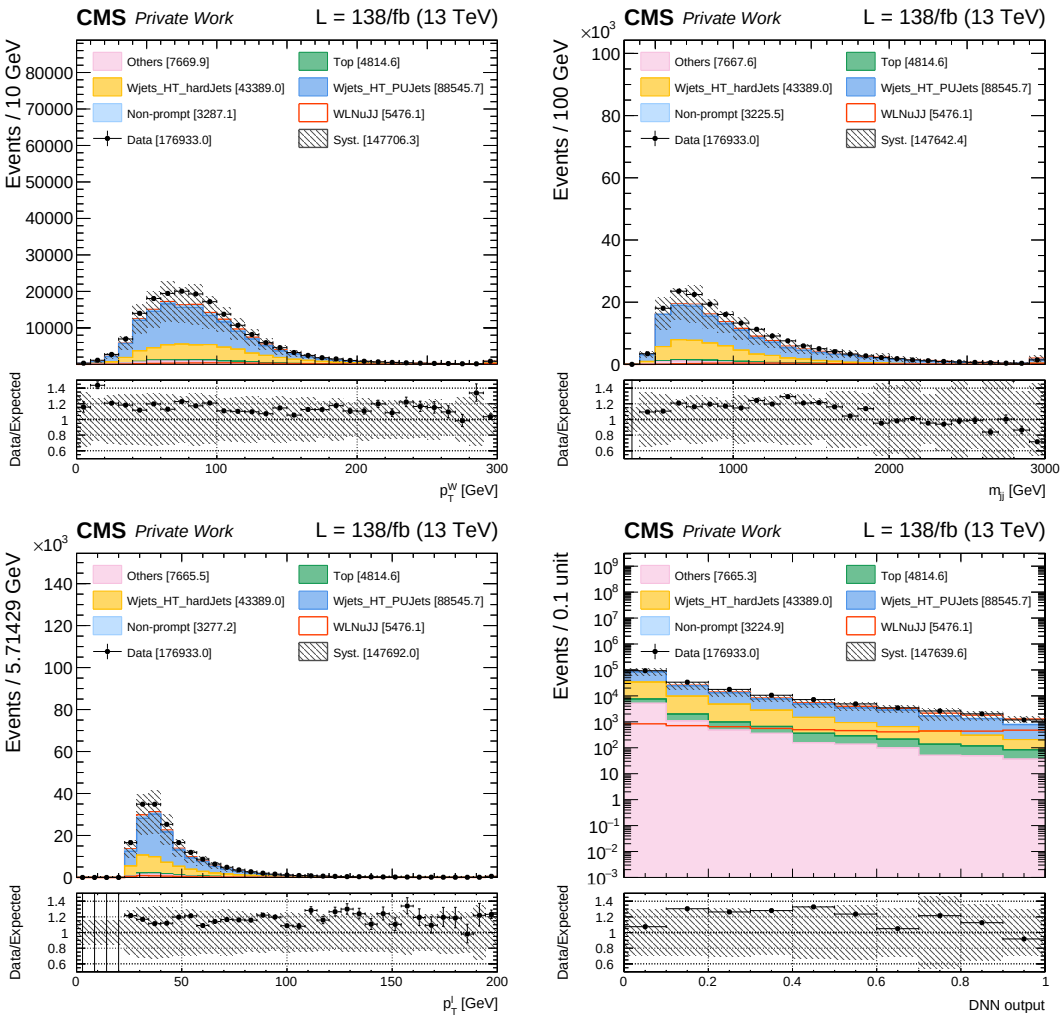


Figure 3.45: Data/MC agreement in the W+1PU Jets CR for the muon final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

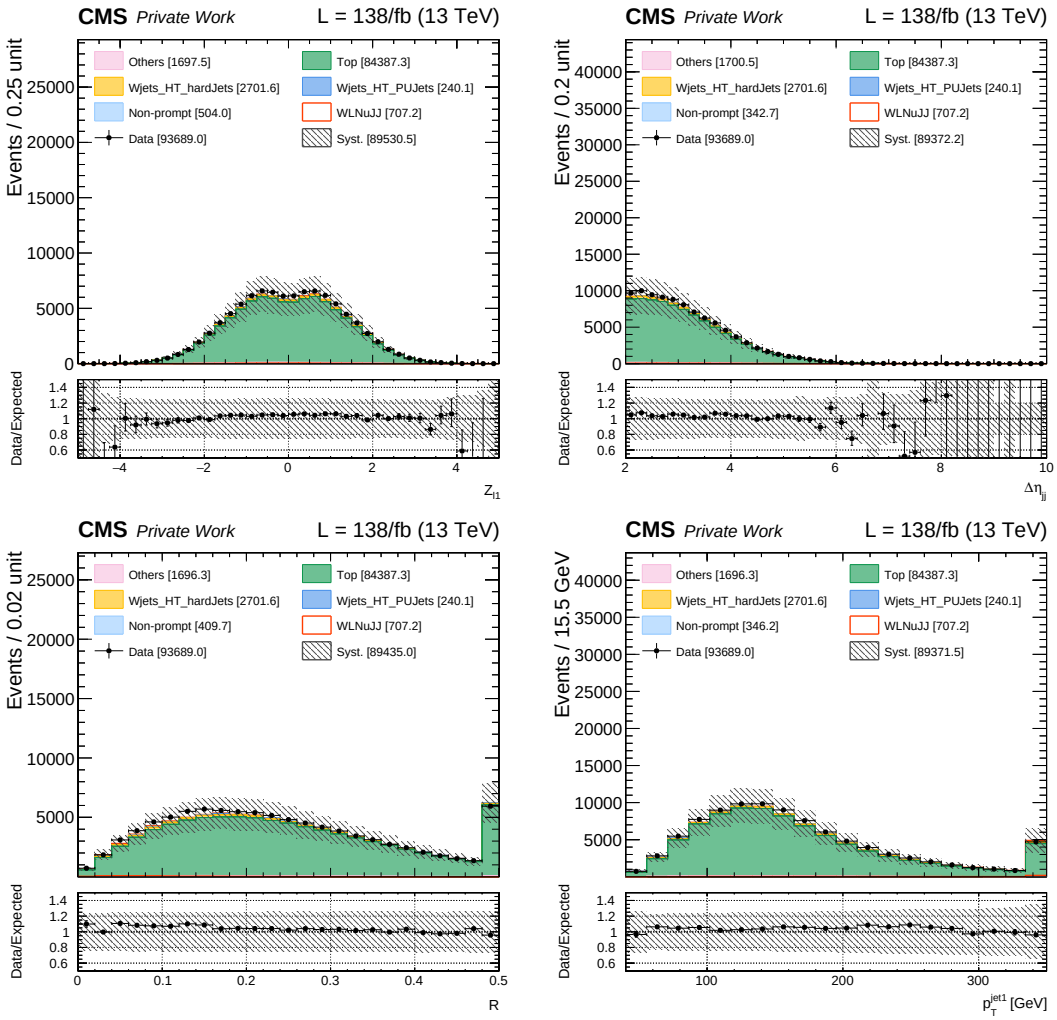


Figure 3.46: Data/MC agreement in the top-enriched CR for the electron final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty.

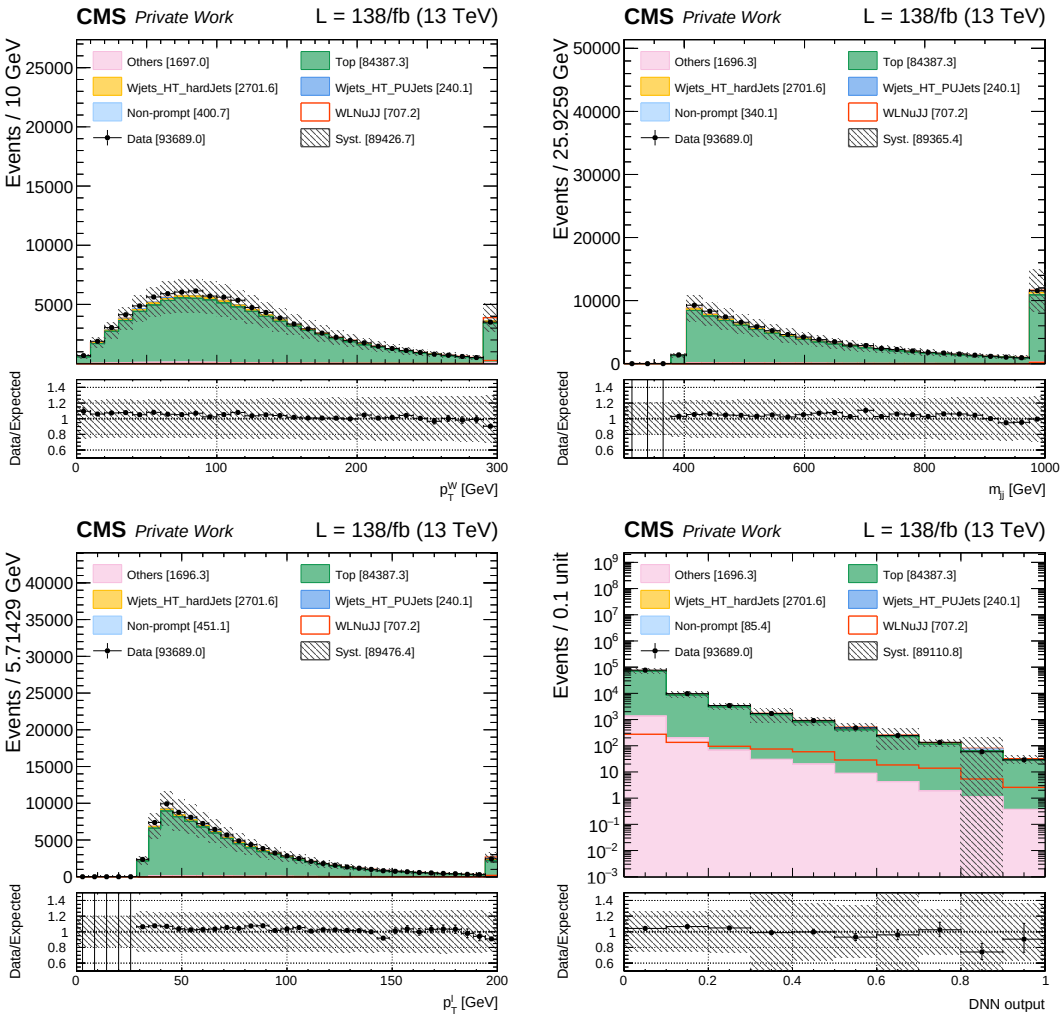


Figure 3.47: Data/MC agreement in the top-enriched CR for the electron final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty.

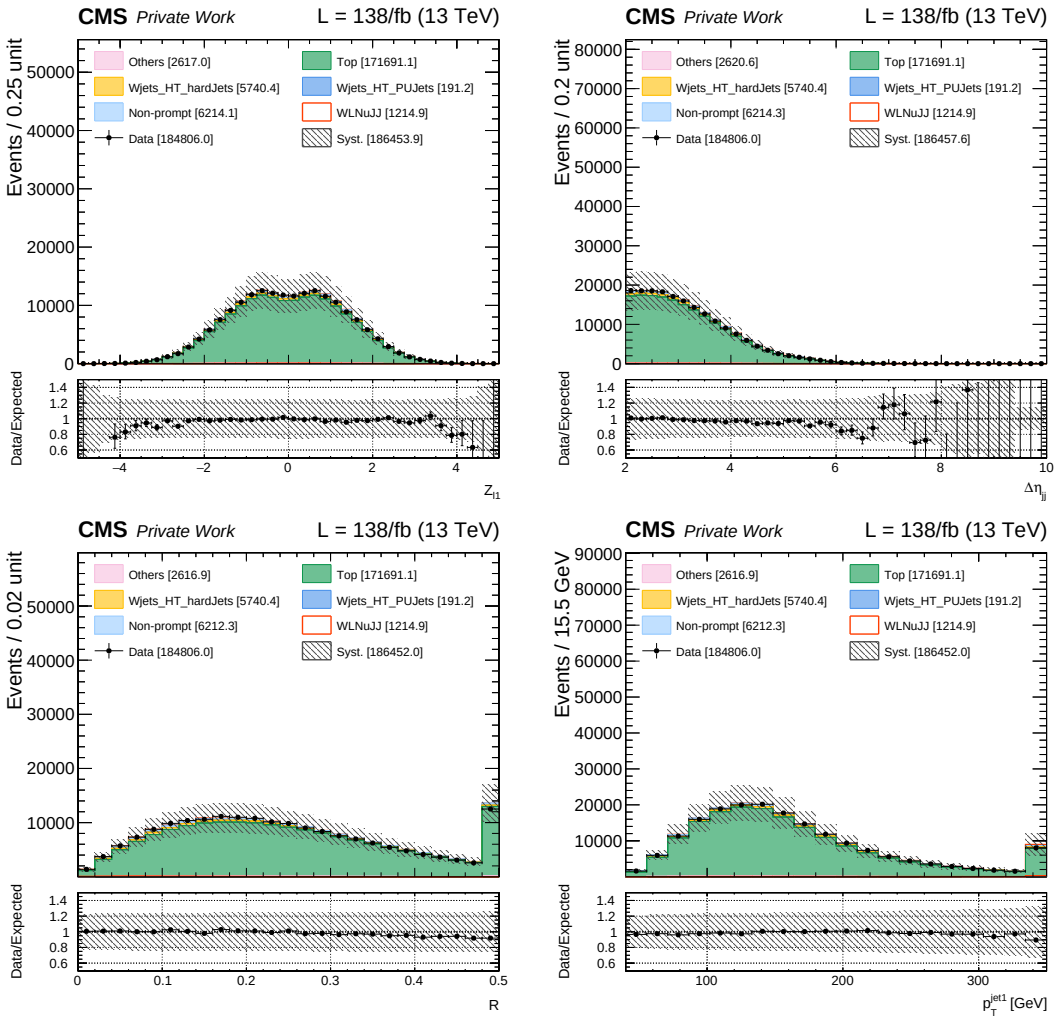


Figure 3.48: Data/MC agreement in the top-enriched CR for the muon final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. The grey-shaded areas represent the total systematic uncertainty.

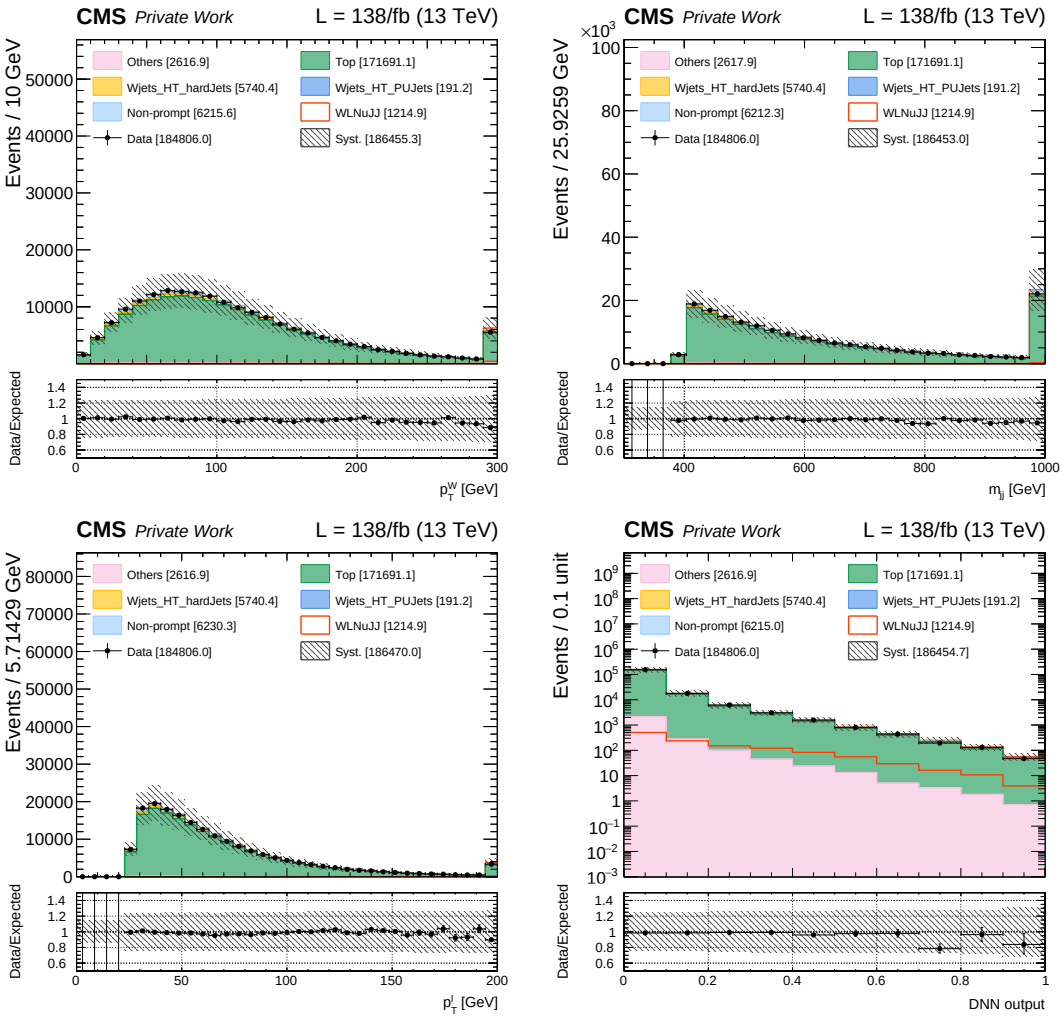


Figure 3.49: Data/MC agreement in the top-enriched CR for the muon final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. The grey-shaded areas represent the total systematic uncertainty.

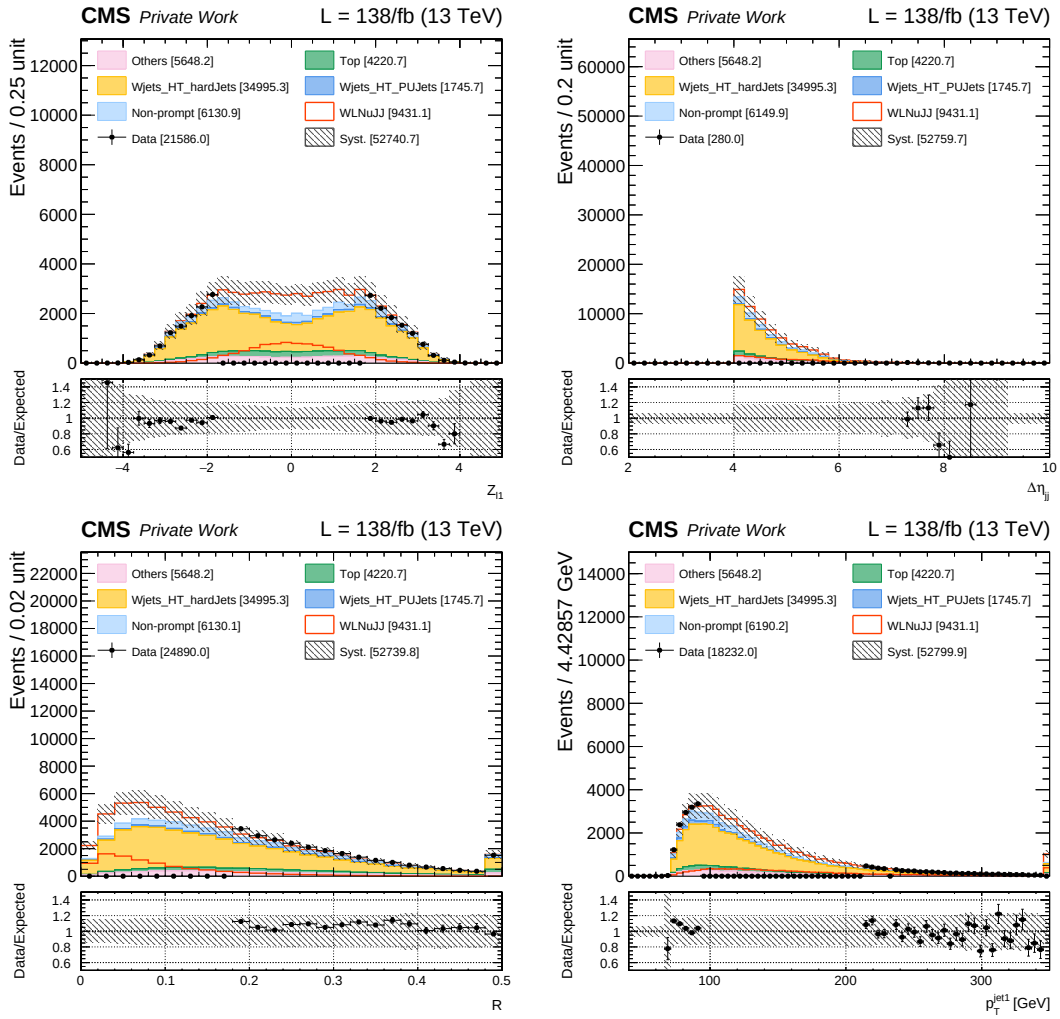


Figure 3.50: Data/MC agreement in the SR for the electron final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. Only bins with $S/\sqrt{B} < 5$ are shown. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

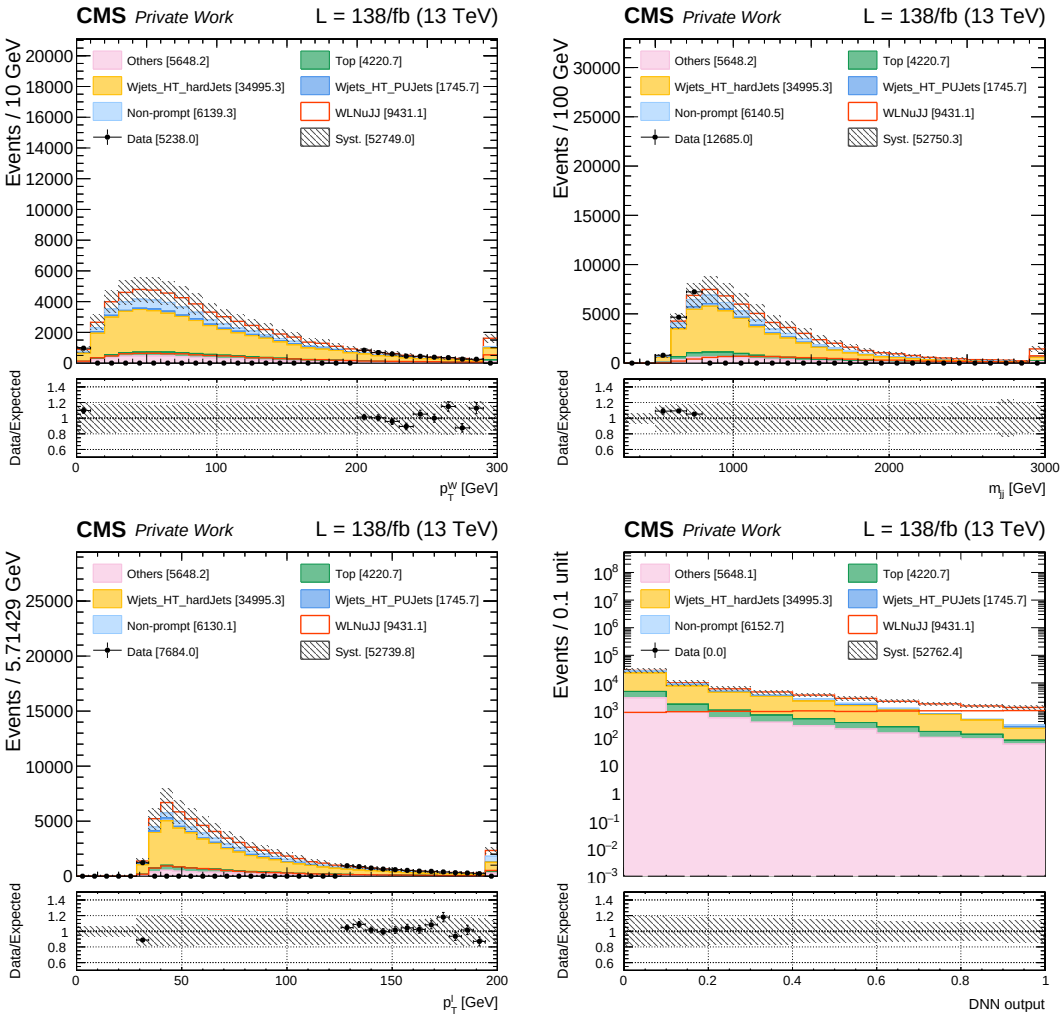


Figure 3.51: Data/MC agreement in the SR for the electron final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. Only bins with $S/\sqrt{B} < 5$ are shown. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

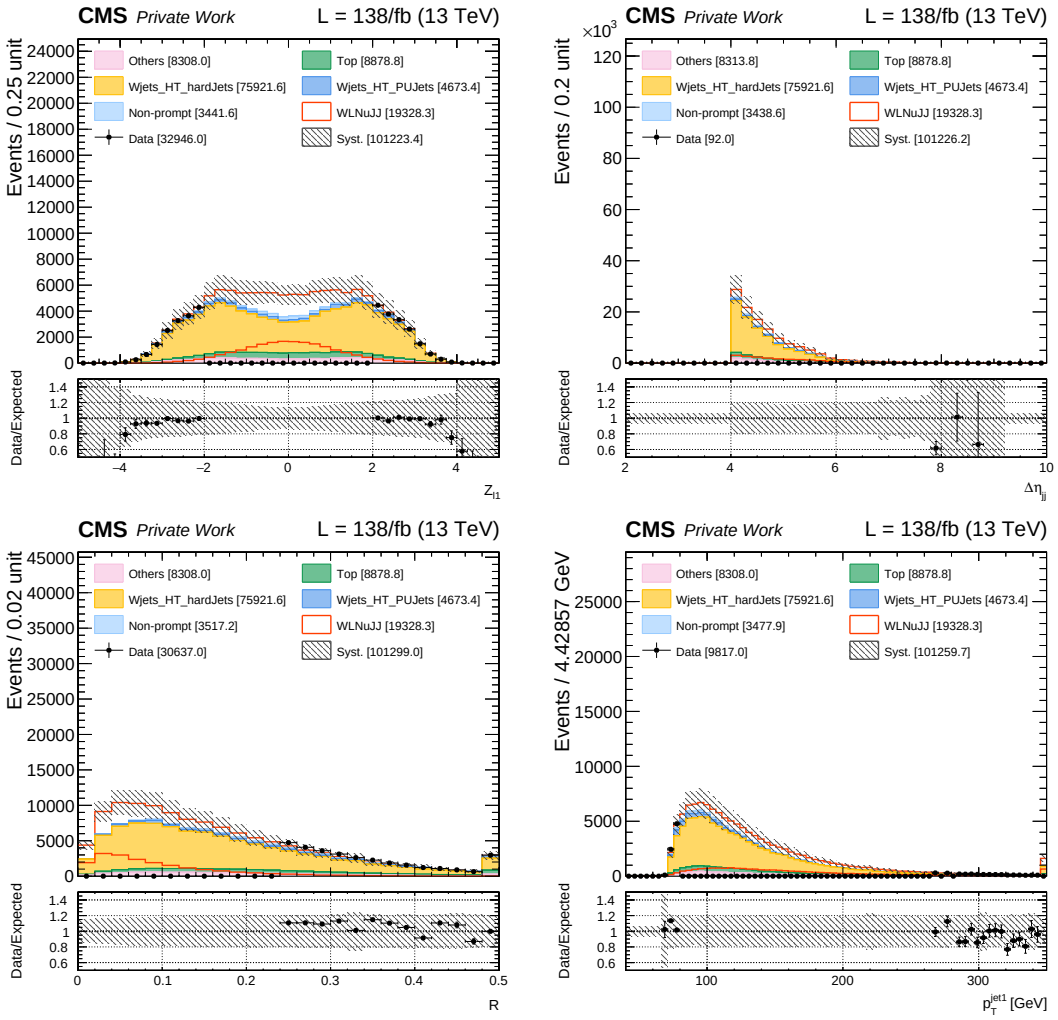


Figure 3.52: Data/MC agreement in the SR for the muon final state. Top row: Zeppenfeld variable of the lepton and $|\Delta\eta_{jj}|$ of the dijet system. Bottom row: total transverse momentum projection R and p_T of the leading jet. Only bins with $S/\sqrt{B} < 5$ are shown. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

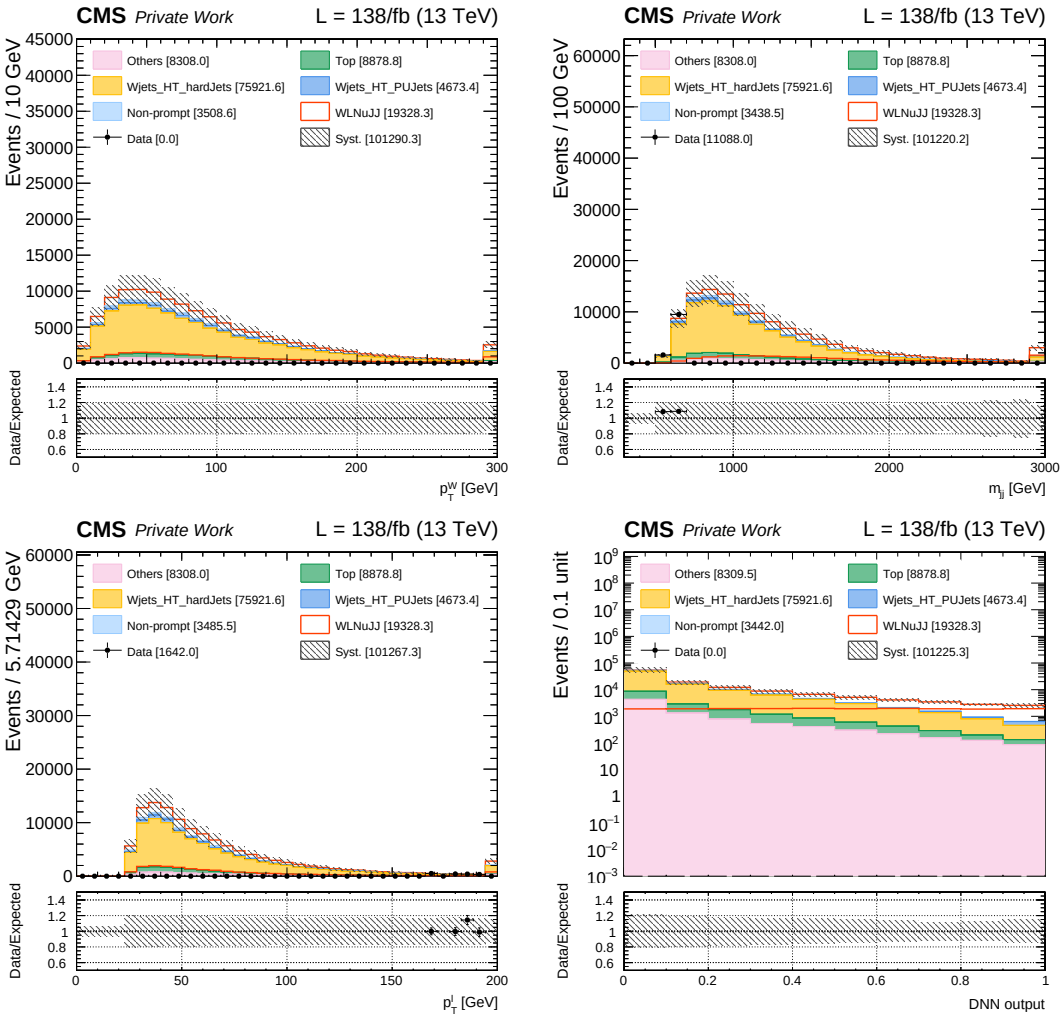


Figure 3.53: Data/MC agreement in the SR for the muon final state. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and DNN output. Only bins with $S/\sqrt{B} < 5$ are shown. The grey-shaded areas represent the total systematic uncertainty, including the post-fit rate parameter uncertainties for the main background processes.

3.8.1 Inclusive cross section measurement

The signal extraction is performed by fitting all control and signal regions simultaneously, using a consistent requirement of DNN > 0.5 across all categories entering the likelihood. This threshold, as discussed in Section 3.6, corresponds to the working point that maximises the expected signal significance in the signal region. In addition, including the low-DNN region would force the fit to constrain several nuisance parameters associated with the dominant backgrounds and with the signal itself. Using a common DNN threshold across all regions ensures both consistent signal- and background-enriched selections and improved stability of the fit.

Figure 3.54 shows the likelihood scan for each individual year and for the full Run II combination. The resulting expected cross section for the EW production of a W boson in association with two jets is:

$$\sigma_{\mathrm{EW}}(W{+}jj) = 23.04^{+0.53}_{-0.48}\,(\mathrm{exp})\,^{+0.97}_{-0.85}\,(\mathrm{cal})\,^{+2.09}_{-1.95}\,(\mathrm{theo})\,\pm 0.51\,(\mathrm{syst})\,\pm 0.48\,(\mathrm{stat})\,\mathrm{pb}.$$

The dominant role of systematic uncertainties is evident both in the impact plot (Figure 3.55) and in the breakdown likelihood scan (Figure 3.54), where groups of uncertainties are progressively frozen. These studies highlight that theoretical sources, in particular the QCD scale variations on the LO signal sample, are the leading contributors to the total uncertainty.

The *theory* group includes variations of the QCD scales (i.e. renormalisation and factorisation scales), as well as uncertainties from parton distribution functions (PDFs) and parton-shower modelling, which affect both the EW signal and dominant backgrounds such as W+jets and top-quark production. The *calibration* group collects effects related to the energy and momentum calibration of reconstructed objects, including the jet energy scale and resolution, lepton momentum scale, missing transverse energy, and b-tagging efficiency. The *experimental* group includes uncertainties on the trigger and reconstruction efficiencies, pileup reweighting, and integrated luminosity determination. Finally, all remaining sources, mainly data-driven background normalisations and limited MC statistics, are grouped under the *systematic* category.

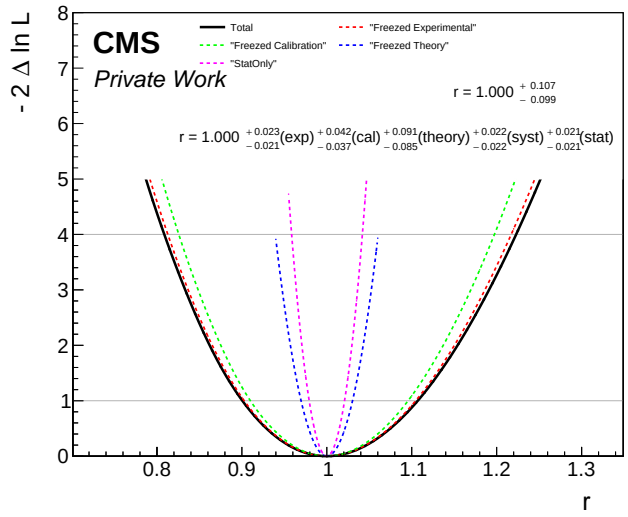


Figure 3.54: Breakdown of the expected uncertainty in the signal strength for the full Run II combination. Groups of systematic uncertainties are progressively frozen in the likelihood scan, illustrating their relative contribution to the total uncertainty. The scan shows that theoretical sources, particularly QCD scale variations, dominate the overall measurement uncertainty.

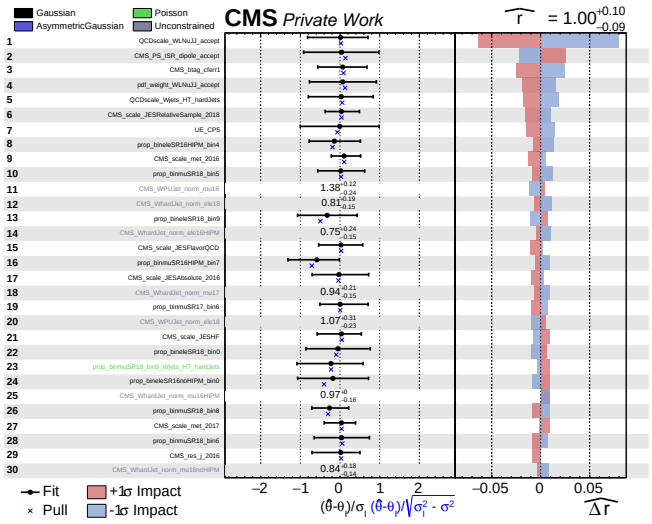


Figure 3.55: Impact plot showing the effect of various nuisance parameters on the signal strength, as defined in Section 1.9. Each horizontal bar represents a different nuisance parameter and its corresponding impact on the central value and uncertainty of the parameter of interest.

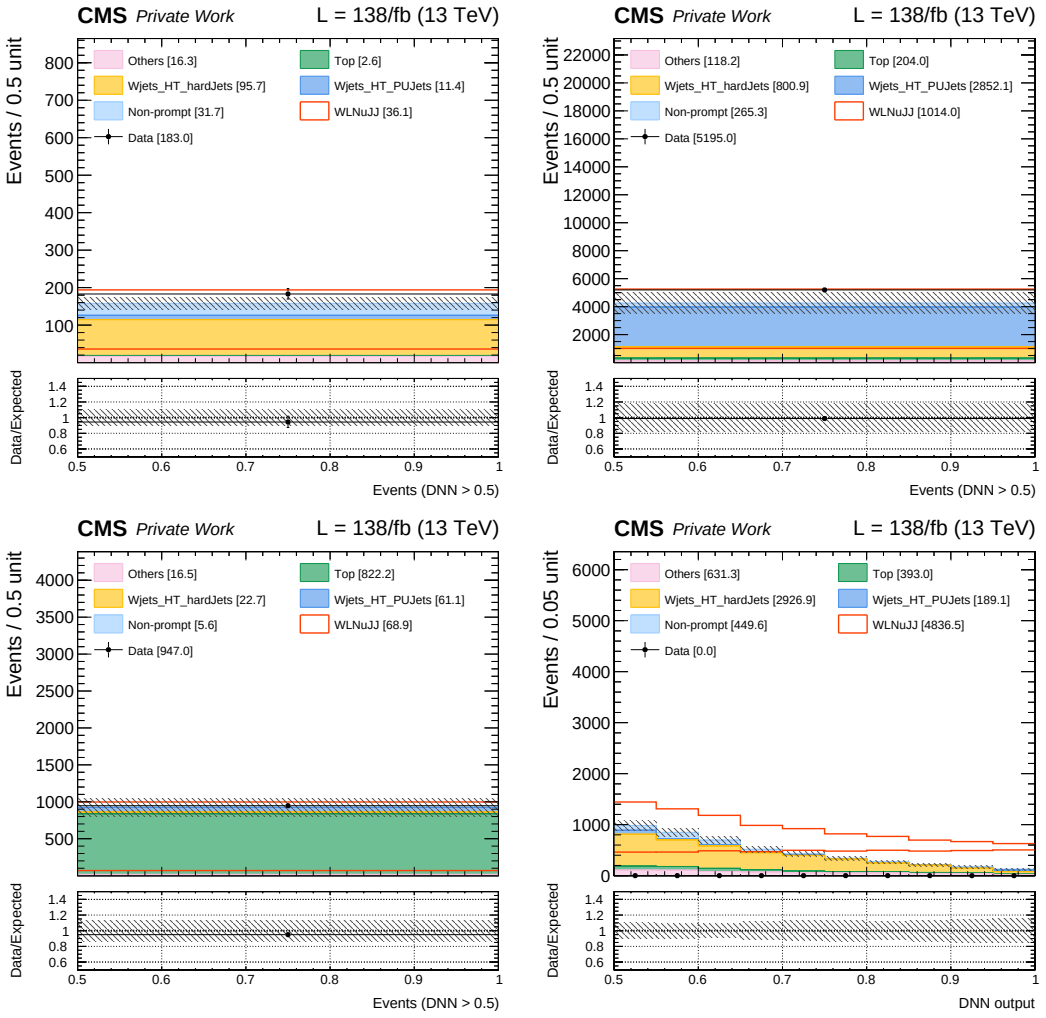


Figure 3.56: Template inputs to the likelihood fit in the electron channel for events with DNN > 0.5. The upper row shows the W+jets control regions, corresponding to the hard-jet (left) and pileup-jet (right) categories. The lower row displays the top-quark control region (bottom left) and the signal region (bottom right), where the latter is shown blinded and binned in ten intervals of the DNN output. In all panels, the EW Wjj signal is shown both stacked within the total expectation and overlaid as an outline to illustrate its relative contribution.

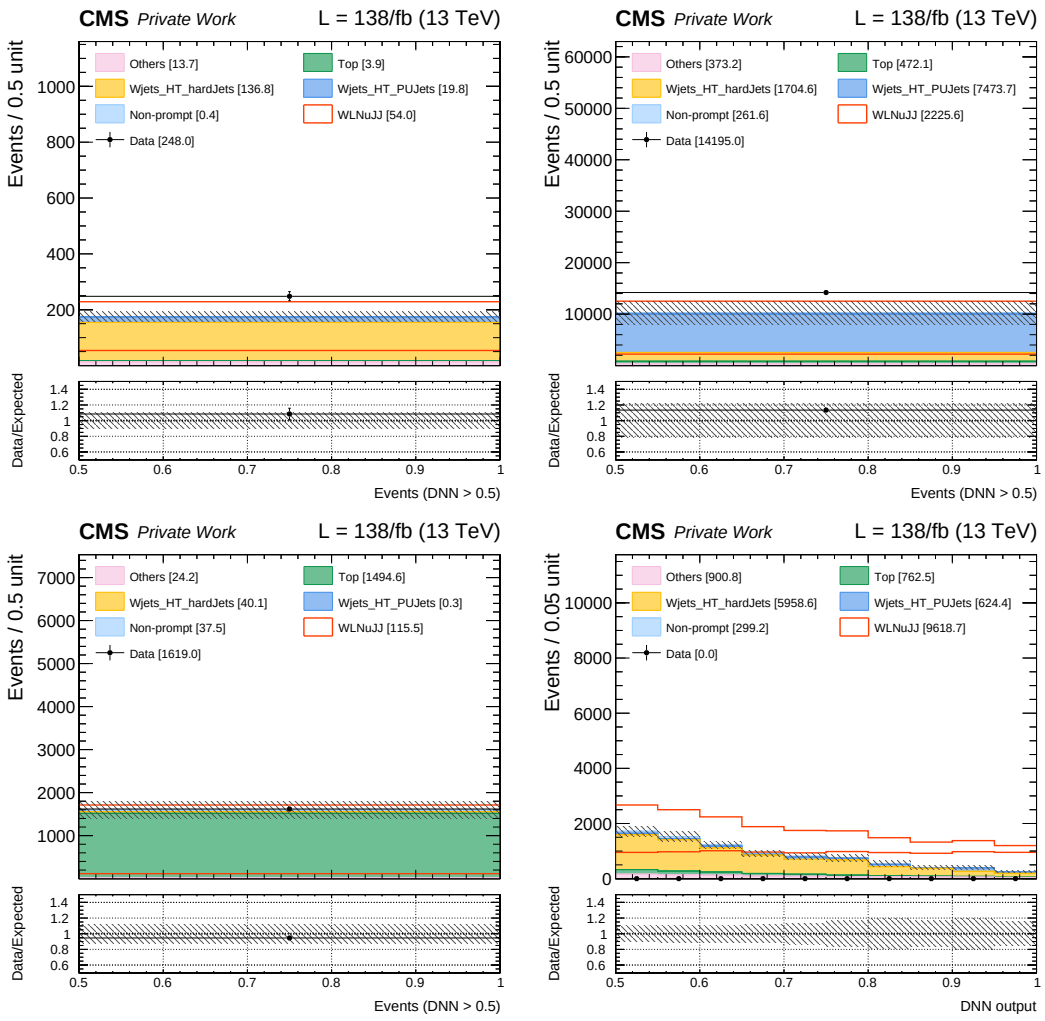


Figure 3.57: Template inputs to the likelihood fit in the muon channel for events with DNN > 0.5. The upper row shows the W+jets control regions, corresponding to the hard-jet (left) and pileup-jet (right) categories. The lower row displays the top-quark control region (bottom left) and the signal region (bottom right), where the latter is shown blinded and binned in ten intervals of the DNN output. In all panels, the EW Wjj signal is shown both stacked within the total expectation and overlaid as an outline to illustrate its relative contribution.

Chapter 4

Fiducial Cross Section Results

This chapter presents the measurement of the EW Wjj production cross section in a restricted, generator-level phase space, referred to as the *fiducial region*. The fiducial approach provides a quasi-model-independent determination of the production rate, avoiding extrapolations to the full phase space and enabling a direct comparison with theoretical predictions. The definition of the fiducial region, designed to emulate the reconstructed-level signal selection while minimising migration effects closely, is detailed in the following section.

The measurement strategy proceeds in two steps. First, an inclusive fiducial cross section is extracted within the defined phase space in Section 4.1. Subsequently, in Section 4.2, differential measurements are performed as a function of several generator-level observables, yielding unfolded particle-level distributions that provide a more detailed characterisation of the underlying dynamics. For the first observable, the dijet invariant mass m_{jj}, discussed in Section 4.2.3, two complementary methodologies are compared: the classical one-dimensional unfolding approach, which exploits the reconstructed distribution of the observable to gain sensitivity to its generator-level spectrum, and an alternative method based on a multiclass neural-network classifier that captures correlations among multiple final-state observables. From Section 4.2.4 onwards, the same classical unfolding technique is applied to all remaining differential observables.

