ETP-KA/2024-10

Karlsruher Institut für Technologie

Searches for the rare tWZ and $tW\gamma$ processes at the LHC using machine learning techniques

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
angenommene

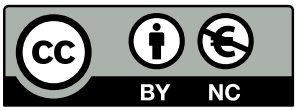
DISSERTATION

von

M.Sc. Michele Mormile

Tag der mündlichen Prüfung: 10. Januar 2025

Referent: Prof. Dr. Ulrich Husemann
Koreferentin: Prof. Dr. Abideh Jafari



Contents

Introduction

Established through decades of experimental discoveries and ultimately developed in its current form in the 1970s, the Standard Model of particle physics (SM) is the widely accepted theory which describes the universe at the microscopical level. The SM explains the behavior of elementary particles, the fundamental constituents of matter, and three of the interactions between them: the electromagnetic, weak, and strong forces. The gravitational force, whose effects are negligible at the microscopical scale, is not included in the SM.

The SM has been extensively tested in the years, and all its predictions have been confirmed by experiments; the most recent one being the discovery of the Higgs boson in 2012 by the ATLAS and CMS collaborations at the Large Hadron Collider (LHC) at CERN in Geneva, Switzerland. The history of experimental discoveries anticipated by theoretical predictions in particle physics is a rich and fruitful one, which dates back to the discovery of the positron by Anderson in 1932 [1], which was predicted by Dirac in 1928 [2].

While the SM is an extremely successful theory, it is known today to be incomplete. In addition to it not describing the gravitational interaction, the SM also does not include explanations for the origin of dark matter and the imbalance between matter and antimatter in the universe, among other phenomena. At the LHC, the SM is tested in several complementary ways, which include the direct search of new particles and processes that are not included in the theory, and the precise measurement of known processes to test the SM predictions. An additional way to test the SM is to search for rare processes that are predicted by the theory, but have not been observed yet. This is the case for the associated production of a single top quark with a W boson and an additional photon ($tW\gamma$) or Z boson (tWZ), whose first searches, and resulting first evidence, are presented in this thesis work.

The $tW\gamma$ and tWZ processes share many similar characteristics, the most important of which is their overlap at Feynman diagram level with the $t\bar{t}\gamma$ and $t\bar{t}Z$ processes at Next-To-Leading Order (NLO) in Quantum chromodynamics (QCD) perturbation theory. This overlap causes the perturbative calculation of the $tW\gamma$ and tWZ processes not to be reliable, unless it is treated with dedicated techniques called overlap removal.

On the experimental level, the higher production cross sections of the $t\bar{t}\gamma$ and $t\bar{t}Z$ processes compared to the $tW\gamma$ and tWZ processes, and the similar final state topology, make the separation of the signal from the background a challenging task. The approach used in this

thesis consists in developing dedicated Machine Learning (ML) algorithms to discriminate the signal from the background, and employ the output of these algorithms to perform a maximum likelihood fit of the simulated signal and background events to the data, in order to extract the statistical significance of the signal.

This thesis is organized as follows. Chapter 1 presents the theoretical framework of the SM, and it describes the properties of the top quark. Chapter 2 describes the studies performed in the context of this thesis work in order to determine a reliable modeling of the $tW\gamma$ and tWZ processes. Particularly, it describes in detail the techniques used to remove the overlap between the $t\bar{t}\gamma$ and $t\bar{t}Z$ processes and the $tW\gamma$ and tWZ processes. Additionally, studies on the agreement between two of the overlap removal schemes are presented.

Chapter 3 is divided in two main parts. Section 3.1 describes the statistical model used to perform the maximum likelihood fit from which the significance is extracted for both the $tW\gamma$ and tWZ processes. It additionally describes the saturated goodness-of-fit tests performed to validate the input features to the ML algorithms used in the search for the $tW\gamma$ process. Section 3.2 describes ML, and the ML algorithms used in both the searches presented in this thesis work. This section also provides an introduction to Graph Neural Networks (GNN), which is the ML algorithm used in the search for the $tW\gamma$ process.

Chapter 4 describes the LHC and the CMS experiment, which is the experiment where the data used in this thesis work was collected. It also describes how particles are reconstructed and identified in the CMS detector, and how their properties are measured.

Finally, Chapter 5 and Chapter 6 present the first searches and first evidence for tWZ and $tW\gamma$ production, respectively. The analysis were performed using data collected by the CMS experiment in 2016, 2017, and 2018, corresponding to an integrated luminosity of 138 fb^{-1}.

Lastly, a summary of the results and some concluding remarks are presented, together with an outlook on future studies that can be performed in the context of the $tW\gamma$ and tWZ processes.