4.1 Inclusive Fiducial Cross Section Measurement

The fiducial region emulates the selections applied at the reconstructed level in the SR, while being slightly more relaxed on kinematic thresholds to reduce migration of events outside the fiducial acceptance. The fiducial phase space is defined by the generator-level selections reported in table 4.1:

At the generator level, physics objects are defined to mirror as closely as possible their reconstructed counterparts. Leptons are defined as *dressed* leptons, obtained by adding to the prompt leptons all photons found within $\Delta R < 0.3$

Region	Requirements
Fiducial Region	$p_T^{\ell} \geq 20$ GeV
	$N_{\text{jets}}(p_T \geq 20 \text{ GeV}) \geq 2$
	$p_T^{j_1} \geq 60$ GeV
	$p_T^{j_2} \geq 60$ GeV
	$m_{jj} \geq 300$ GeV
	$\lvert\Delta\eta_{jj}\rvert \geq 4.0$
	$\lvert\eta_{\ell}\rvert < 2.0$
	$\lvert\eta_{j_1}\rvert < 4.7$
	$\lvert\eta_{j_2}\rvert < 4.7$

Table 4.1: Generator-level selections defining the fiducial phase space. These cuts emulate the SR reconstruction-level selections while being slightly more relaxed on kinematic thresholds.

of their direction. This procedure reproduces the way lepton energies are reconstructed in CMS, where electrons and muons collect both true final-state radiation and photons emitted through interactions with the detector material (e.g. bremsstrahlung in the ECAL). Using dressed leptons avoids introducing artificial biases between generator- and reconstruction-level definitions and ensures that the migration between the two remains small.

Jets are then built from stable particles using the anti-k_T algorithm with $R = 0.4$, excluding neutrinos and the dressed leptons defined above inside a cone of $\Delta R < 0.3$. This yields a generator-level jet collection that is consistent with the PF-based AK4 jet definition used in the analysis and avoids double-counting energy deposits between jets and leptons.

These cuts select events that closely resemble the SR selections at the reconstruction level (Table 3.9), while allowing slightly more relaxed thresholds on energetic variables such as jet, dijet, and lepton kinematics. Events failing these requirements are treated as background-controlled in the fit and assigned a dedicated rate parameter, similarly to what is done for the main background processes.

The strategy for measuring the fiducial cross section is similar to the standard analysis in the SR. Events in the SR are required to satisfy a DNN output threshold of 0.5. The DNN output is then directly used as the observable in the fit, which simultaneously constrains the signal yield and the main backgrounds.

Several optimisation studies were performed to validate the choices made in the DNN-based event selection. As discussed in Section 3.6, a threshold of DNN > 0.5 was found to provide the optimal sensitivity at the nominal level. To ensure that this conclusion remains valid once all systematic uncertainties are taken into account, the full expected likelihood scans were repeated for different threshold values between 0.5 and 0.8. As shown in Figure 4.1, a threshold of 0.5 still provides the best expected performance. It was further verified that this behaviour is driven by the larger available statistics at lower thresholds, rather

than by tighter post-fit constraints on the nuisance parameters (see Figure 4.4). Finally, the binning of the observable used in the fit was varied: beyond ten bins, the expected sensitivity saturates, indicating that additional binning brings negligible improvement. Consequently, the final fit configuration employs a DNN $>$ 0.5 selection and uses a maximum of ten bins, consistent with the configuration adopted in the inclusive cross section measurement.

The dominant role of theoretical uncertainties is highlighted both by the impact study (Figure 4.4) and by the breakdown of the expected likelihood scan (Figure 4.5), where groups of systematic uncertainties are progressively frozen. The two studies consistently indicate that theoretical sources, such as parton shower modelling, PDFs, and QCD scales, drive the overall sensitivity of the measurement.

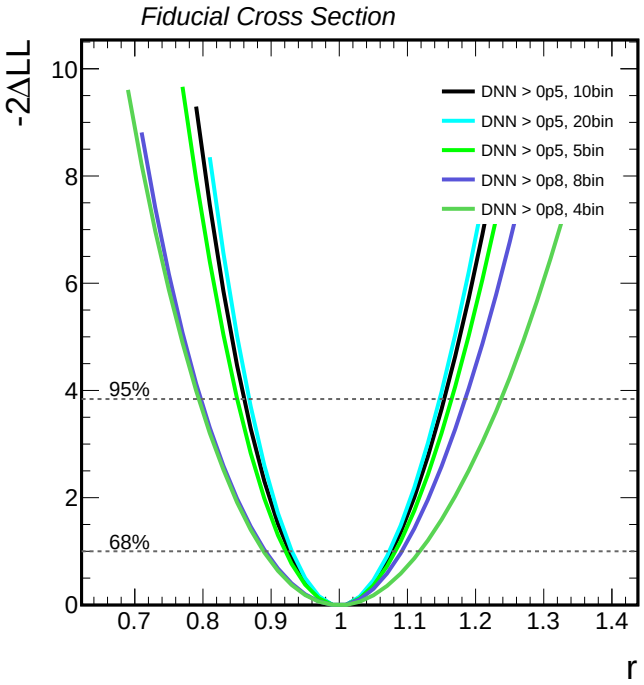


Figure 4.1: Expected likelihood scans for different choices of the DNN threshold and number of bins in the fit. The left panel shows the effect of varying the DNN cut between 0.5 and 0.8, illustrating that a lower threshold preserves more statistics while maintaining adequate purity. The right panel shows the impact of varying the number of bins, with the expected limits saturating beyond 10 bins, motivating the choice of a maximum of 10 bins for the final fit.

In the CRs, the same DNN threshold of 0.5 is applied. This ensures that CR events resemble SR events, allowing for consistent normalisation of the main backgrounds while not overconstraining the uncertainties. The template inputs used for the likelihood fit are shown in Figures 4.2 and 4.3 for the electron and muon channels, respectively.

The fiducial cross section is extracted via a simultaneous fit of the SR and CR, using the DNN output as the observable in SR and the number of events in the

various CRs. Backgrounds are constrained by the CRs, while the signal is extracted within the fiducial phase space. Systematic uncertainties affecting both the shape and the normalisation are included in the fit. All the uncertainties described in Section 3.7 are applied to this measurement as well. The theoretical uncertainties (PS, PDF, and QCD scale) are normalised in the fiducial region, and therefore account only for the acceptance component for the signal. The correlation scheme follows that described in Tables 3.13 and 3.14.

The expected fiducial cross section is:

$$\sigma_{\mathrm{EW}}^{\mathrm{fid}}(W{+}jj) = 1.80^{+0.045}_{-0.036}\,(\mathrm{exp})\,^{+0.104}_{-0.101}\,(\mathrm{theo})\,\pm 0.056\,(\mathrm{cal})\,\pm 0.038\,(\mathrm{syst})\,\pm 0.036\,(\mathrm{stat})\,\mathrm{pb}.$$

The uncertainty breakdown follows the same classification scheme as in the inclusive measurement (Section 3.8.1). Theoretical uncertainties remain the dominant contribution, primarily driven by QCD scale variations. Their overall impact, however, is reduced compared to the inclusive case, as the fiducial phase space is more restricted and closely matches the reconstructed selection, limiting the sensitivity of the result to modelling differences and extrapolations into poorly constrained kinematic regions.

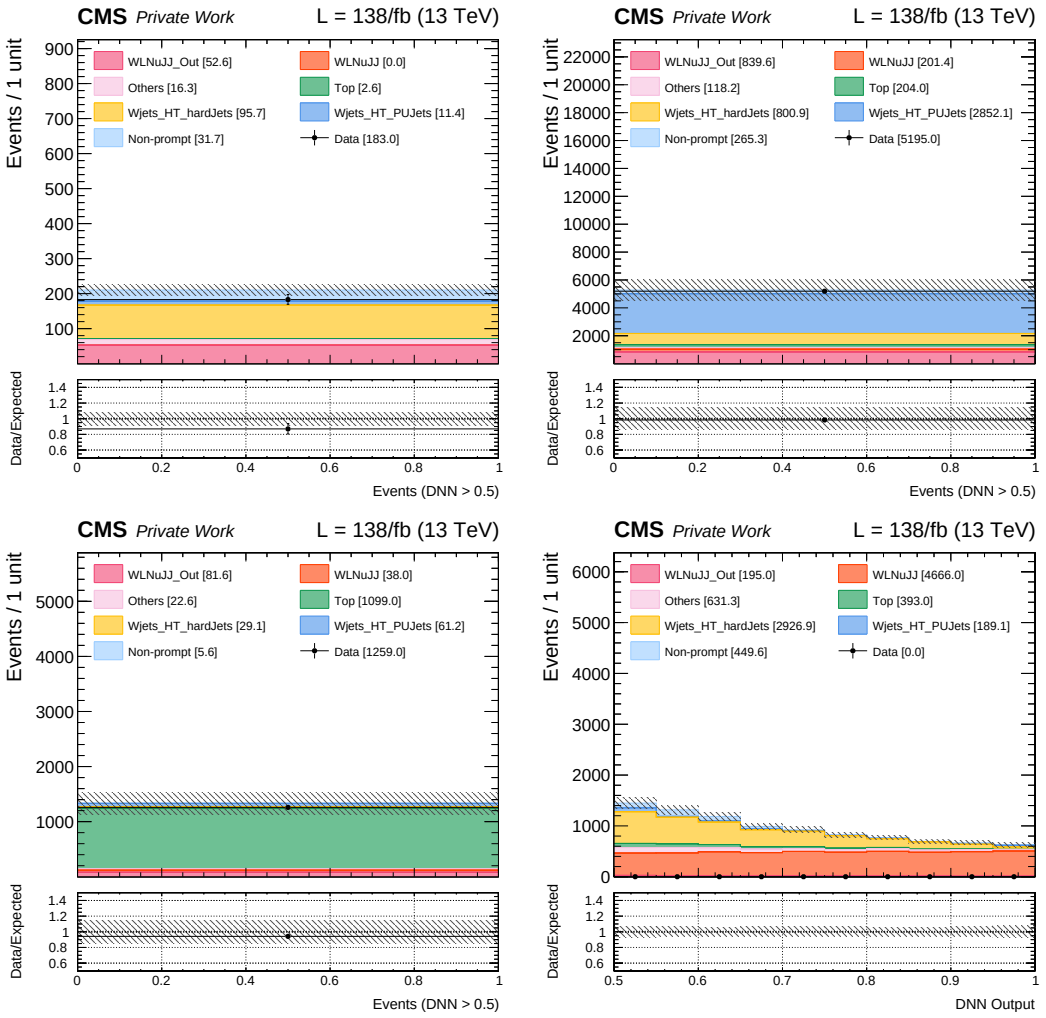


Figure 4.2: Template inputs for the likelihood fit in the electron channel, with DNN > 0.5. On top: W+jets control regions (hard and PU jets) left and right, respectively. Top CR is shown in the bottom left and SR (blind) with DNN binned in 10 bins in the bottom right.

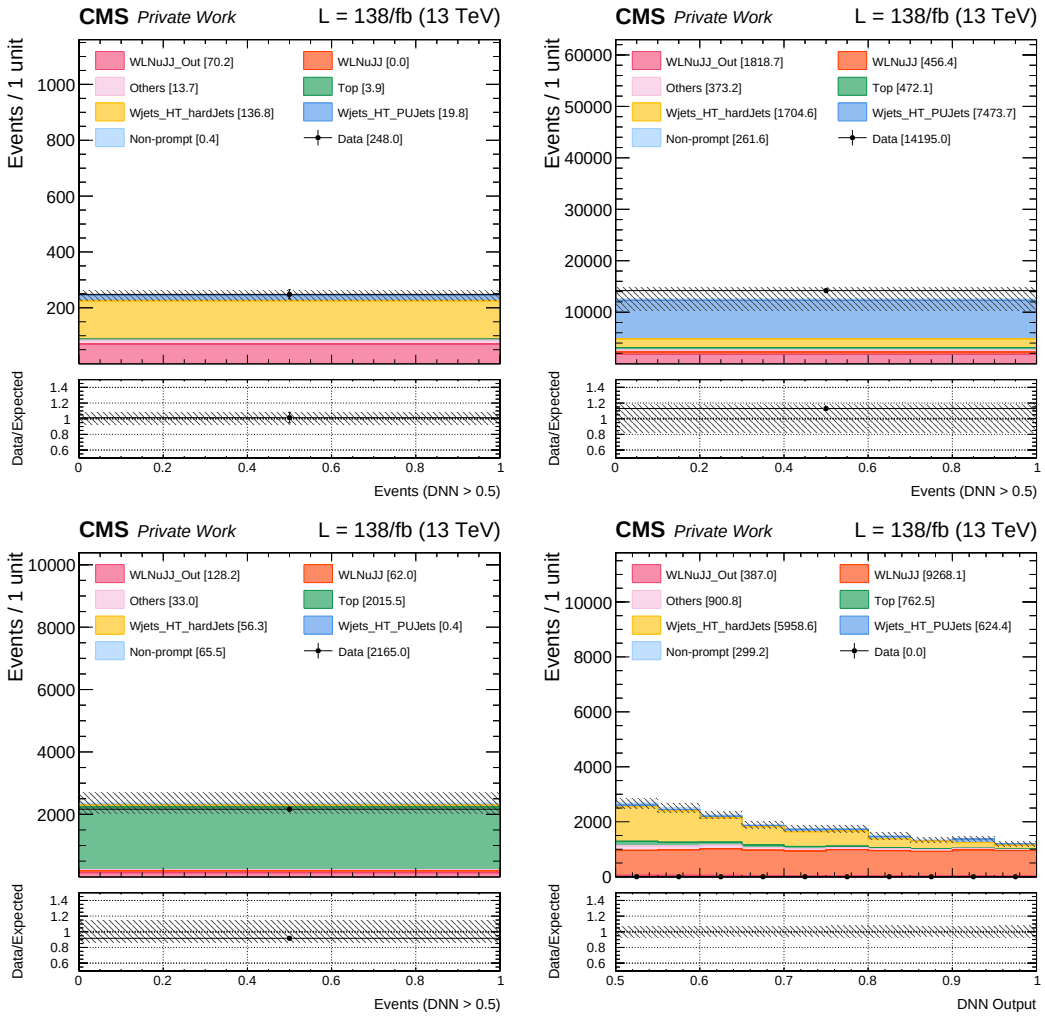


Figure 4.3: Template inputs for the likelihood fit in the muon channel, with DNN > 0.5. On top: W+jets control regions (hard and PU jets) left and right, respectively. Top CR is shown in the bottom left and SR (blind) with DNN binned in 10 bins in the bottom right.

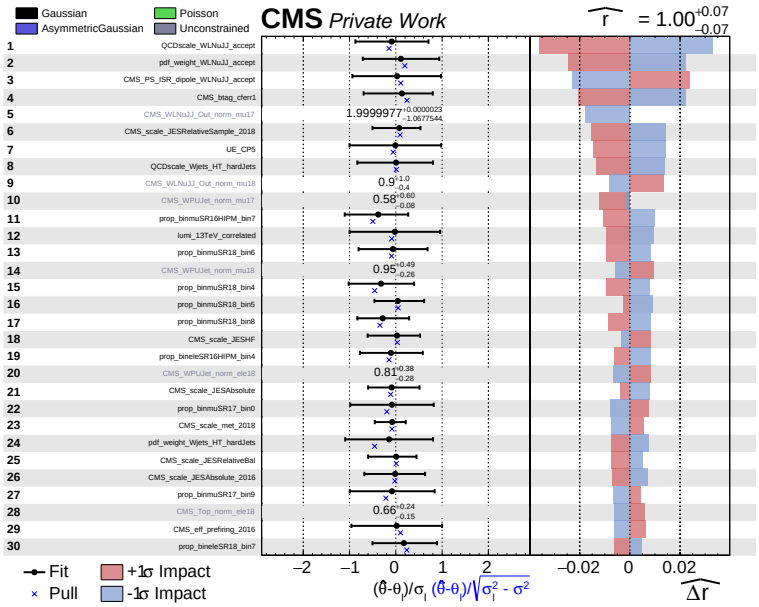


Figure 4.4: Post-fit expected impact of the main systematic uncertainties on the measured fiducial cross section. The dominant contributions arise from theoretical sources, such as PS modelling, PDF variations, and QCD scale uncertainties. A non-negligible contribution is also observed from the b-tagging uncertainty in c-quark jets, which can be understood considering that about 30% of the signal events contain a charm quark in the final state.

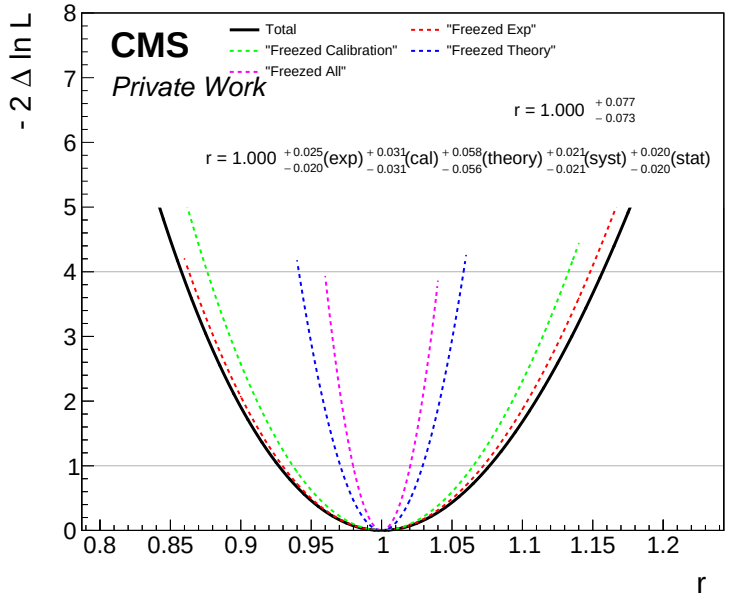


Figure 4.5: Breakdown of the expected likelihood scan obtained by progressively freezing groups of systematic uncertainties. The plot highlights that the dominant limitation to the sensitivity arises from theoretical sources, in agreement with the post-fit impact shown in Figure 4.4.

4.2 Differential Cross Section Measurement

In this section, the results of the unfolding analysis are presented. The method relies on the definition of generator-level bins used as independent signal templates in the fit. The binning was determined through an iterative procedure aimed at balancing resolution, statistical precision, and interpretability.

For each observable, the number of generator bins was chosen between five and ten, depending on the reconstruction-to-generator resolution ratio: observables with a smaller reco/gen ratio—i.e. better resolution—were assigned a finer binning. Once the number of bins was fixed, the bin edges were optimised to produce an approximately flat signal distribution in the SR. The resulting edges were then rounded to convenient values to ease the physical interpretation of the results.

After preliminary studies, it was observed that some adjacent bins exhibited strong correlations, indicating limited independent sensitivity. These bins were therefore merged to improve the overall stability of the fit. The final binning configuration adopted for each observable is summarised in Table 4.2.

Observable	**Generator-level bin edges**
$\Delta\eta_{jj}$	[4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.2, 5.5, 6.0, 10.0]
$\Delta\phi_{jj}$	[0.0, 2.0, 2.4, 2.6, 2.8, 2.9, 3.0, 3.14]
m_{jj}	[300.0, 850.0, 1000.0, 1200.0, 1400.0, 1600.0, 2000.0, 2500.0, 10000.0]
p_T^{jj}	[0.0, 30.0, 50.0, 75.0, 100.0, 130.0, 180.0, 500.0]
$p_T^{j_1}$	[60.0, 110.0, 140.0, 180.0, 230.0, 1000.0]
$p_T^{j_2}$	[60.0, 80.0, 100.0, 130.0, 170.0, 230.0, 1000.0]
N_{jets}	[1.5, 2.5, 3.5, 10.5]
p_T^{ℓ}	[0.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 100.0, 150.0, 1000.0]
p_T^{W}	[0.0, 30.0, 60.0, 90.0, 120.0, 180.0, 500.0]

Table 4.2: Generator-level binning configuration adopted for each observable used in the unfolding. The bin edges were optimised to ensure a flat signal distribution in the Signal Region while preserving sufficient statistical precision. The binning for m_{jj} corresponds to the multiclass implementation.

In the following, the unfolded results are presented separately for each observable. Except for m_{jj}, for which two independent strategies are compared, all other observables are unfolded using the same methodology and share an identical structure. This uniform presentation reflects the common analysis framework adopted for all variables and facilitates a direct comparison of the corresponding differential results.

4.2.1 Unfolding strategy

The unfolding procedure follows the same analysis strategy adopted for the fiducial measurement. The same event selections and control regions are used, with the classifier requirement DNN > 0.5 applied uniformly across all regions

to enhance the signal purity. The motivation for this choice has been discussed in Section 3.6.

In this configuration, the DNN output used previously in the fit is replaced by the reconstructed observable corresponding to the generator-level quantity to be unfolded, such as m_{jj}, $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, p_T^{jj}, p_T^W, or p_T^ℓ. This approach allows the fit to directly capture the detector response for each observable and extract the corresponding generator-level distribution. The reconstructed binning is defined with a granularity typically about three times finer than the generator-level binning, providing an accurate description of migration effects while maintaining sufficient statistical precision.

The fit is performed simultaneously in the SR and in the dedicated CRs, implemented as single-bin templates. This setup ensures that the main backgrounds are well constrained and that the unfolding remains sensitive to the observable of interest.

This strategy maintains full consistency with the fiducial measurement while enabling the extraction of differential results for all observables under study.

4.2.2 Alternative multiclass approach

As a complementary cross-check, an alternative unfolding strategy based on multiclass categorisation has been explored. In this approach, the classifier is trained to predict directly the generator-level bin of each event, rather than providing a single binary discriminator between signal and background. The SR is then partitioned into mutually exclusive categories, each enriched in events corresponding to a specific generator-level bin. The unfolding fit is subsequently performed simultaneously across these categories, allowing the analysis to exploit the additional information provided by the multiclass output.

This strategy has been tested on several observables. In general, the improvement with respect to the simpler threshold-based approach is limited, since most observables are only weakly correlated with the classifier score, and the increased complexity does not translate into a significant gain in sensitivity. The notable exception is m_{jj}, which shows a strong correlation with the binary classifier. In this case, the multiclass approach leads to a slight improvement in performance, particularly in the low-m_{jj} region where the background contamination is highest. For this reason, the multiclass strategy is retained only for the unfolding of m_{jj}, while for all other observables, the threshold-based approach remains the baseline, with multiclass results serving primarily as a validation of the method.

It is also worth noting that maintaining this strategy is considerably more demanding. First, a dedicated multiclass classifier must be trained for each observable under study, in contrast to the threshold-based approach, which can be applied more universally. In addition, every change in the generator-level binning requires retraining the corresponding classifier from scratch. This makes

the approach particularly resource-intensive and less practical for general use, especially in the context of multiple observables and iterative binning optimisations. For these reasons, and given that the observed improvements in sensitivity are only slight, the multiclass method is not considered worthy of systematic application, and is retained exclusively for the unfolding of m_{jj} where it provides a clear added value.

Multiclass DNN Architecture and Inputs

The multiclass classifier used in this analysis is designed to identify the most likely generator-level signal bin for each event, while also distinguishing signal from background. Compared to earlier binary classifiers, this network solves a more complex task and therefore uses a larger architecture and an extended set of input variables to improve separation power and capture subtle features in the data.

Input Variables The network takes as input a vector of event-level features, including high-level kinematic quantities, event shape observables, and object-level information sensitive to the underlying signal topology. The full list of variables is reported in Table 4.3. The network outputs a vector of probabilities p_i with $i = 0, \ldots, n$, where n is the number of gen-level signal bins and $i = 0$ corresponds to the background class. These outputs are mutually exclusive and collectively exhaustive, allowing a straightforward probabilistic interpretation.

Architecture and Training The network consists of an input layer matching the number of features, followed by several hidden layers with 128 and 64 neurons, each using ReLU activation. Dropout and batch normalization layers are applied after each hidden layer to improve regularization and stability. The output layer uses softmax activation:

$$p_i = \frac{e^{z_i}}{\sum_{j=0}^{n} e^{z_j}}, \tag{4.1}$$

where z_i are the pre-activation outputs.

The network is trained with the Adam optimizer (learning rate 0.001) using the sparse categorical cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N} \sum_{k=1}^{N} w_k \log p_{y_k}^{(k)}, \tag{4.2}$$

where w_k are per-event weights, adjusted such that the sum of weights in each signal class is equal and the events from different data-taking years are reweighted according to their integrated luminosity. Early stopping is applied with patience of 20 epochs, and the maximum number of epochs is 150 with a batch size of 1024. Accuracy is monitored on the validation set:

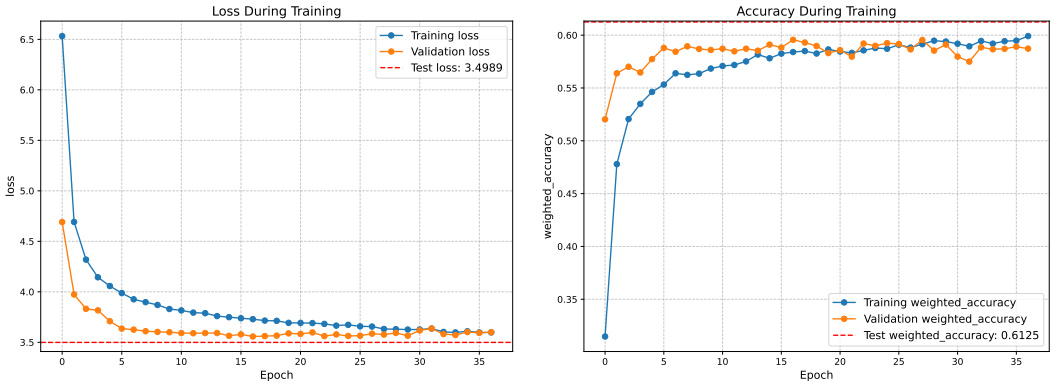


Figure 4.6: Evolution of the multiclass DNN performance during training. The left panel shows the loss function evaluated on the training and validation samples, demonstrating good convergence and absence of overfitting thanks to dropout and early stopping. The right panel displays the classification accuracy on the validation and test sets as a function of training epoch, reaching a stable plateau after approximately 80 epochs.The test sample shows slightly higher accuracy, as dropout is disabled during inference.

$$\text{Accuracy} = \frac{1}{N_{\text{val}}} \sum_{k=1}^{N_{\text{val}}} \mathbb{1}\{\hat{y}_k = y_k\}. \tag{4.3}$$

Interpretability with SHAP Feature importance is assessed using SHAP (SHapley Additive exPlanations). Both class-specific beeswarm plots and a global average over all classes are produced, providing a compact overview of the most influential features without biasing toward background or any individual signal class. The global SHAP summary is shown in Figure 4.7.

Event Categorisation and Fitting Strategy

To make use of the classifier output in the unfolding measurement, each event in the SR is assigned to a unique class based on the argmax of its output vector. Since the multiclass classifier produces, for each event, a vector of n+1 values, each representing the probability that the event belongs to one of the n signal generator-level bins or to the background, the entry with the highest probability is interpreted as the predicted class. In other words, the classifier "decides" that the event most likely belongs to the class corresponding to the index of the largest output value.

This procedure naturally partitions the SR into n+1 mutually exclusive categories, each corresponding to a specific gen-level signal bin (for indices 1 to n) or to the background (index 0). Notably, the subset of events for which the background class (index 0) has the highest score defines a data-driven control region, referred to as "class 0." Since these events originate from the same SR selection

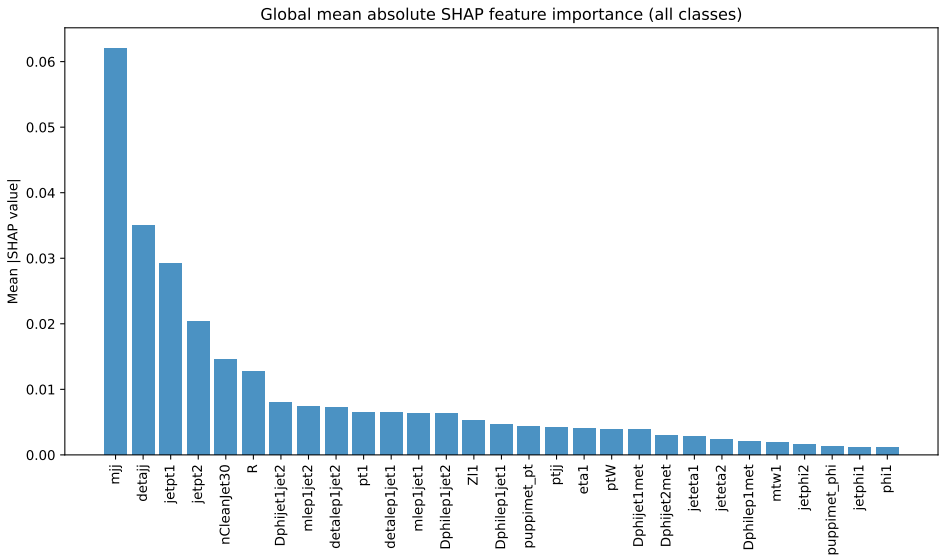


Figure 4.7: Global SHAP feature importance averaged over all signal and background classes. The y-axis represents the mean absolute SHAP value for each feature, $\langle|\text{SHAP}|\rangle$, which quantifies the average contribution of that feature to the classifier output across all events. Features with higher values have a stronger influence on the network's decision-making process, while lower values indicate less impact. This plot provides a compact summary of which input variables drive the multiclass DNN predictions.

and are classified purely based on the learned kinematic patterns, this control region is kinematically very close to the signal-enriched categories, making it a particularly useful handle for background validation and modelling.

Within each class, we perform a template fit using a discriminating variable derived from the classifier output. Specifically, for a given signal class i, the fit variable is defined as:

$$\frac{p_i}{p_i + p_0}$$

where p_i is the classifier-assigned probability for the signal class i, and p_0 is the background probability. This quantity is then flattened using signal quintiles to improve fit performance and sensitivity. The use of this transformed variable allows for improved separation from background while preserving class-specific signal features. Figure 4.9 shows the distributions of the transformed classifier variable $p_i/(p_i + p_0)$ for each signal class, before rebinning into signal quintiles, illustrating the separation between signal and background within each class.

Alternative definitions for the fit variable, including raw classifier probabilities or ratios involving neighbouring class probabilities, have been explored. However, no single option demonstrated consistently superior performance across all scenarios. The selected form represents a balanced choice in terms of interpretability, stability, and discriminating power.

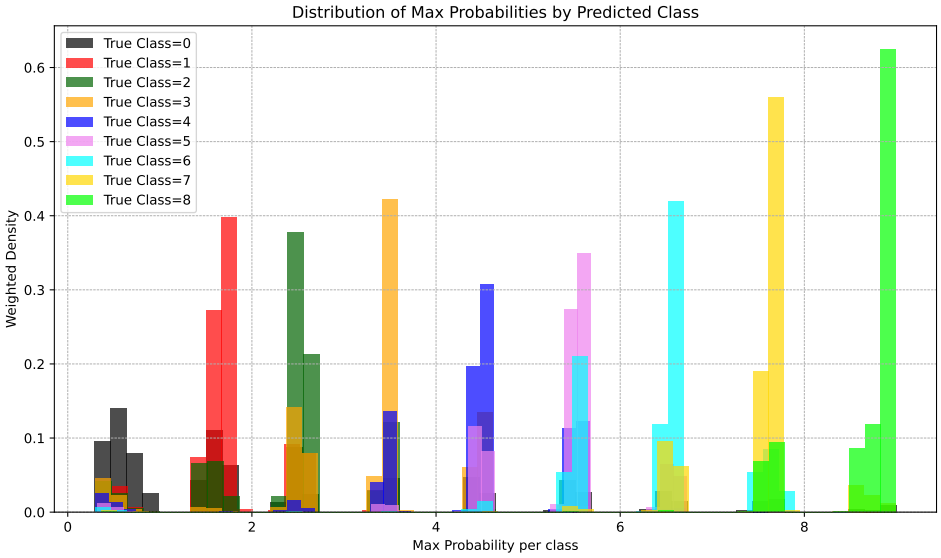


Figure 4.8: Distribution of the classifier output probability P(i) for each predicted class i, shown on the x-axis. Each bin (labelled "class 0", "class 1", ..., "class n") includes only events assigned to that class via the argmax of the classifier output. Within each bin, the histogram shows the normalised distribution of P(i), i.e., the classifier's confidence for the predicted class. As expected, class 0 is dominated by background events, while classes 1 to n are primarily populated by signal events from the corresponding gen-level bin. Note that all histograms are normalised independently and do not reflect the relative population sizes of each class.

The overall fitting strategy proceeds as follows. Each of the n signal classes (1 to n), effectively subregions of the SR, is fitted using the transformed variable described above, with the number of bins optimised per class to ensure good sensitivity while maintaining fit stability. To allow a consistent comparison with the standard unfolding strategy, the transformed variable is divided into three bins per class, corresponding to the typical granularity used in the reco-level fits, where three reconstructed bins are used per generator-level bin. For the top and W+1 PU jet control regions, class 0, the background-enriched category defined by the multiclass classifier, is included as a single-bin template. Since class 0 corresponds to events with a low classifier score, it is effectively background-dominated, making it a suitable template for constraining these control regions. In addition, for events assigned to class 0, the sum of all signal probabilities, $\mathrm{Sum}(P_{\mathrm{signal}})$, is required to be greater than 0.6 to select events that are more signal-like even within this background-dominated category. This additional selection effectively emulates the DNN > 0.5 requirement applied in the control regions of the standard strategy, ensuring a consistent definition of signal-like events across both approaches. This configuration ensures a coherent and data-driven strategy across all phases of the fit, while preserving a comparable level of binning granularity with respect to the standard unfolding approach.

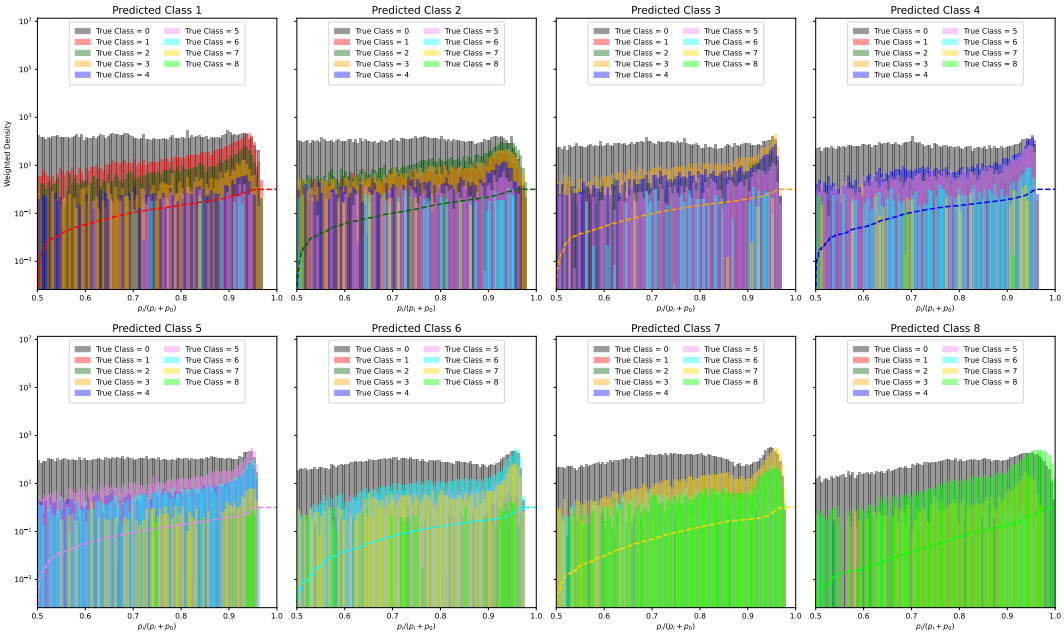


Figure 4.9: Distributions of the transformed classifier variable $p_i/(p_i + p_0)$ for each signal class, shown before rebinning into signal quintiles. Each histogram illustrates how the classifier separates signal events from background within the corresponding class, providing the input for the subsequent template fit. Histograms are weighted according to their physical weight and log scale is shown on y-axis.

Variable	Description
$p_T(l)$	Transverse momentum of the lepton
$\eta(l)$	Pseudorapidity of the lepton
$\phi(l)$	Azimuthal angle of the lepton
$p_T(j_1)$	Transverse momentum of the leading jet
$\eta(j_1)$	Pseudorapidity of the leading jet
$\phi(j_1)$	Azimuthal angle of the leading jet
$p_T(j_2)$	Transverse momentum of the subleading jet
$\eta(j_1)$	Pseudorapidity of the subleading jet
$\phi(j_2)$	Azimuthal angle of the subleading jet
MET	Magnitude of the missing transverse momentum (PUPPI)
MET_ϕ	Azimuthal angle of the missing transverse momentum (PUPPI)
$p_T(W)$	Transverse momentum of the W boson (lepton + p_T^{miss})
$M_T(W)$	Transverse mass of the W boson
m_{jj}	Invariant mass of the two leading jets
$\Delta\eta_{jj}$	Pseudorapidity separation between the two leading jets
$p_T(jj)$	Transverse momentum of the dijet system
$\Delta\phi(j_1, MET)$	Azimuthal separation between leading jet and MET
$\Delta\phi(j_2, MET)$	Azimuthal separation between subleading jet and MET
$\Delta\phi(j_1, j_2)$	Azimuthal separation between leading and subleading jets
$\Delta\phi(\ell, j_1)$	Azimuthal separation between lepton and leading jet
$\Delta\phi(\ell, j_2)$	Azimuthal separation between lepton and subleading jet
$\Delta\phi(\ell, MET)$	Azimuthal separation between lepton and MET
$\Delta\eta(\ell, j_1)$	Pseudorapidity separation between lepton and leading jet
$\Delta\eta(\ell, j_2)$	Pseudorapidity separation between lepton and subleading jet
$m(\ell, j_1)$	Invariant mass of the lepton and the leading jet
$m(\ell, j_2)$	Invariant mass of the lepton and the subleading jet
R	Centrality ratio: $\frac{\lvert\vec{p_T}(jj)+\vec{p_T}(W)\rvert}{\lvert\vec{p_T}(j_1)\rvert+\lvert\vec{p_T}(j_2)\rvert+\lvert\vec{p_T}(W)\rvert}$
Z_l	Centrality of the lepton: $\eta_l - \frac{1}{2}(\eta_{j_1} + \eta_{j_2})$
n_{Jet}	Number of reco jets in the event (with $p_T > 30$ GeV)

Table 4.3: Input variables used for the multiclass DNN training.

4.2.3 Unfolded m_{jj} distribution

MultiClass strategy

Figures 4.10 and 4.11 show the distributions used to build the templates that are ultimately used in the fit, illustrating the overall strategy.

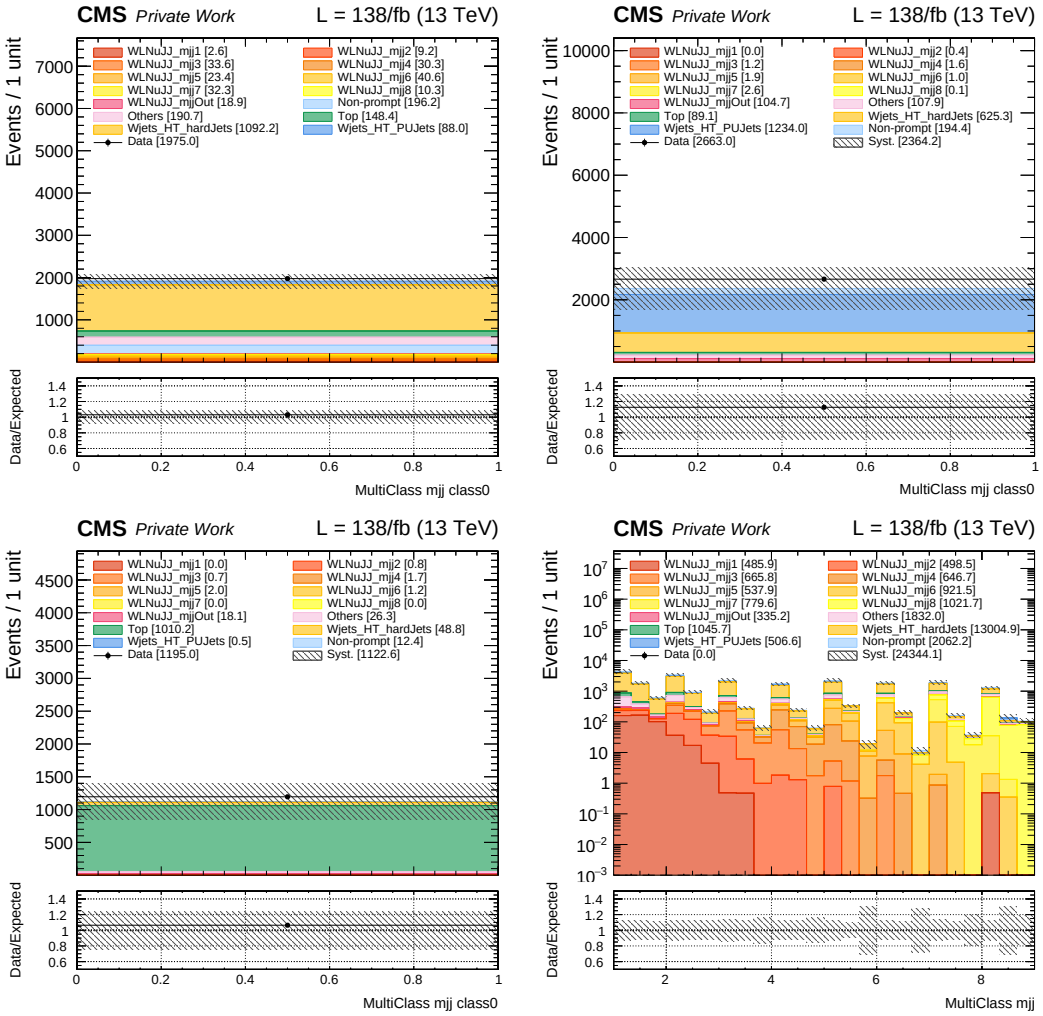


Figure 4.10: Multiclass strategy for the m_{jj} distribution in the electron channel. On top: W+jets control regions (hard and PU jets) left and right, respectively. Top CR is shown in the bottom left, and SR (blind) with multiclass distribution divided into three bins per class in the bottom right. The Hard W+jets sample is controlled in the SR class0.

Reco variable-based strategy

Figures 4.12 and 4.13 show the distributions used to build the templates that are ultimately used in the fit, illustrating the overall strategy.

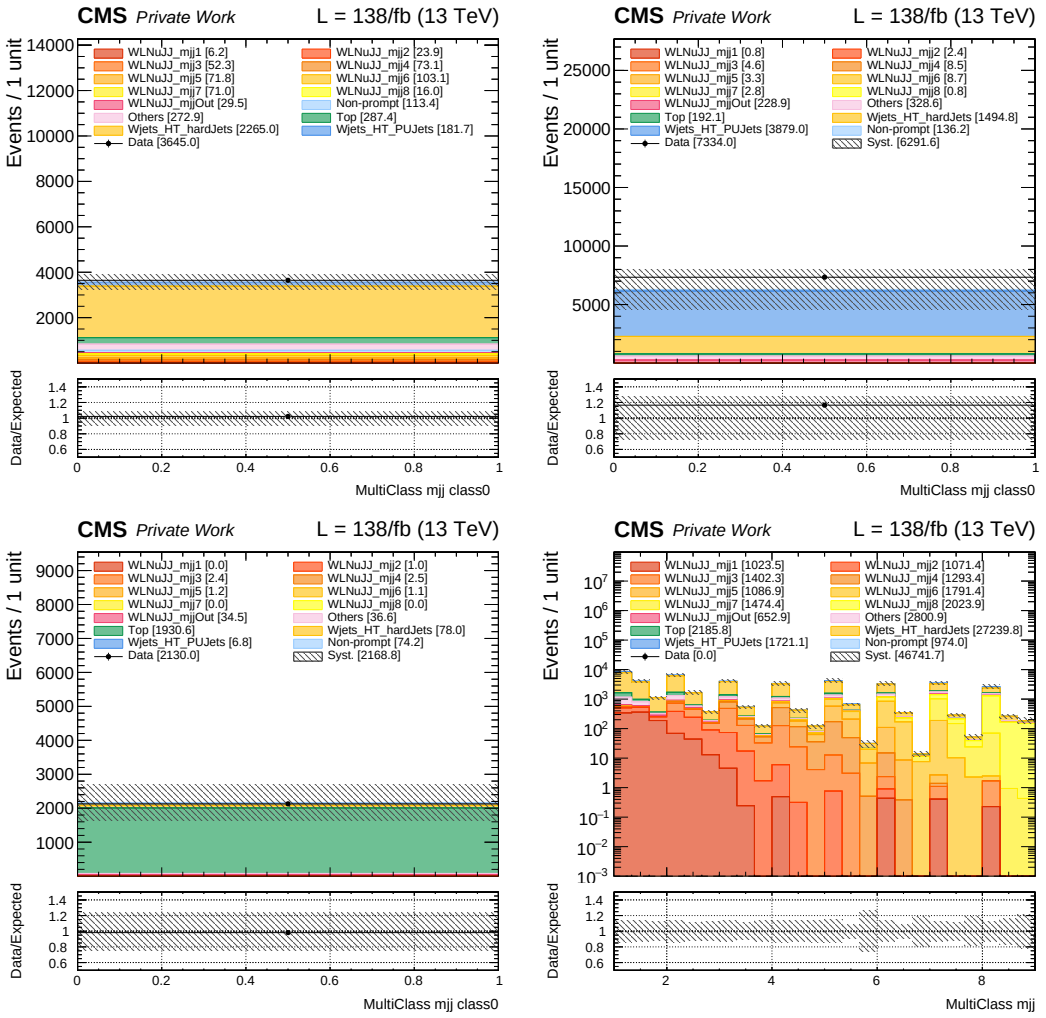


Figure 4.11: Multiclass strategy for the m_{jj} distribution in the muon channel. On top: W+jets control regions (hard and PU jets) left and right, respectively. Top CR is shown in the bottom left, and SR (blind) with multiclass distribution divided into three bins per class in the bottom right. The Hard W+jets sample is controlled in the SR class0

Results

The results of the measurement, presented as a function of the dijet invariant mass, are outlined in this section. All one-dimensional likelihood scans correspond to *profiled fits*, in which a single parameter of interest (POI) is scanned while all remaining POIs and nuisance parameters are left free to vary. This procedure incorporates correlations among bins and propagates the impact of shared systematic uncertainties. A global comparison of the one-dimensional likelihood scans obtained in each m_{jj} bin is shown in Figure 4.14, for both the reco-variable and the Multiclass discriminants. This summary plot illustrates the sensitivity achieved with the two approaches and their mutual consistency. The Multiclass strategy exhibits higher sensitivity in the low-score region while maintaining comparable performance elsewhere.