1 The Standard Model of particle physics

This chapter provides an overview of the Standard Model of particle physics (SM), which is the theoretical framework that describes the fundamental particles and their interactions. The SM is a quantum field theory (QFT) based on the principles of quantum mechanics and special relativity. The theory has been developed over the course of the 20th century in order to account for the experimental evidence coming from numerous experiments. It has made a number of predictions that have been confirmed by experiments, the latest one being the discovery of the Higgs boson by the CMS and ATLAS experiments at the LHC in 2012 [3, 4]. Although the SM is a very successful theory, which has still not been contradicted by any experimental evidence, it is not a complete theory of particle physics. It does not include a description of gravitational interaction, and it does not account for the existence of dark matter, which make up most of the matter in the universe [5]. The SM also does not include a source of CP violation large enough to justify the asymmetry between baryonic and antibaryonic matter in the universe.

This chapter is structured as follows: Section 1.1 provides an overview of the particles and interactions considered by the SM, with the Lagrangian formalism of the theory described in Section 1.2. Sections 1.3 to 1.6 detail the quantum field theories that describe the interactions of the SM. Sections 1.7 and 1.8 introduce Feynman diagrams, which are a graphical representation of the interactions between particles in a perturbative approach, and the renormalization procedure, which is used to remove the divergences that arise in perturbative calculations. Section 1.9 describes hadronic collisions and how they are modelled. Finally, Section 1.10 provides an overview of the top quark, the heaviest known elementary particle, which is central in the discussion of the analyses presented in this thesis.

The units used in this chapter and throughout this thesis are natural units, for which $\hbar = c = 1$. Moreover, the Einstein summation convention is used, where repeated indices are summed over.

The discussions in this chapter are inspired by Refs. [6–9].

Table 1.1: Properties of the quarks in the Standard Model, from Ref. [14].

Quark	**Quark type**	**Electric charge (e**	**Mass** (GeV)
up	up	$+\frac{2}{3}$	$(2.16 \pm 0.07) \cdot 10^{-3}$
down	down	$-\frac{1}{3}$	$(4.70 \pm 0.07) \cdot 10^{-3}$
charm	up	$+\frac{2}{3}$	1.2730 ± 0.0046
strange	down	$-\frac{1}{3}$	$(93.5 \pm 0.8) \cdot 10^{-3}$
top	up	$+\frac{2}{3}$	172.57 ± 0.29
bottom	down	$-\frac{1}{3}$	4.183 ± 0.007

Table 1.2: Properties of the leptons in the Standard Model, from Ref. [14].

Lepton	**Electric charge (e)**	**Mass** (MeV)
electron neutrino	0	$< 8 \cdot 10^{-4}$
electron	-1	0.511
muon neutrino	0	$< 8 \cdot 10^{-4}$
muon	-1	105.7
tau neutrino	0	$< 8 \cdot 10^{-4}$
tau	-1	1776.93 ± 0.09

1.1 The Standard Model

The SM is the established QFT that describes elementary particle physics and their interactions. The interactions considered by the SM are the strong, weak, and electromagnetic interactions. The fourth fundamental force of nature, gravity, is not included in the SM, as it is not yet described by a quantum field theory.

All the particles included in the model can be categorized into fermions and bosons, according to their spin. Fermions are particles with half-integer spin which obey the Fermi-Dirac statistics [10, 11] and the Pauli exclusion principle: two fermions cannot occupy the same quantum state. Bosons, on the other hand, have integer spin and follow the Bose-Einstein statistics [12]: any number of bosons can occupy the same quantum state.

Fermions make up the ordinary matter of the universe and all elementary fermions carry a spin of 1/2. They are further divided according to the interactions they experience. Quarks are fermions which interact through all the fundamental forces of the SM and are always collected in bound states called hadrons. They are divided in up-type and down-type quarks, according to their electric charge. The properties of quarks are shown in Table 1.1. Leptons are fermions that do not experience the strong force, they are divided in charged leptons, which interact through the electromagnetic and weak forces, and neutrinos, which only interact through the weak force. Neutrinos are assumed to be massless in the SM, but experimental evidence coming from neutrino oscillation experiments has shown that neutrinos must have a mass, although it is very small [13]. The properties of the leptons are shown in Table 1.2. For each fermion, there is a corresponding antiparticle with the same mass and opposite charge.

Table 1.3: Properties of the mediator bosons in the Standard Model, from Ref. [14].

Boson	**Elecric charge (e)**	**Mass** (GeV)
photon	0	0
gluon	0	0
$W^{\pm}$	± 1	80.37 ± 0.01 GeV
Z	0	91.188 ± 0.002 GeV

Fermions come in three generations, each containing an up-type quark, a down-type quark, a charged lepton, and a neutrino. The generations are different in terms of the mass of the particles, as they become heavier with each generation. The other characteristics of fermions are essentially the same for all generations.