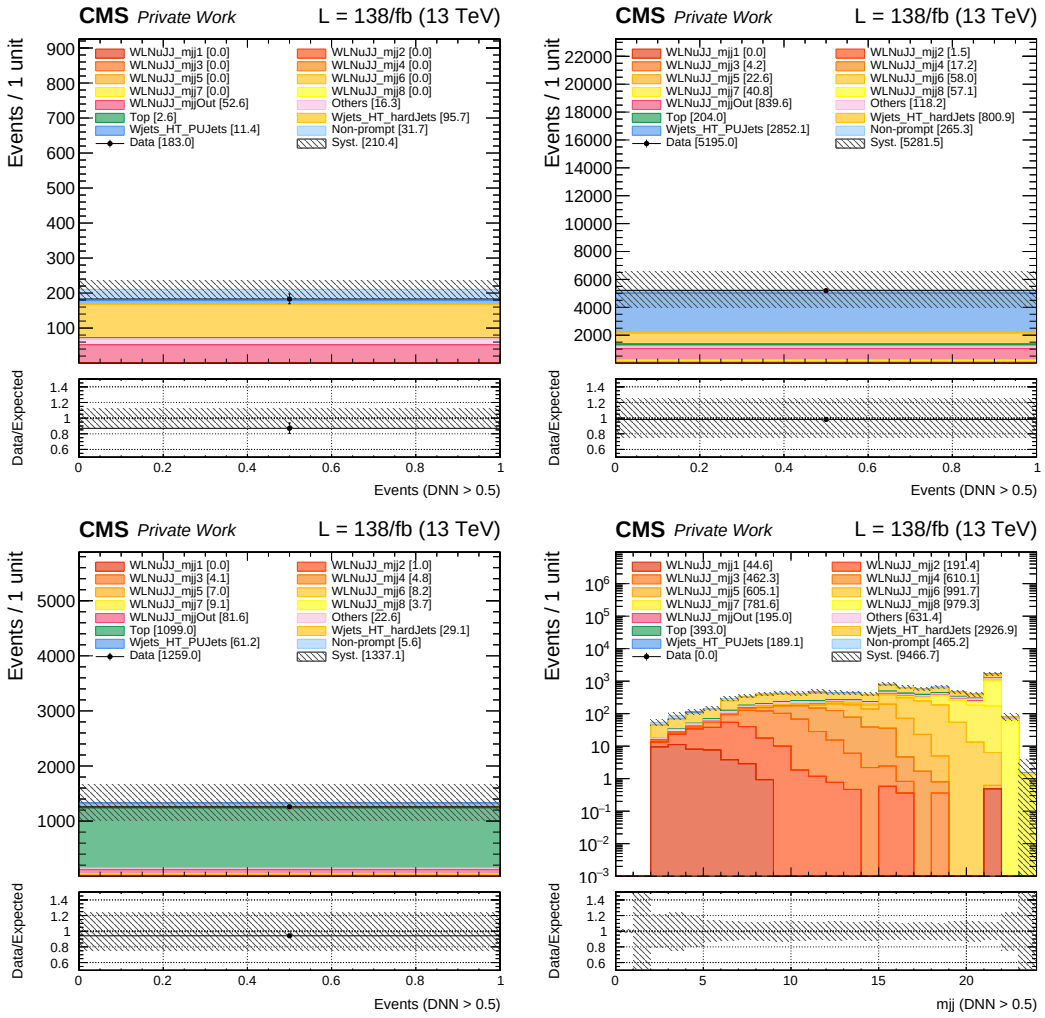


Figure 4.12: Reco variable-based strategy for the m_{jj} distribution in the electron channel, with DNN > 0.5. On top: W+jets control regions (hard and PU jets) left and right, respectively. Top CR is shown in the bottom left, and SR (blind) with the reco variable distribution binned with three bin per generator bin is shown bottom right.

To further investigate intrinsic correlations among the POIs, 2D likelihood scans are produced with all systematic uncertainties and remaining POIs fixed at their nominal values. Unlike the one-dimensional scans, these are not profiled and simply illustrate the expected correlation pattern between pairs of POIs as determined by the likelihood structure. Figure 4.15 shows the resulting correlation coefficients, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties is evaluated using a data-Asimov dataset (see Section 1.9), in which pseudo-data are generated from the post-fit model while preserving the best-fit nuisance parameters. Dedicated post-fit impact plots, produced for all bins and for both strategies, are reported in Appendix D.1. The most significant systematic effects originate from neighboring POIs, reflecting intrinsic correlations among adjacent m_{jj} bins. These correlations arise

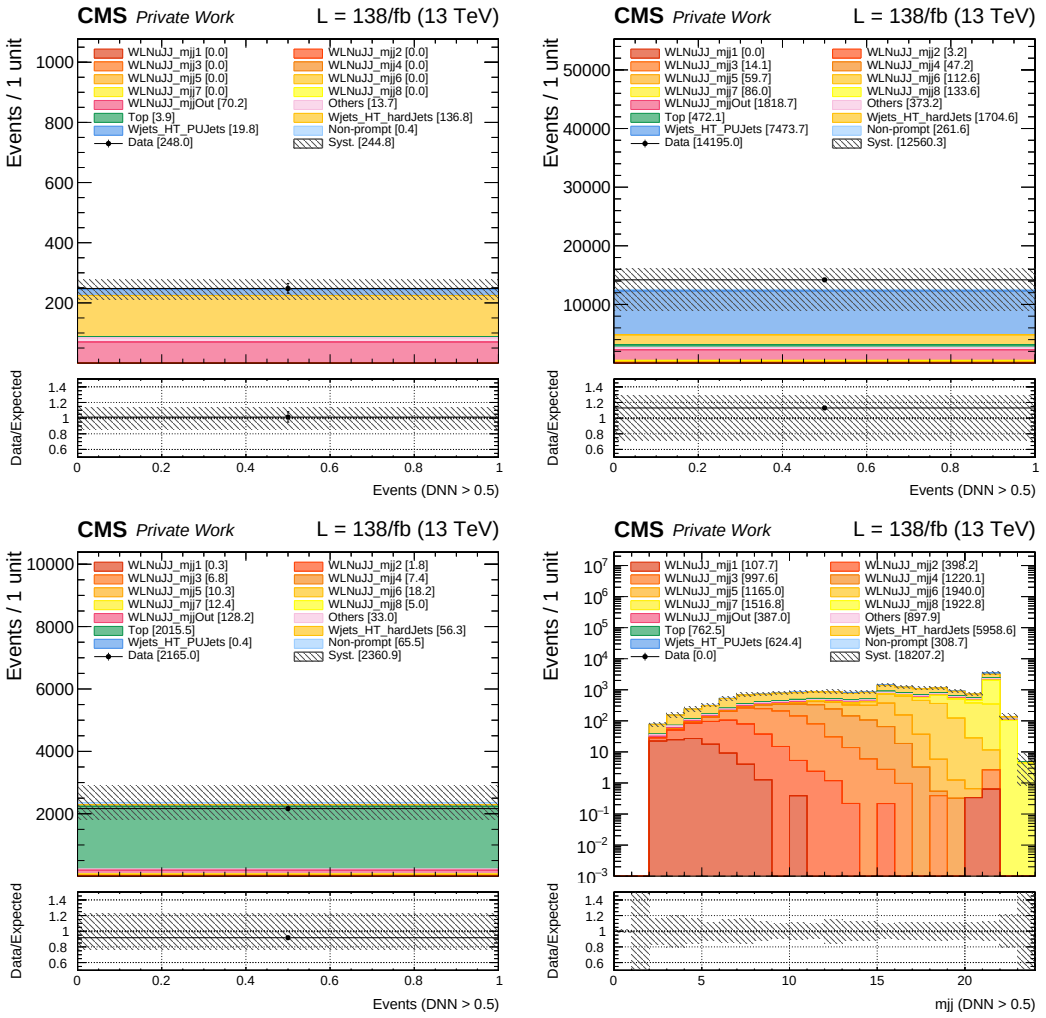


Figure 4.13: Reco variable-based strategy for the m_{jj} distribution in the muon channel, with DNN > 0.5. On top: W+jets control regions (hard and PU jets) left and right, respectively. Top CR is shown in the bottom left, and SR (blind) with the reco variable distribution binned with three bin per generator bin is shown bottom right.

mainly from the limited dijet-mass resolution, which causes migrations between nearby generator-level bins and partially overlapping reconstructed distributions.

Finally, the expected unfolded distribution is shown in Figure 4.16. The measurements are presented together with total uncertainties and compared to SM predictions. Multiple theoretical calculations are overlaid for comparison, including SHERPA (LO, multi-jet merged) and POWHEG (NLO) interfaced with PYTHIA8, allowing direct assessment of the agreement between data and theory across different perturbative orders and parton-shower models.

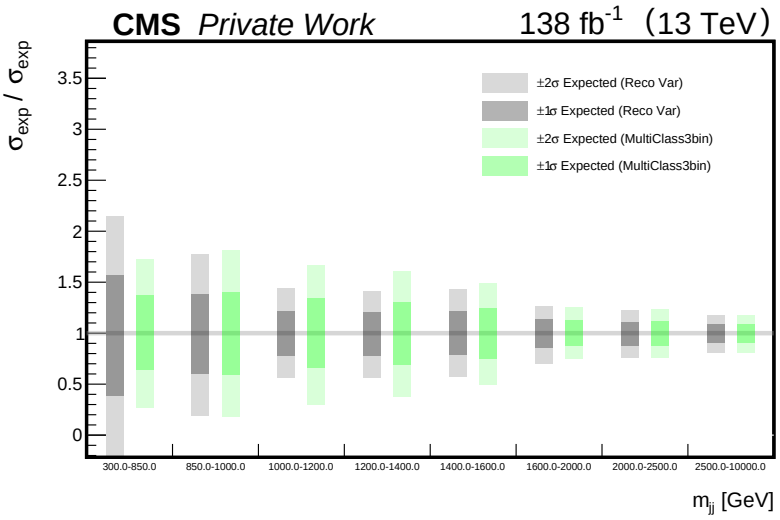


Figure 4.14: Expected profile likelihood scans for the measurement in the m_{jj} bins, comparing the reco variable-based method (black) and the Multiclass approach (green). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity and the consistency between the two methods.

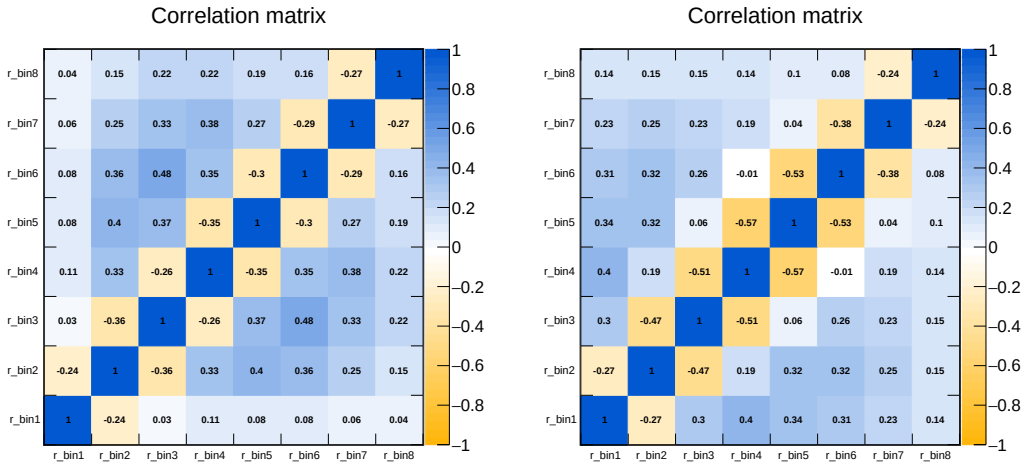


Figure 4.15: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. The left panel shows the correlations obtained with the reco variable-based method, while the right panel corresponds to the multiclass approach. Only correlations at 1σ deviations are shown, illustrating the intrinsic correlations induced by the fit construction.

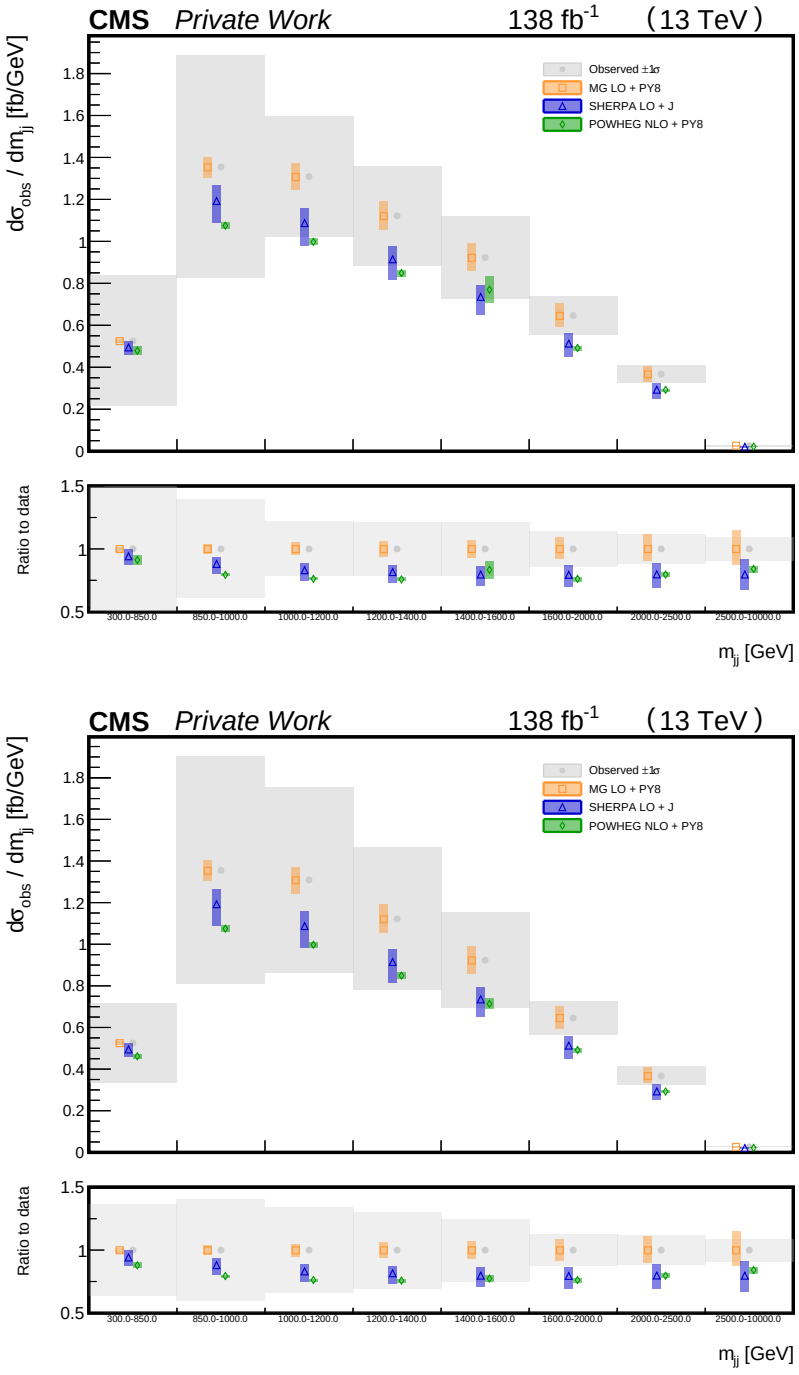


Figure 4.16: Expected unfolded cross sections in the m_{jj} bins, with total uncertainties, compared to SM predictions. The upper panel shows the results obtained using the reconstructed-variable method, while the lower panel presents those from the multiclass neural network unfolding. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged). The two methods are found to be in good agreement within the quoted uncertainties.

4.2.4 Unfolded $p_T(jj)$ distribution

Figures 4.17 and 4.18 show the distributions used to build the templates that are ultimately used in the fit, illustrating the overall strategy.

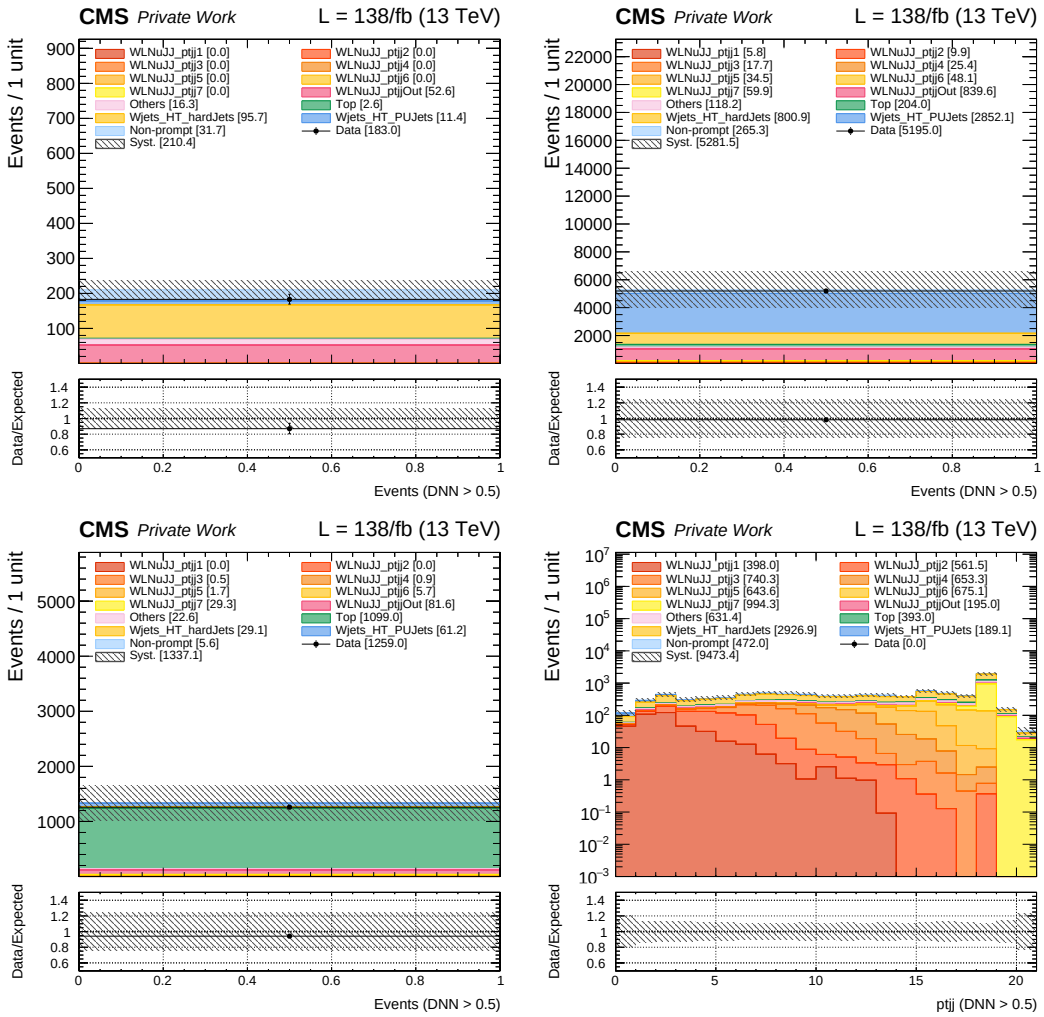


Figure 4.17: Distribution of $p_T(jj)$ in the electron channel, with DNN > 0.5. On top, the W+jets control regions with hard and PU jets are shown on the left and right, respectively. The bottom-left panel shows the top control region, while the SR (blind) on the bottom right displays the reco variable distribution binned with three reconstructed bins per generator-level bin.

Results

The results of the measurement, presented as a function of the dijet transverse momentum $p_T(jj)$, are outlined in this section. All one-dimensional likelihood scans correspond to profiled fits, in which a single parameter of interest (POI) is scanned while all remaining POIs and nuisance parameters are left free to vary in the fit. This procedure incorporates correlations among bins and propagates the impact of shared systematic uncertainties. Figure 4.19 provides an overview

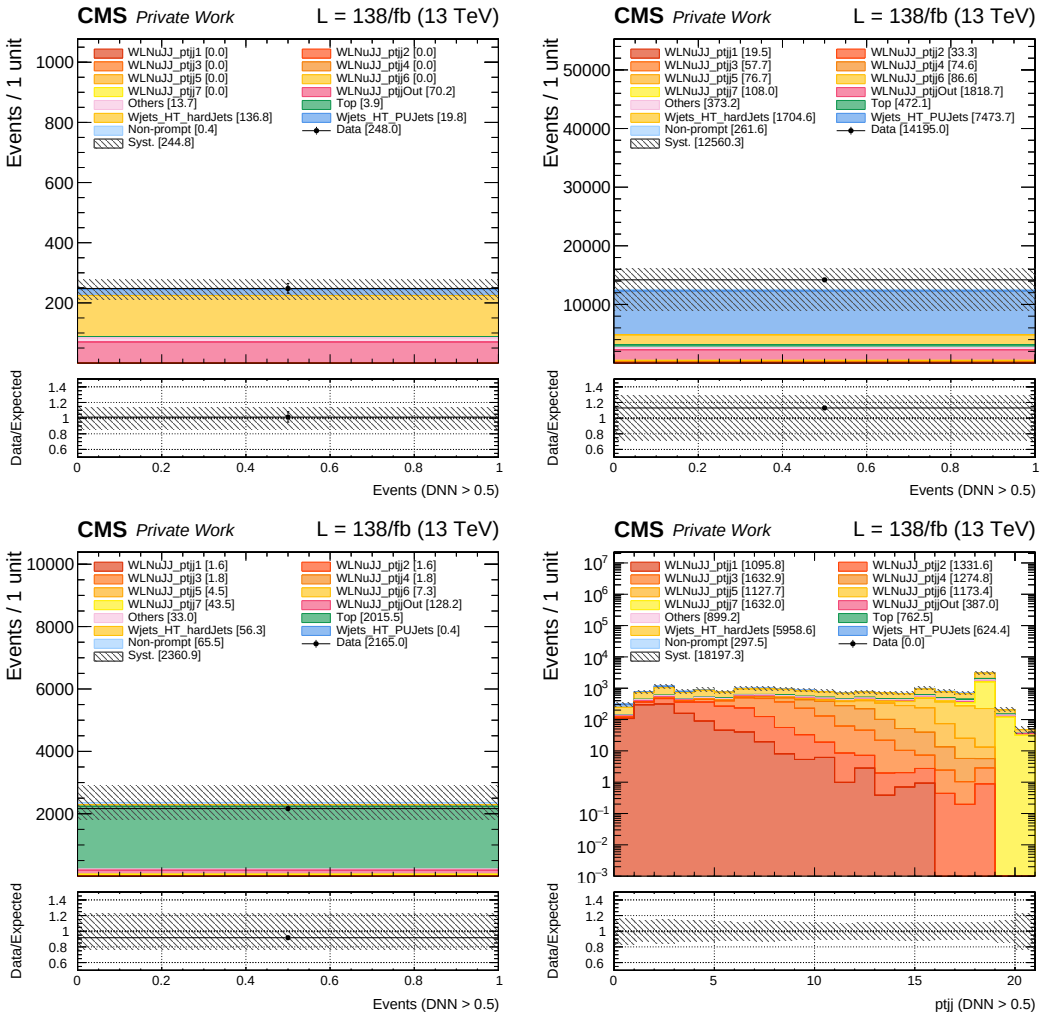


Figure 4.18: Distribution of $p_T(jj)$ in the muon channel, with DNN > 0.5. On top, the W+jets control regions with hard and PU jets are shown on the left and right, respectively. The bottom-left panel shows the top control region, while the SR (blind) on the bottom right displays the reco variable distribution binned with three reconstructed bins per generator-level bin.

of the resulting likelihood scans in each $p_T(jj)$ bin, obtained from fits performed using the reconstructed variable. This summary plot illustrates the sensitivity achieved and the overall consistency of the measurement across bins.

To further investigate the intrinsic correlations among the parameters of interest, two-dimensional likelihood scans are produced using Monte Carlo simulations only, with all systematic uncertainties and remaining POIs fixed at their nominal values. Unlike the one-dimensional scans, these 2D scans are not profiled and are used to illustrate the expected correlation pattern between pairs of POIs as determined by the structure of the likelihood model. Figure 4.20 shows the resulting correlation coefficients, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties is evaluated using a data-Asimov dataset (see Section 1.9), in which pseudo-data are generated from the post-fit model

while preserving the best-fit values of the nuisance parameters. Dedicated post-fit impact plots, produced for all $p_T(jj)$ bins, are reported in Appendix D.2. The most significant systematic effects originate from neighboring POIs, reflecting the intrinsic correlations among adjacent bins caused by the limited dijet-momentum resolution.

Finally, the expected unfolded distribution is shown in Figure 4.21. The unfolded measurements are presented together with their total uncertainties and compared to SM predictions. Multiple theoretical calculations are overlaid for comparison, including SHERPA (LO, multi-jet merged) and POWHEG (NLO) interfaced with PYTHIA8. Theoretical uncertainties are shown as bands around the predictions, allowing a direct assessment of the agreement between the measured spectrum and different perturbative orders and parton-shower models.

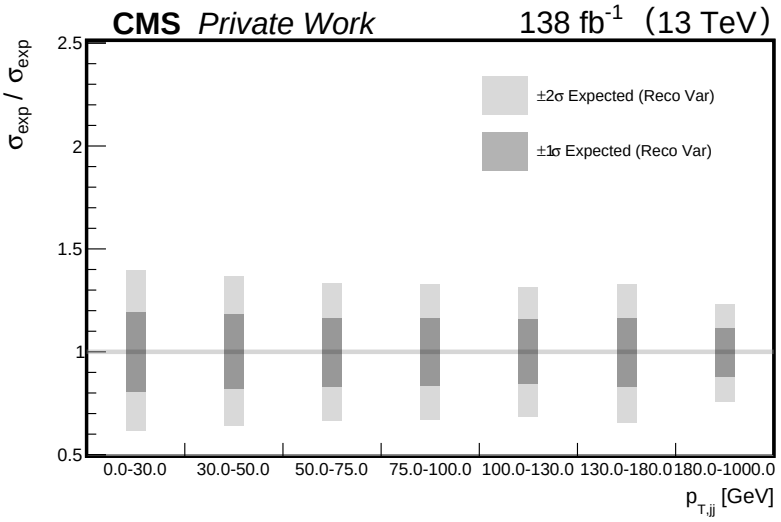


Figure 4.19: Expected profile likelihood scans for the measurement in the $p_T(jj)$ bins using the reco variable-based discriminant (black). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity achieved with the reco variable method.

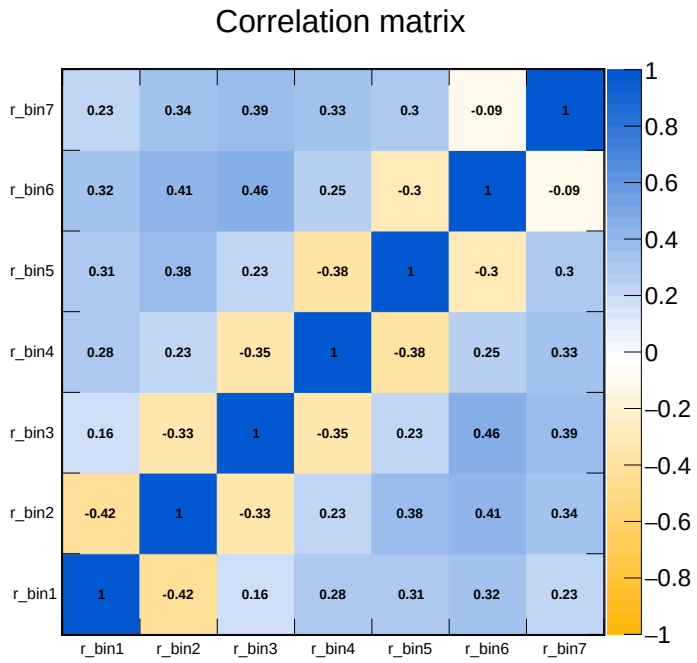


Figure 4.20: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. Only the correlations at 1σ deviations are shown, highlighting the intrinsic correlations induced by the fit construction.

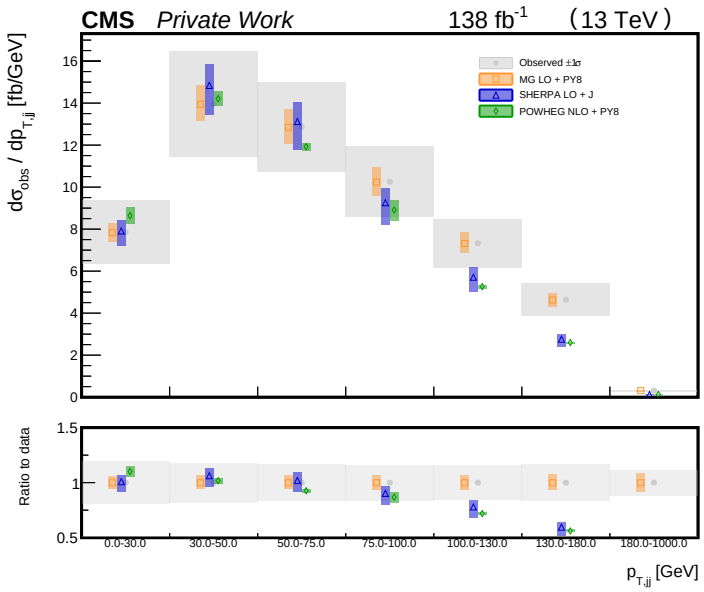


Figure 4.21: Expected unfolded cross sections in the $p_T(jj)$ bins, shown with total uncertainties and compared to SM predictions. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged).

4.2.5 Unfolded $\Delta\eta_{jj}$ distribution

Figure 4.22 shows the distributions used to construct the SR templates that enter the fit. Since the control regions remain identical to those introduced in Section 4.2.4, only the SR is shown, to allow the differences in signal composition and the correlation among signal components to be clearly seen.

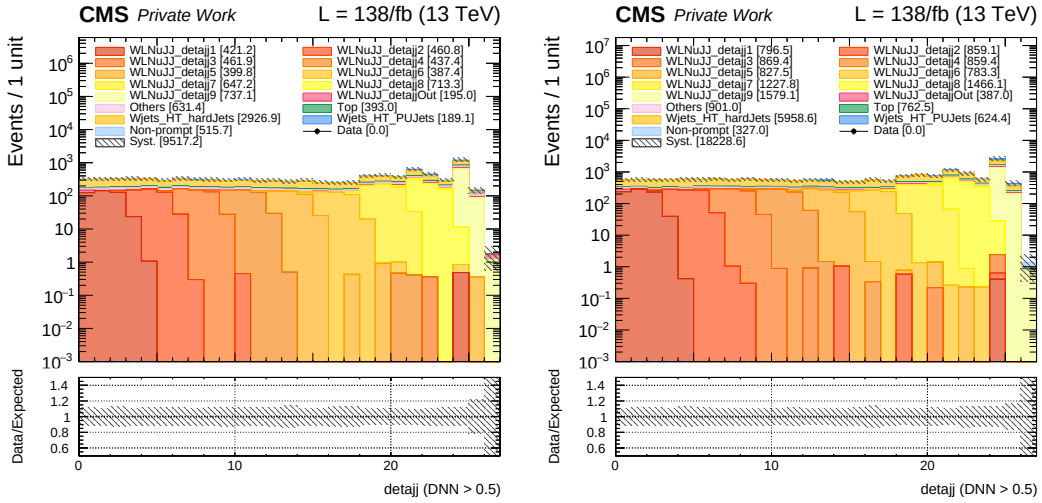


Figure 4.22: Distribution of $\Delta\eta_{jj}$ in the signal region for the electron (left) and muon (right) channels, with DNN > 0.5. Control regions are unchanged with respect to those defined in Section 3.4.5 and are omitted for brevity. The plots show the reconstructed $\Delta\eta_{jj}$ distribution binned with three reconstructed bins per generator-level bin, which serve as inputs to the unfolded fit.

Results

The results of the measurement, presented as a function of the dijet pseudorapidity separation $\Delta\eta_{jj}$, are outlined in this section. All one-dimensional likelihood scans correspond to *profiled fits*, in which a single parameter of interest (POI) is scanned while all remaining POIs and nuisance parameters are left free to vary in the fit. This procedure incorporates correlations among bins and propagates the impact of shared systematic uncertainties. Figure 4.23 provides an overview of the resulting likelihood scans in each $\Delta\eta_{jj}$ bin, obtained from fits performed using the reconstructed variable. This summary plot illustrates the sensitivity achieved and the overall consistency of the measurement across bins.

To further investigate the intrinsic correlations among the parameters of interest, two-dimensional likelihood scans are produced using Monte Carlo simulations only, with all systematic uncertainties and remaining POIs fixed at their nominal values. Unlike the one-dimensional scans, these 2D scans are not profiled and are used to illustrate the expected correlation pattern between pairs of POIs as determined by the structure of the likelihood. Figure 4.24 shows the resulting correlation coefficients, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties is evaluated using a data-Asimov dataset (see Section 1.9), in which pseudo-data are generated from the post-fit model while preserving the best-fit values of the nuisance parameters. Dedicated post-fit impact plots, produced for all $\Delta\eta_{jj}$ bins, are reported in Appendix D.3. The most significant systematic effects originate from neighboring POIs, reflecting intrinsic correlations among adjacent $\Delta\eta_{jj}$ bins.

Finally, the expected unfolded distribution is shown in Figure 4.25. The unfolded measurements are presented together with their total uncertainties and compared to SM predictions. Multiple theoretical calculations, including SHERPA (LO, multi-jet merged) and POWHEG (NLO) interfaced with PYTHIA8, are overlaid together with their theoretical uncertainties, allowing a direct comparison between data and predictions based on different perturbative orders and parton-shower models.

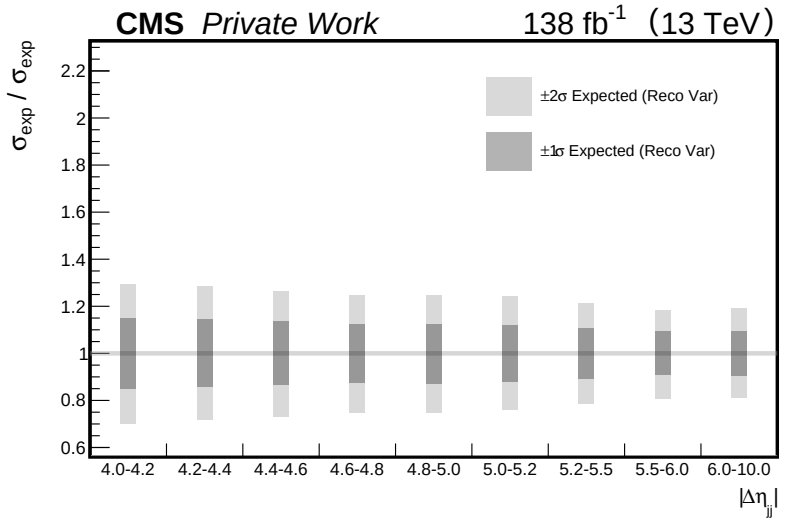


Figure 4.23: Expected profile likelihood scans for the measurement in the $\Delta\eta_{jj}$ bins using the reco variable-based discriminant (black). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity achieved with the reco variable method.

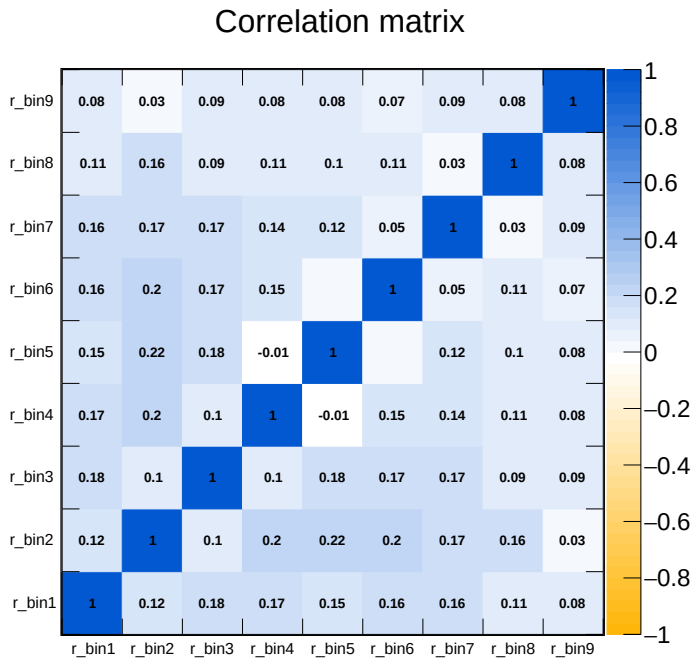


Figure 4.24: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. Only the correlations at 1σ deviations are shown, highlighting the intrinsic correlations induced by the fit construction.

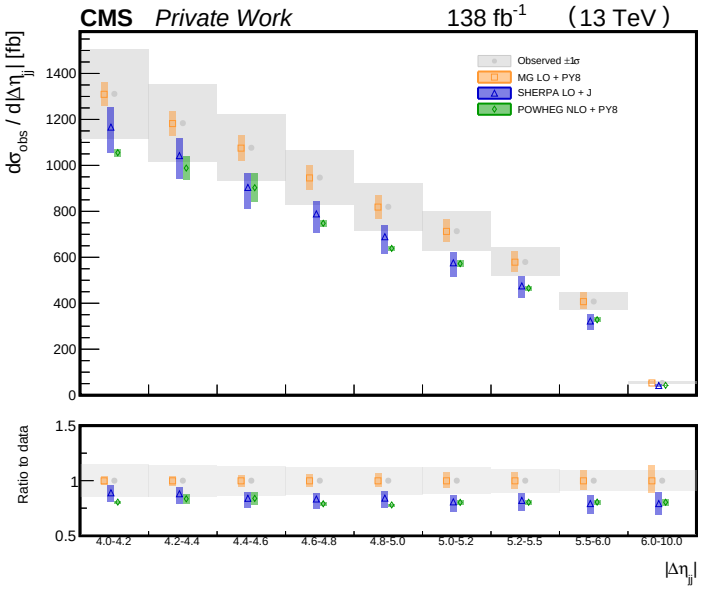


Figure 4.25: Expected unfolded cross sections in the $\Delta\eta_{jj}$ bins, shown with total uncertainties and compared to SM predictions. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged).

4.2.6 Unfolded $\Delta\phi_{jj}$ distribution

Figure 4.26 shows the distributions used to construct the SR templates entering the fit. As the control regions are identical to those introduced in Section 4.2.4, only the signal region is shown here, allowing the differences in signal composition and the correlation among the signal components.

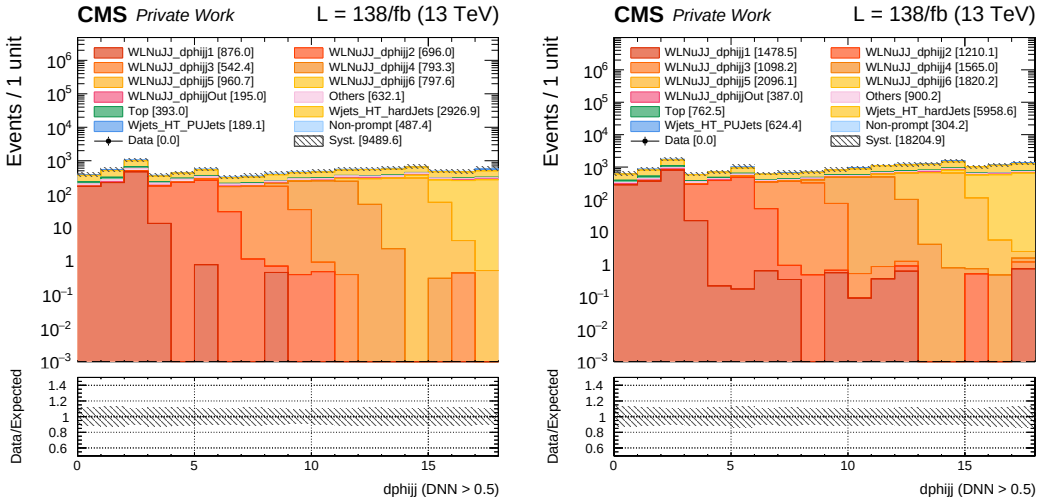


Figure 4.26: Distribution of $\Delta\phi_{jj}$ in the electron channel, with DNN > 0.5. On top, the W+jets control regions with hard and PU jets are shown on the left and right, respectively. The bottom-left panel shows the top control region, while the SR (blind) on the bottom right displays the reco variable distribution binned with three reconstructed bins per generator-level bin.

Results

The results of the measurement, presented as a function of the dijet pseudorapidity separation $\Delta\eta_{jj}$, are outlined in this section. All one-dimensional likelihood scans correspond to *profiled fits*, in which a single parameter of interest (POI) is scanned while all remaining POIs and nuisance parameters are left free to vary in the fit. This procedure incorporates correlations among bins and propagates the impact of shared systematic uncertainties. Figure 4.23 provides an overview of the resulting likelihood scans in each $\Delta\eta_{jj}$ bin, obtained from fits performed using the reconstructed variable. This summary plot illustrates the sensitivity achieved and the overall consistency of the measurement across bins.

To further investigate the intrinsic correlations among the parameters of interest, two-dimensional likelihood scans are produced using Monte Carlo simulations only, with all systematic uncertainties and remaining POIs fixed at their nominal values. Unlike the one-dimensional scans, these 2D scans are not profiled and are used to illustrate the expected correlation pattern between pairs of POIs as determined by the structure of the likelihood model. Figure 4.24 shows the resulting correlation coefficients, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties is evaluated using a data-Asimov dataset (see Section 1.9), in which pseudo-data are generated from the post-fit model while preserving the best-fit values of the nuisance parameters. Dedicated post-fit impact plots, produced for all $\Delta\eta_{jj}$ bins, are reported in Appendix D.3. The most significant systematic effects originate from neighboring POIs, reflecting intrinsic correlations among adjacent $\Delta\eta_{jj}$ bins.

Finally, the expected unfolded distribution is shown in Figure 4.25. The unfolded measurements are presented together with their total uncertainties and compared to SM predictions. Multiple theoretical calculations, including SHERPA (LO, multi-jet merged) and POWHEG (NLO) interfaced with PYTHIA8, are overlaid together with their theoretical uncertainties, allowing a direct comparison between data and predictions based on different perturbative orders and parton-shower models.

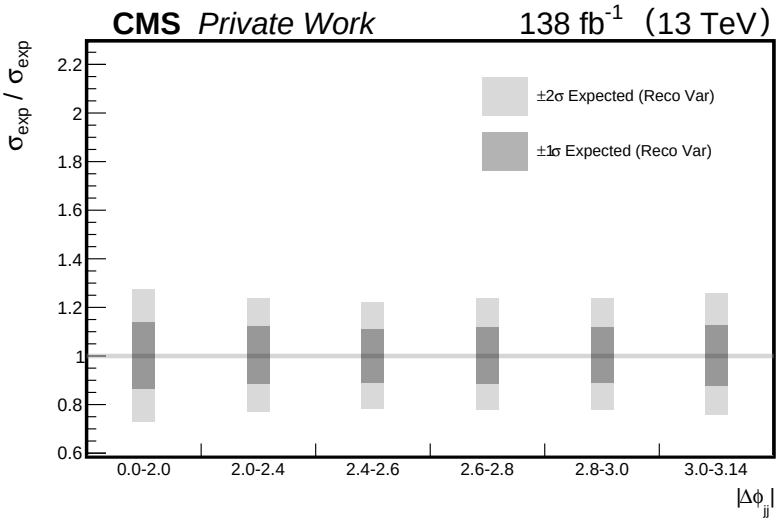


Figure 4.27: Expected profile likelihood scans for the measurement in the $\Delta\phi_{jj}$ bins using the reco variable-based discriminant (black). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity achieved with the reco variable method.

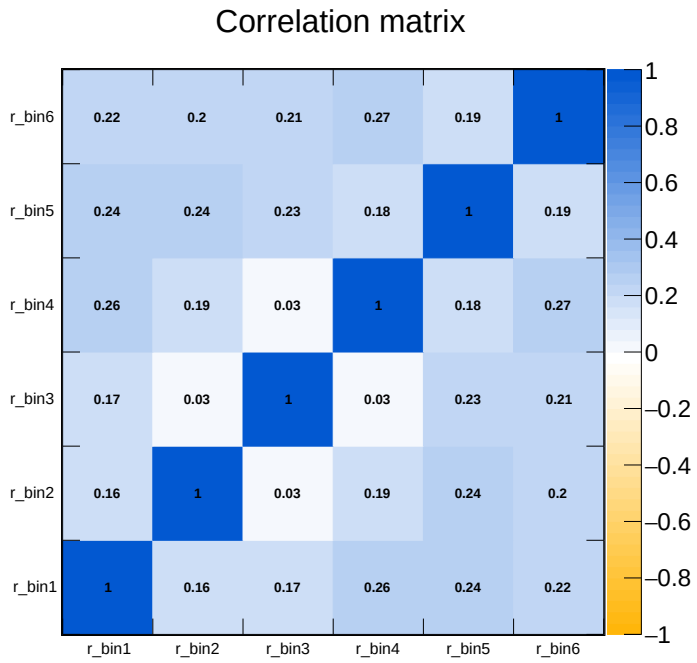


Figure 4.28: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. Only the correlations at 1σ deviations are shown, highlighting the intrinsic correlations induced by the fit construction.

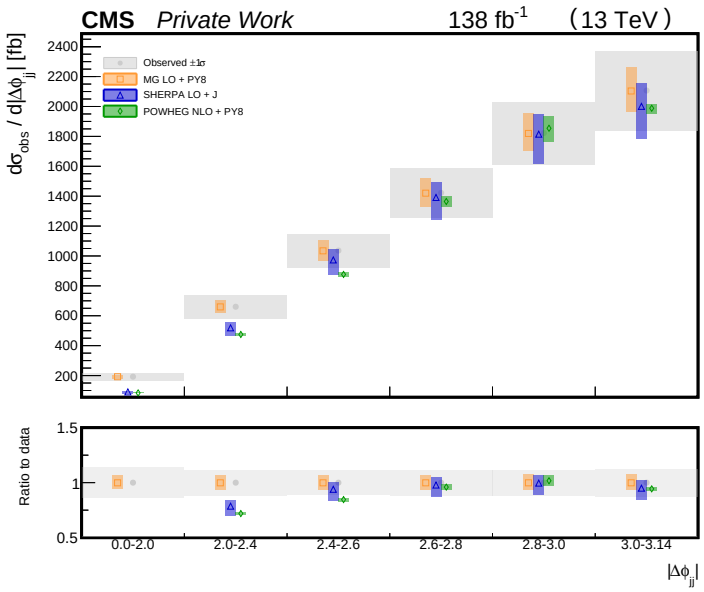


Figure 4.29: Expected unfolded cross sections in the $\Delta\phi_{jj}$ bins, shown with total uncertainties and compared to SM predictions. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged).

4.2.7 Unfolded $p_T(j1)$ distribution

Figures 4.30 show the distributions used to construct the SR templates that enter the fit. Since the control regions remain identical to those introduced in Section 4.2.4, only the SR distributions are shown, to illustrate the different signal composition and the correlations among processes in this observable.

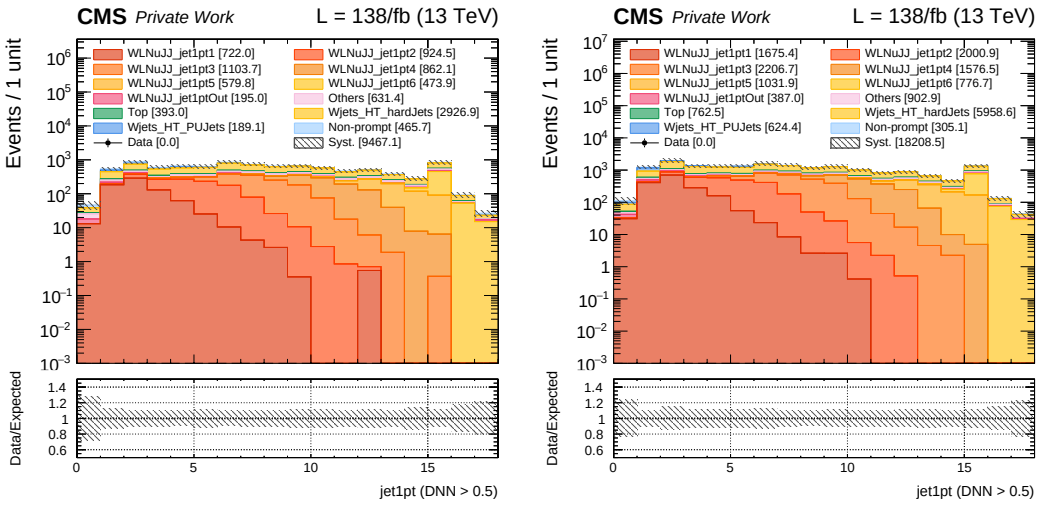


Figure 4.30: Distributions of $p_T(j_1)$ in the electron (left) and muon (right) Signal Regions, for events with DNN > 0.5. Only the SR is shown, highlighting the reconstructed observable used as input to the fit and the variation in signal composition across bins. The variable is binned with three reconstructed bins per generator-level bin.

Results

The results of the measurement, presented as a function of the leading jet transverse momentum $p_T(j_1)$, are outlined in this section. All one-dimensional likelihood scans discussed here correspond to *profiled fits*, in which a single parameter of interest (POI) is scanned while all remaining POIs and nuisance parameters are left free to vary. This approach naturally incorporates correlations among bins and propagates the impact of shared systematic uncertainties across all fitted parameters.

Figure 4.31 provides an overview of the likelihood scans profiled in each $p_T(j_1)$ bin, obtained from fits performed using the reconstructed observable. This summary plot illustrates the sensitivity achieved in each bin and the overall consistency of the measurement.

To explore the intrinsic correlations among the POIs, two-dimensional likelihood scans are produced using Monte Carlo simulations only, with all systematic uncertainties and the remaining POIs fixed at their nominal values. Unlike the one-dimensional scans, these 2D scans are not profiled and are used to visualise the expected correlation pattern between pairs of bins as determined

by the structure of the likelihood model itself. Figure 4.32 shows the correlation coefficients derived from these scans, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties on each POI is evaluated using a data-Asimov dataset (see Section 1.9), in which pseudo-data are generated from the post-fit model while preserving the best-fit values of the nuisance parameters. Dedicated post-fit impact plots, produced for all bins, are reported in Appendix D.5. The most significant effects also in this case are found to originate from neighboring POIs.

Finally, the expected unfolded distribution is shown in Figure 4.33. The unfolded measurements are presented together with their total uncertainties and compared to SM predictions. Multiple theoretical predictions, obtained with different generators and parton-shower models at various perturbative orders, are overlaid to assess the agreement between the measurement and theoretical expectations.

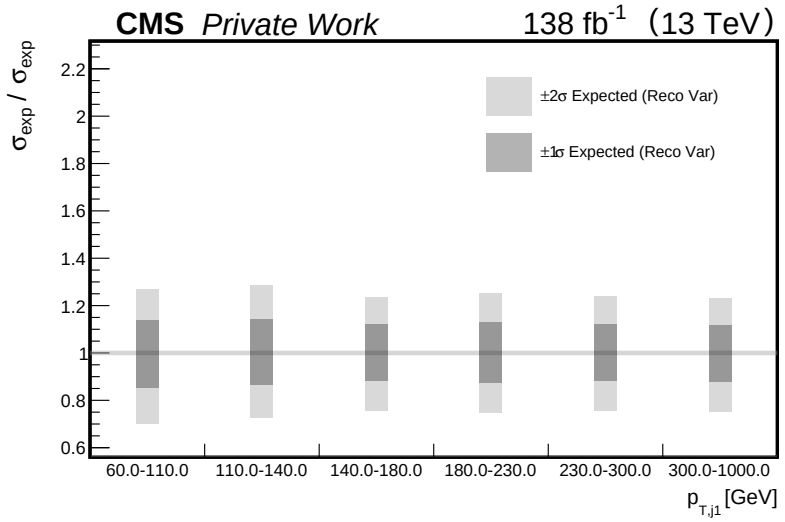


Figure 4.31: Expected profile likelihood scans for the measurement in the $p_T(j1)$ bins using the reco variable-based discriminant (black). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity achieved with the reco variable method.

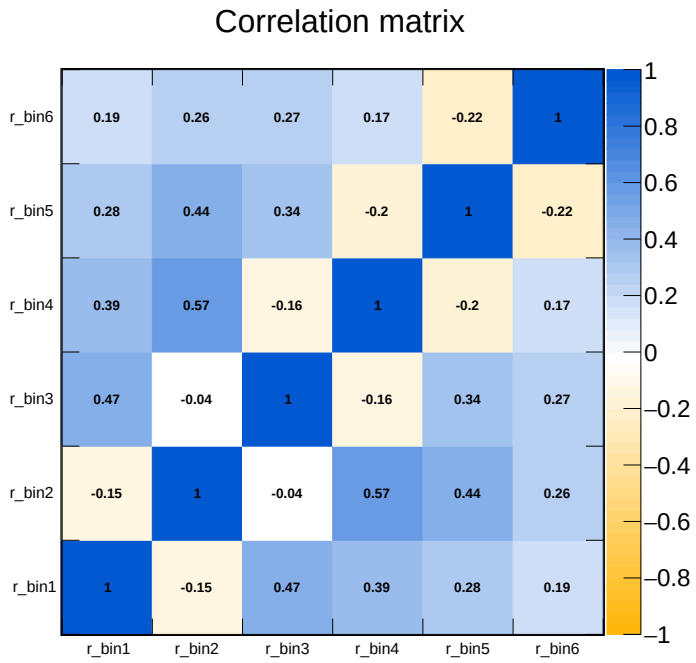


Figure 4.32: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. Only the correlations at 1σ deviations are shown, highlighting the intrinsic correlations induced by the fit construction.

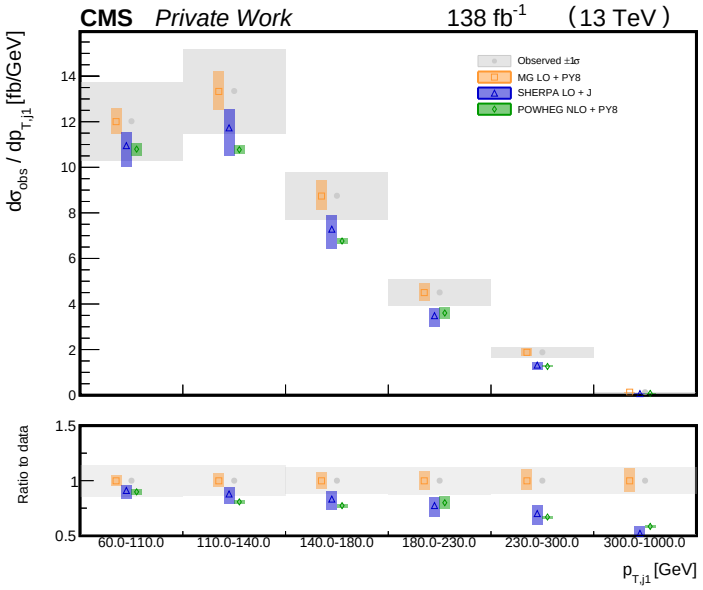


Figure 4.33: Expected unfolded cross sections in the $p_T(j_1)$ bins, shown with total uncertainties and compared to SM predictions. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged).

4.2.8 Unfolded $p_T(j2)$ distribution

Figures 4.34 show the distributions used to construct the SR templates entering the fit. Since the control regions remain identical to those defined in Section 4.2.4, only the SR is shown here, as it provides the most direct insight into the variable's behaviour and the correlation among signal components.

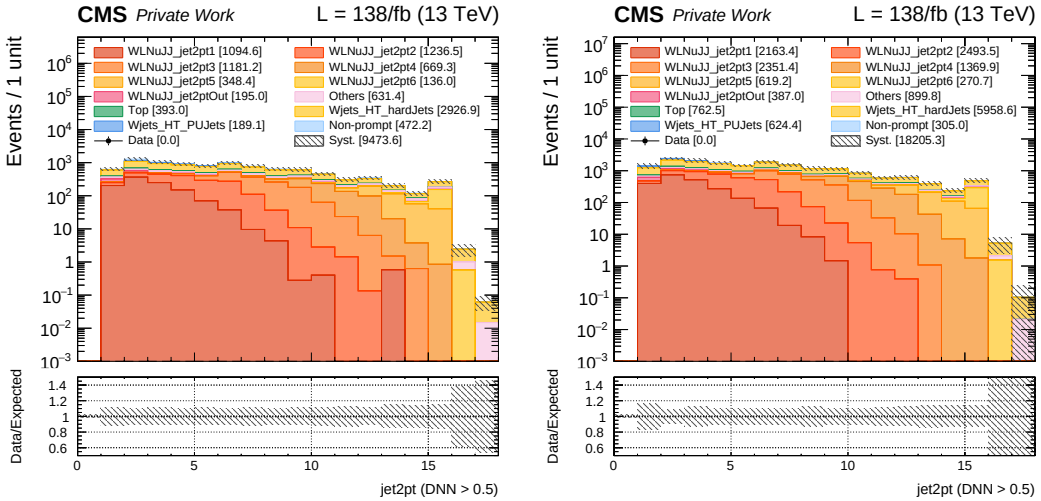


Figure 4.34: Distributions of $p_T(j_2)$ in the electron (left) and muon (right) Signal Regions, for events with DNN > 0.5. Only the SR is shown, highlighting the reconstructed observable used as input to the fit and illustrating differences in signal composition and correlations across bins. The variable is binned with three reconstructed bins per generator-level bin.

Results

The results of the measurement, presented as a function of the subleading jet transverse momentum $p_T(j_2)$, are outlined in this section. All one-dimensional likelihood scans discussed here correspond to *profiled fits*, in which a single parameter of interest (POI) is scanned while all remaining POIs and nuisance parameters are left free to vary in the fit. This procedure naturally incorporates correlations among bins and propagates the effects of shared systematic uncertainties across all parameters.

Figure 4.35 provides an overview of the likelihood scans profiled in each $p_T(j_2)$ bin, obtained from fits performed using the reconstructed observable. This summary plot illustrates the sensitivity achieved in each bin and the overall consistency of the measurement.

To explore the intrinsic correlations among the POIs, two-dimensional likelihood scans are produced using Monte Carlo simulations only, with all systematic uncertainties and the remaining POIs fixed at their nominal values. Unlike the one-dimensional scans, these 2D scans are not profiled and are used to visualise the expected correlation pattern between pairs of bins as determined

by the structure of the likelihood model itself. Figure 4.36 shows the correlation coefficients derived from these scans, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties on each POI is evaluated using a data-Asimov dataset (see Section 1.9), in which pseudo-data are generated from the post-fit model while preserving the best-fit values of the nuisance parameters. Dedicated post-fit impact plots, produced for all bins, are reported in Appendix D.6. The most significant effects are found to originate from neighboring POIs, reflecting the correlations among adjacent $p_T(j_2)$ bins.

Finally, the expected unfolded distribution is shown in Figure 4.37. The unfolded measurements are presented together with their total uncertainties and compared to SM predictions. Multiple theoretical predictions are overlaid for comparison, including SHERPA (LO, multi-jet merged) and POWHEG (NLO) interfaced with PYTHIA8. These comparisons allow us to assess the level of agreement between the measured spectrum and theoretical expectations across different perturbative orders and parton-shower models.

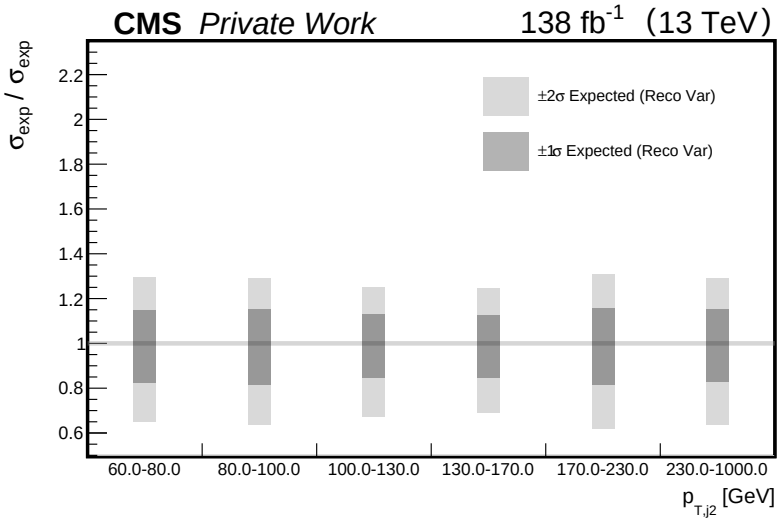


Figure 4.35: Expected profile likelihood scans for the measurement in the $p_T(j2)$ bins using the reco variable-based discriminant (black). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity achieved with the reco variable method.

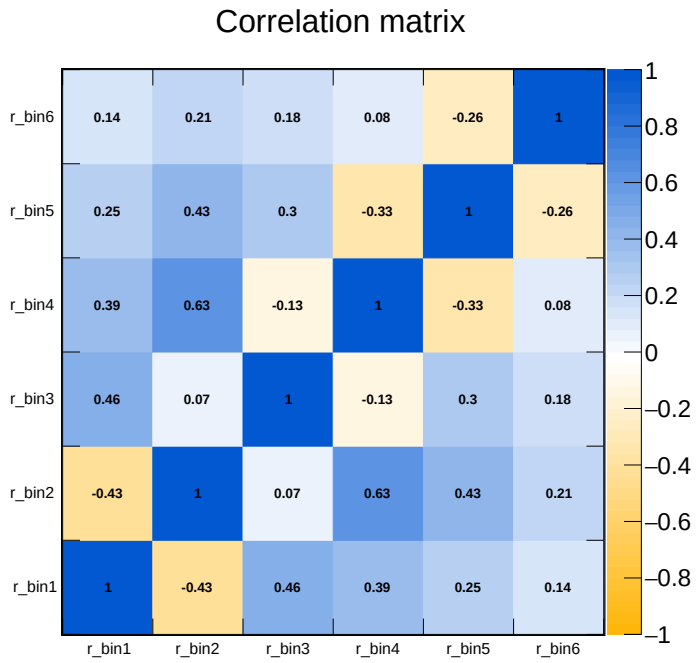


Figure 4.36: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. Only the correlations at 1σ deviations are shown, highlighting the intrinsic correlations induced by the fit construction.

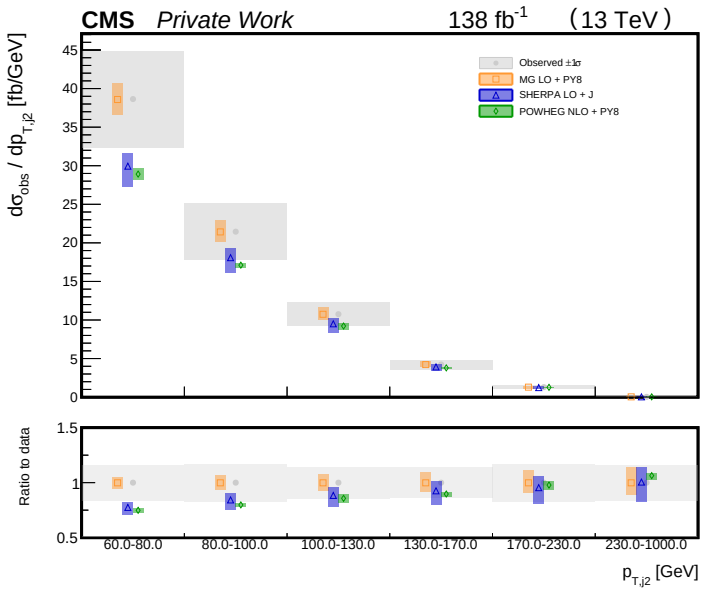


Figure 4.37: Expected unfolded cross sections in the $p_T(j_2)$ bins, shown with total uncertainties and compared to SM predictions. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged).

4.2.9 Unfolded $p_T(l)$ distribution

Figures 4.38 show the distributions used to construct the SR templates that enter the fit. Since the control regions remain identical to those defined in Section 4.2.4, only the SR is shown here, as it is the part of the analysis where the observable changes and allows one to visualise differences in signal composition and correlations among processes.

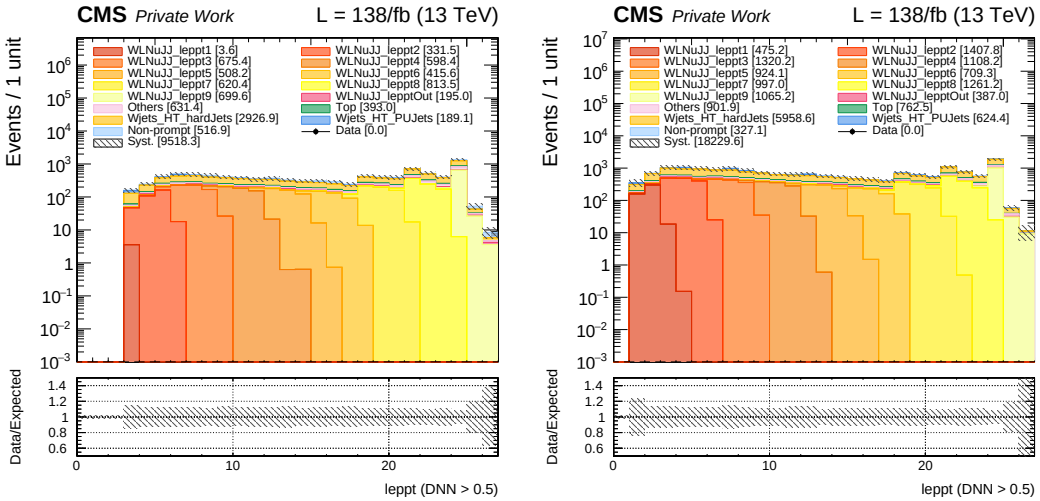


Figure 4.38: Distributions of the lepton transverse momentum $p_T(\ell)$ in the electron (left) and muon (right) Signal Regions, for events with DNN > 0.5. Only the SR is shown, as it highlights the reconstructed observable used as input to the fit and illustrates variations in signal composition and inter-bin correlations. The variable is binned with three reconstructed bins per generator-level bin.

Results

The results of the measurement, presented as a function of the lepton transverse momentum $p_T(\ell)$, are outlined in this section. All one-dimensional likelihood scans discussed here correspond to *profiled fits*, in which a single parameter of interest (POI) is scanned while all remaining POIs and nuisance parameters are left free to vary in the fit. This procedure naturally incorporates correlations among bins and propagates the effects of shared systematic uncertainties across all parameters.

Figure 4.39 provides an overview of the likelihood scans profiled in each $p_T(\ell)$ bin, obtained from fits performed using the reconstructed observable. This summary plot illustrates the sensitivity achieved in each bin and the overall consistency of the measurement.

To explore the intrinsic correlations among the POIs, two-dimensional likelihood scans are produced using Monte Carlo simulations only, with all systematic uncertainties and the remaining POIs fixed at their nominal values. Unlike

the one-dimensional scans, these 2D scans are not profiled and are used to visualise the expected correlation pattern between pairs of bins as determined by the structure of the likelihood model itself. Figure 4.40 shows the correlation coefficients derived from these scans, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties on each POI is evaluated using a data-Asimov dataset (see Section 1.9), in which pseudo-data are generated from the post-fit model while preserving the best-fit values of the nuisance parameters. Dedicated post-fit impact plots, produced for all bins, are reported in Appendix D.7. The most significant effects are found to originate from neighboring POIs, reflecting correlations among adjacent $p_T(\ell)$ bins.

Finally, the expected unfolded distribution is shown in Figure 4.41. The unfolded measurements are presented together with their total uncertainties and compared to Standard Model predictions. Multiple theoretical predictions, including SHERPA (LO, multi-jet merged) and POWHEG (NLO) interfaced with PYTHIA8, are overlaid for comparison, allowing a direct assessment of the agreement between data and different perturbative orders and parton-shower models.

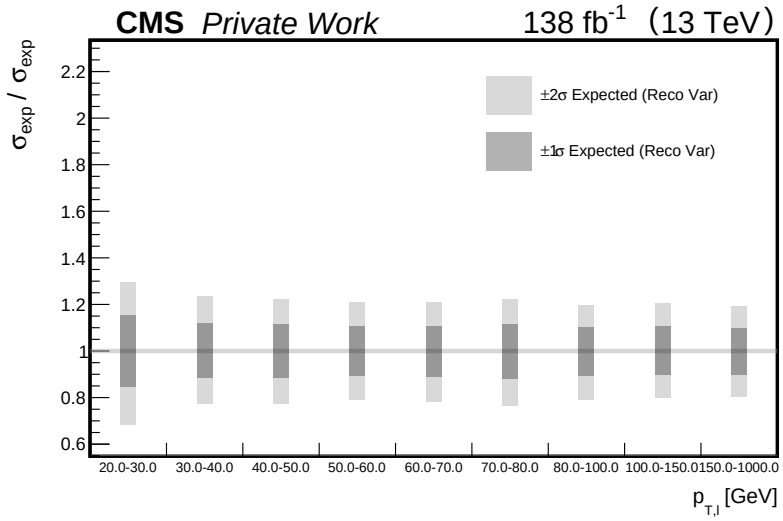


Figure 4.39: Expected profile likelihood scans for the measurement in the $p_T(l)$ bins using the reco variable-based discriminant (black). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity achieved with the reco variable method.

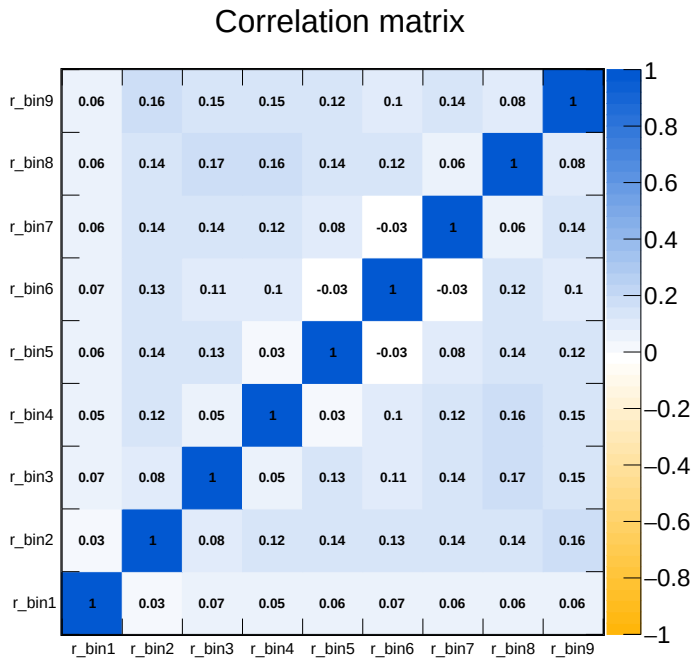


Figure 4.40: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. Only the correlations at 1σ deviations are shown, highlighting the intrinsic correlations induced by the fit construction.

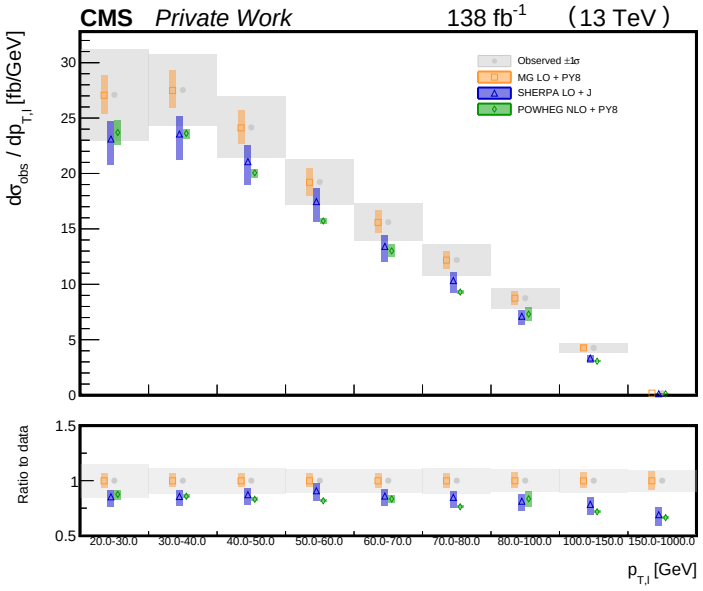


Figure 4.41: Expected unfolded cross sections in the $p_T(\ell)$ bins, shown with total uncertainties and compared to Standard Model predictions. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged).

4.2.10 Unfolded $p_T(W)$ distribution

Figure 4.42 shows the distributions used to construct the SR templates that enter the fit. Since the control regions remain identical to those introduced in Section 4.2.4, they are not shown here. Only the SR is displayed, as it provides the most relevant view of the variable's behaviour and of the correlations among signal and background components.

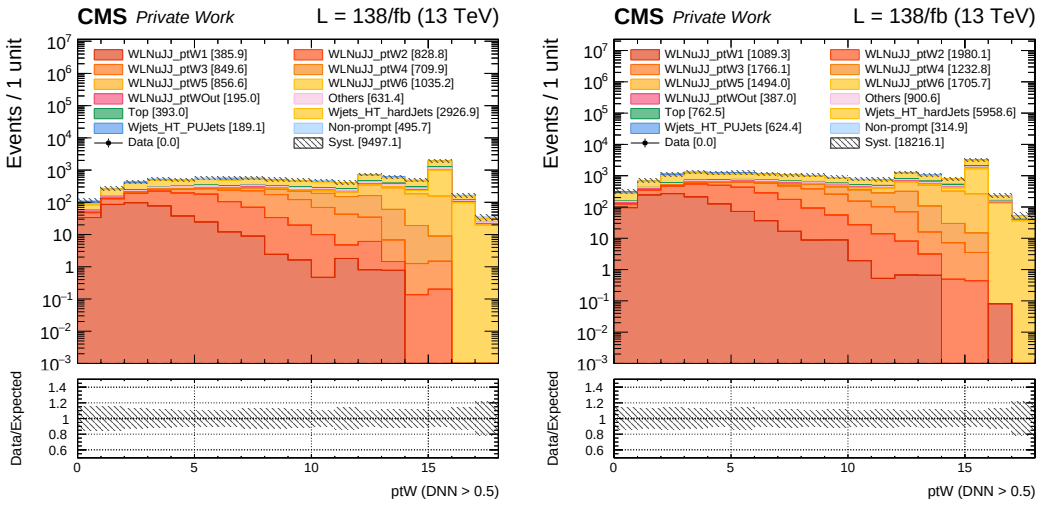


Figure 4.42: Distribution of the reconstructed $p_T(W)$ in the Signal Region, with DNN > 0.5. The templates are shown for the electron (left) and muon (right) channels, binned in three reconstructed bins per generator-level bin. The signal contributions are shown both stacked and superimposed to illustrate their relative correlations.

Results

The results of the measurement, presented as a function of the transverse momentum of the $W \to \ell\nu$ system, $p_T(W)$, are outlined in this section.

All one-dimensional likelihood scans correspond to *profiled fits*, where each parameter of interest (POI) is scanned while the remaining POIs and nuisance parameters are left free to vary in the fit. Figure 4.43 provides an overview of the resulting scans in each $p_T(W)$ bin, summarising the achieved sensitivity and illustrating the overall consistency of the measurement across bins.

To better understand the intrinsic correlations among the parameters of interest, two-dimensional likelihood scans are produced using Monte Carlo simulations only, with all systematic uncertainties and remaining POIs kept fixed at their nominal values. Unlike the one-dimensional scans, these 2D scans are not profiled; they illustrate the correlation patterns among $p_T(W)$ bins arising purely from the structure of the likelihood. Figure 4.44 shows the correlation coefficients derived from these scans, evaluated at the 1σ deviation of each POI.

The impact of systematic uncertainties is evaluated using a data-Asimov dataset (see Section 1.9), where pseudo-data are generated from the post-fit model with all nuisance parameters fixed to their best-fit values from data. Dedicated post-fit impact plots, produced for all POIs and reported in Appendix D.8, quantify the effect of each uncertainty in the measurement. As for other observables, the most significant systematic effects originate from neighbouring POIs, reflecting correlations among adjacent $p_T(W)$ bins.

Finally, the expected unfolded distribution is shown in Figure 4.45. The unfolded measurements are presented together with their total uncertainties and compared to SM predictions. Multiple theoretical calculations, obtained with different generators and parton-shower models at varying perturbative orders, are overlaid to assess the agreement between the measurement and theoretical expectations.

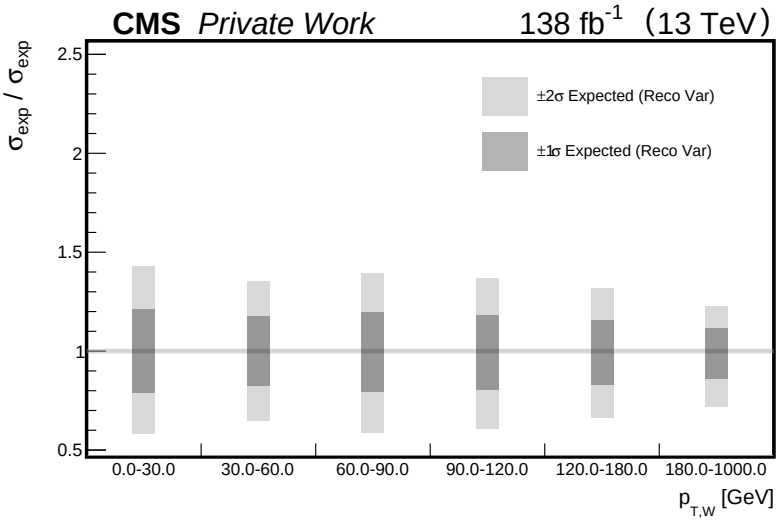


Figure 4.43: Expected profile likelihood scans for the measurement in the $p_T(W)$ bins using the reco variable-based discriminant (black). The inner and outer bands correspond to the 1σ and 2σ intervals, respectively. This summary illustrates the overall sensitivity achieved with the reco variable method.

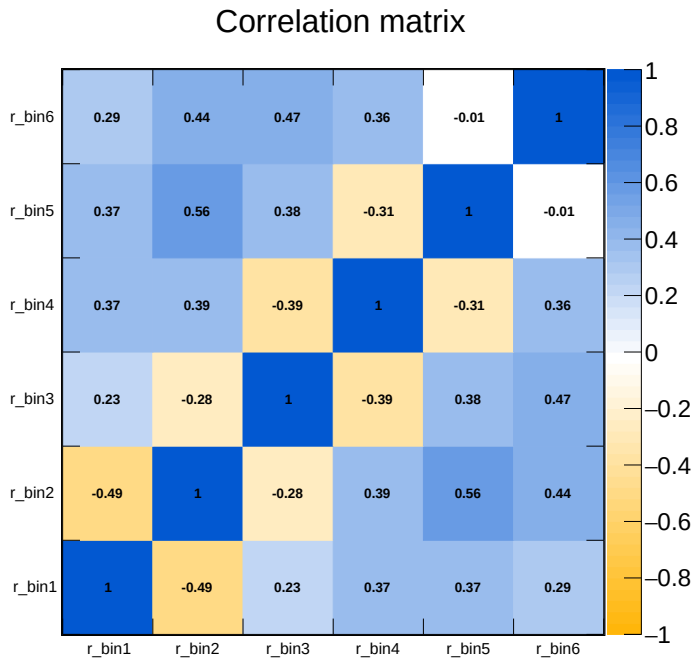


Figure 4.44: Correlation coefficients between pairs of parameters of interest, obtained from two-dimensional likelihood scans with systematic uncertainties frozen. Only the correlations at 1σ deviations are shown, highlighting the intrinsic correlations induced by the fit construction.

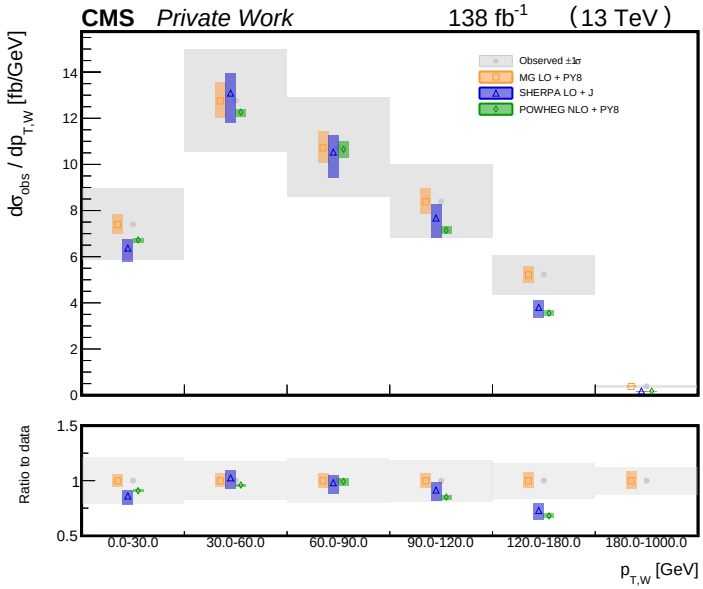


Figure 4.45: Expected unfolded cross sections in the $p_T(W)$ bins, shown with total uncertainties and compared to SM predictions. Theoretical predictions from different generators and perturbative orders are overlaid, including POWHEG (NLO) +Pythia8 and Sherpa (LO, multi-jet merged).

Chapter 5

Probing New Physics in Electroweak W+Jets Production

In this chapter, the EW $W \to \ell\nu jj$ analysis is extended beyond the SM measurement to search for possible signs of new physics within the framework of the Standard Model Effective Field Theory (SMEFT). The aim is to exploit the same analysis strategy to constrain higher-dimensional operators that could modify the EW interactions.

After introducing the relevant operators and signal modelling in Sections 5.1–5.2, the statistical implementation of the EFT model in COMBINE and its application to the single Wjj process are described in Sections 5.2.1–5.3. Finally, Section 5.4 presents the global combination of multiple EW channels, providing a coherent EFT interpretation and highlighting the complementary role of the Wjj process in the overall picture.

5.1 Effective Field Theory Interpretation

After establishing the SM measurement, the same statistical framework is extended to an interpretation in the context of the SMEFT. In this framework, potential deviations from the SM are parametrised in terms of higher-dimensional operators that encode the low-energy effects of heavy new physics. Their contributions are expressed through dimensionless Wilson coefficients c_i, suppressed by powers of the new-physics scale Λ, as introduced in Section 1.8.1.

The EW production of a W boson in association with two jets is particularly sensitive to such effects, since several dimension-six operators can modify both its kinematic distributions and overall cross section. Among the large set of operators defined in the Warsaw basis [47], only a subset contributes at leading order to the $pp \to Wjj$ process. The operators relevant to this analysis are listed in Table 5.1.

Operator	Expression
$\mathcal{O}_{H\ell}^{(3)}$	$\left(H^\dagger i \overleftrightarrow{D}^i_\mu H\right)\left(\bar{\ell}_p \sigma^i \gamma^\mu \ell_r\right)$
$\mathcal{O}_W$	$\epsilon^{ijk} W^{i\nu}_\mu W^{j\rho}_\nu W^{k\mu}_\rho$
$\mathcal{O}_{\widetilde{W}}$	$\epsilon^{ijk} \widetilde{W}^{i\nu}_\mu W^{j\rho}_\nu W^{k\mu}_\rho$
$\mathcal{O}_{\ell\ell}$	$\left(\bar{\ell}_p \gamma_\mu \ell_r\right)\left(\bar{\ell}_s \gamma^\mu \ell_t\right)$
$\mathcal{O}_{Hj}^{(1)}$	$\left(H^\dagger i \overleftrightarrow{D}_\mu H\right)\left(\bar{q} \gamma^\mu q\right)$
$\mathcal{O}_{Hj}^{(3)}$	$\left(H^\dagger i \overleftrightarrow{D}_\mu H\right)\left(\bar{q} \sigma^i \gamma^\mu q\right)$
$\mathcal{O}_{HWB}$	$H^\dagger \sigma^i H W^i_{\mu\nu} B^{\mu\nu}$
$\mathcal{O}_{\widetilde{HWB}}$	$H^\dagger \sigma^i H \widetilde{W}^i_{\mu\nu} B^{\mu\nu}$
$\mathcal{O}_{HDD}$	$\left(D^\mu H^\dagger H\right)\left(H^\dagger D_\mu H\right)$

Table 5.1: Dimension-six SMEFT operators relevant to the $pp \to Wjj$ analysis.

5.2 EFT Signal Modelling

The signal samples, including effective operators, are modelled using the SMEFTsim package [145, 146], which provides a consistent implementation of the SMEFT Lagrangian in the MadGraph5_aMC@NLO framework. This enables the generation of parton-level events that include both the SM and dimension-six contributions, propagated through the same simulation and reconstruction chain used for the SM measurement.

In the presence of higher-dimensional operators, the scattering amplitude for EW Wjj production can be written as

$$|\mathcal{A}|^2 = |\mathcal{A}_{\text{SM}}|^2 + 2\,\text{Re}\left(\mathcal{A}_{\text{SM}}\mathcal{A}^\dagger_{\text{EFT}}\right) + |\mathcal{A}_{\text{EFT}}|^2, \tag{5.1}$$

leading to a quadratic dependence of the event yield on each Wilson coefficient $x = \mathcal{C}^{(6)}_\alpha/\Lambda^2$:

$$N_\alpha(x) = \text{SM} + x \cdot \text{Lin}_\alpha + x^2 \cdot \text{Quad}_\alpha. \tag{5.2}$$

The three components — SM, linear, and quadratic — are extracted algebraically from samples generated at $x = 0, \pm 1$:

$$\text{SM} = N(0), \quad \text{Lin} = \tfrac{1}{2}\left[N(1) - N(-1)\right], \quad \text{Quad} = \tfrac{1}{2}\left[N(1) + N(-1) - 2N(0)\right]. \tag{5.3}$$

To optimise production, only the $x = 1$ sample is explicitly generated, while all other hypotheses are obtained via the built-in reweighting feature of MadGraph5_aMC@NLO. Roughly ten million events are generated per data-taking year (2016HIPM, 2016noHIPM, 2017, and 2018). A large number of events ensures negligible statistical uncertainties in the reweighted templates, which are fully correlated across hypotheses and therefore not explicitly modelled in the likelihood. The generation setup mirrors that of the signal samples used for the SM analysis, except for a dedicated reweight card and a parton-level requirement $m_{jj} > 300$ GeV to enhance generation efficiency, consistent with the $m_{jj} > 400$ GeV cut applied at reconstruction level.

A closure test was performed to validate the private EFT production. To ensure a like-for-like comparison between the private SMEFT samples and the SM NanoAOD signal sample, the two samples are compared after applying an identical set of particle-level (LHE) selections. These selections are chosen to be the most restrictive common subset of the requirements of the two productions. For example, the private sample includes a generator-level requirement of $m_{jj} > 300$ GeV, while the official sample does not; this condition is applied to both samples when performing the closure test. This approach guarantees that any observed differences arise from physics modelling rather than from differing LHE phase-space definitions.

A second subtlety concerns the syntax used in Madgraph5_aMC@NLO for generating the two samples. The official SM sample is generated with `generate p p > l+ vl j j / t t  h QCD=0`, where intermediate top quarks and Higgs bosons are explicitly excluded from the amplitude. For the SMEFT production, however, the recommended syntax is `generate p p > l+ vl j j QCD=0`, which omits the exclusions. This choice is required to ensure a consistent SMEFT expansion and to avoid unitarity-violating artefacts that may arise if individual diagrams are selectively removed. In SMEFT, amplitudes grow with energy because of higher-dimensional operators, and excluding only a subset of diagrams (such as those involving a top quark or a Higgs boson) while retaining their SMEFT-modified counterparts can lead to an inconsistent truncation of the theory and unphysical high-energy behaviour. Using the inclusive syntax avoids this issue and ensures a gauge-invariant and unitary amplitude throughout the relevant phase space.

Unlike the official SM sample, the SMEFT sample is generated without excluding top- or Higgs-mediated diagrams, in order to preserve a gauge-invariant and unitary amplitude. The LHE record does not store intermediate resonances, but only the final-state partons. As a consequence, a direct veto on top- or Higgs-mediated contributions cannot be applied at the particle level. To recover an equivalent phase space to the official SM sample, a particle-level veto is therefore applied on all events containing at least one b quark in the final state. In genuine electroweak Wjj production, this removes only a negligible fraction of events (mostly from PDF-induced $g \to b\bar{b}$ splittings), while efficiently rejecting top-quark processes, which are excluded by construction in the official sample. In principle, one would also veto Higgs-mediated contributions, but implementing such a veto at the LHE level is nontrivial and their expected rate is extremely low.