Mediator bosons are the carriers of the fundamental forces of the SM and are responsible for the interactions between the fermions. All mediator bosons are spin-1 particles. The photon is the carrier of the electromagnetic force, it is massless and does not have electric charge. The gluon is the carrier of the strong force; it is massless and carries color charge, it can therefore interact with itself. The SM includes eight gluons, corresponding to the eight possible combinations of color charge. The $W^{\pm}$ and Z bosons are the carriers of the weak force, they are the only massive interaction carriers in the SM. The $W^{\pm}$ bosons are charged and are the only bosons known to change the flavor of a fermion. The Z boson is neutral. The properties of the bosons are shown in Table 1.3.

The last particle of the SM is the Higgs boson, discovered in 2012 by the ATLAS and CMS collaborations at the Large Hadron Collider (LHC). It is the only scalar boson in the SM, as it have a spin of 0. The Higgs boson has been predicted to exist as it arises from the Higgs mechanism, which is responsible for the generation of the masses of the $W^{\pm}$ and Z bosons [15, 16]. Ref. [14] reports the mass of the Higgs boson as 125.20 ± 0.11 GeV.

1.2 The Lagrangian formalism of the Standard Model

The particles considered in the SM are treated as quantum fields $\psi(x)$, where the x are the four spacetime coordinates. The properties of the fields are described through the Lagrangian density $\mathcal{L}$, which is a function of the fields and their gradients. The Euler-Lagrange equations are used to derive the dynamics of the fields from the Lagrangian density:

$$\frac{\partial \mathcal{L}}{\partial \psi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \right) = 0 \ . \tag{1.1}$$

The Lagrangian density that describes a free fermion field is given by the Dirac Lagrangian:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \ , \tag{1.2}$$

where $\bar{\psi}$ is the adjoint of the field ψ: $\bar{\psi} = \psi^\dagger \gamma^0$ and m is the mass of the fermion. The γ^μ are the Dirac matrices, first introduced by Dirac in 1928 [2], together with the Dirac equation. The Dirac equation is obtained by evaluating Eq. (1.1) for Eq. (1.2). The first

term on the right-hand side can be interpreted as the kinetic term of the fermion field, while the second term represents its mass.

The interaction between fermions is described as the exchange of mediator bosons. They are introduced in the Lagrangian density through the gauge principle, which states that the Lagrangian must be invariant under local gauge transformations. The SM is therefore a local gauge theory. Quantum electrodynamics (QED) is described in the next section as an example of a local gauge theory.

1.3 Quantum electrodynamics

Quantum electrodynamics (QED) is the quantum field theory that describes the electromagnetic interaction. It is based on the premise that the Lagrangian density of the theory must be invariant under a local $\mathbf{U}(1)$ gauge transformation. $\mathbf{U}(1)$ is the group of all complex numbers with absolute value 1 and a local transformation is a transformation that depends on the spacetime coordinates.

The $\mathbf{U}(1)$ group is responsible for the transformation:

$$\psi(x) \to \psi'(x) = e^{iq\alpha(x)}\psi(x) \ , \tag{1.3}$$

where q is the charge of the fermion field and $\alpha(x)$ is a scalar function of the spacetime coordinates. For the gauge principle, a new field A_μ needs to be introduced in order to keep the Lagrangian density invariant. The new field has to transform under the action of the $\mathbf{U}(1)$ group as:

$$A^\mu \to A'^\mu = A^\mu - \partial^\mu \alpha(x) \ . \tag{1.4}$$

The field A_μ is identified as the electromagnetic vector potential, which respects the Maxwell equations:

$$\partial^\mu \partial_\mu A^\nu = 0 \ . \tag{1.5}$$

The corresponding Lagrangian density for the electromagnetic field is:

$$\mathcal{L}_{\text{EM}} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} \ , \tag{1.6}$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the electromagnetic field strength tensor, which describes the electromagnetic field in the spacetime coordinates.

It is possible to define a gauge invariant Lagrangian density for QED by replacing the partial derivative in the Dirac Lagrangian with the covariant derivative, defined as:

$$D_\mu = \partial_\mu - iqA_\mu \ . \tag{1.7}$$

The QED Lagrangian density is then:

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu A_\mu \psi - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \ . \tag{1.8}$$

The third term on the right-hand side of Eq. (1.8) represents the interaction between the fermion field and the electromagnetic field. The charge q of the fermion field corresponds to the strength of the coupling between the fermion and the electromagnetic field. According to Noether's theorem [17], the invariance of the Lagrangian density under a global $\mathbf{U}(1)$ transformation implies the conservation of the electric charge.