Figures 5.2 and 5.1 show the comparison between the SM prediction obtained from the reweighted private sample ($x = 0$) and the official NanoAOD sample in the SM signal region, after applying the harmonised LHE selections discussed above. The agreement is excellent across all main kinematic observables and for the DNN discriminant, with the residual differences fully covered by statistical uncertainties. In particular, the high-DNN region, responsible for most of the EFT sensitivity, shows very good closure.

5.2.1 Implementation in Combine

The interpretation of the measurement in the EFT framework is performed using the COMBINE tool, following the formalism introduced in Section 1.9.7. In this implementation, the total signal prediction is decomposed into three separate template components corresponding to the terms of Eq. 5.2: the SM contribution, the linear interference term, and the pure quadratic contribution. The Wilson coefficient $x = \mathcal{C}_\alpha^{(6)}/\Lambda^2$ is then treated as a parameter of interest and extracted through a likelihood fit, employing the same statistical methodology used for the SM cross section measurement. When applied to the Asimov dataset, this procedure yields the expected precision and sensitivity to each operator.

A direct implementation of the SMEFT expansion can lead to negative binomial contents in the interference term, as local destructive effects may occur relative to the SM prediction. Since COMBINE employs Poisson-based likelihoods, which require all input templates to be positive-definite, an algebraic reformulation of the model is applied to preserve this property. The transformation, originally developed by the Milano CMS group [79], redefines the event yield as:

$$N = (1 - x)\,\mathrm{SM} + x\,(\mathrm{SM} + \mathrm{Lin}_\alpha + \mathrm{Quad}_\alpha) + (x^2 - x)\,\mathrm{Quad}_\alpha, \tag{5.4}$$

which guarantees positive yields for any value of x.

With this parameterization, the fit relies on three physically meaningful templates, summarised in Table 5.2:

Template	**Description**
$\lvert\mathcal{A}_{\mathrm{SM}}\rvert^2$	Standard Model component
$\lvert\mathcal{A}_{\mathrm{SM}} + \mathcal{A}_{\mathrm{EFT}}\rvert^2$	SM + interference + quadratic contributions
$\lvert\mathcal{A}_{\mathrm{EFT}}\rvert^2$	Pure EFT contribution

Table 5.2: Template components entering the EFT fit as implemented in COMBINE.

These templates are constructed algebraically from the simulated samples according to Eq. 5.3, allowing COMBINE to perform a positive-definite, fully consistent likelihood fit of the signal yield as a function of the Wilson coefficients. This setup provides a unified statistical treatment of EFT effects within the same framework used for the SM measurement, forming the basis for the extraction of both one- and two-dimensional constraints on the operators considered in this analysis.

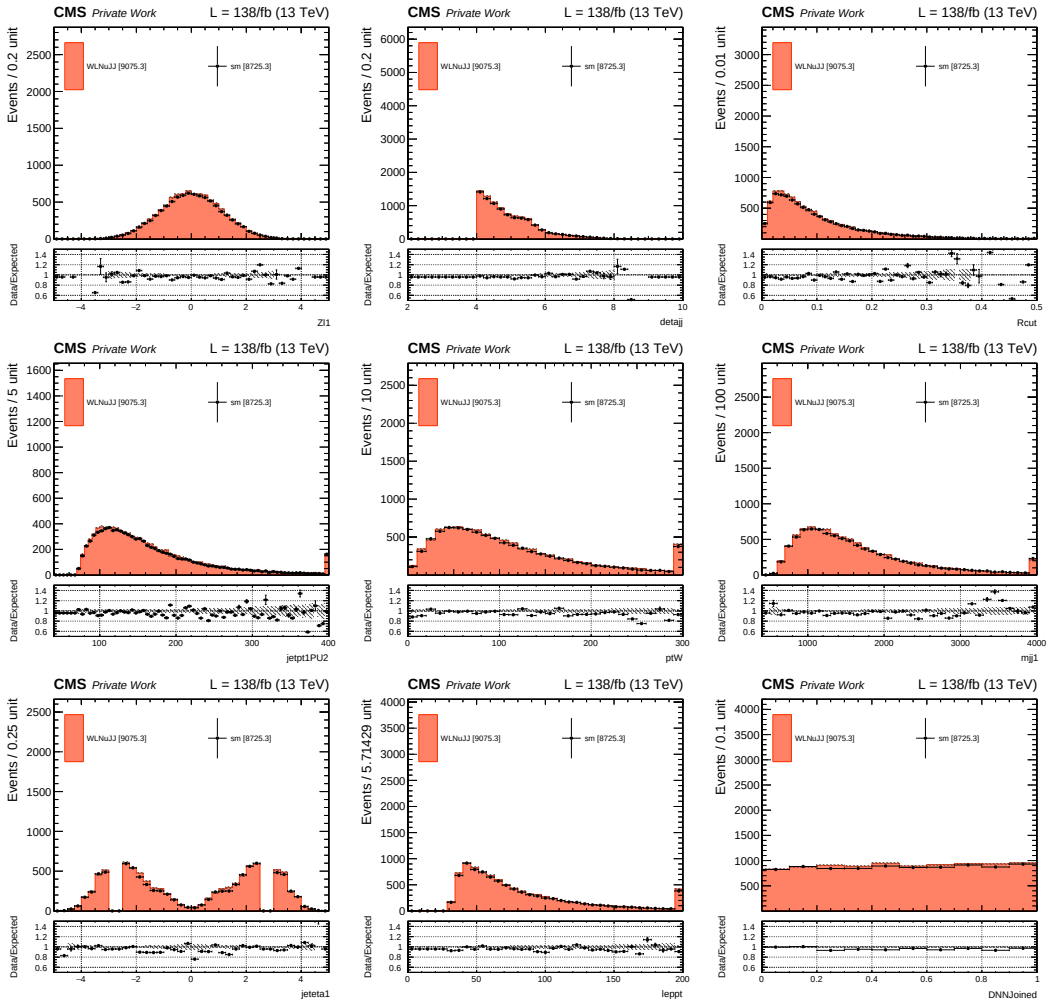


Figure 5.1: Validation of the private EFT signal generation in the electron channel. The SM component obtained from the reweighted MadGraph5_aMC@NLO LO sample (black markers) is compared to the reference signal sample used in the SM analysis (red) in the signal region. Good agreement is observed across all key observables for the full Run II dataset, confirming the reliability of the private EFT production and reweighting procedure.

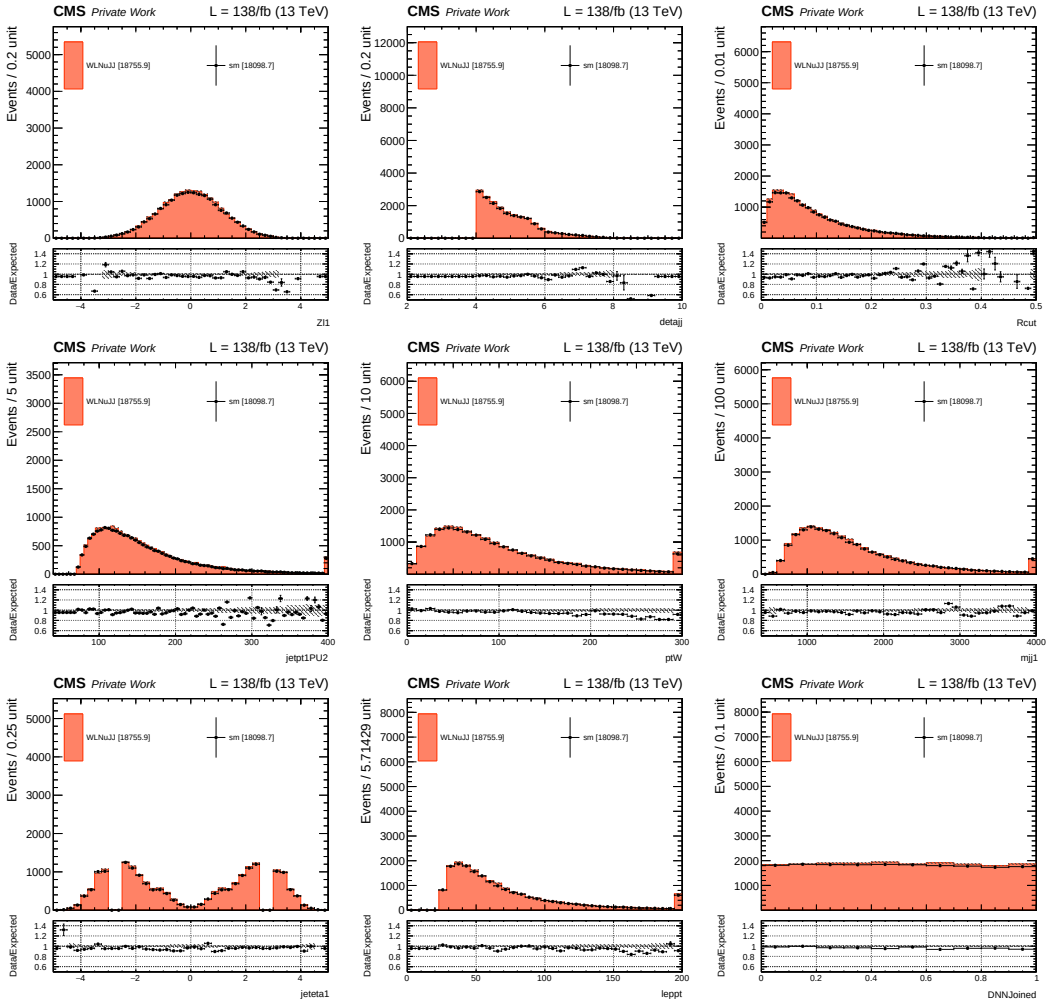


Figure 5.2: Validation of the private EFT signal generation in the muon channel. The SM component obtained from the reweighted MADGRAPH5_aMC@NLO LO sample (black markers) is compared to the reference signal sample used in the SM analysis (red) in the signal region. Good agreement is observed across all key observables for the full Run II dataset, confirming the consistency of the private EFT production and reweighting procedure.

5.3 Single-channel EFT interpretation

Building upon the framework and implementation of the SMEFT model described in the previous section, this part presents its application to the EW production of a W boson in association with two jets. The analysis follows the same structure as the SM measurement, extending it to extract constraints on the dimension-six Wilson coefficients within the same likelihood formalism. The overall strategy and fit configuration are discussed below, followed by the presentation of the one- and two-dimensional results.

5.3.1 Strategy

The fitting strategy follows the same philosophy adopted in the inclusive and differential SM measurements, namely defining a sufficiently signal-enriched phase space through the DNN > 0.5 requirement applied uniformly across the signal and control regions. This ensures that events entering the fit are already dominated by EW-like topologies and reduces the impact of large background extrapolations.

Within this common phase-space definition, the signal region is further exploited by performing a one-dimensional fit in the azimuthal separation between the two leading jets, $\Delta\phi_{jj}$, identified as the most sensitive observable to EFT-induced effects. Although different EFT operators can alter various kinematic observables through modifications of different vertices or Lorentz structures, a single observable must be used consistently in a global fit where multiple coefficients are fitted simultaneously. This guarantees that correlations among operators are properly captured within a common likelihood structure. Among the observables studied, $\Delta\phi_{jj}$ proved particularly effective, being highly sensitive to modifications of the triple gauge coupling induced by operators such as $\mathcal{O}_W$, while still retaining discriminating power for other dimension-six contributions. For this reason, it has been adopted as the reference observable for the SMEFT fit across all operators.

Ten uniform bins in $\Delta\phi_{jj}$ are used in the signal region, while the control regions enter the likelihood only as event yields (single-bin templates). Control regions are included to constrain the main backgrounds, in particular $t\bar{t}$ and QCD-induced W+jets. Theoretical uncertainties on the signal, such as PDF and QCD scale variations, are kept unnormalized to fully propagate their impact on the fit. All detector- and modelling-related systematic uncertainties defined for the SM analysis are consistently propagated to the SMEFT templates and are treated as fully correlated between the SM and EFT components, as they originate from the same underlying simulation samples.

The input distributions entering the likelihood scan, signal and control regions across all years and channels, are shown in Figs. 5.3 and 5.4. For illustration purposes, only the pure quadratic contributions of each operator are displayed. In the actual fit, however, both the linear interference with the SM amplitude

and the cross-interference among operators are included, as explicitly written in Eq. 5.1.

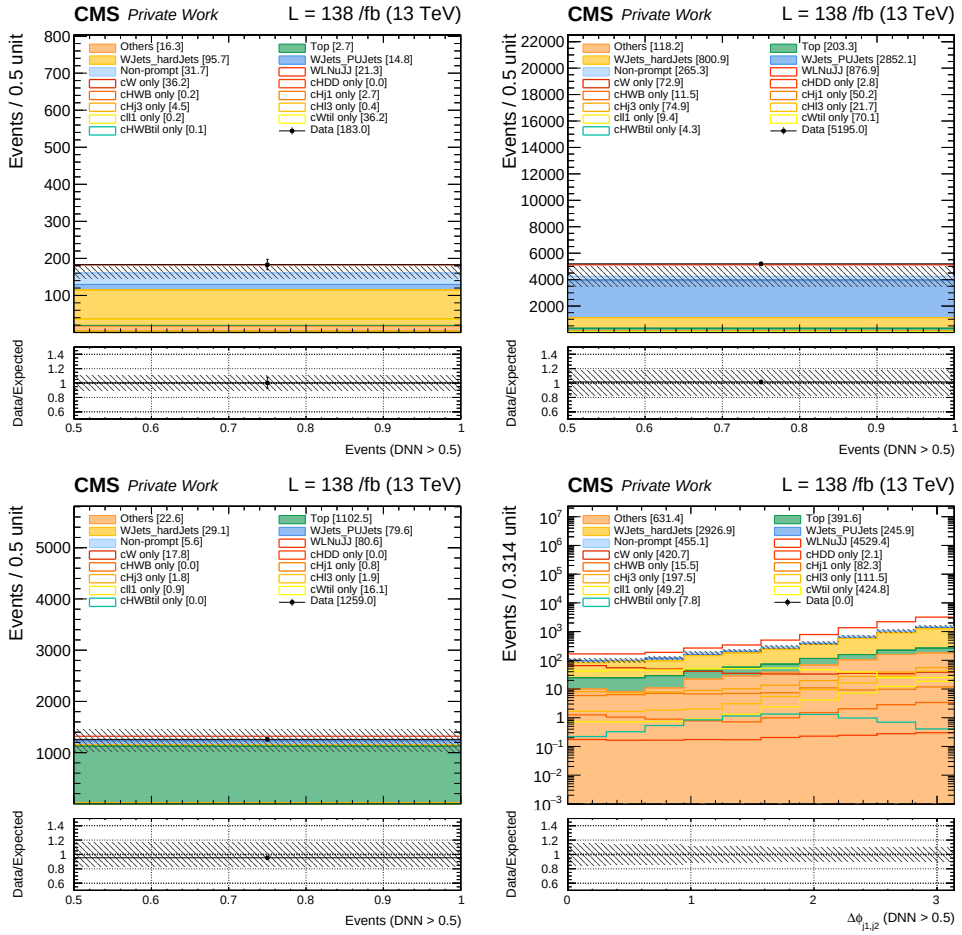


Figure 5.3: Inputs to the EFT fit in the electron channel. The SM EW component is shown stacked, while pure EFT contributions (quadratic terms) are overlaid.

5.3.2 Results

Figure 5.5 presents the expected one-dimensional likelihood scans for each operator, obtained from the combined fit over the full Run II dataset. Figure 5.6 shows the two-dimensional likelihood scans for selected pairs of operators, illustrating the correlations among the corresponding Wilson coefficients. The numerical values of the expected 68% and 95% confidence intervals derived from these scans are summarised in Table 5.3.

The corresponding post-fit impact plots, quantifying the effect of systematic uncertainties on each operator, are reported in Appendix E.

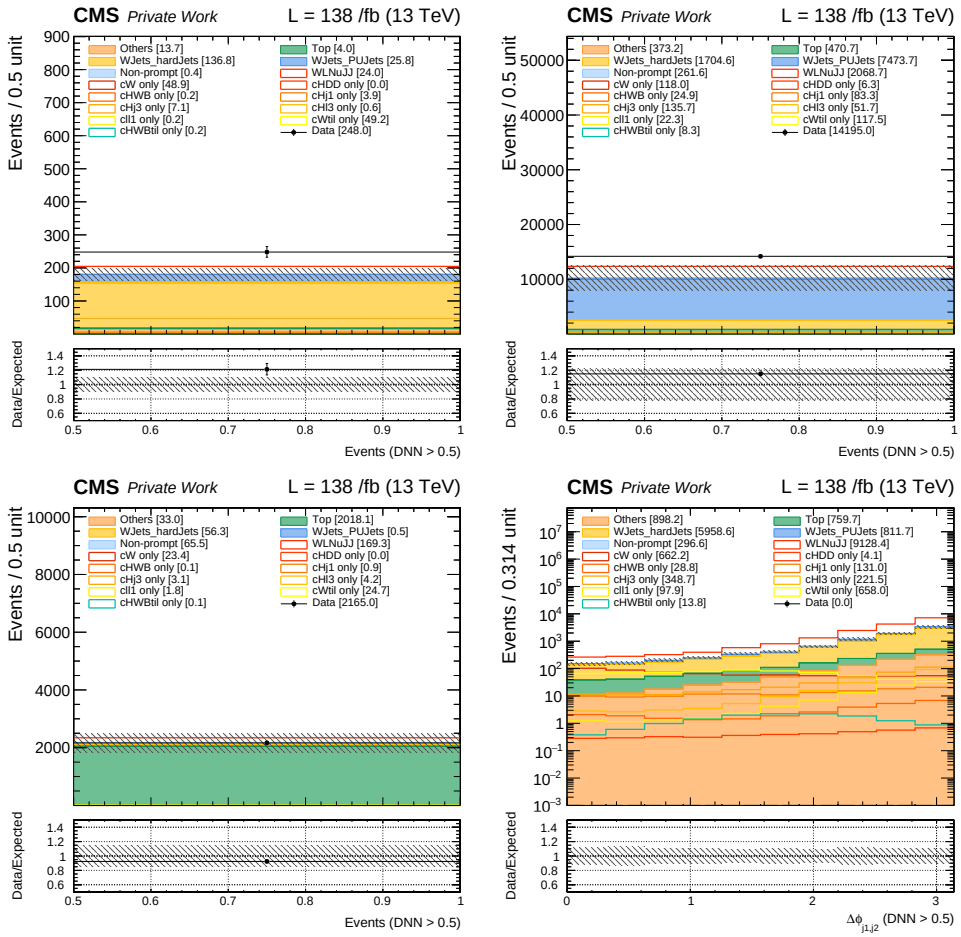


Figure 5.4: Inputs to the EFT fit in the muon channel. The SM EW component is shown stacked, while pure EFT contributions (quadratic terms) are overlaid.

Operator	68% CL interval	95% CL interval
$\mathcal{O}_W$	$[-0.12, 0.11]$	$[-0.25, 0.22]$
$\mathcal{O}_{HWB}$	$[-1.39, 1.91]$	$[-2.52, 5.39]$
$\mathcal{O}_{HDD}$	$[-1.86, 1.87]$	$[-3.76, 4.22]$
$\mathcal{O}_{Hj}^{(1)}$	$[-0.94, 1.59]$	$[-1.48, 2.21]$
$\mathcal{O}_{Hj}^{(3)}$	$[-0.23, 0.25]$	$[-0.44, 0.51]$
$\mathcal{O}_{Hl}^{(3)}$	$[-0.42, 0.42]$	$[-0.82, 0.86]$
$\mathcal{O}_{\ell\ell}^{(1)}$	$[-0.48, 0.48]$	$[-0.97, 0.95]$
$\mathcal{O}_{\widetilde{W}}$	$[-0.47, 0.46]$	$[-0.67, 0.66]$
$\mathcal{O}_{\widetilde{HWB}}$	$[-2.27, 2.26]$	$[-4.15, 4.43]$

Table 5.3: Expected 68% and 95% confidence level (CL) intervals on the Wilson coefficients obtained from the individual $W \to \ell\nu jj$ EFT interpretation.

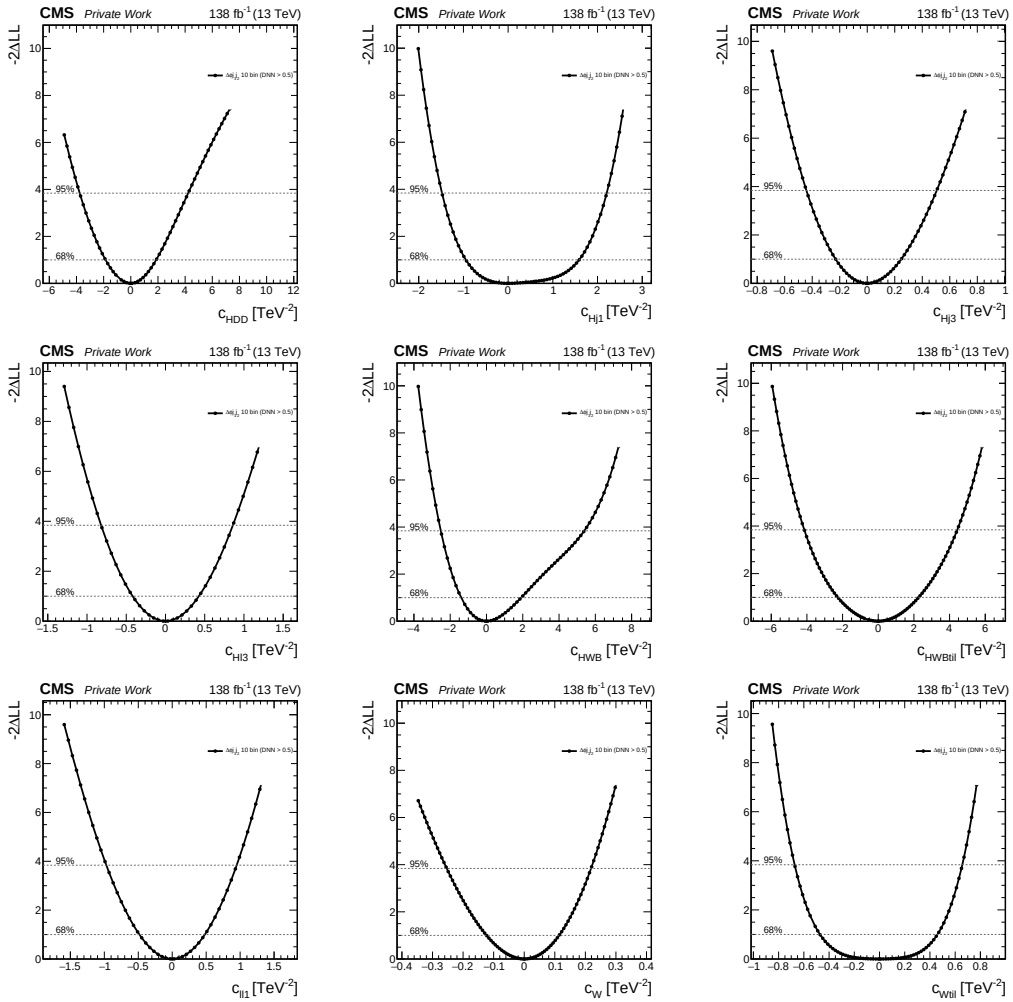


Figure 5.5: Expected one-dimensional likelihood scans for each EFT operator, obtained from the combined fit over all data-taking years. Each curve represents the profile likelihood as a function of a single Wilson coefficient, with all other operators fixed to their SM values.

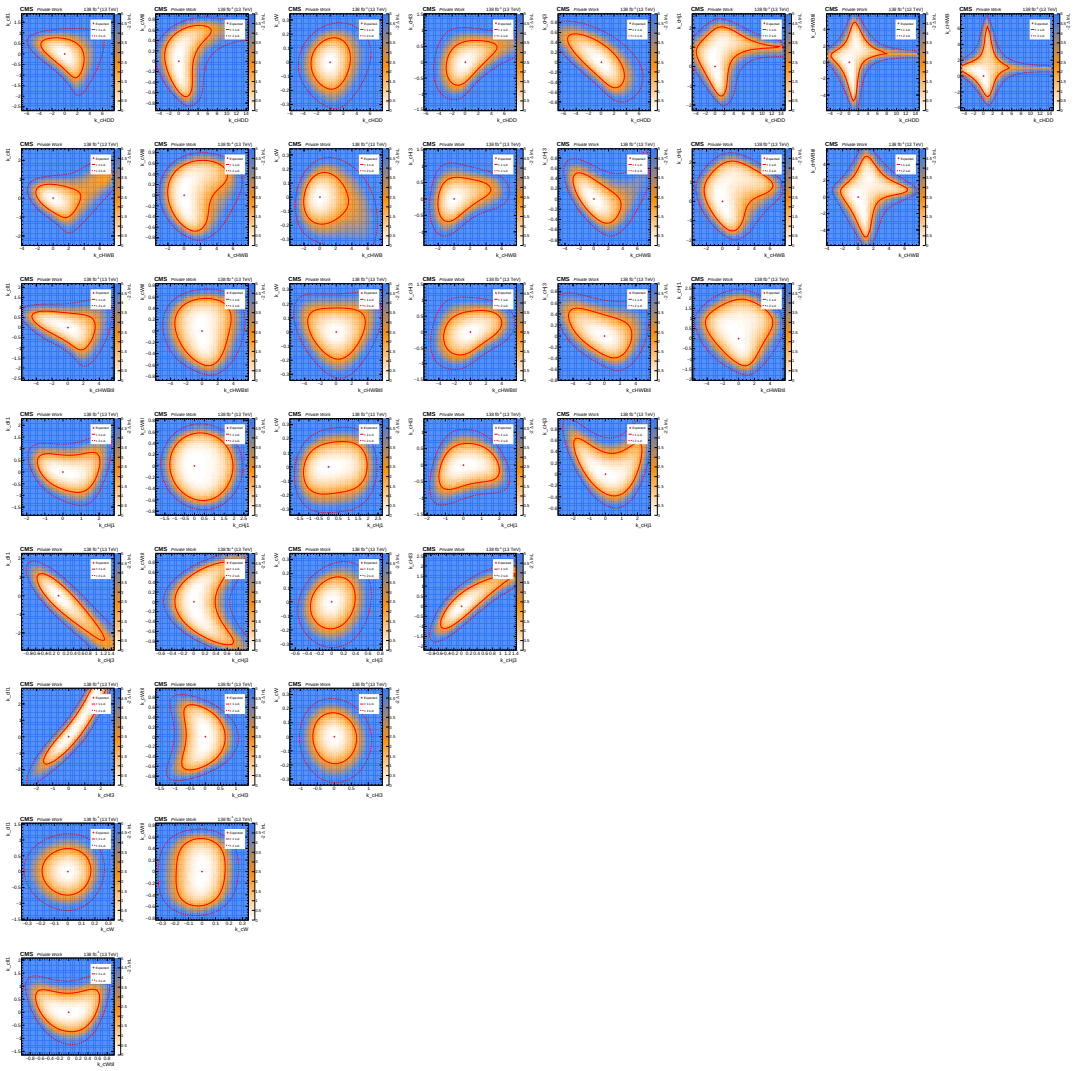


Figure 5.6: Two-dimensional likelihood scans for selected pairs of EFT operators. The scans are arranged in a matrix format, with each panel corresponding to the pair of operators defined by the row and column labels. Only the upper-triangular part of the matrix is shown to avoid duplication. The contours represent the expected confidence levels at 1,2 sigmas.

5.4 Multi-channel EFT interpretation

The single-channel measurement described in the previous section is further extended through a global combination of multiple EW analyses interpreted within the SMEFT framework. This combined approach exploits the complementary sensitivities of different processes to anomalous gauge couplings, providing tighter and more robust constraints on the dimension-six bosonic operators. The following subsections summarise the motivation, inputs, and results of this combined interpretation, highlighting the specific contribution of the Wjj channel.

5.4.1 Motivation for a global approach

Individual measurements of EW boson scattering and fusion processes provide complementary information on the gauge sector of the SM and possible deviations induced by new physics. In SMEFT, these deviations are encoded in higher-dimensional operators whose effects can manifest simultaneously in multiple vertices of the EW Lagrangian. As a consequence, a single operator can contribute coherently to several production mechanisms and decay modes, affecting observables across different analyses and final states. Since new physics at higher energy scales is expected to modify several couplings at once, a comprehensive interpretation therefore requires a global treatment in which multiple operators and processes are fitted together.

However, within a single measurement, the sensitivity is typically limited to a subset of operators, and strong correlations or redundancies can arise among them. This makes it challenging to perform a fully general simultaneous fit in a single channel. A statistical combination of multiple EW measurements provides a practical and physically motivated solution: it exploits the complementarity of different final states and observables, extends the explored phase space, and improves the overall precision and stability of the SMEFT interpretation.

5.4.2 Inputs of the combination

The combination encompasses nine CMS analyses based on the full Run II dataset of proton–proton collisions at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}. The input channels include vector-boson scattering (VBS), vector-boson fusion (VBF), diboson, and triboson processes. Each provides detector-level templates parameterized in terms of six CP-even bosonic operators in the Warsaw basis:

All signal processes are simulated at leading order using the SMEFTsim model, ensuring consistent operator definitions and coupling conventions across analyses. The statistical interpretation is performed with the CMS Combine frame-

Analysis	Reference	Final state	Relevant operators
VBS-SSWW(e,μ)	[147]	$\ell^\pm\ell^\pm\nu\nu jj$	$\mathcal{O}_{HD}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{H\Box}$, $\mathcal{O}_W$
VBS-SSWW(τ_h)	[148]	$\ell^\pm\tau_h^\pm\nu\nu jj$	$\mathcal{O}_{HD}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{H\Box}$, $\mathcal{O}_W$
VBS-OSWW	[149]	$\ell^\pm\ell^\mp\nu\nu jj$	$\mathcal{O}_{HB}$, $\mathcal{O}_{HD}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{H\Box}$, $\mathcal{O}_W$
VBS-WZ	[147]	$\ell^+\ell^-\ell^\pm\nu jj$	$\mathcal{O}_{HD}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{H\Box}$, $\mathcal{O}_W$
VBS-WV	[150], [151]	$\ell^\pm\nu jjjj$	$\mathcal{O}_{HB}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{H\Box}$, $\mathcal{O}_W$
VBS-ZV	[152]	$\ell^+\ell^-jjjj$	$\mathcal{O}_{HD}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{H\Box}$, $\mathcal{O}_W$
VBF-W	SMP-24-018	$\ell^\pm\nu jj$	$\mathcal{O}_{HD}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_W$
VBF-Z	SMP-24-016	$\ell^+\ell^-jj$	$\mathcal{O}_{HD}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_W$
WW	SMP-24-008	$\ell^+\ell^-\nu\nu$	$\mathcal{O}_{HD}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_W$
VVV/γ	[153]	$6j$, $\ell^\pm\nu jjjj$, $\ell^+\ell^-jjjj$, $\ell^+\ell^-\nu\nu jj$, $\ell^\pm\ell^\pm\nu\nu jj$	$\mathcal{O}_{HB}$, $\mathcal{O}_{HD}$, $\mathcal{O}_{HW}$, $\mathcal{O}_{HWB}$, $\mathcal{O}_{H\Box}$, $\mathcal{O}_W$

Table 5.4: Summary of the analyses entering the bosonic EFT combination, targeting VBS, VBF, diboson, and triboson production. All measurements use the full Run II dataset of 138 fb^{-1}. Final states are specified where ℓ denotes a light lepton (e,μ) and τ_h a hadronically decaying τ lepton. The last column lists the dimension-six operators, in the Warsaw basis, contributing to each process.

work, following the standard profile-likelihood approach introduced in Sec 1.9. Systematic uncertainties are consistently treated across all input channels, and dedicated studies have shown that the impact of inter-channel correlations on the final results is negligible.

The combination includes processes sensitive to both trilinear and quartic gauge couplings, with complementary kinematic regimes: VBS channels probe the high-energy tails of diboson scattering, while VBF channels provide enhanced sensitivity to EW vertex corrections at lower jet multiplicities.

5.4.3 Contribution from the $W \to \ell\nu jj$ channel

The definition of the analysis regions used as input to the combination follows the same strategy described in Section 3.4.5, with a small modification introduced to ensure orthogonality with the VBS WV channel. Events containing an AK8 jet with a mass compatible with a hadronically decaying W or Z boson are vetoed. AK8 jets are reconstructed with a larger distance parameter ($R = 0.8$) compared to the standard AK4 jets ($R = 0.4$) described in Section 3.4, and are typically used to identify hadronically decaying, highly boosted vector bosons. In the signal definition of the EW production of the W + jets, the W boson always decays leptonically, so this veto simply rejects events in which an additional large-radius jet accidentally falls in the W/Z mass window ($70 < m_J < 150$ GeV). The selection has a negligible impact on the signal efficiency (about 98%) but guarantees that the event sample is statistically independent from that used in the VBS WV analysis.

The inputs provided to the combination follow the same categorisation scheme adopted in the stand–alone analysis. Events in the CRs are used to constrain the

background normalisations and systematic uncertainties, while events in the SR enter the SMEFT interpretation. In the SR, the templates are constructed as two-dimensional histograms in the plane of the DNN output and the azimuthal separation of the VBF jets, $\Delta\phi_{jj}$, with a (5×5) binning that is common across all years and lepton flavours. This representation preserves the discriminating power of the DNN while retaining the sensitivity of $\Delta\phi_{jj}$ to the angular structure of the EW interaction.

The $W \to \ell\nu jj$ channel enters the combination through a dedicated likelihood constructed using the same control regions defined in 3.4.5 and a two-dimensional template in (DNN, $\Delta\phi_{jj}$) for the signal region. This configuration retains the event categorisation of the SM analysis while providing enhanced sensitivity to possible SMEFT-induced deviations. The input distributions entering the combination are shown in Figs. 5.7 and 5.8.

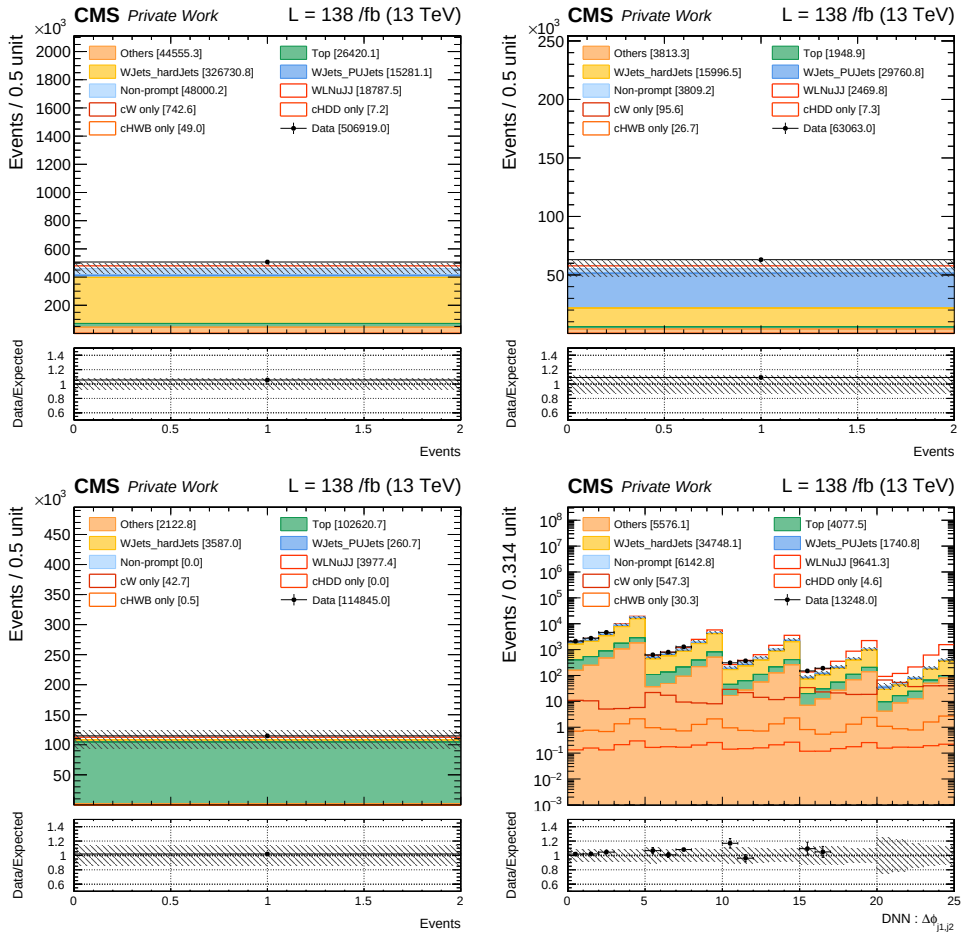


Figure 5.7: Inputs to the EFT fit in the electron channel. The SM EW component is shown stacked, pure EFT contributions (quadratic terms) are overlaid.

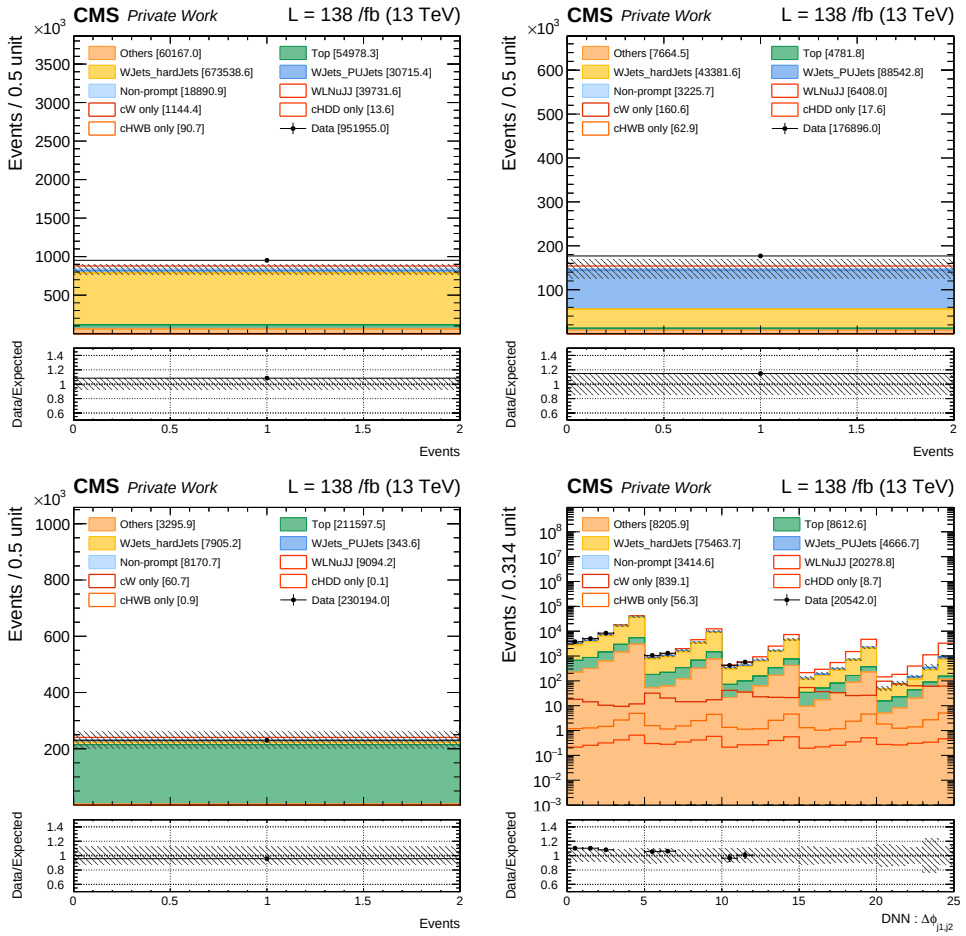


Figure 5.8: Inputs to the EFT fit in the muon channel. The SM EW component is shown stacked, pure EFT contributions (quadratic terms) are overlaid.

5.4.4 Results

The statistical interpretation is performed through a profile-likelihood fit using the templates described in the previous section. The fit is carried out both individually for each Wilson coefficient and simultaneously for all six operators included in the bosonic EFT model. The results are presented in terms of one-dimensional (1D) and two-dimensional (2D) likelihood scans, as well as 1D profiled scans in which one operator is probed while all others are profiled.

In addition to the nominal quadratic SMEFT model, corresponding fits are also performed using a linearised formulation, in which the expansion of the Lagrangian is truncated at $\mathcal{O}(1/\Lambda^2)$. In this approximation, only the interference terms between the SM and the dimension-six operators are retained, while pure quadratic contributions, which scale as $\mathcal{O}(1/\Lambda^4)$, are neglected. This approach is theoretically motivated, as quadratic terms are of the same formal order as the contributions from dimension-eight operators and should ideally be treated consistently within a complete higher-order expansion.

Figures 5.9, 5.10, and 5.11 summarise the corresponding 1D, 2D, and profiled

likelihood scans for the six bosonic operators considered in the combination. All scans are performed using the Asimov dataset under the SM hypothesis, with confidence intervals corresponding to 68% and 95% CL.

In the one-dimensional scans, the VBF W analysis provides the dominant contribution to the overall sensitivity in the three operators to which it is directly sensitive, namely c_{HDD}, c_{HWB}, and c_W. In particular, it almost entirely drives the constraints on c_{HDD} and c_{HWB}, and contributes significantly to c_W, with its minimum located close to that of the global fit.

The two-dimensional scans reveal a strong correlation between c_{HDD} and c_{HWB} within the VBF W channel alone, a feature that is partially resolved in the global combination thanks to the inclusion of other EW processes, notably VBF Z and WW production. These additional channels constrain complementary directions in the parameter space, effectively reducing the degeneracy observed in the single-channel fits and leading to narrower confidence regions. As a result, the profiled one-dimensional scans exhibit only a moderate degradation in sensitivity with respect to the individual 1D fits—unlike in standalone analyses, where degenerate directions in the likelihood often produce substantially weaker constraints.

Finally, the linearised model shows performance comparable to the full quadratic SMEFT formulation in regions where the sensitivity is dominated by VBF processes. This behaviour is expected, since in the VBF Wjj topology, the EFT sensitivity arises mainly from interference terms rather than quadratic contributions. This can be understood by noting that the typical momentum transfers are well below the new-physics scale Λ, making the linear term dominant over the quadratic one, which scales as $\mathcal{O}(1/\Lambda^4)$. In contrast, in genuine VBS processes, the higher energies involved enhance the relative importance of quadratic contributions.

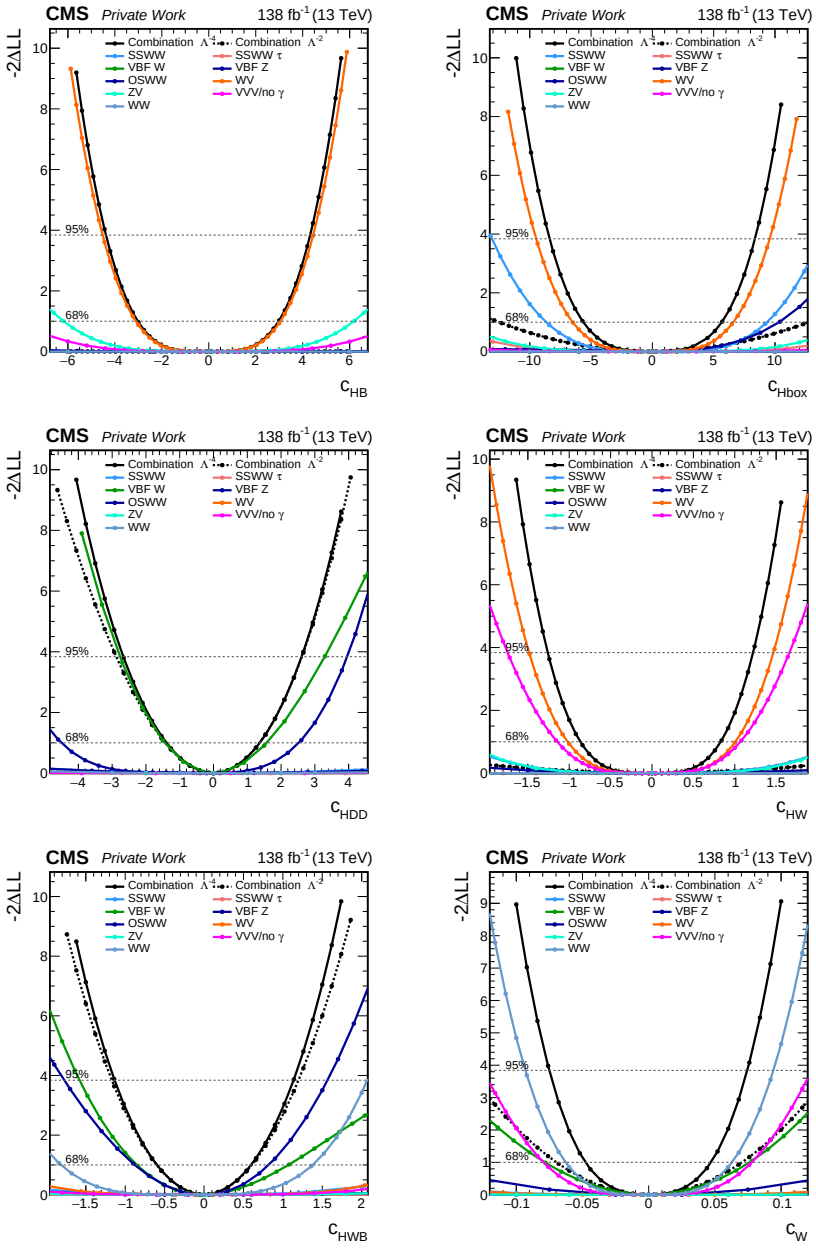


Figure 5.9: Expected one-dimensional likelihood scans for each EFT operator, obtained from the combined fit over all data-taking years. Each curve represents the profile likelihood as a function of a single Wilson coefficient, with all other operators fixed to their SM values.

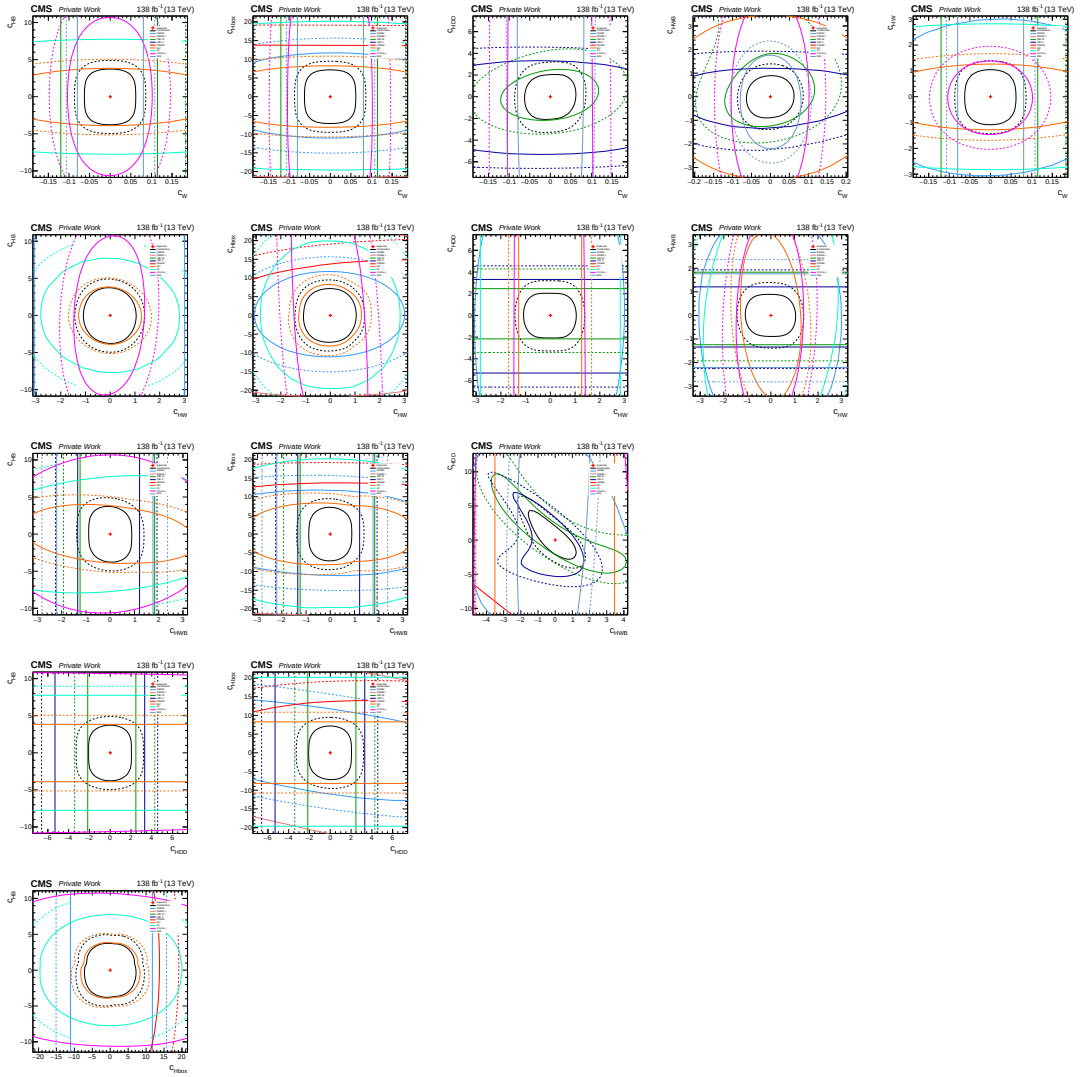


Figure 5.10: Two-dimensional likelihood scans for selected pairs of EFT operators. The scans are arranged in a matrix format, with each panel corresponding to the pair of operators defined by the row and column labels. Only the upper-triangular part of the matrix is shown to avoid duplication. The contours represent the expected confidence levels at 1,2 sigmas.

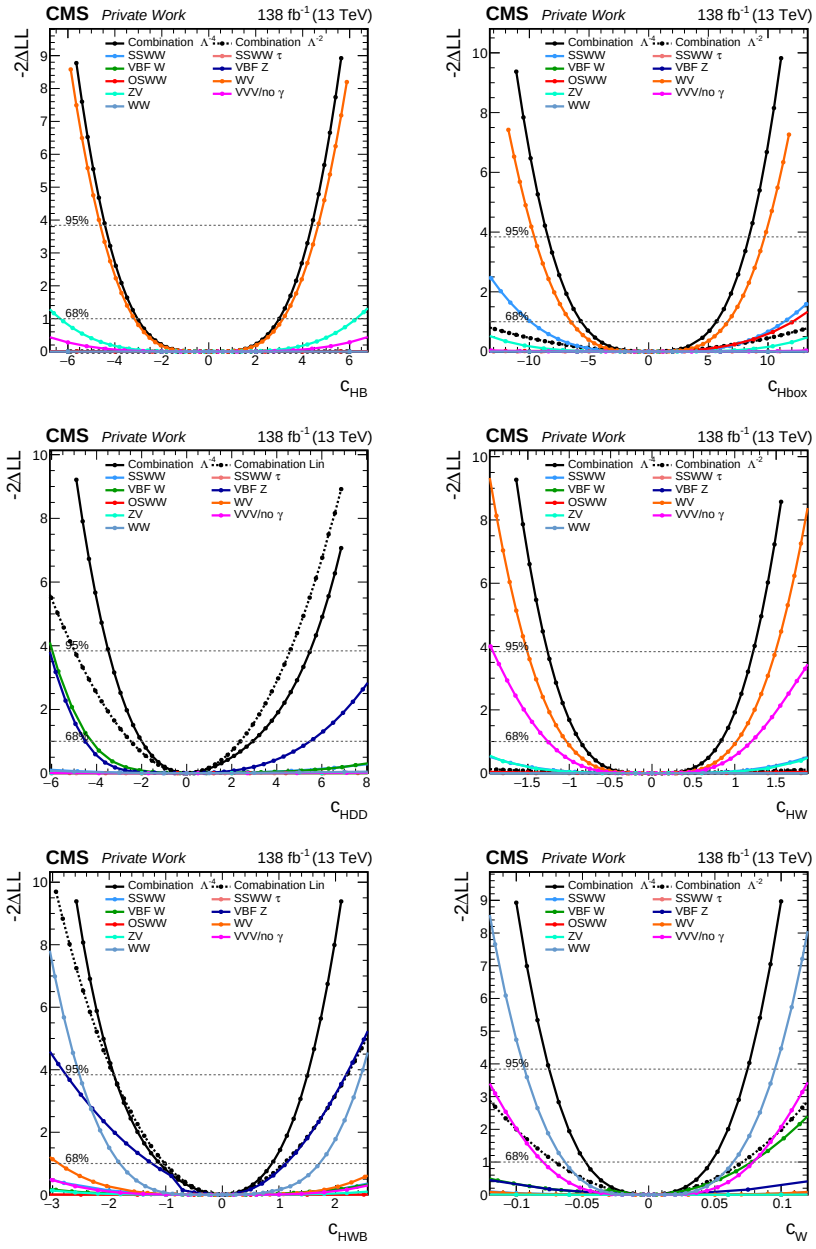


Figure 5.11: Expected one-dimensional likelihood scans for each EFT operator, obtained from the combined fit over all data-taking years. Each curve represents the profile likelihood as a function of a single Wilson coefficient, with all other operators fixed to their SM values.

Chapter 6

Conclusions

This work presents a precise measurement of the purely EW production of a W boson in association with two jets, performed using the full Run II dataset collected by the CMS experiment at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}.

The inclusive EW $W{+}jj$ cross section is measured to be

$$\begin{aligned}\sigma_{\text{EW}}(W{+}jj) &= 23.0 \pm 2.1\ (\text{theo}) \pm 1.1\ (\text{exp}) \pm 0.5\ (\text{stat}) \\ &= 23.0 \pm 2.4\ (\text{syst}) \pm 0.5\ (\text{stat})\ \text{pb}.\end{aligned}$$

Although the result is referred to as an "inclusive" EW $W{+}jj$ cross section, the process is only well defined within the generator-level phase space used in the gridpack. The EW $W{+}jj$ simulation imposes minimal LHE requirements to avoid infrared and collinear divergences, namely $p_T^j > 10$ GeV, $p_T^\gamma > 10$ GeV, $m_{jj} > 200$ GeV, and $m(\ell\nu) > 50$ GeV. These conditions are part of the theoretical definition of the process and not analysis selections. The measurement should therefore be interpreted as inclusive within this generator-level phase space.

The measurement achieves a total precision of about 10%, dominated by theoretical uncertainties from QCD renormalisation and factorisation scales, while experimental and calibration effects play a secondary role.

A fiducial measurement, defined within the phase space reported in Table 4.1, yields

$$\begin{aligned}\sigma_{\text{EW}}^{\text{fid}}(W{+}jj) &= 1.80 \pm 0.10\ (\text{theo}) \pm 0.07\ (\text{exp}) \pm 0.04\ (\text{stat}) \\ &= 1.80 \pm 0.12\ (\text{syst}) \pm 0.04\ (\text{stat})\ \text{pb}.\end{aligned}$$

Beyond the inclusive results, differential cross sections have been extracted for observables sensitive to the EW dynamics: the dijet invariant mass (m_{jj}), the transverse momentum of the dijet system (p_T^{jj}), the transverse momenta of the two leading jets (p_T^{j1} and p_T^{j2}), their pseudorapidity and azimuthal separations ($\Delta\eta_{jj}$ and $\Delta\phi_{jj}$), the W-boson transverse momentum (p_T^W), and the charged-lepton transverse momentum (p_T^ℓ). For the dijet invariant mass, most

strongly correlated with the DNN discriminant, an alternative unfolding strategy based on a multiclass neural network was developed and compared to the standard reconstructed-variable approach, which is used for all other observables. The unfolded spectra are compared to theoretical predictions from MADGRAPH5_aMC@NLO, HERWIG7, and SHERPA.

Following the SM measurement, the analysis was extended to a search for possible deviations from the Standard Model within the framework of the Standard Model Effective Field Theory (SMEFT). To provide a more comprehensive and global picture of EW interactions, the $W \to \ell\nu jj$ channel was included in the first CMS-wide bosonic EFT combination, in which multiple EW processes are fitted simultaneously using fully simulated EFT samples with a consistent treatment of theoretical and experimental uncertainties. The resulting 68% CL profiled limits on the dimension-six operators are reported in Table 6.1, together with their corresponding interpretation in terms of the lower bounds on the new-physics scale Λ.

Operator	**Profiled limits**	**Lower bound on Λ [TeV]**		
	68% CL	$\lvert c_i\rvert = 0.1$	$\lvert c_i\rvert = 1$	$\lvert c_i\rvert = 4\pi$
$\mathcal{O}_{HB}$	$[-3.04, 3.00]$	0.18	0.57	2.03
$\mathcal{O}_{H\Box}$	$[-5.66, 5.68]$	0.13	0.42	1.49
$\mathcal{O}_{HDD}$	$[-1.99, 2.93]$	0.18	0.58	2.07
$\mathcal{O}_{HW}$	$[-0.87, 0.83]$	0.34	1.07	3.81
$\mathcal{O}_{HWB}$	$[-1.11, 0.88]$	0.30	0.95	3.37
$\mathcal{O}_{W}$	$[-0.048, 0.047]$	1.45	4.58	16.2

Table 6.1: Expected 68% CL profiled limits on the dimension-six operators from the CMS bosonic EFT combination, and corresponding lower limits on the scale of new physics Λ under three benchmark coupling scenarios: weakly coupled ($|c_i| = 0.1$), order–one ($|c_i| = 1$), and strongly coupled ($|c_i| = 4\pi$).

Under the natural assumption of order–one couplings, the combination constrains new EW interactions up to scales of a few TeV, reaching beyond 4 TeV for the most precisely determined operator $\mathcal{O}_W$. The Wjj channel provides a significant contribution to the global EW combination, which itself represents an important step toward a more complete and unified understanding of the EW sector.

Bibliography

[1] Jakob Schwichtenberg. *Physics from Symmetry*. Undergraduate Lecture Notes in Physics. Cham: Springer International Publishing, 2018. ISBN: 978-3-319-66630-3, 978-3-319-66631-0. DOI: 10.1007/978-3-319-66631-0.

[2] William Fulton and Joe Harris. "Lie Algebras and Lie Groups". In: *Representation Theory: A First Course*. New York, NY: Springer New York, 2004, pp. 104–120. ISBN: 978-1-4612-0979-9. DOI: 10.1007/978-1-4612-0979-9_8. URL: https://doi.org/10.1007/978-1-4612-0979-9_8.

[3] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-201-50397-5.

[4] Armand Borel. "Essays in the history of Lie groups and algebraic groups". In: AMS and London Mathematical Society, 2001. ISBN: 978-0-201-50397-5. URL: https://api.semanticscholar.org/CorpusID:41987826.

[5] Albert Einstein. "On the Electrodynamics of Moving Bodies". In: *Annalen der Physik* (1905), pp. 891–921.

[6] Bruce D. Popp. "On the Dynamics of the Electron". In: *Henri Poincaré: Electrons to Special Relativity: Translation of Selected Papers and Discussion*. Cham: Springer International Publishing, 2020, pp. 45–102. ISBN: 978-3-030-48039-4. DOI: 10.1007/978-3-030-48039-4_5. URL: https://doi.org/10.1007/978-3-030-48039-4_5.

[7] Howard Stein. "On Einstein–Minkowski Space–Time". In: *The Journal of Philosophy* 65.1 (1968), pp. 5–23. ISSN: 0022362X. URL: http://www.jstor.org/stable/2024512 (visited on 07/31/2023).

[8] L. D. Landau and E. M. Lifschits. *The Classical Theory of Fields*. Vol. Volume 2. Course of Theoretical Physics. Oxford: Pergamon Press, 1975. ISBN: 978-0-08-018176-9.

[9] Steven Weinberg. *The Quantum Theory of Fields*. Vol. 1. Cambridge University Press, 1995. DOI: 10.1017/CBO9781139644167.

[10] Hendrik B. G. Casimir. "Rotation of a Rigid Body in Quantum-mechanics". In: *Nature* 129 (1932), pp. 780–780. URL: https://api.semanticscholar.org/CorpusID:4103911.

[11] Taeseung Choi and Sam Young Cho. *Spin operators and representations of the Poincaré group*. 2020. arXiv: 1807.06425 [physics.gen-ph].

[12] Sergey I Kryuchkov, Nathan A Lanfear, and Sergei K Suslov. "The role of the Pauli–Lubań ski vector for the Dirac, Weyl, Proca, Maxwell and Fierz–Pauli equations". In: *Physica Scripta* 91.3 (Feb. 2016), p. 035301. DOI: 10.1088/0031-8949/91/3/035301. URL: https://doi.org/10.1088%2F0031-8949%2F91%2F3%2F035301.

[13] Paul A. M. Dirac. "The Lagrangian in quantum mechanics". In: *Phys. Z. Sowjetunion* 3 (1933), pp. 64–72.

[14] Richard Phillips Feynman and Albert Roach Hibbs. *Quantum mechanics and path integrals*. International series in pure and applied physics. New York, NY: McGraw-Hill, 1965. URL: https://cds.cern.ch/record/100771.

[15] Julian Seymour Schwinger. "Quantum Kinematics And Dynamics". In: W.A. Benjamin Inc. Publisher, 1970. URL: https://api.semanticscholar.org/CorpusID:117850892.

[16] Richard Phillips Feynman. "The Principle of Least Action in Quantum Mechanics." PhD thesis. Princeton University, New Jersey, May 1942.

[17] E. Noether. "Invariante Variationsprobleme". ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: http://eudml.org/doc/59024.

[18] Kenneth G. Wilson. "Confinement of quarks". In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI: 10.1103/PhysRevD.10.2445. URL: https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[19] Collectif. "Les connexions infinitésimales dans un espace fibré différentiable". fr. In: *Séminaire Bourbaki : années 1948/49 - 1949/50 - 1950/51, exposés 1-49*. Séminaire Bourbaki 1. talk:24. Société mathématique de France, 1952. URL: http://www.numdam.org/item/SB_1948-1951__1__153_0/.

[20] Paul A. M. Dirac. "The quantum theory of the electron". In: *Proc. Roy. Soc. Lond. A* 117 (1928), pp. 610–624. DOI: 10.1098/rspa.1928.0023.

[21] P. A. M. Dirac. "A Theory of Electrons and Protons". In: *Proc. Roy. Soc. Lond. A* 126.801 (1930), pp. 360–365. DOI: 10.1098/rspa.1930.0013.

[22] Al. Proca. "Particules libres photons et particules " charge pure "". In: *J. Phys. Radium* 8.1 (1937), pp. 23–28. DOI: 10.1051/jphysrad:019370080102300. URL: https://hal.science/jpa-00233468.

[23] Al. Proca. "Sur la théorie ondulatoire des électrons positifs et négatifs". In: *J. Phys. Radium* 7.8 (1936), pp. 347–353. DOI: 10.1051/jphysrad:0193600708034700. URL: https://hal.science/jpa-00233436.

[24] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[25] Gerard 't Hooft and M. J. G. Veltman. "Regularization and Renormalization of Gauge Fields". In: *Nucl. Phys. B* 44 (1972), pp. 189–213. DOI: 10.1016/0550-3213(72)90279-9.

[26] Gerard 't Hooft. "Renormalization of Massless Yang-Mills Fields". In: *Nucl. Phys. B* 33 (1971), pp. 173–199. DOI: 10.1016/0550-3213(71)90395-6.

[27] Murray Gell-Mann. "A Schematic Model of Baryons and Mesons". In: *Phys. Lett.* 8 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3.

[28] 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.

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

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

[31] J. H. Christenson et al. "Evidence for the 2π Decay of the K_2^0 Meson". In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[32] Sheldon Lee Glashow. "The renormalizability of vector meson interactions". In: *Nuclear Physics* 10 (1959), pp. 107–117. URL: https://api.semanticscholar.org/CorpusID:123163322.

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

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

[35] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[36] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[37] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[38] G. Altarelli. *The Standard Electroweak Theory and Beyond*. 2000. arXiv: hep-ph/0011078 [hep-ph].

[39] Sidney Coleman and Erick Weinberg. "Radiative Corrections as the Origin of Spontaneous Symmetry Breaking". In: *Phys. Rev. D* 7 (6 Mar. 1973), pp. 1888–1910. DOI: 10.1103/PhysRevD.7.1888. URL: https://link.aps.org/doi/10.1103/PhysRevD.7.1888.

[40] J. Goldstone. "Field Theories with Superconductor Solutions". In: *Nuovo Cim.* 19 (1961), pp. 154–164. DOI: 10.1007/BF02812722.

[41] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[42] Thomas Appelquist and J. Carazzone. "Infrared singularities and massive fields". In: *Phys. Rev. D* 11 (10 May 1975), pp. 2856–2861. DOI: 10.1103/PhysRevD.11.2856. URL: https://link.aps.org/doi/10.1103/PhysRevD.11.2856.

[43] 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.

[44] Ilaria Brivio and Michael Trott. "The standard model as an effective field theory". In: *Physics Reports* 793 (Feb. 2019), pp. 1–98. DOI: 10.1016/j.physrep.2018.11.002. URL: https://doi.org/10.1016%2Fj.physrep.2018.11.002.

[45] Steven Weinberg. "Baryon- and Lepton-Nonconserving Processes". In: *Phys. Rev. Lett.* 43 (21 Nov. 1979), pp. 1566–1570. DOI: 10.1103/PhysRevLett.43.1566. URL: https://link.aps.org/doi/10.1103/PhysRevLett.43.1566.

[46] Elizabeth E. Jenkins, Aneesh V. Manohar, and Peter Stoffer. "Low-energy effective field theory below the electroweak scale: operators and matching". In: *Journal of High Energy Physics* 2018.3 (Mar. 2018). DOI: 10.1007/jhep03(2018)016. URL: https://doi.org/10.1007%2Fjhep03%282018%29016.

[47] B. Grzadkowski et al. "Dimension-six terms in the Standard Model Lagrangian". In: *Journal of High Energy Physics* 2010.10 (Oct. 2010). DOI: 10.1007/jhep10(2010)085. URL: https://doi.org/10.1007%2Fjhep10%282010%29085.

[48] Gitte Elgaard-Clausen and Michael Trott. "On expansions in neutrino effective field theory". In: *Journal of High Energy Physics* 2017.11 (Nov. 2017). DOI: 10.1007/jhep11(2017)088. URL: https://doi.org/10.1007%2Fjhep11%282017%29088.

[49] L. F. Abbott and Mark B. Wise. "Effective Hamiltonian for nucleon decay". In: *Phys. Rev. D* 22 (9 Nov. 1980), pp. 2208–2212. DOI: 10.1103/PhysRevD.22.2208. URL: https://link.aps.org/doi/10.1103/PhysRevD.22.2208.

[50] Landon Lehman. "Extending the standard model effective field theory with the complete set of dimension-7 operators". In: *Physical Review D* 90.12 (Dec. 2014). DOI: 10.1103/physrevd.90.125023. URL: https://doi.org/10.1103%2Fphysrevd.90.125023.

[51] Landon Lehman and Adam Martin. "Low-derivative operators of the Standard Model effective field theory via Hilbert series methods". In: *Journal of High Energy Physics* 2016.2 (Feb. 2016). DOI: 10.1007/jhep02(2016)081. URL: https://doi.org/10.1007%2Fjhep02%282016%29081.

[52] Brian Henning et al. *2, 84, 30, 993, 560, 15456, 11962, 261485, ...: Higher dimension operators in the SM EFT*. 2019. arXiv: 1512.03433 [hep-ph].

[53] K. Hagiwara et al. "Low energy effects of new interactions in the electroweak boson sector". In: *Phys. Rev. D* 48 (5 Sept. 1993), pp. 2182–2203. DOI: 10.1103/PhysRevD.48.2182. URL: https://link.aps.org/doi/10.1103/PhysRevD.48.2182.

[54] Gian Francesco Giudice et al. "The strongly-interacting light Higgs". In: *Journal of High Energy Physics* 2007.06 (June 2007), pp. 045–045. DOI: 10.1088/1126-6708/2007/06/045. URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F06%2F045.

[55] Joan Elias-Miró et al. "Scaling and tuning of EW and Higgs observables". In: *Journal of High Energy Physics* 2014.5 (May 2014). DOI: 10.1007/jhep05(2014)019. URL: https://doi.org/10.1007%2Fjhep05%282014%29019.

[56] Nicola Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531. URL: https://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[57] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[58] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. "Remarks on the unified model of elementary particles". In: *Prog. Theor. Phys.* 28 (1962), pp. 870–880. DOI: 10.1143/PTP.28.870.

[59] B. Pontecorvo. "Inverse beta processes and nonconservation of lepton charge". In: *Zh. Eksp. Teor. Fiz.* 34 (1957), p. 247.

[60] Ilaria Brivio. "SMEFTsim 3.0 — a practical guide". In: *Journal of High Energy Physics* 2021.4 (Apr. 2021). DOI: 10.1007/jhep04(2021)073. URL: https://doi.org/10.1007%2Fjhep04%282021%29073.

[61] Ilaria Brivio and Michael Trott. "Scheming in the SMEFT. . . and a reparameterization invariance!" In: *Journal of High Energy Physics* 2017.7 (July 2017). DOI: 10.1007/jhep07(2017)148. URL: https://doi.org/10.1007%2Fjhep07%282017%29148.

[62] Ansgar Denner and Stefan Dittmaier. "Electroweak radiative corrections for collider physics". In: *Physics Reports* 864 (June 2020), pp. 1–163. DOI: 10.1016/j.physrep.2020.04.001. URL: https://doi.org/10.1016%2Fj.physrep.2020.04.001.

[63] Ilaria Brivio et al. "Electroweak input parameters". In: (Nov. 2021). arXiv: 2111.12515 [hep-ph].

[64] Rodrigo Alonso et al. "Renormalization Group Evolution of the Standard Model Dimension Six Operators III: Gauge Coupling Dependence and Phenomenology". In: *JHEP* 04 (2014), p. 159. DOI: 10.1007/JHEP04(2014)159. arXiv: 1312.2014 [hep-ph].

[65] Darius A. Faroughy et al. "Flavour symmetries in the SMEFT". In: *JHEP* 08 (2020), p. 166. DOI: 10.1007/JHEP08(2020)166. arXiv: 2005.05366 [hep-ph].

[66] J. M. Gerard. "FERMION MASS SPECTRUM IN SU(2)-L x U(1)". In: *Z. Phys. C* 18 (1983), p. 145. DOI: 10.1007/BF01572477.

[67] Alexander L. Kagan et al. "General Minimal Flavor Violation". In: *Phys. Rev. D* 80 (2009), p. 076002. DOI: 10.1103/PhysRevD.80.076002. arXiv: 0903.1794 [hep-ph].

[68] G. D'Ambrosio et al. "Minimal flavour violation: an effective field theory approach". In: *Nuclear Physics B* 645.1-2 (Nov. 2002), pp. 155–187. DOI: 10.1016/s0550-3213(02)00836-2. URL: https://doi.org/10.1016%2Fs0550-3213%2802%2900836-2.

[69] Gianluca Blankenburg, Gino Isidori, and Joel Jones-Pérez. "Neutrino masses and LFV from minimal breaking of U(3)5 and U(2)5 flavor symmetries". In: *The European Physical Journal C* 72.8 (Aug. 2012). DOI: 10.1140/epjc/s10052-012-2126-7. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-012-2126-7.

[70] Riccardo Barbieri et al. "Flavour physics from an approximate U(2)3 symmetry". In: *Journal of High Energy Physics* 2012.7 (July 2012). DOI: 10.1007/jhep07(2012)181. URL: https://doi.org/10.1007%2Fjhep07%282012%29181.

[71] D. Barducci et al. "Interpreting top-quark LHC measurements in the standard-model effective field theory". In: (Feb. 2018). Ed. by Juan Antonio Aguilar-Saavedra et al. arXiv: 1802.07237 [hep-ph].

[72] The CMS Collaboration. "The Combine tool for the interpretation of LHC results". In: *J. Phys. G* 52 (2025), p. 094001. DOI: 10.1088/1361-6471/ad4f3a. arXiv: 2404.06614 [hep-ex].

[73] J. Neyman and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philos. Trans. Roy. Soc. Lond. A* 231 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.

[74] 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].

[75] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.

[76] A. L. Read. "Presentation of search results: the CL(s) technique". In: *J. Phys. G* 28 (2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[77] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling". In: *eConf* C0303241 (2003). Ed. by L. Lyons and Muge Karagoz, MOLT007. arXiv: physics/0306116.

[78] Lorenzo Moneta et al. "The RooStats Project". In: *PoS* ACAT2010 (2010). Ed. by T. Speer et al., p. 057. DOI: 10.22323/1.093.0057. arXiv: 1009.1003 [physics.data-an].

[79] *Analytic Anomalous Coulings GitHub repository*. URL: https://github.com/amassiro/AnalyticAnomalousCoupling.

[80] "LHC Machine". In: *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

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

[82] W. Adam et al. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16.02 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304 [physics.ins-det].

[83] "CMS Technical Design Report for the Pixel Detector Upgrade". In: (Sept. 2012). Ed. by David Aaron Matzner Dominguez et al. DOI: 10.2172/1151650.

[84] "The CMS electromagnetic calorimeter project: Technical Design Report". In: (1997).

[85] P. Adzic et al. "Energy resolution of the barrel of the CMS electromagnetic calorimeter". In: *JINST* 2 (2007), P04004. DOI: 10.1088/1748-0221/2/04/P04004.

[86] Serguei Chatrchyan et al. "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JINST* 8 (2013), P09009. DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex].

[87] S. Abdullin et al. "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c". In: *Eur. Phys. J. C* 60 (2009). [Erratum: Eur.Phys.J.C 61, 353–356 (2009)], pp. 359–373. DOI: 10.1140/epjc/s10052-009-0959-5.

[88] "The CMS muon project: Technical Design Report". In: (1997).

[89] G. L. Bayatian et al. “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. In: (2006).

[90] Georges Charpak et al. “High Accuracy Localization of Minimum Ionizing Particles Using the Cathode Induced Charge Center of Gravity Readout”. In: *Nucl. Instrum. Meth.* 167 (1979). Ed. by Georges Charpak, p. 455. DOI: 10.1016/0029-554X(79)90227-1.

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

[92] “CMS Technical Design Report for the Level-1 Trigger Upgrade”. In: (June 2013). Ed. by A. Tapper and Darin Acosta.

[93] Alexandre Zabi et al. “The Phase-2 Upgrade of the CMS Level-1 Trigger”. In: (2020).

[94] A. Zabi et al. “The CMS Level-1 Calorimeter Trigger for the LHC Run II”. In: *JINST* 12.01 (2017), p. C01065. DOI: 10.1088/1748-0221/12/01/C01065.

[95] A. Svetek et al. “The Calorimeter Trigger Processor Card: the next generation of high speed algorithmic data processing at CMS”. In: *Journal of Instrumentation* 11.02 (Feb. 2016), p. C02011. DOI: 10.1088/1748-0221/11/02/C02011. URL: https://dx.doi.org/10.1088/1748-0221/11/02/C02011.

[96] D. Acosta et al. “The CMS Modular Track Finder boards, MTF6 and MTF7”. In: *JINST* 8 (2013), p. C12034. DOI: 10.1088/1748-0221/8/12/C12034.

[97] Vardan Khachatryan et al. “The CMS trigger system”. In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[98] “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”. In: (Apr. 2009).

[99] G. ’t Hooft and M. Veltman. “Regularization and renormalization of gauge fields”. In: *Nuclear Physics B* 44.1 (1972), pp. 189–213. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(72)90279-9. URL: https://www.sciencedirect.com/science/article/pii/0550321372902799.

[100] *Measurement of the cross section of the electroweak production of a W boson with two jets in pp collisions at sqrt(s) = 8 TeV*. Tech. rep. Geneva: CERN, 2015. URL: https://cds.cern.ch/record/2033044.

[101] “Measurement of electroweak production of a W boson and two forward jets in proton-proton collisions at $\sqrt{s} = 8$ TeV”. In: *Journal of High Energy Physics* (2016/11/24). DOI: https://doi.org/10.1007/JHEP11(2016)147.

[102] The ATLAS collaboration. "Measurements of electroweak Wjj production and constraints on anomalous gauge coupling with the ATLAS detector". In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5007-2. URL: http://dx.doi.org/10.1140/epjc/s10052-017-5007-2.

[103] "Measurement of electroweak production of a W boson in association with two jets in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *The European Physical Journal C* (2020/01/18). DOI: 10.1140/epjc/s10052-019-7585-7.

[104] *Common analysis objects definitions and trigger efficiencies for the WW and HWW analyses using the Ultra Legacy reconstruction of the full Run-II dataset*. CMS Analysis Note CMS-AN-2022/169. 2022. URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2022/169.

[105] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: https://www.sciencedirect.com/science/article/pii/S0010465515000442.

[106] "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". In: (2018).

[107] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[108] 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.

[109] J. Alwall et al. "MadGraph 5: going beyond". In: *Journal of High Energy Physics* 2011.6 (2011). DOI: 10.1007/JHEP06(2011)128. URL: https://doi.org/10.1007/JHEP06(2011)128.

[110] Alessandro Ballestrero et al. "Precise predictions for same-sign W-boson scattering at the LHC". In: *The European Physical Journal C* 78.8 (2018), p. 671. DOI: 10.1140/epjc/s10052-018-6136-y. URL: https://doi.org/10.1140/epjc/s10052-018-6136-y.

[111] Michelangelo L. Mangano et al. "ALPGEN, a generator for hard multiparton processes in hadronic collisions". In: *JHEP* 07 (2003), p. 001. DOI: 10.1088/1126-6708/2003/07/001. eprint: hep-ph/0206293.

[112] *Search for the EW production of a VW pair plus two jets in the semi-leptonic $l\nu jj$ channel with full Run-II data*. CMS Analysis Note CMS-AN-2019/239. 2019. URL: https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/239.

[113] Simone Alioli et al. "NLO vector-boson production matched with shower in POWHEG". In: *Journal of High Energy Physics* 2008.07 (July 2008), p. 060. DOI: 10.1088/1126-6708/2008/07/060. URL: https://dx.doi.org/10.1088/1126-6708/2008/07/060.

[114] Paolo Nason. "A new method for combining NLO QCD with shower Monte Carlo algorithms". In: *Journal of High Energy Physics* 2004.11 (Dec. 2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. URL: https://dx.doi.org/10.1088/1126-6708/2004/11/040.

[115] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. URL: https://dx.doi.org/10.1088/1126-6708/2007/11/070.

[116] S. Alioli, P. Nason, and C. et al. Oleari. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (June 2010). DOI: https://doi.org/10.1007/JHEP06(2010)043.

[117] 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].

[118] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). ISSN: 1029-8479. DOI: 10.1007/jhep12(2012)061. URL: http://dx.doi.org/10.1007/JHEP12(2012)061.

[119] Y. Gao et al. "Spin determination of single-produced resonances at hadron colliders". In: *Phys. Rev. D* 81 (2010), p. 075022. DOI: 10.1103/PhysRevD.81.075022. arXiv: 1001.3396 [hep-ph].

[120] I. Anderson et al. "Constraining anomalous HVV interactions at proton and lepton colliders". In: *Phys. Rev. D* 89 (2014), p. 035007. DOI: 10.1103/PhysRevD.89.035007. arXiv: 1309.4819 [hep-ph].

[121] *Reweighting recipe to emulate Level 1 ECAL and Muon prefiring.* URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe.

[122] *Performance of electron reconstruction and selection with 2017 data.* Tech. rep. CMS-PAS-EGM-17-001. CERN, 2017.

[123] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. URL: https://dx.doi.org/10.1088/1126-6708/2008/04/063.

[124] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. DOI: 10.1088/1748-0221/15/12/p12012. URL: https://doi.org/10.1088%2F1748-0221%2F15%2F12%2Fp12012.

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

[126] BTV POG, ed. *b-Tagging recommendations for Ultra Legacy*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation#UltraLegacy%5C_scale%5C_factor%5C_uncerta.

[127] Daniele Bertolini et al. "Pileup per particle identification". In: *Journal of High Energy Physics* 2014.10 (2014), p. 59. DOI: 10.1007/JHEP10(2014)059. URL: https://doi.org/10.1007/JHEP10(2014)059.

[128] *Noise Filter Recommendations for Run II Run III*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2#Moriond_2018.

[129] Scott M. Lundberg et al. "From local explanations to global understanding with explainable AI for trees". In: *Nature Machine Intelligence* 2.1 (2020), pp. 2522–5839.

[130] A.M. Sirunyan, A. Tumasyan, W. Adam, et al. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: *The European Physical Journal C* (2021). DOI: 10.1140/epjc/s10052-021-09538-2. URL: https://doi.org/10.1140/epjc/s10052-021-09538-2.

[131] "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV". In: (2018).

[132] "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV". In: (2019).

[133] Lumi POG, ed. *Luminosity recommendations for Run 2 analyses*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/LumiRecommendationsRun2.

[134] *MET uncertainty correlations among years*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/MissingETRun2Corrections#Uncertainty_correlations_among_y.

[135] *Recommendations on uncertainty correlations for JEC, JER for Run-2*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources#Run2_JEC_uncertainty_correlation.

[136] *Recommended Jet Energy Corrections and Uncertainties for Run-2*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC#Jet_Energy_Corrections_in_Run2.

[137] *Introduction to Jet Energy Corrections at CMS*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/IntroToJEC.

[138] *Jet energy scale uncertainty sources*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources.

[139] *Recommended Jet Energy Corrections and Uncertainties For Data and MC*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC#Jet_Energy_Corrections_in_Run2.

[140] *Jet Energy Resolution and Corrections - JERC Group*. URL: https://cms-jerc.web.cern.ch/Recommendations/#uncertainites-and-correlations.

[141] BTV POG, ed. *Apply b-tag scale factors*. URL: https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods.

[142] BTV POG, ed. *UltraLegacy scale factor uncertainties and correlations across years*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation#UltraLegacy_scale_factor_uncerta.

[143] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Jan. 2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. URL: https://dx.doi.org/10.1088/0954-3899/43/2/023001.

[144] "The CMS Statistical Analysis and Combination Tool: Combine". In: *Computing and Software for Big Science* 8.1 (Nov. 2024). ISSN: 2510-2044. DOI: 10.1007/s41781-024-00121-4. URL: http://dx.doi.org/10.1007/s41781-024-00121-4.

[145] Ilaria Brivio, Yun Jiang, and Michael Trott. "The SMEFTsim package, theory and tools". In: *Journal of High Energy Physics* 2017.12 (Dec. 2017). ISSN: 1029-8479. DOI: 10.1007/jhep12(2017)070. URL: http://dx.doi.org/10.1007/JHEP12(2017)070.

[146] Ilaria Brivio. "SMEFTsim 3.0 — a practical guide". In: *Journal of High Energy Physics* 2021.4 (Apr. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep04(2021)073. URL: http://dx.doi.org/10.1007/JHEP04(2021)073.

[147] 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.

[148] The CMS Collaboration. "Study of same-sign W boson scattering and anomalous couplings in events with one tau lepton from pp collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2025.10 (2025), p. 219. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2025)219. URL: https://doi.org/10.1007/JHEP10(2025)219.

[149] The CMS Collaboration. "Observation of electroweak W+W- pair production in association with two jets in proton-proton collisions at s=13TeV". In: *Physics Letters B* 841 (2023), p. 137495. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2022.137495. URL: https://www.sciencedirect.com/science/article/pii/S0370269322006293.

[150] The CMS Collaboration. "Evidence for WW/WZ vector boson scattering in the decay channel qq produced in association with two jets in proton-proton collisions at s=13 TeV". In: *Physics Letters B* 834 (2022), p. 137438. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2022.137438. URL: https://www.sciencedirect.com/science/article/pii/S037026932200572X.

[151] The CMS Collaboration. "Study of same-sign W boson scattering and anomalous couplings in events with one tau lepton from pp collisions at $\sqrt{s} = 13$ TeV". In: (2025). arXiv: 2510.00118 [hep-ex]. URL: https://cds.cern.ch/record/2944855/files/CMS-SMP-22-011-arXiv.pdf.

[152] The CMS Collaboration. *Study of same-sign W boson scattering and anomalous couplings in events with one tau lepton from pp collisions at* $sqrts = 13$ *,TeV*. 2025. arXiv: 2510.00118 [hep-ex]. URL: https://arxiv.org/abs/2510.00118.

[153] "Search for new physics in triple boson production at $\sqrt{s} = 13$ TeV using the effective field theory approach". In: (2025). URL: https://cds.cern.ch/record/2946572.

Appendix A

W + Jets Data Driven Correction

A.1 Reweight Through Subcategory Normalizations

A second, alternative approach has been tested as a cross-check of the nominal functional reweighting method. Following the strategy outlined in AN-19-239 [112], the modeling of the W+Jets background is corrected through a discrete reweighting scheme based on subcategory normalizations, rather than a smooth analytical function.

In this method, the p_T of the most energetic jet is divided into several bins (or subcategories), and an independent normalization factor is assigned to each subcategory. These factors are treated as floating rate parameters in the statistical fit and are constrained using data from the W+Jets control region. In this way, both the shape and normalization of the W+Jets background are effectively adjusted within the fit itself, without introducing an explicit functional dependence.

The method has been tested on the 2018 dataset to evaluate its performance relative to the functional reweighting approach. The two methods yield compatible results within statistical uncertainties, confirming the robustness of the W+Jets modeling correction adopted in the nominal analysis.

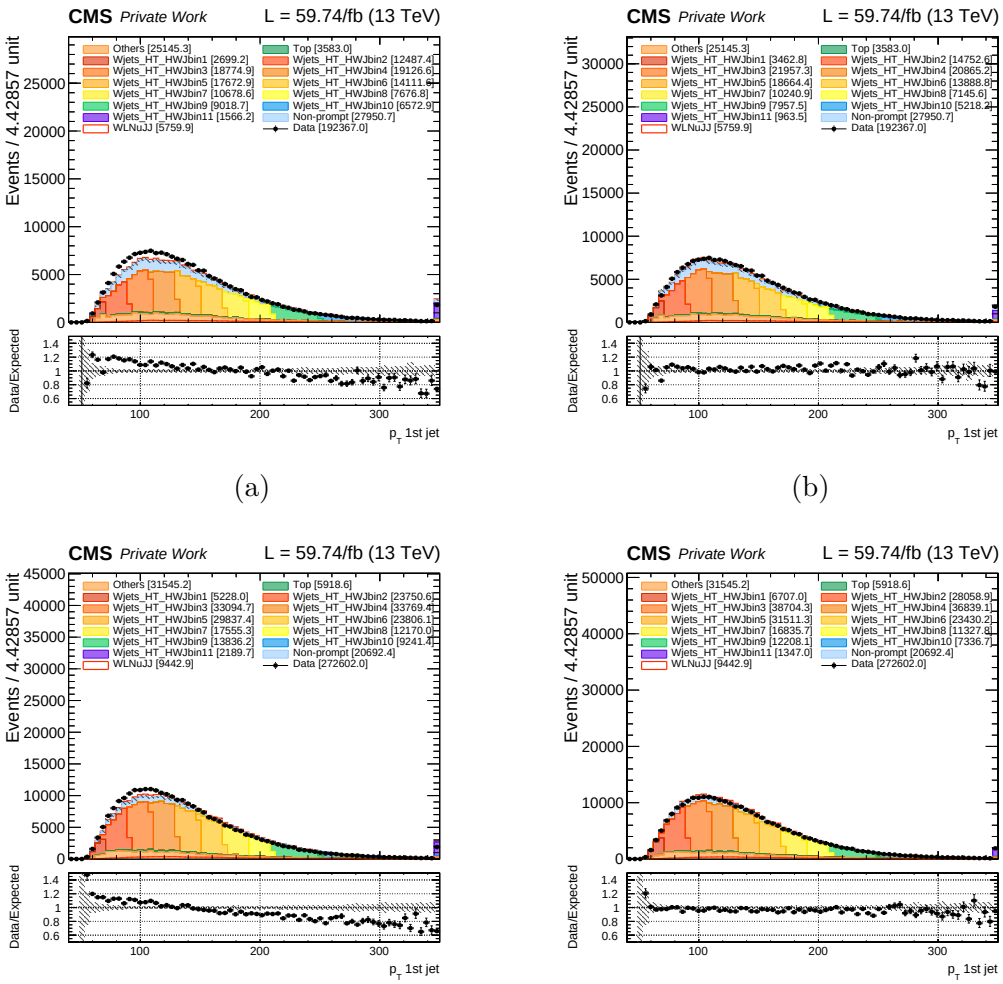


Figure A.1: Data/MC agreement in W+Jets CR. The grey band represents the statistical uncertainty of the MC samples. No systematic uncertainty is shown. Left: p_T of the most energetic jet before correction. Right: p_T of the most energetic jet after correction. Top: electron channel. Bottom: muon channel.

A.2 Closure Test for The Two Methods

A closure test was performed by conducting a likelihood scan of the signal strength modifier with the background estimated using the previously described methods. The results showed very similar performances.

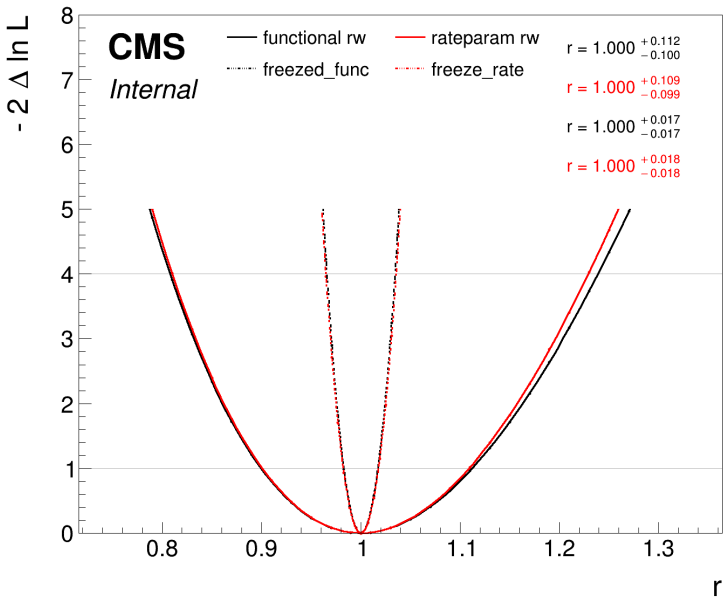


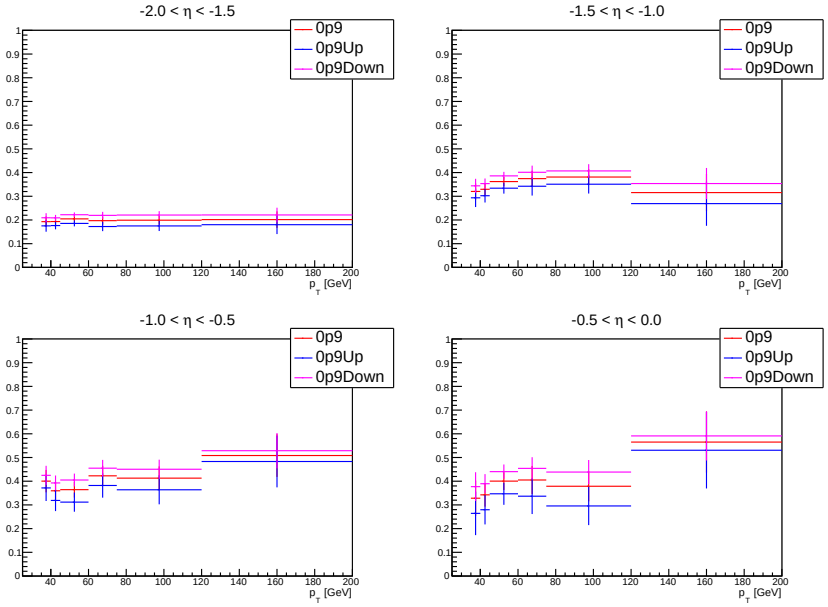
Figure A.2: Closure test fitting the signal strength modifier with (continuous line) and without (dashed line) systematic uncertainties. The fit is performed on the DNN output in the signal region, combining both the electron and muon channels. The only difference is that the mixed W+Jets background is corrected using either a reweight function (black) or subcategory normalisation (red).