1.4 The electroweak interaction

A first approach to describe the weak interaction was given by Fermi in 1934 [18], who proposed a four-fermion contact interaction to describe the beta decay of the neutron. This theory was successful in describing the weak interaction at low energies and is today seen as a low-energy approximation of the electroweak theory. Among the first proposals of a unified electroweak theory was the one by Schwinger in 1957 [19], who envisioned the electroweak interaction as mediated by three bosons, one of which being the photon mediating the neutral weak interaction and two $W^{\pm}$ bosons mediating the charged weak interaction, which changes the flavor of the fermions. This theory fell short of explaining parity violation in the weak interaction, which was observed by Wu in 1957 [20]. More experimental evidence then proved that the weak charged current maximally violates parity, i.e., it only couples to left-handed fermionic chiral states and right-handed antifermionic chiral states [21].

Chirality is a property of the fermions that is described using the chirality operator $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. The left-handed and right-handed chiral states are the eigenstates of the chirality operator with eigenvalues -1 and $+1$, respectively. These can be obtained by projecting the fermion field onto the left-handed and right-handed chiral states using the projection operators

$$P_L = \frac{1}{2}(1-\gamma^5) \quad \text{and} \quad P_R = \frac{1}{2}(1+\gamma^5)\ . \tag{1.9}$$

The accepted electroweak unification theory was described by the Glashow-Salam-Weinberg (GSW) model [22–24], which is a local gauge theory based on the $\mathbf{SU}(2)_L \times \mathbf{U}(1)_Y$ group. The L subscript in $SU(2)_L$ stands for left-handed, as the weak interaction only couples to left-handed fermions, while the Y subscript in $\mathbf{U}(1)_Y$ stands for hypercharge, which is a combination of the weak isospin T and the electric charge. The weak isospin is the quantity that is conserved under the weak interaction. Left-handed up-type quarks (U) and charged leptons (ℓ^-) are grouped into weak isospin doublets with the down-type quark (D') and the neutrino (ν_ℓ) of the same generation:

$$\begin{pmatrix} U \\ D' \end{pmatrix}_L \quad \text{and} \quad \begin{pmatrix} \nu_\ell \\ \ell^- \end{pmatrix}_L\ , \tag{1.10}$$

where D' is a linear combination of the down-type quarks through the Cabibbo-Kobayashi-Maskawa (CKM) matrix [25].

The doublets have a weak isospin of 1/2, with the up-type quark and the neutrino having a value of the third component of the weak isospin $T_3 = +1/2$ and the down-type quark and the charged lepton having $T_3 = -1/2$. Right-handed fermions are weak isospin singlets, as they do not interact through the weak force.

The weak hypercharge Y is defined as:

$$Y = 2(Q - T_3) \ , \tag{1.11}$$

where Q is the electric charge of the fermion. The weak hypercharge of the left-handed doublets is $Y = -1$ and the weak hypercharge of the right-handed singlets is $Y = -2$.

By repeating the procedure of Section 1.3, it is possible to introduce four new bosons in the Lagrangian density of the electroweak theory. From the $\mathbf{U}(1)_Y$ group, a new field B_μ is introduced in the Lagrangian density in exactly the same way as the electromagnetic vector potential A_μ was introduced in QED, with the difference that the electric charge q is replaced by $g' \frac{Y}{2}$, where g' is a new coupling constant.

The $\mathbf{SU}(2)_L$ group is responsible for the gauge transformation of the weak isospin doublets ψ_L:

$$\psi_L \rightarrow \psi'_L = e^{ig\frac{\vec{\sigma}}{2}\cdot\vec{\alpha}(x)}\psi_L \ , \tag{1.12}$$

where g is the weak coupling constant, $\vec{\sigma}$ are the three Pauli matrices, and $\vec{\alpha}(x)$ is a vector function of the spacetime coordinates.

To respect the gauge principle, three new fields $\vec{W}_{1,2,3}$ need to be introduced in the Lagrangian density, which transform under the action of the $\mathbf{SU}(2)_L$ group as:

$$W_k^\mu \rightarrow W_k'^\mu = W_k^\mu - \partial^\mu \alpha_k - g\varepsilon_{ijk}\alpha_i W_j^\mu \ , \tag{1.13}$$

where ε_{ijk} is the Levi-Civita tensor and $i, j, k = 1, 2, 3$. The Levi-Civita tensor appears as a consequence of the non-abelian nature of the $\mathbf{SU}(2)$ group, i.e., the sigma matrices do not commute with each other.