Appendix B

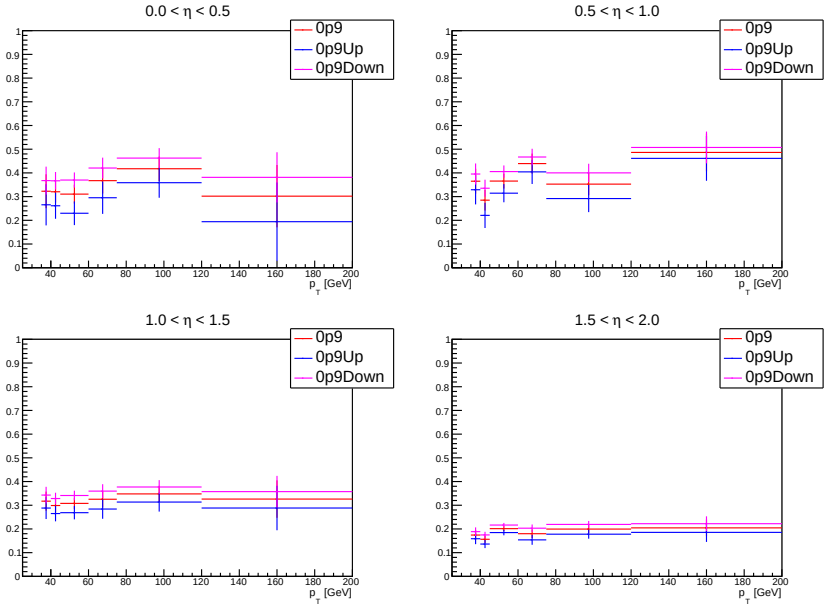
Fake and Prompt Rates

B.1 Fake and Prompt Rate – 2017 only

This section shows plots similar to those presented in Section 3.5.3 for the 2017 dataset. Figures B.1 and B.2 illustrate the systematic uncertainties considered in the Fake Rate measurement.

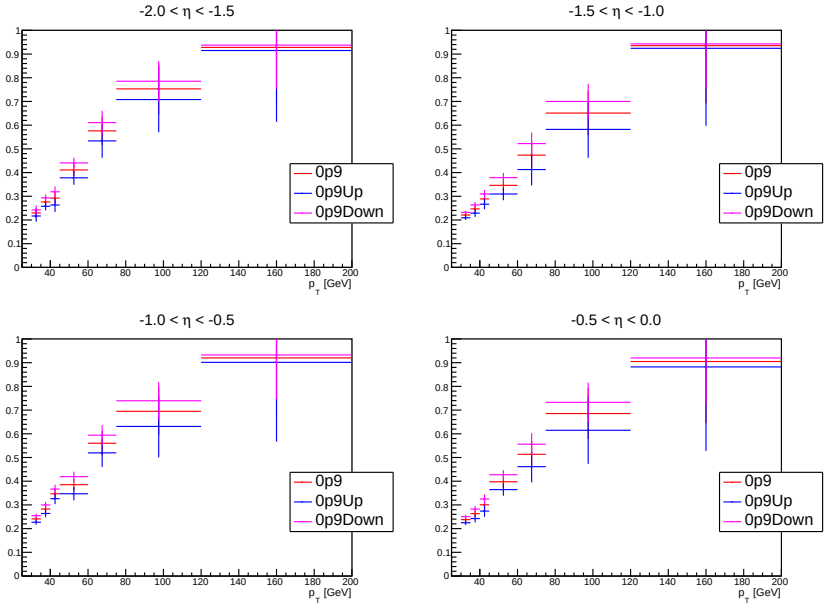


(a)

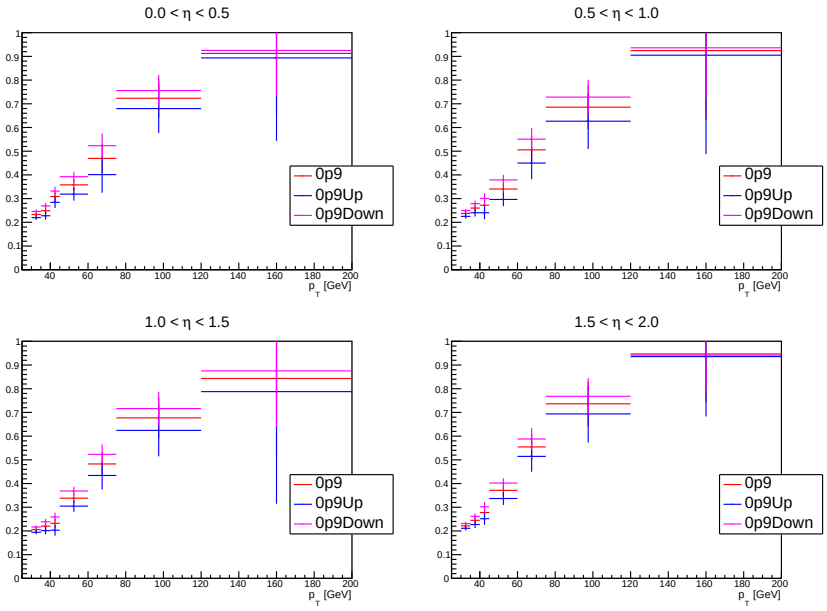


(b)

Figure B.1: Electron fake rate (FR) electroweak scale factor (SF) variations for 2017. A 10% variation up/down is applied to the unique SF used to subtract the residual prompt contamination from the QCD-enriched region (Eq. 3.6). The nominal SF corresponds to the value measured in the DY process.



(a)

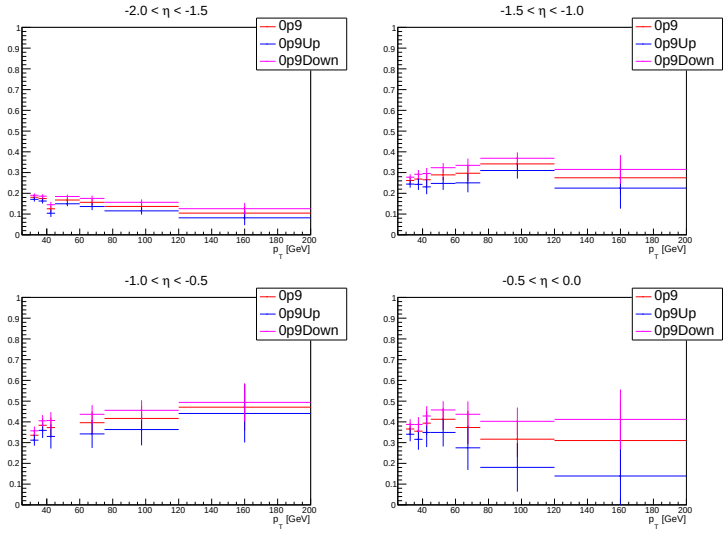


(b)

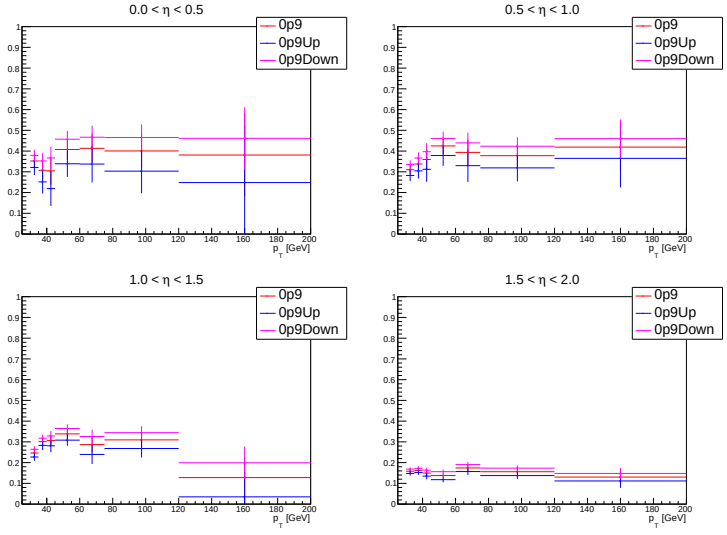
Figure B.2: Muon fake rate (FR) electroweak scale factor (SF) variations for 2017. A 10% variation up/down is applied to the unique SF used to subtract the residual prompt contamination from the QCD-enriched region (Eq. 3.6). The nominal SF corresponds to the value measured in the DY process.

B.2 Fake and Prompt Rate – 2016HIPM

This section shows plots similar to those presented in Section 3.5.3 for the 2016HIPM dataset. Figures B.5 and B.6 illustrate the systematic uncertainties considered in the Fake Rate measurement.

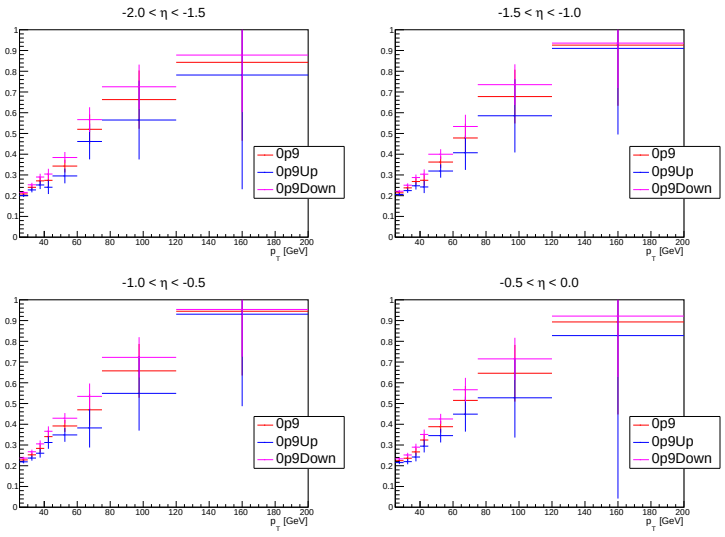


(a)

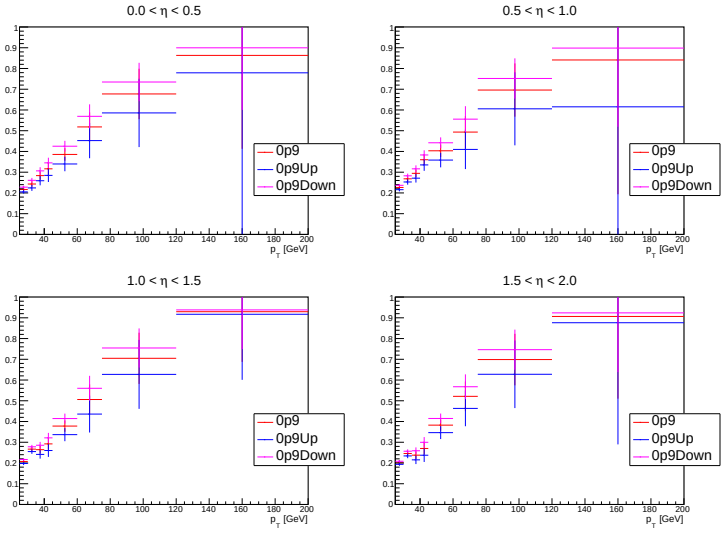


(b)

Figure B.3: Electron fake rate (FR) electroweak scale factor (SF) variations for 2016HIPM. A 10% variation up/down is applied to the unique SF used to subtract the residual prompt contamination from the QCD-enriched region (Eq. 3.6). The nominal SF corresponds to the value measured in the DY process.



(a)

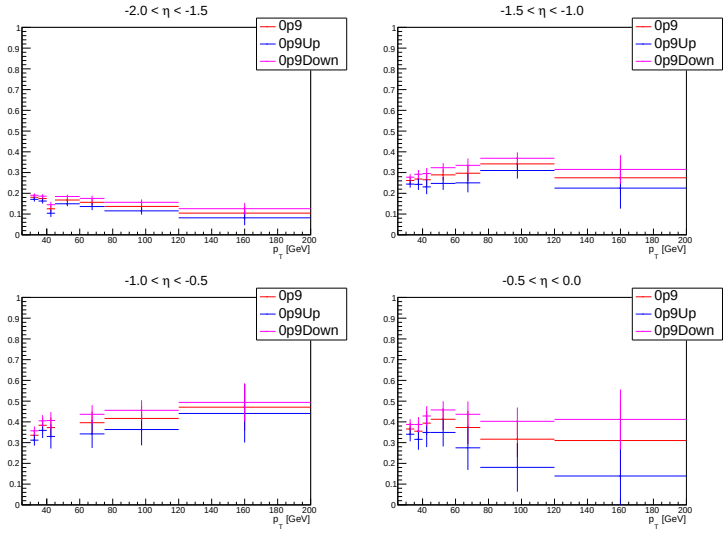


(b)

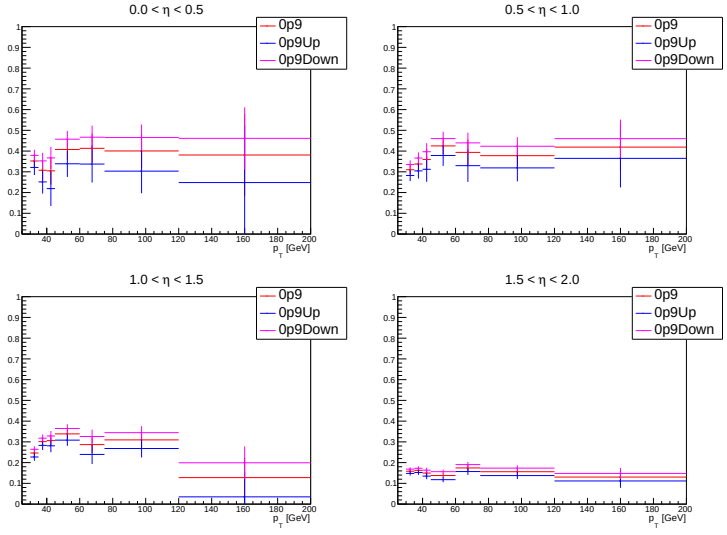
Figure B.4: Muon fake rate (FR) electroweak scale factor (SF) variations for 2016HIPM. A 10% variation up/down is applied to the unique SF used to subtract the residual prompt contamination from the QCD-enriched region (Eq. 3.6). The nominal SF corresponds to the value measured in the DY process.

B.3 Fake and Prompt Rate – 2016HIPM

This section shows plots similar to those presented in Section 3.5.3 for the 2016HIPM dataset. Figures B.5 and B.6 illustrate the systematic uncertainties considered in the Fake Rate measurement.

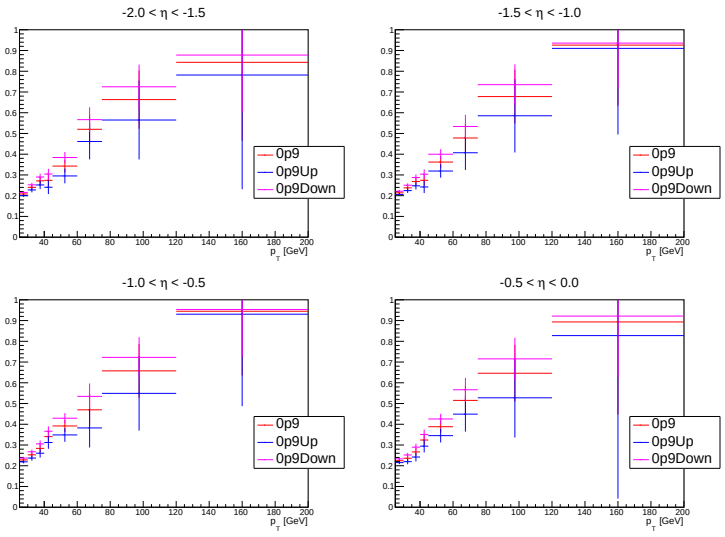


(a)

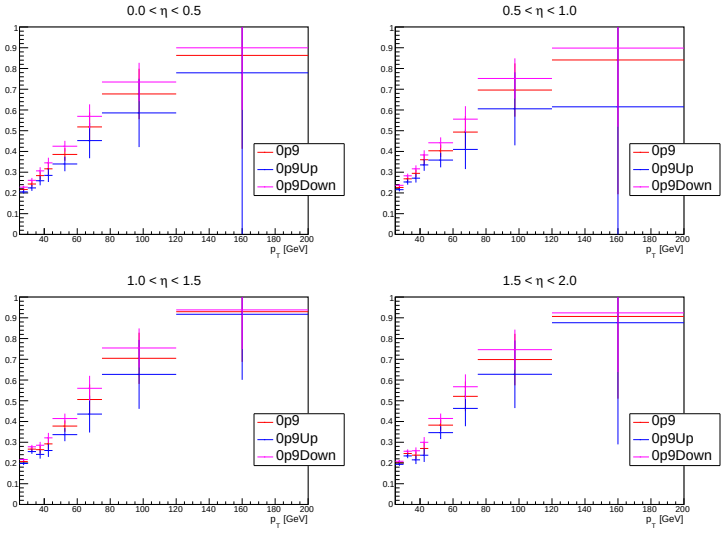


(b)

Figure B.5: Electron fake rate (FR) electroweak scale factor (SF) variations for 2016HIPM. A 10% variation up/down is applied to the unique SF used to subtract the residual prompt contamination from the QCD-enriched region (Eq. 3.6). The nominal SF corresponds to the value measured in the DY process.



(a)

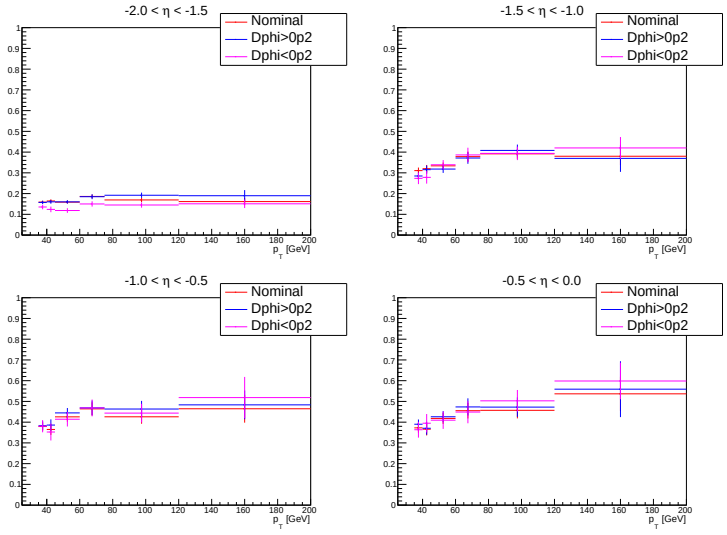


(b)

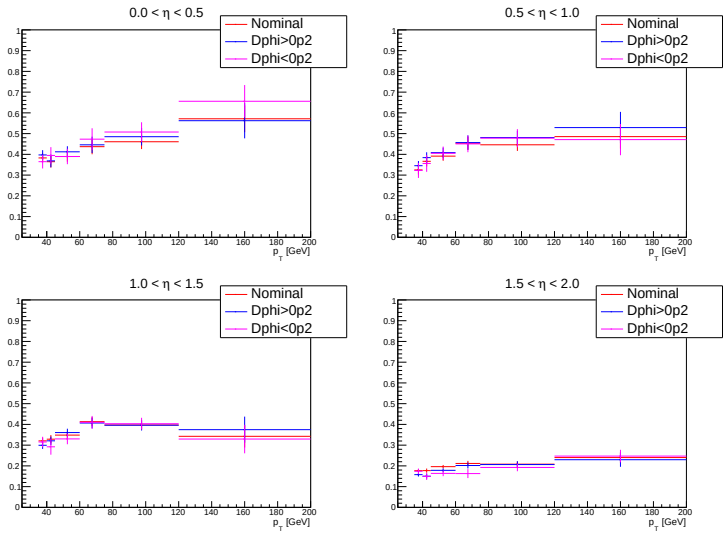
Figure B.6: Muon fake rate (FR) electroweak scale factor (SF) variations for 2016HIPM. A 10% variation up/down is applied to the unique SF used to subtract the residual prompt contamination from the QCD-enriched region (Eq. 3.6). The nominal SF corresponds to the value measured in the DY process.

B.4 Fake and Prompt Rate – $\Delta\phi(\ell_1, \text{MET})$-binned

Several cross-checks have been performed to validate the stability of the fake rate estimation. Figures B.7 and B.8 show the results obtained when splitting the measurement into two phase spaces defined by $\Delta\phi(\ell_1, \text{MET}) \gtrless 0.2$, compared to the inclusive (nominal) measurement. The threshold value of 0.2 is chosen to ensure comparable statistics between the two regions.

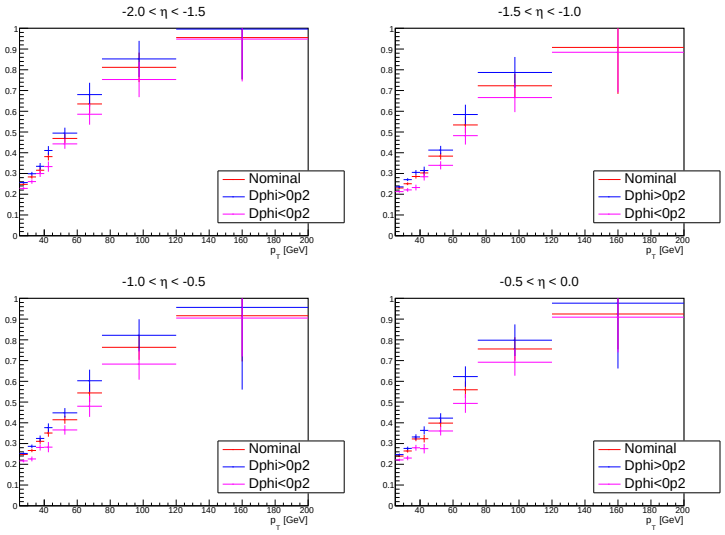


(a)

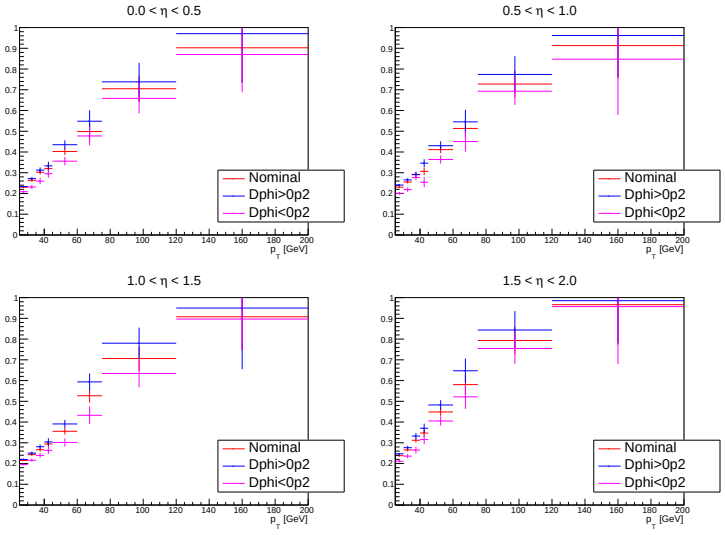


(b)

Figure B.7: Electron fake rate (FR) measurement in 2018, performed in two separate $\Delta\phi(\ell_1, \text{MET})$ regions. Within uncertainties, the fake rate is consistent with the inclusive measurement, showing no significant dependence on $\Delta\phi(\ell_1, \text{MET})$.



(a)



(b)

Figure B.8: Muon fake rate (FR) measurement in 2018, performed in two separate $\Delta\phi(\ell_1, \text{MET})$ regions. Within uncertainties, the fake rate is consistent with the inclusive measurement, showing no significant dependence on $\Delta\phi(\ell_1, \text{MET})$.

Appendix C

Input Variables to the binary DNN

This chapter presents the normalised distributions of the input variables used in the binary DNN described in Sec. 3.6. For each observable, the signal and background shapes are shown using the full Run II dataset, obtained by combining the 2016, 2017, and 2018 samples. These distributions illustrate the discriminating power of the variables and motivate their inclusion in the training.

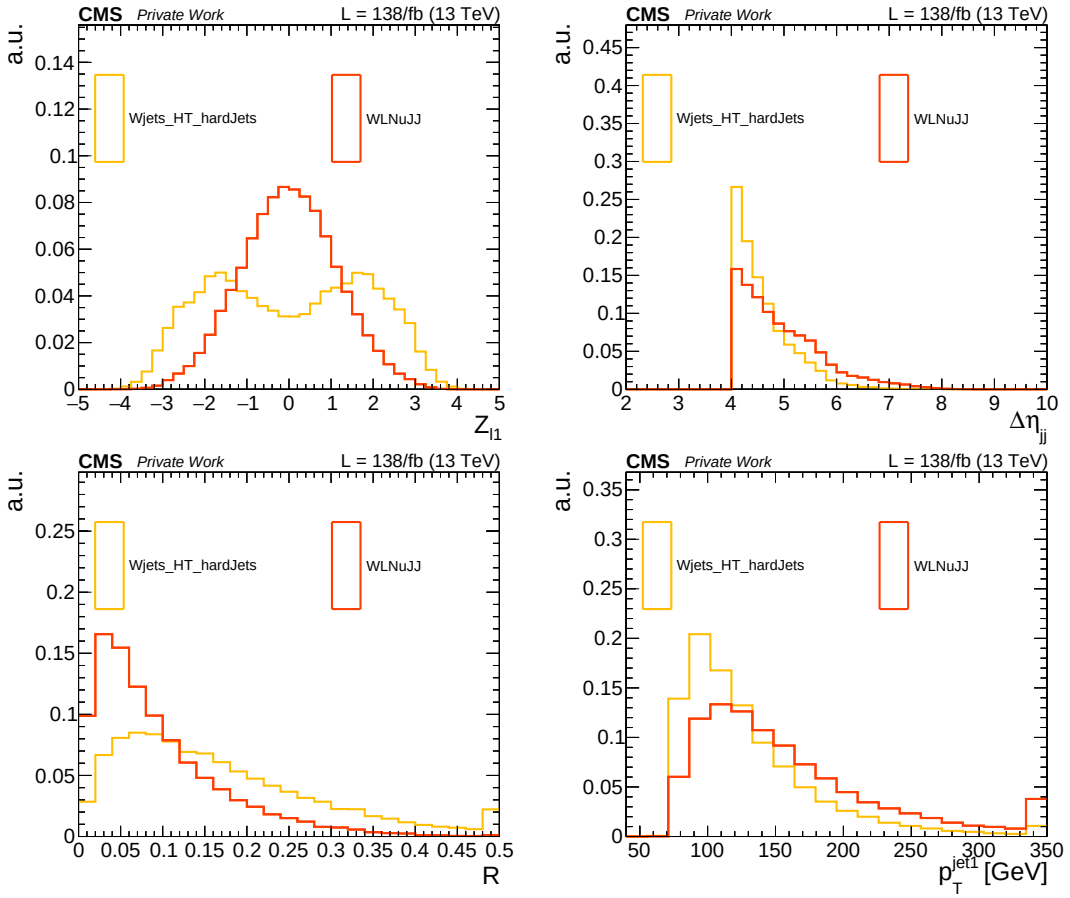


Figure C.1: Normalised signal and background distributions for four of the input variables used in the DNN training. These examples illustrate the differences in shape between the two processes: (top) the lepton Zeppenfeld variable and the dijet $|\Delta\eta_{jj}|$ separation; (bottom) the R transverse momentum projection and the leading-jet p_T. All distributions are area-normalised to highlight their discriminating power and to support the choice of variables included in the network.

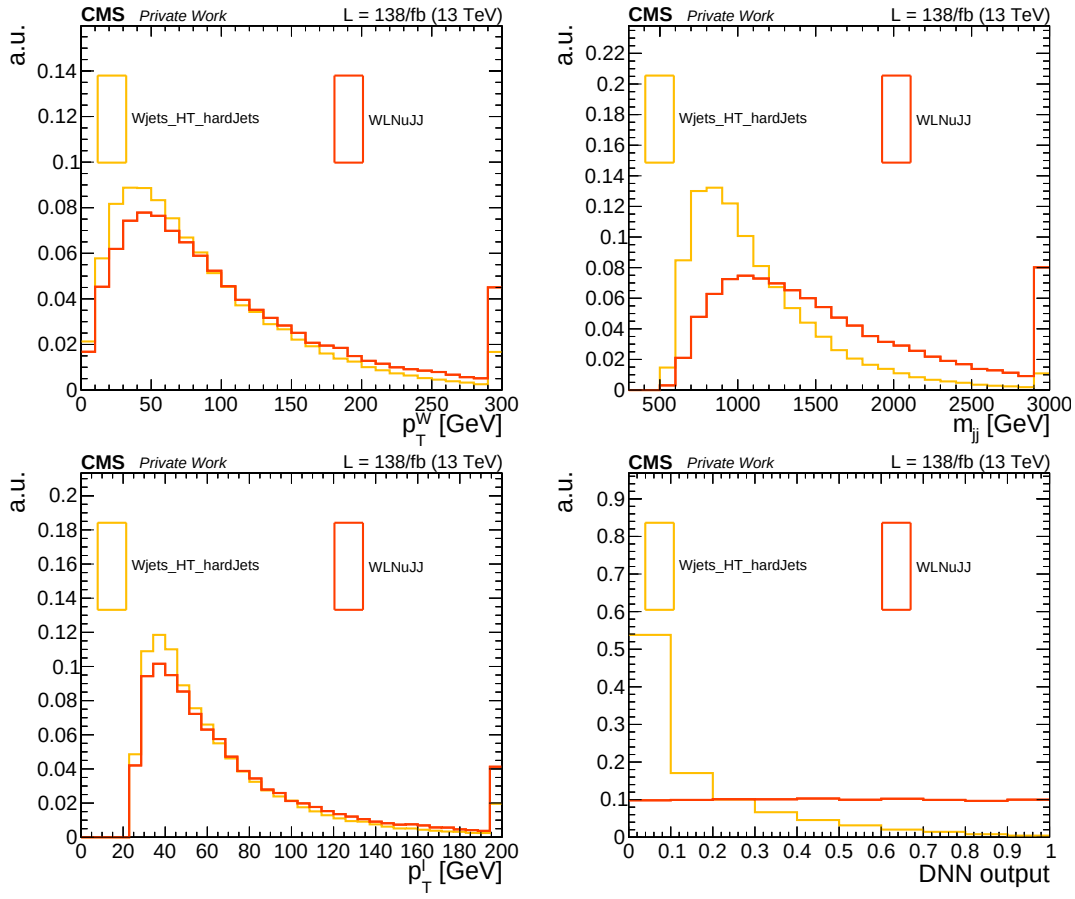


Figure C.2: Normalised signal and background distributions for four representative observables used in the analysis. Top row: transverse momentum of the W boson and dijet invariant mass m_{jj}. Bottom row: lepton transverse momentum and the DNN output. All distributions are area-normalised to emphasise the shape differences between signal and background and to illustrate their discriminating power. The DNN output, shown in the bottom-right panel, is the final classifier response, flattened in signal quantiles, and is not used as an input to the DNN.

Appendix D

Differential Measurement Impact Plots

This appendix collects the post-fit impact plots corresponding to the differential measurements presented in Chapter 4.2. Each section reports the breakdown of systematic uncertainties for the parameters of interest associated with the corresponding observable, evaluated using the data-Asimov dataset described in Section 1.9.

D.1 m_{jj} impacts

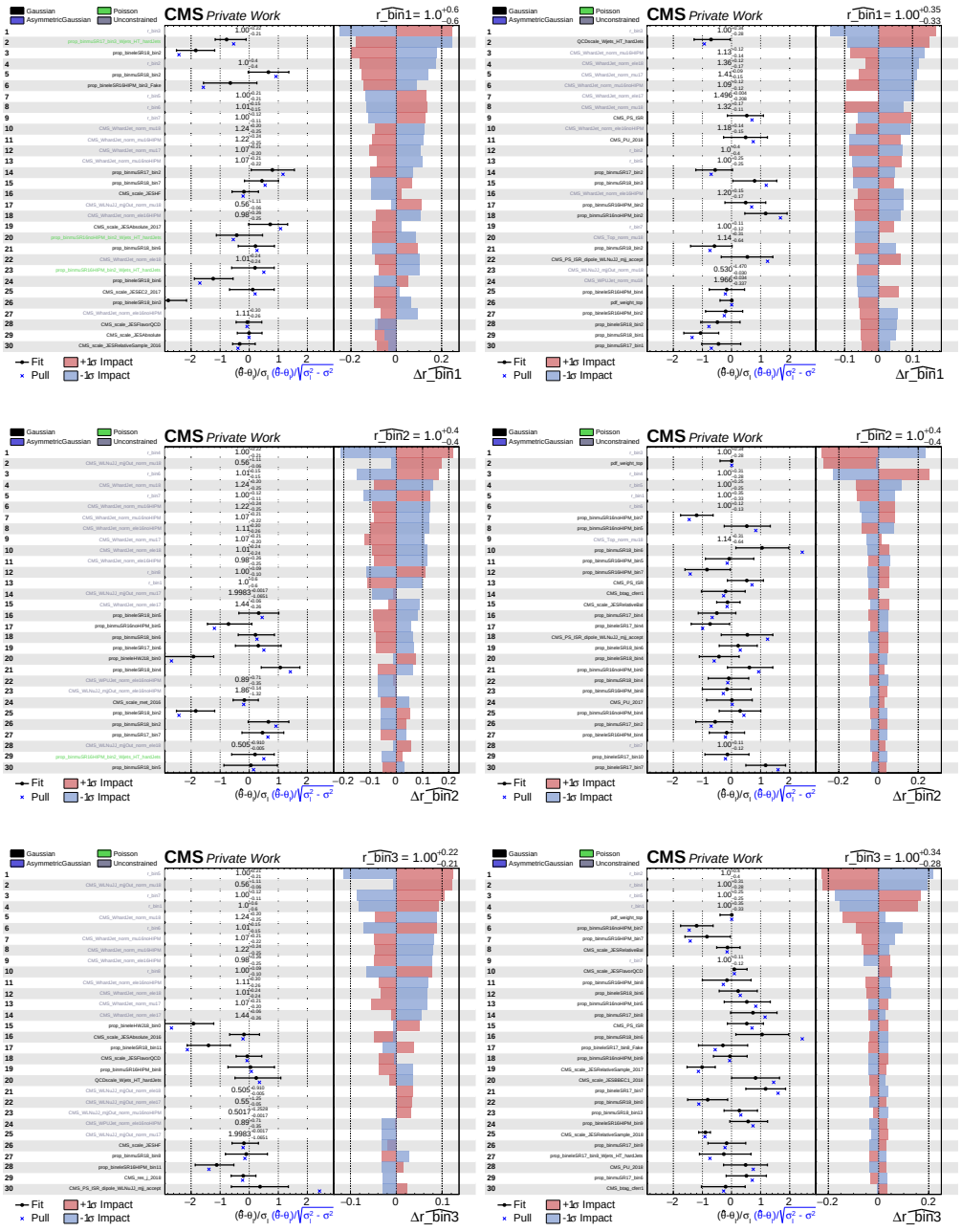


Figure D.1: Impacts plot for all gen-bin of m_{jj}. Left: reconstructed variable-based method. Right: Multi-class-based method

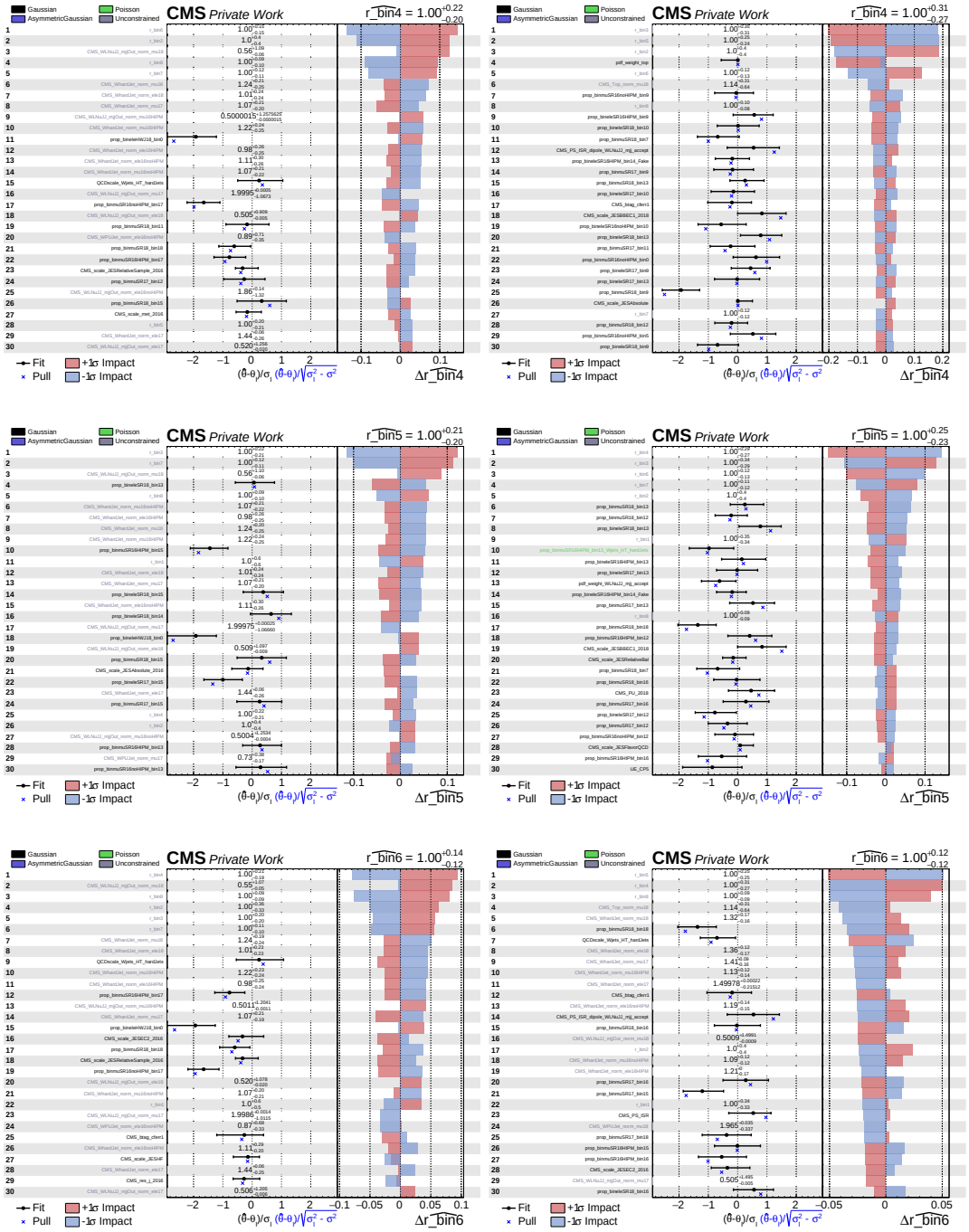


Figure D.2: Impacts plot for all gen-bin of m_{jj}. Left: reconstructed variable-based method. Right: Multi-class-based method

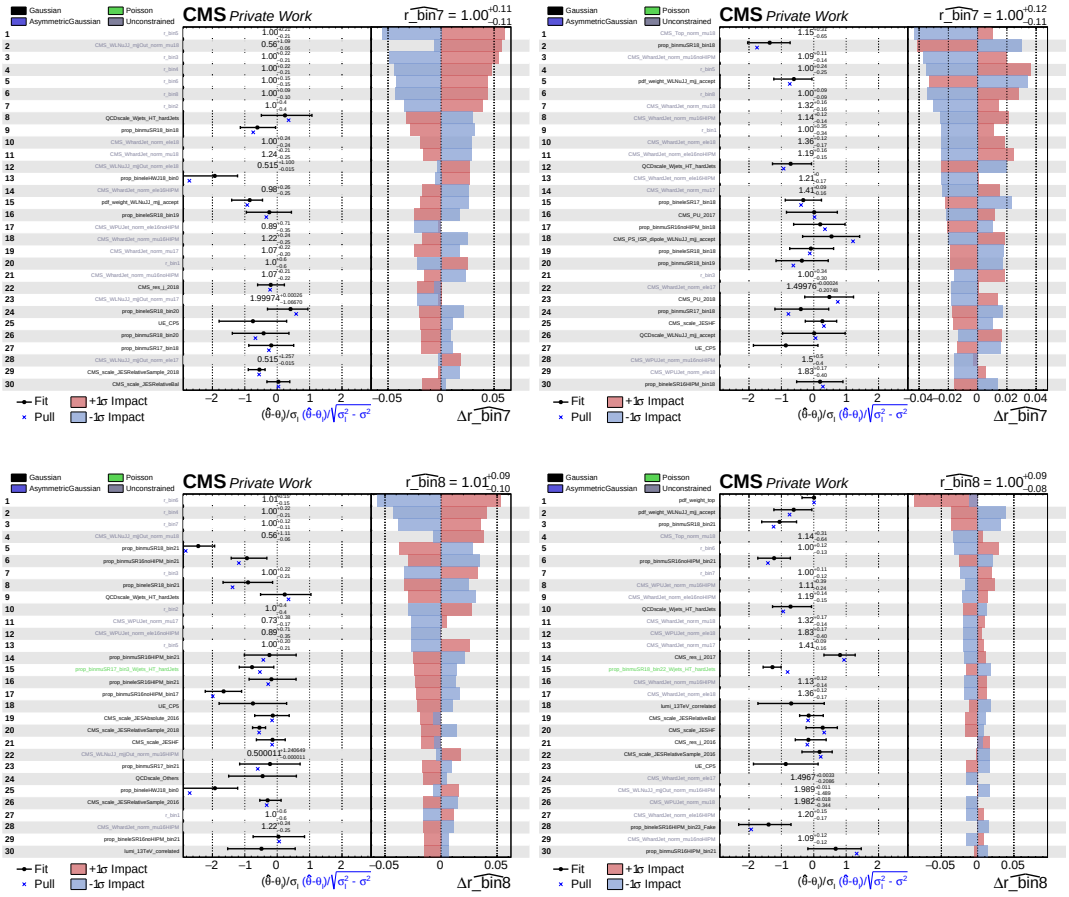
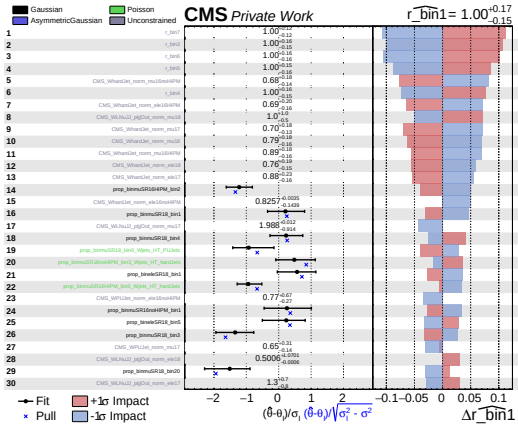
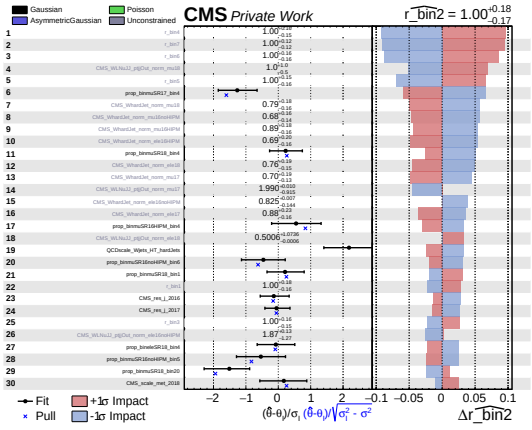
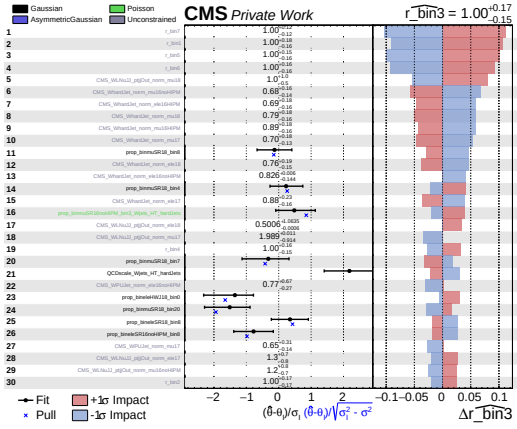