The interaction between the fermions and the $W_{1,2,3}$ bosons is then described by the Lagrangian density:

$$\mathcal{L}_\mathrm{W} = -g\bar{\psi}_L\gamma^\mu\frac{\vec{\sigma}}{2}\cdot\vec{W}_\mu\psi_L \ , \tag{1.14}$$

obtained by replacing the partial derivative in the Dirac Lagrangian with the covariant derivative $D_\mu = \partial_\mu - ig\frac{\vec{\sigma}}{2}\cdot\vec{W}_\mu$.

From Eq. (1.14), the physical $W^\pm$ bosons can be identified as the linear combinations of the $\mathrm{W}_{1,2}$ bosons:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_{1\mu} \mp iW_{2\mu}) \ . \tag{1.15}$$

The remaining physical bosons are the Z boson and the photon, which are the linear combinations of the B_μ and $W_{3\mu}$ bosons:

$$A_\mu = B_\mu \cos\theta_W + W_{3\mu} \sin\theta_W \ , \tag{1.16}$$

$$Z_\mu = -B_\mu \sin\theta_W + W_{3\mu} \cos\theta_W \ , \tag{1.17}$$

where θ_W is the weak mixing angle. To reproduce the observed couplings of QED, the weak mixing angle is defined as:

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \quad \text{and} \quad \sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \ . \tag{1.18}$$

At the time the GSW model was proposed, the existence of weak neutral currents was not yet observed. The first evidence of the weak neutral currents was provided by the Gargamelle experiment at CERN in 1973 [26], which also measured for the first time the weak mixing angle.

Massive spin-1 need to respect the Proca equations, which are the generalization of the Maxwell equations for massive fields. The corresponding Lagrangian density for a massive vector field is:

$$\mathcal{L}_{\text{Proca}} = -\frac{1}{4}(\partial^\mu W_i^\nu - \partial^\nu W_i^\mu)(\partial_\mu W_\nu^i - \partial_\nu W_\mu^i) + \frac{1}{2} m^2 W_i^\mu W_\mu^i \ , \tag{1.19}$$

where m is the mass of the vector boson and the index i runs over the three components of the vector boson. The first term on the right hand side of Eq. (1.19) is the kinetic term of the vector boson, while the second term is the mass term. Both terms are not invariant under local gauge transformations. The kinetic term can be made invariant by introducing additional terms in the Lagrangian:

$$\mathcal{L}_{\text{int}} = \frac{1}{2} g_W \varepsilon_{ijk} (\partial^\mu W_i^\nu - \partial^\nu W_i^\mu) W_{j\mu} W_{k\nu} - \frac{1}{4} g_W^2 \varepsilon_{ijk} \varepsilon_{imn} W_j^\mu W_k^\nu W_{m\mu} W_{n\nu} \ . \tag{1.20}$$

The terms in Eq. (1.20) are the trilinear and quadrilinear self-interaction terms of the vector bosons, which causes the weak bosons to interact with each other.

A similar approach is not possible for the mass term, which therefore breaks the gauge invariance of the Lagrangian. The solution to this issue is the incorporation of the Higgs mechanism in the SM, which is responsible for the generation of the masses of the $W^\pm$ and Z bosons, and justifies the introduction of the θ_W mixing angle in the electroweak theory.

1.5 The Higgs mechanism

The Higgs mechanism is the mechanism of spontaneous gauge symmetry breaking that generates the masses of the $W^\pm$ and Z bosons in the SM. It is based on the introduction

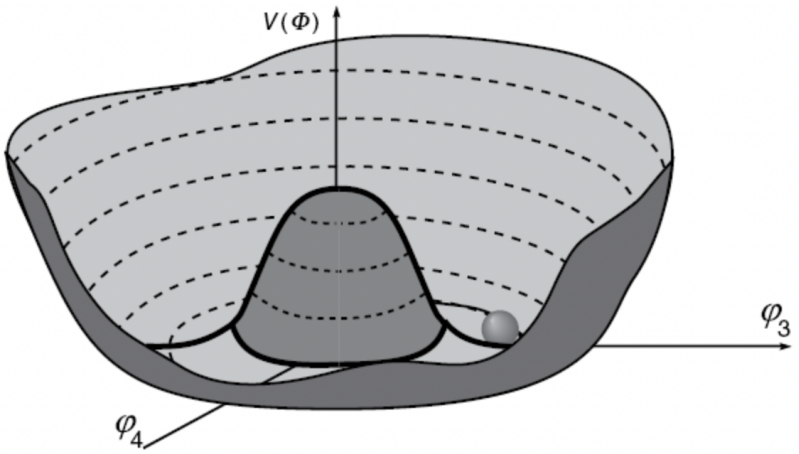


Figure 1.1: Shape of the Higgs potential for $\mu^2 < 0$, from Ref. [7]. The ball *rolled down* to one minimum of the infinitely many on the circle in the $\phi_3 - \phi_4$ plane of the potential represents the spontaneous symmetry breaking.