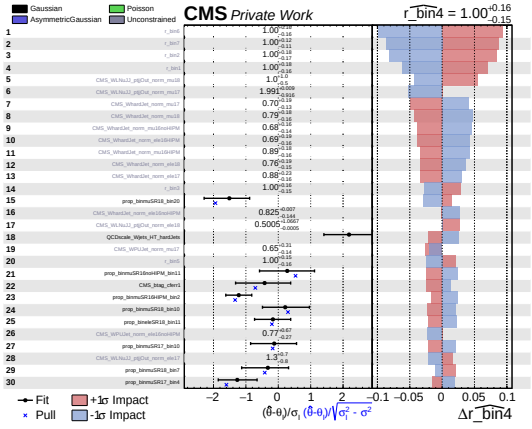
Figure D.3: Impacts plot for all gen-bin of m_{jj}. Left: reconstructed variable-based method. Right: Multi-class-based method

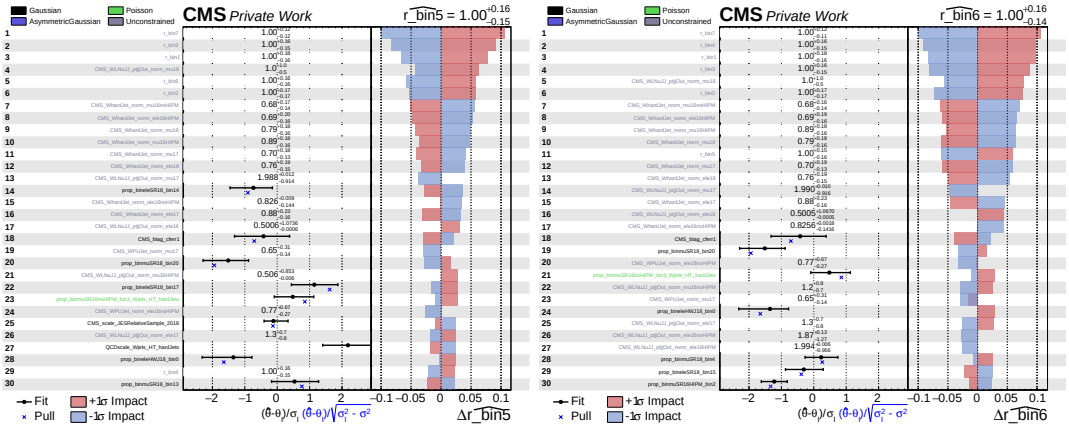
D.2 $p_T(jj)$ impacts

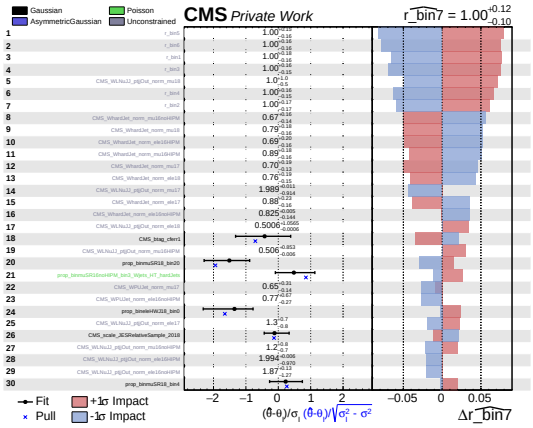




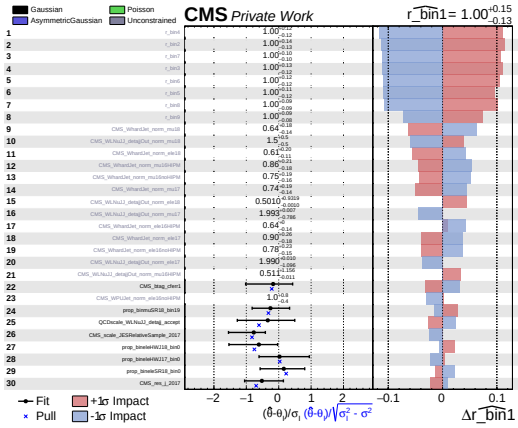


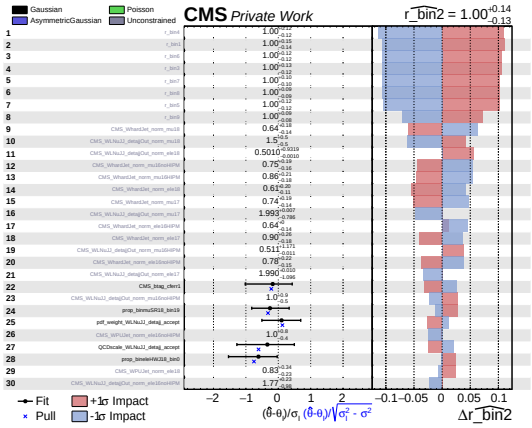


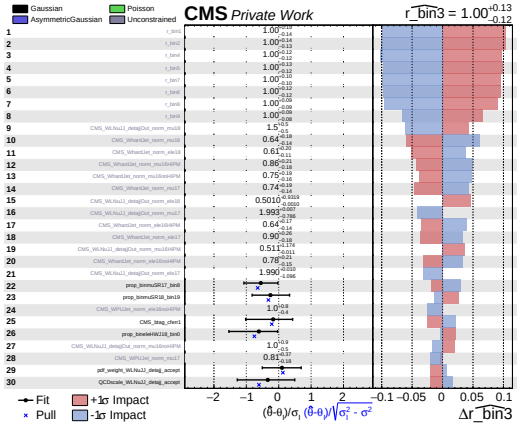

CMS Private Work
r_bin5 = 1.00 +0.16 −0.15
Gaussian
Poisson
AsymmetricGaussian
Unconstrained
Fit
Pull
+1σ Impact
-1σ Impact
(θ-θ0)/σi
(θ-θ0)/√σi² - σ²
Δr_bin5
CMS Private Work
r_bin6 = 1.00 +0.16 −0.14
Δr_bin6

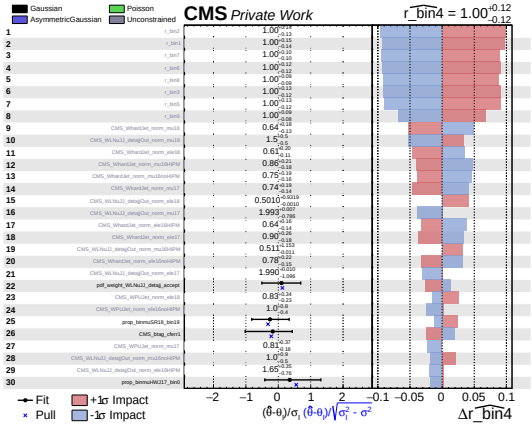

CMS Private Work
r_bin7 = 1.00 +0.12 −0.10
Gaussian
Poisson
AsymmetricGaussian
Unconstrained
Fit
Pull
+1σ Impact
-1σ Impact
(θ-θ0)/σi
(θ-θ0)/√σi² - σ²
Δr_bin7

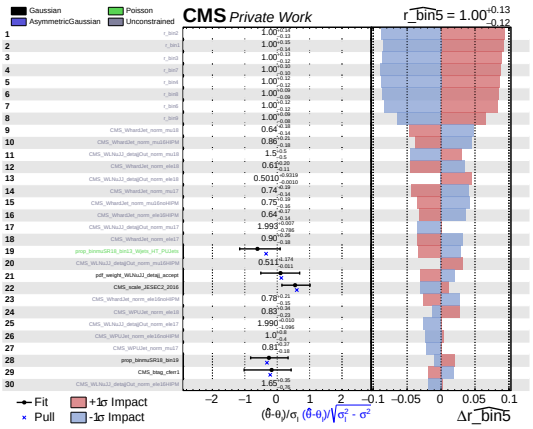
D.3 $\Delta\eta_{jj}$ impacts

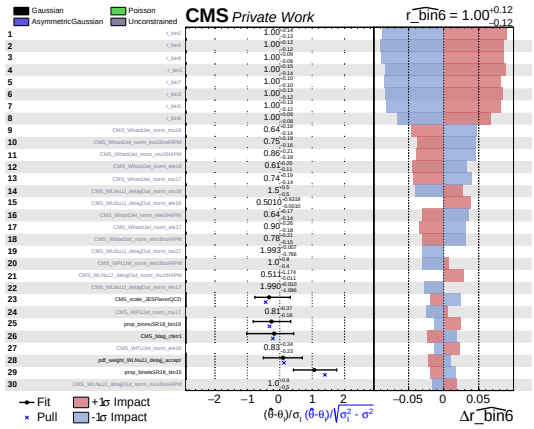


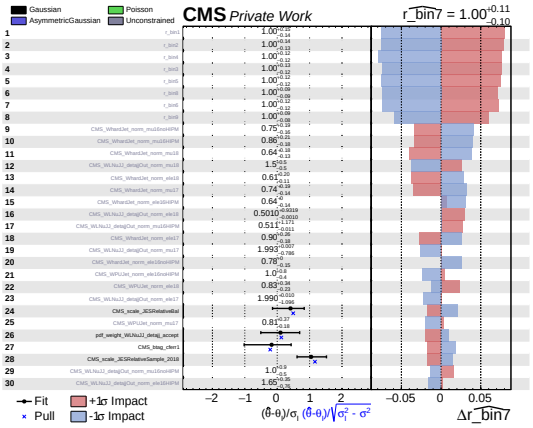


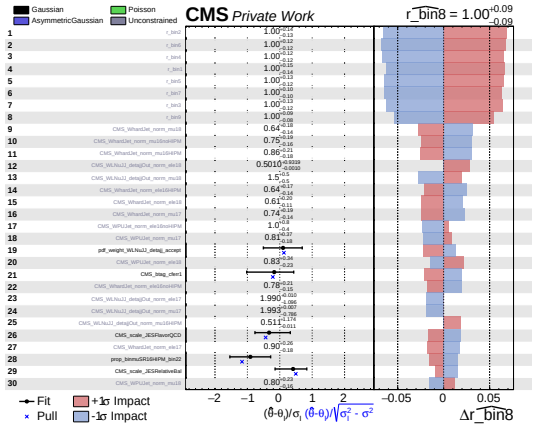


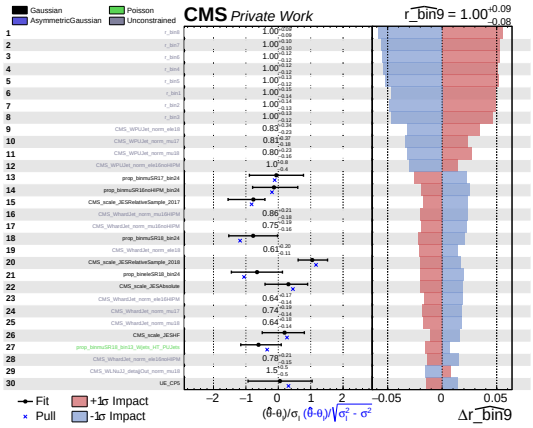



Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin5 = 1.00^{+0.13}_{-0.12}
Fit
Pull
+1σ Impact
-1σ Impact
Δr_bin5

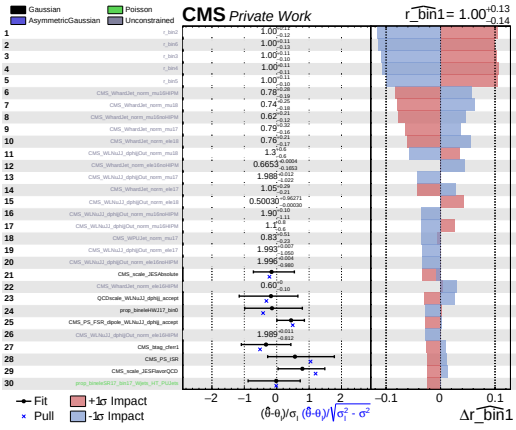

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin6 = 1.00^{+0.12}_{-0.12}
Fit
Pull
+1σ Impact
-1σ Impact
Δr_bin6

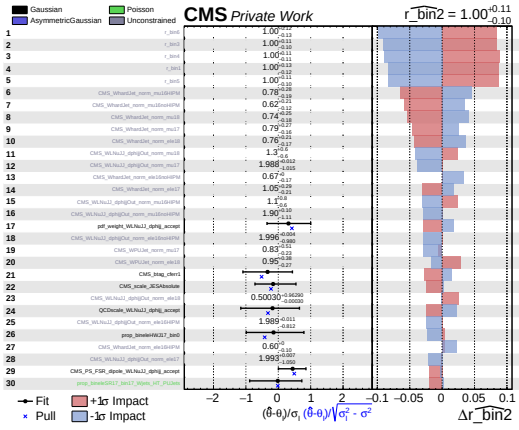

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin7 = 1.00^{+0.11}_{-0.10}
Fit
Pull
+1σ Impact
-1σ Impact
Δr_bin7

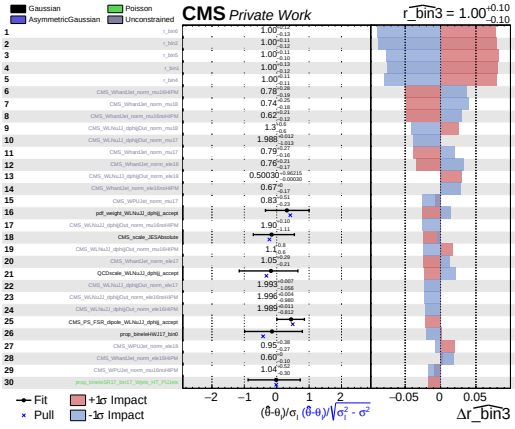

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin8 = 1.00^{+0.09}_{-0.09}
Fit
Pull
+1σ Impact
-1σ Impact
Δr_bin8

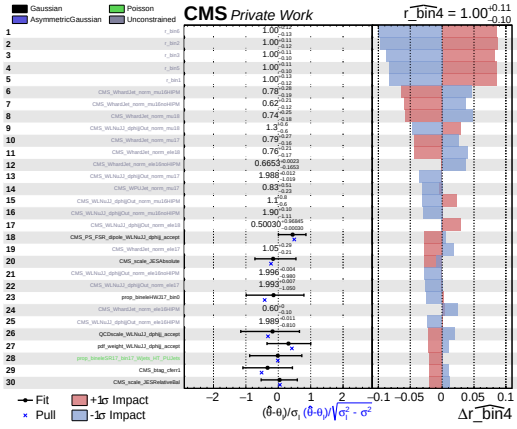

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin9 = 1.00^{+0.09}_{-0.08}
Fit
Pull
+1σ Impact
-1σ Impact
Δr_bin9

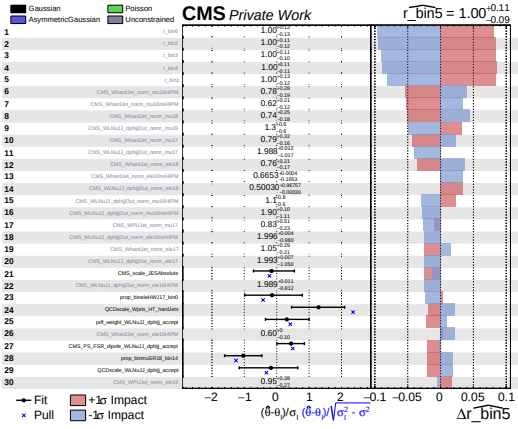
D.4 $\Delta\phi_{jj}$ impacts

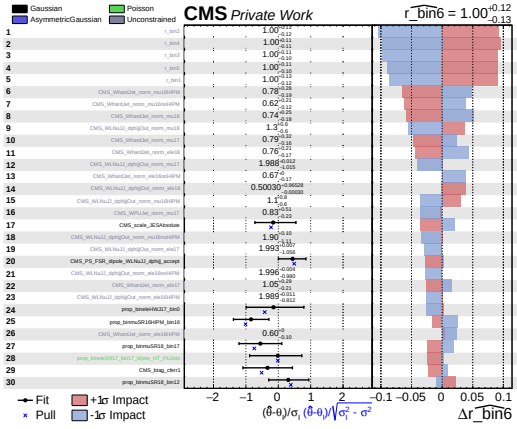




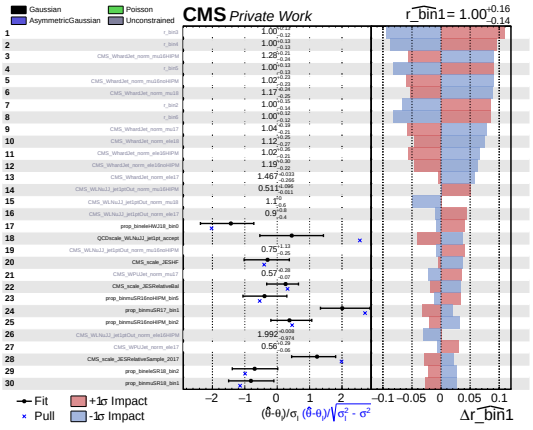


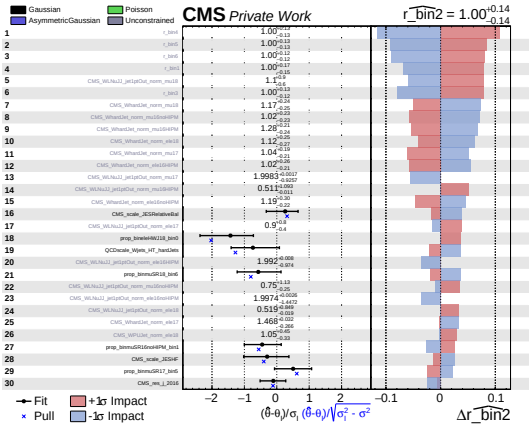


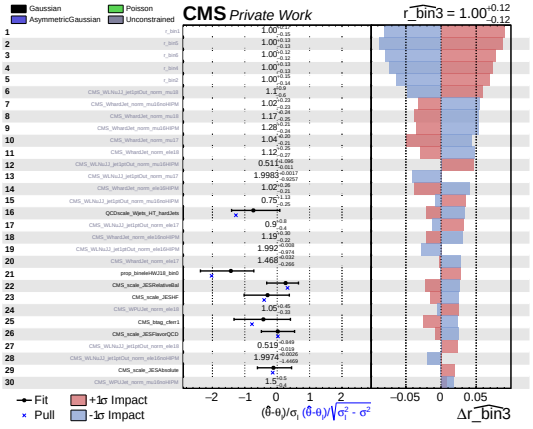


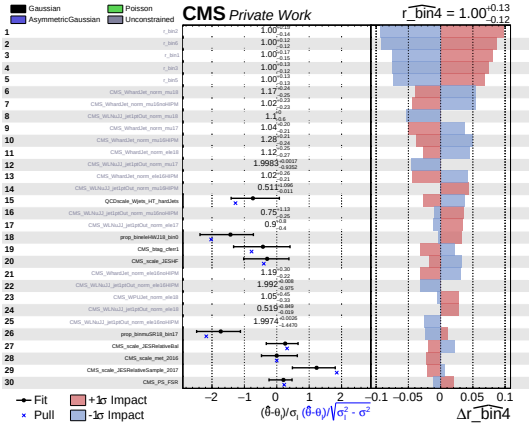


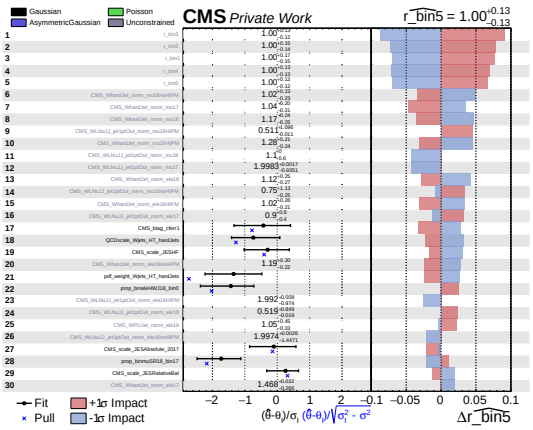
D.5 $p_T(j_1)$ impacts

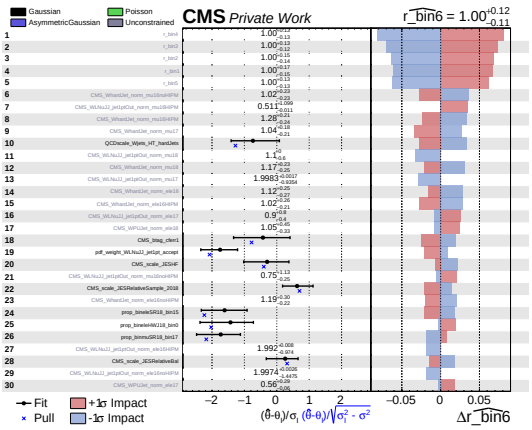




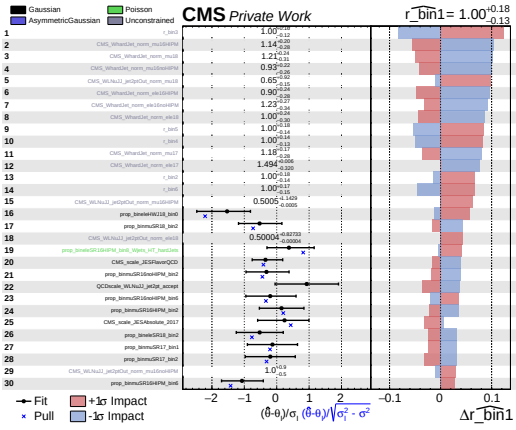


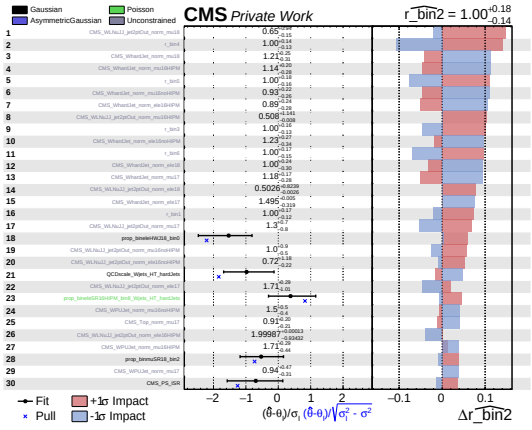


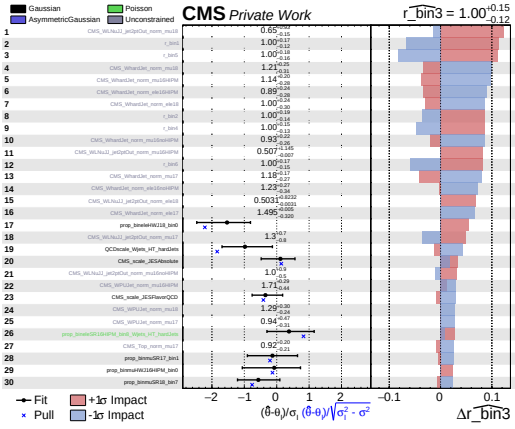


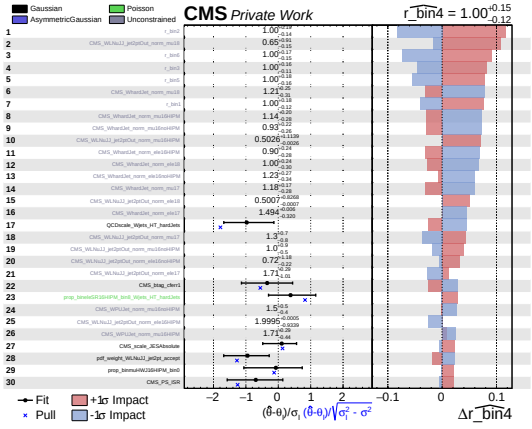


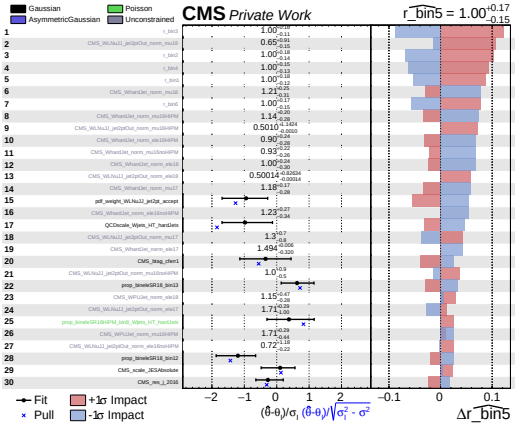
D.6 $p_T(j_2)$ impacts

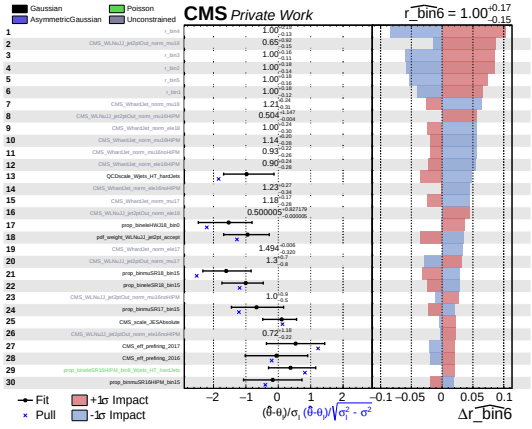




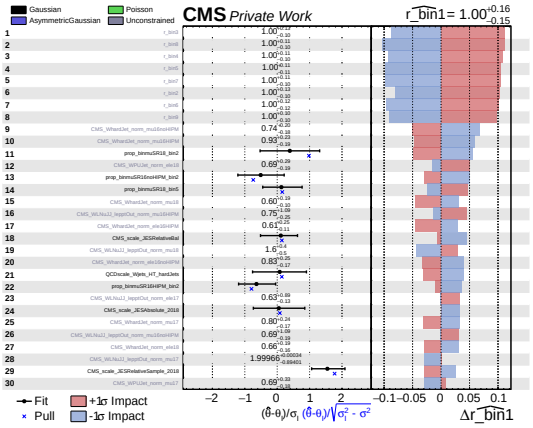


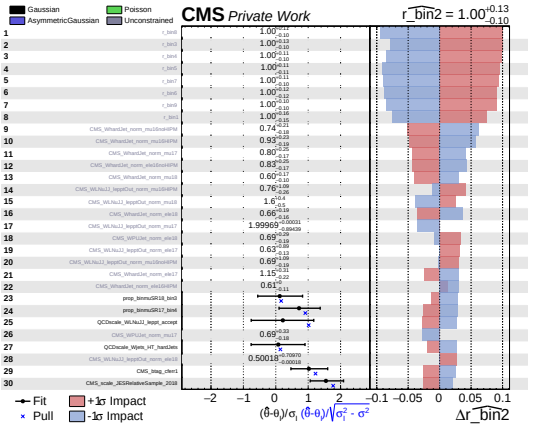


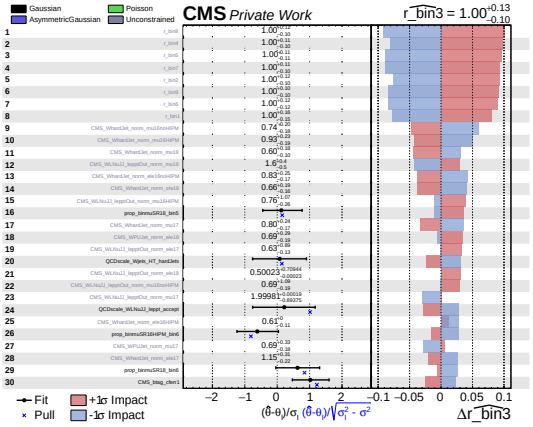


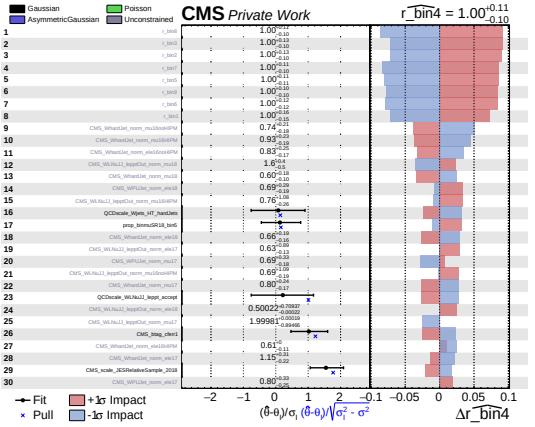


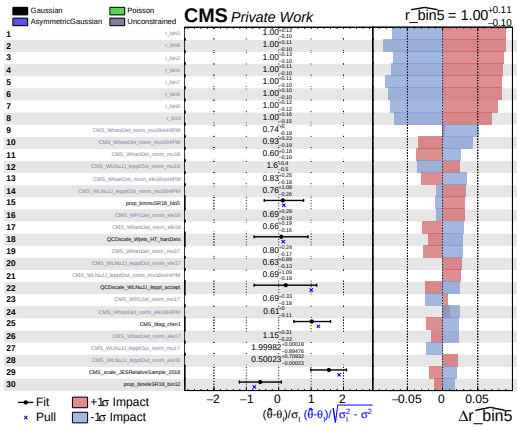
D.7 $p_T(\ell)$ impacts

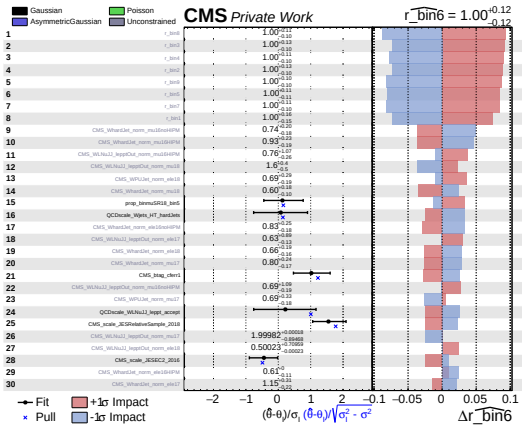


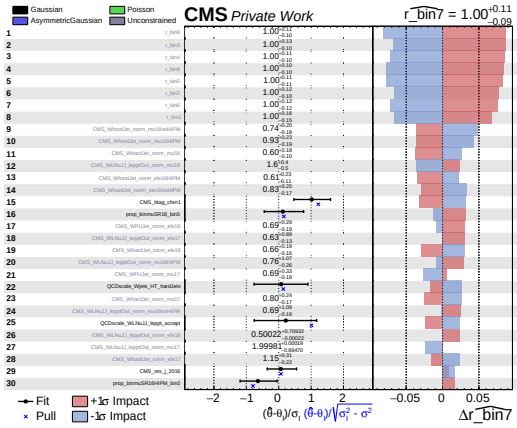


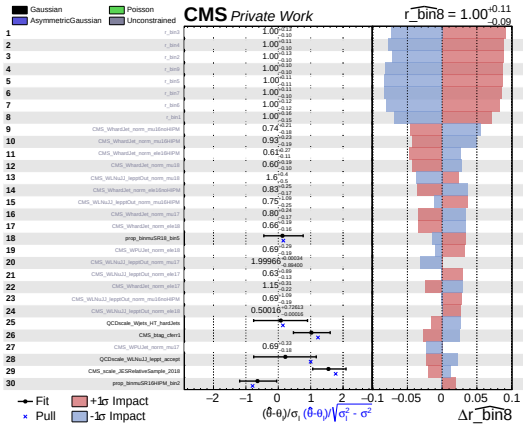


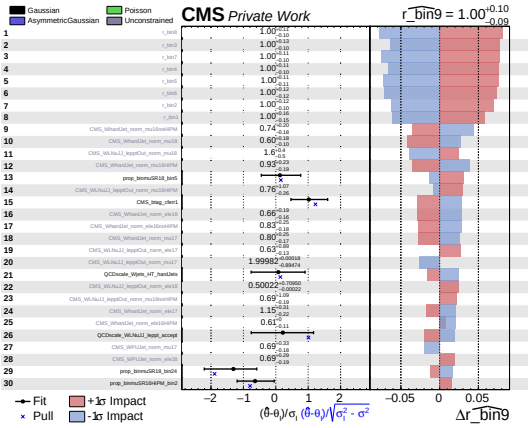



Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin5 = 1.00+0.11−0.10
Fit
Pull
+1σ Impact
-1σ Impact
(θ-θ0)/σ
(θ-θ0)/√σ2 - σ2
Δr_bin5

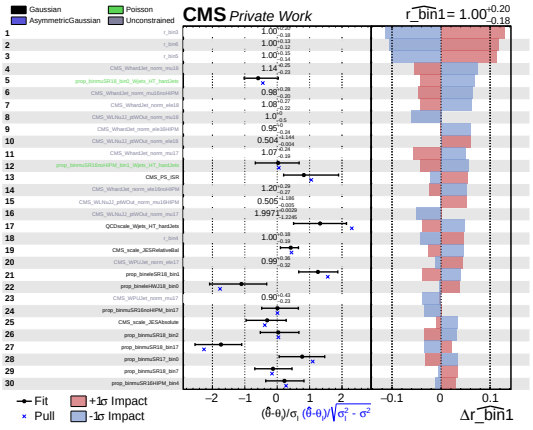

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin6 = 1.00+0.12−0.12
Fit
Pull
+1σ Impact
-1σ Impact
(θ-θ0)/σ
(θ-θ0)/√σ2 - σ2
Δr_bin6

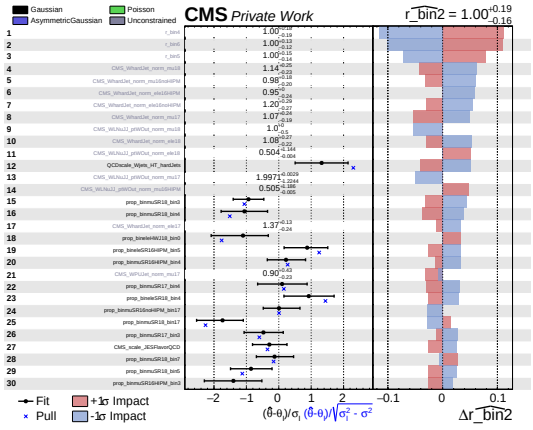

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin7 = 1.00+0.11−0.09
Fit
Pull
+1σ Impact
-1σ Impact
(θ-θ0)/σ
(θ-θ0)/√σ2 - σ2
Δr_bin7

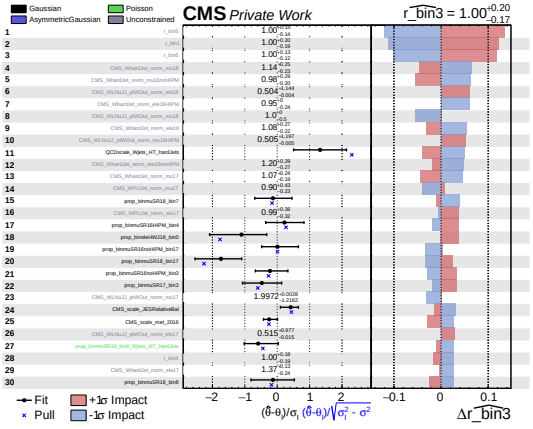

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin8 = 1.00+0.11−0.09
Fit
Pull
+1σ Impact
-1σ Impact
(θ-θ0)/σ
(θ-θ0)/√σ2 - σ2
Δr_bin8

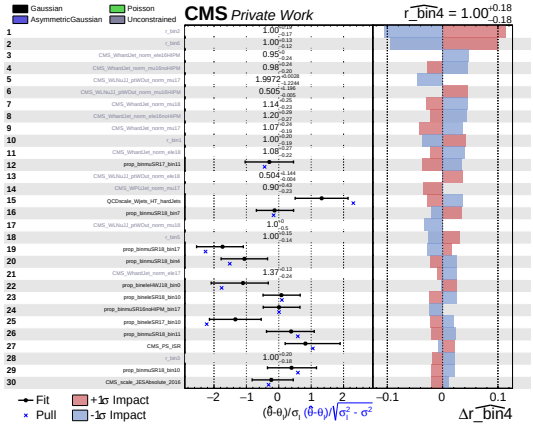

Gaussian
Poisson
AsymmetricGaussian
Unconstrained
CMS Private Work
r_bin9 = 1.00+0.10−0.09
Fit
Pull
+1σ Impact
-1σ Impact
(θ-θ0)/σ
(θ-θ0)/√σ2 - σ2
Δr_bin9

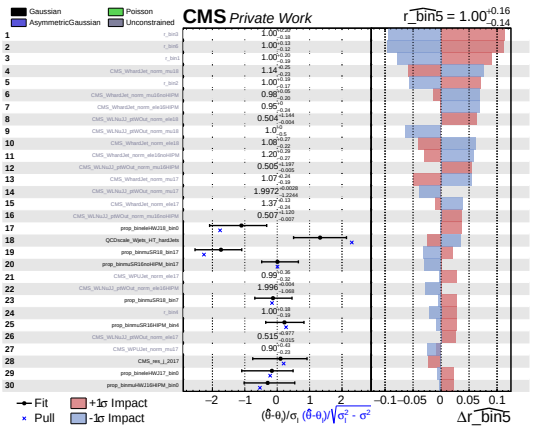
D.8 $p_T(W)$ impacts

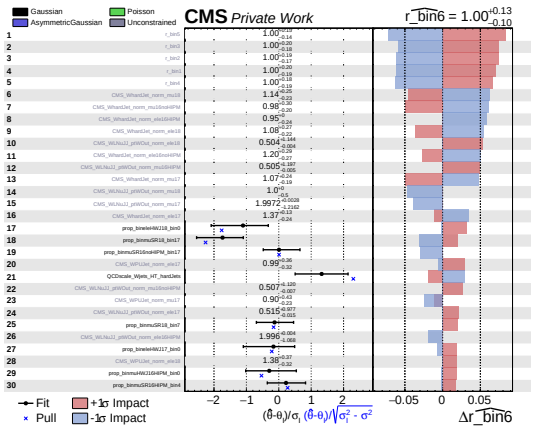






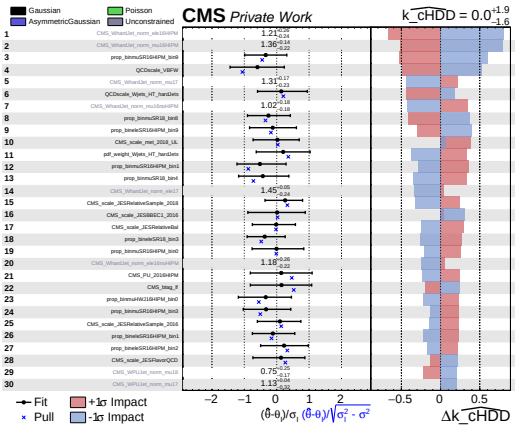


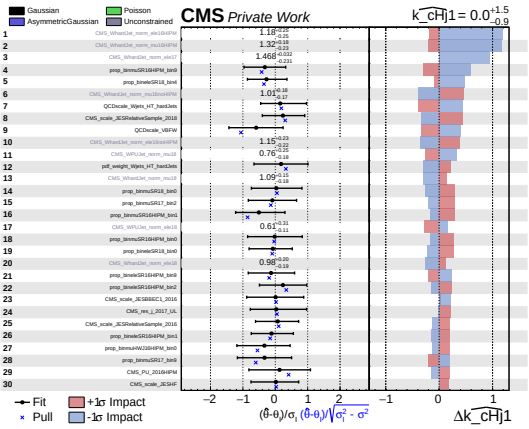


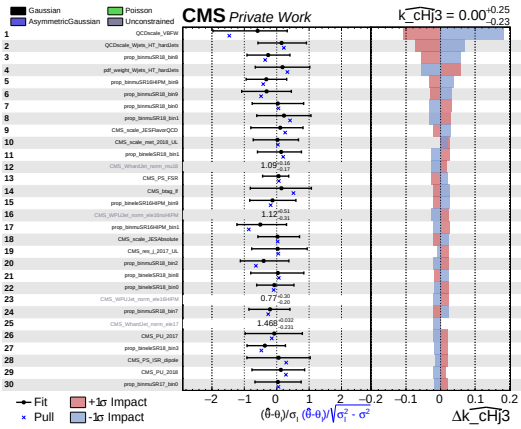


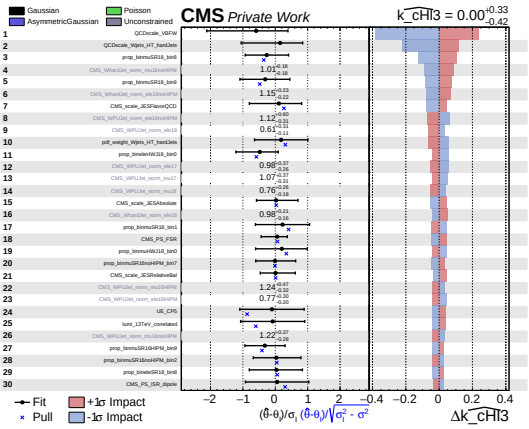
Appendix E

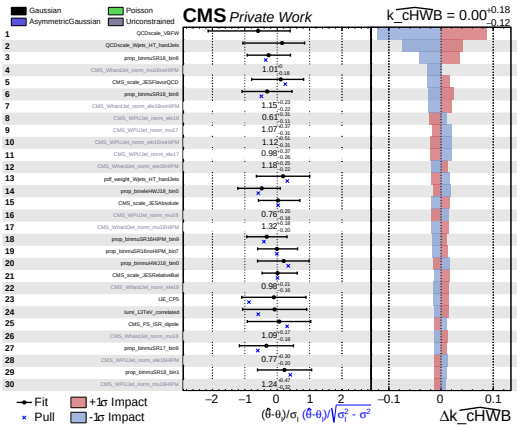
EFT Impact plots

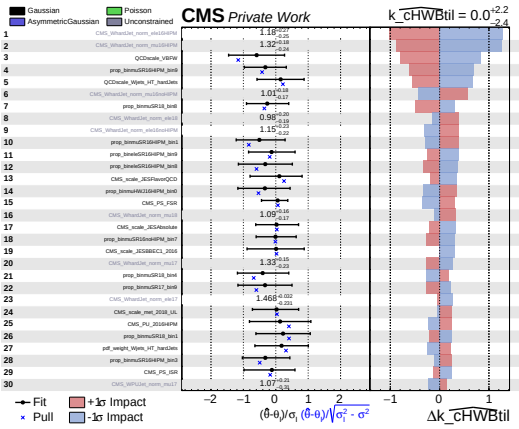


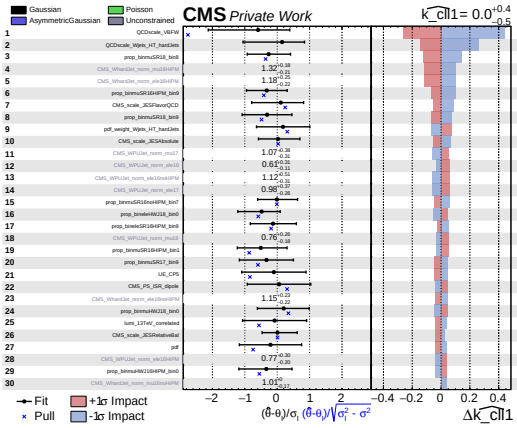


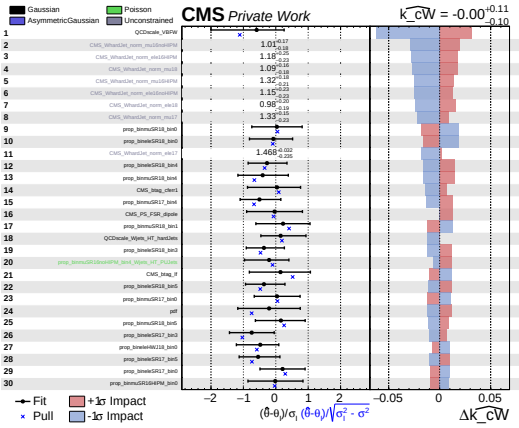


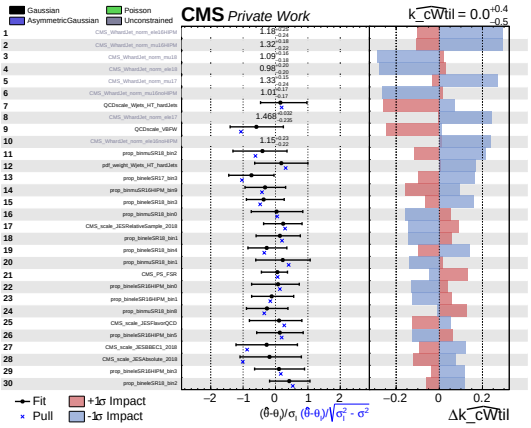



CMS Private Work
k_cHWB = 0.00 +0.18 −0.12
Δk_cHWB


CMS Private Work
k_cHWBtil = 0.0 +2.2 −2.4
Δk_cHWBtil


CMS Private Work
k_cll1 = 0.0 +0.4 −0.5
Δk_cll1


CMS Private Work
k_cW = -0.00 +0.11 −0.10
Δk_cW


CMS Private Work
k_cWtil = 0.0 +0.4 −0.5
Δk_cWtil