of two complex scalar fields placed in a weak isospin doublet:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} . \tag{1.21}$$

The Lagrangian density of the Higgs field is:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} &= (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi) \\ &= (D^\mu \Phi)^\dagger (D_\mu \Phi) - \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 , \end{aligned} \tag{1.22}$$

where D_μ is the covariant derivative of the $\mathbf{SU}(2)_L \times \mathbf{U}(1)_Y$ group:

$$D_\mu = \partial_\mu - ig\frac{\vec{\sigma}}{2} \cdot \vec{W}_\mu - ig'\frac{Y}{2} B_\mu . \tag{1.23}$$

Figure 1.1 shows the shape of the Higgs potential $V(\phi)$ in the $\phi_3 - \phi_4$ plane with $\mu^2 < 0$. The potential is originally symmetrical, and therefore gauge invariant, but the system is in an unstable equilibrium.

A small perturbation can cause the system to fall into any of the degenerate minima of the potential, breaking the gauge symmetry. The minima are defined by:

$$\phi^\dagger \phi = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda} , \tag{1.24}$$

and the Higgs field is said to have a non-zero vacuum expectation value (VEV) v. Without any loss of generality, the Higgs field can be expanded around the minimum of the potential:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + \phi_3 + i\phi_4 \end{pmatrix} . \tag{1.25}$$

Via a convenient choice of the gauge, the Higgs field can be written as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} , \tag{1.26}$$

where h represents the physical Higgs boson.

By substituting Eq. (1.26) into Eq. (1.22) and expanding the covariant derivative, the mass terms of the $W^{\pm}$ and Z bosons can be obtained as:

$$m_W = \frac{1}{2} v g \quad \text{and} \quad m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2} = \frac{m_W}{\cos \theta_W}. \tag{1.27}$$

This result links together the masses of the $W^{\pm}$ and Z bosons, and the weak mixing angle θ_W. Additionally, the Higgs mechanism produces terms which represent the interaction between the Higgs boson and the $W^{\pm}$ and Z bosons, as well as the trilinear and quadrilinear self-interaction terms, and the mass term, of the Higgs boson.

Finally, the mass term of the Dirac Lagrangian for fermions

$$m\bar{\psi}\psi = m(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L) \tag{1.28}$$

is not invariant under a local $\mathbf{SU}(2)_L \times \mathbf{U}(1)_Y$ gauge transformation, because of the different transformation properties of the left-handed and right-handed components. Weinberg's solution to this issue [23] is the introduction of a Yukawa interaction [27] of the fermionic field and the Higgs field:

$$\mathcal{L}_{\text{Yukawa}} = \frac{1}{\sqrt{2}} f(\bar{\psi}_L \Phi \psi_R + \bar{\psi}_R \Phi \psi_L) , \tag{1.29}$$

where f is the Yukawa coupling constant. The Yukawa interaction generates the masses of the fermions, which are proportional to the Yukawa coupling constant and the VEV of the Higgs field: $m_f = \frac{1}{\sqrt{2}} f v$. The same lagragian density also accounts for the interaction between the Higgs boson and the fermions, the strength of which is proportional to the Yukawa coupling constant, and therefore to the mass of the fermion.

Lastly, it should be noticed that Eq. (1.22) only accounts for the masses of the fermions in the lower element of the isospin doublets, given the representation chosen for the Higgs field. The masses of the fermions in the upper element of the doublet are generated by means of an equivalent Yukawa interaction with the replacement of ϕ with its charge conjugate doublet

$$\phi_c = -i\sigma_2 \phi^* = \begin{pmatrix} v + h \\ 0 \end{pmatrix} . \tag{1.30}$$

1.6 Quantum chromodynamics

Quantum chromodynamics (QCD) is the quantum field theory that describes the strong interaction. The strong interaction is responsible for the interactions between quarks, and

their confinement into hadrons. It also binds together the protons and neutrons in the atomic nuclei.

QCD is a local gauge theory based on the **SU**(3) group: the group of all 3×3 unitary matrices with determinant 1. It can be introduced in the same way of QED, by starting from considering the transformation of the quark fields under the action of the **SU**(3) group:

$$\psi(x) \to \psi'(x) = e^{ig_s \frac{\lambda_i}{2} \alpha_i(x)} \psi(x) \ , \tag{1.31}$$

where g_s is the strong coupling constant, λ_i are the eight Gell-Mann matrices [28], and $\alpha_i(x)$ are eight functions of the spacetime coordinates. Since the Gell-Mann matrices are 3×3 matrices, the quark fields need to be considered as three-component vectors, where all the elements of one vector have the same quark composition but different color charge. Color (green, red, blue) is the charge of the strong interaction.

Eight new fields G^i_μ, representing the gluons, are introduced in the definition of the covariant derivative to ensure the gauge invariance of the Dirac Lagrangian density:

$$D_\mu = \partial_\mu - i g_s \frac{\lambda_i}{2} G^i_\mu \ . \tag{1.32}$$

The gluon fields need to transform under the action of the **SU**(3) group as:

$$G^i_\mu \to G'^i_\mu = G^i_\mu - \partial_\mu \alpha_i - g_s f_{ijk} \alpha_j G^k_\mu \ , \tag{1.33}$$

where f_{ijk} are the structure constants of the **SU**(3) group. As for the Levi-Civita symbols in Eq. (1.13), the f_{ijk} appear as a consequence of the non-abelian nature of the **SU**(3) group.

For the same reason, the kinetic term of the gluons in the Lagrangian density needs to be added together with self-interaction terms in order to respect the gauge invariance of the theory, in complete analogy to what is done in Eq. (1.2). The self-interaction terms include trilinear and quadrilinear couplings of the gluon fields with each other.

1.7 Feynman diagrams and perturbation theory

The results of particle physics are usually expressed in terms of decay rates and cross sections, which represent the probability of a certain transition between fields to occur. The transition rate Γ_{fi} from an initial state i to a final state f is calculated using Fermi's golden rule [29]

$$\Gamma_{fi} = 2\pi |T_{fi}|^2 \rho(E_f), \tag{1.34}$$

where $\rho(E_f)$ is the density of states for the final state at energy E_f, and T_{fi} is the transition matrix element between the initial and final states. The transition matrix element can be calculated using a perturbative method, where the interaction between the fields is treated as a perturbation to the free fields:

$$T_{fi} = \sum_{j \neq i} \frac{\langle f | \mathcal{L}_{\text{int}} | j \rangle \langle j | \mathcal{L}_{\text{int}} | i \rangle}{E_i - E_j} \ + \ldots , \tag{1.35}$$

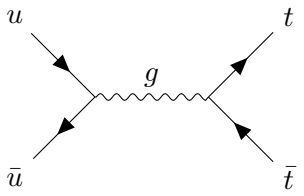


Figure 1.2: Feynman diagram representing the transition from a pair of up quark and anti-quark to a pair of top quark and anti-quark through the exchange of a gluon.

where $\mathcal{L}_{\text{int}}$ is the Lagrangian density of the interaction between the fields, and E_i and E_j are the energies of the initial and final states, respectively.

Eq. (1.35) represents a transition between two states through an intermediate state j, and is defined as Leading Order (LO). In general, it is necessary to consider all possible intermediate states to calculate the transition matrix element, as this is the only way to ensure that the T_{fi} is gauge invariant. Taking only the LO contribution as T_{fi} yields an error of the order of the squared coupling constant of the interaction. Higher order contributions can be calculated by considering more intermediate states in the transition matrix element, and are defined as Next-to-Leading Order (NLO), Next-to-Next-to-Leading Order (NNLO), and so on.

The transition matrix elements are usually calculated using Feynman diagrams [30], which are graphical representations of the terms in Eq. (1.35). An example of a Feynman diagram is given in Figure 1.2, which represents the transition from an initial state of a pair of up quark and anti-quark to a final state of a pair of top quark and anti-quark through the exchange of a gluon. This transition is made possible by the term

$$\mathcal{L}_{\text{int}} = g_s \bar{\psi} \gamma^\mu \frac{\lambda_i}{2} G^i_\mu \psi, \tag{1.36}$$

of the QCD Lagrangian density, which couples the quark fields to the gluon fields.

1.8 Renormalization

Renormalization is the process of removing the unphysical divergences that appear in the perturbative calculations of the transition matrix elements. This section describes the renormalization of interaction vertices. A further discussion, in addition to the description of renormalization for the masses of particles, and for fermionic fields can be found in Ref. [9]. The origin of the divergences can be understood looking at the diagram shown in Figure 1.3, which represents a possible diagram for the scattering of an electron and a positron. The process proceeds through the exchange of a photon, which creates a *loop* of a virtual electron-positron pair. The loop enters the perturbative calculation of the transition matrix as an integral over the momentum of the virtual particles in the loop, which is a divergent quantity.

Renormalizing the theory consists in absorbing the loops of the Feynman diagrams into the definition of the coupling constant of the theory, in this case the electric charge e. The

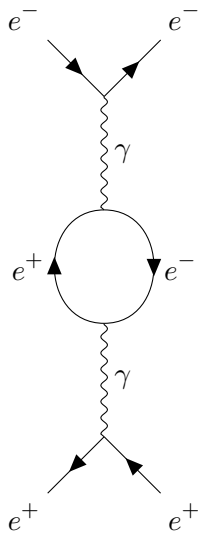


Figure 1.3: Feynman diagram representing the scattering of an electron and a positron proceeding through the exchange of a virtual photon which creates a virtual electron-positron pair.

treatment consists in considering the electric charge $e(q^2)$ as a function of the momentum q of the exchanged photon, i.e., a running coupling constant. Intuitively, this dependency can be understood by considering a free electron as always surrounded by a cloud of virtual fermion-antifermion pairs which screen the bare charge of the electron. The electric charge of the electron can be measured through its interaction with a photon. The higher the energy of the photon, the shorter the distance it can penetrate the cloud of virtual pairs, effectively interacting with a higher net charge.

The running of the coupling constant is usually stated in terms of $\alpha = e^2/4\pi$, which for QED is the fine structure constant. The variation of α with the energy scale is described by the equation of the renormalization group, which links the value $\alpha(q^2)$ at a certain energy scale q^2 to the value $\alpha(\mu)$ measured at a reference energy scale μ:

$$\alpha(q^2) = \frac{\alpha(\mu)}{1 - \frac{\alpha(\mu)}{3\pi} \log\left(\frac{q^2}{\mu^2}\right)} \,. \tag{1.37}$$

For QED, the value of α increases with the energy scale from $\alpha(q^2 \approx 0) = 1/137$.

Renormalization is necessary for QCD as well, as the theory presents the same type of divergences as QED. Due to the self-interaction of the gluons, the loops of the Feynman diagrams also include loops of gluons, additionally to the fermion loops already present in QED. The running of the strong coupling constant $\alpha_s = g_s^2/4\pi$ is described by the

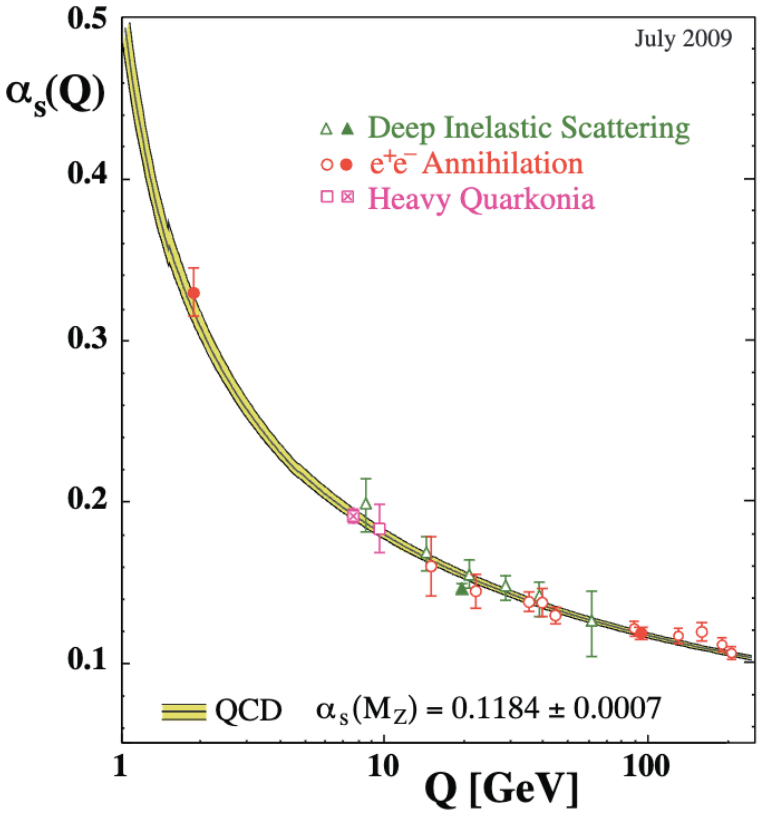


Figure 1.4: Summary of measurements of α_s at different energy scales Q, from Ref. [31].

equation:

$$\alpha_s(q^2) = \frac{\alpha_s(\mu)}{1 + \beta\alpha_s(\mu)\log\left(\frac{q^2}{\mu^2}\right)} , \tag{1.38}$$

$$\beta = \frac{11N_c - 2N_f}{12\pi} , \tag{1.39}$$

where N_c is the number of colors of the theory, and N_f is the number of flavors of quarks accessible at the energy scale considered. In the SM, $N_c = 3$ and $N_f \leq 6$, and therefore the value of β is always positive, meaning that the value of α_s decreases with the energy scale, differently from the running of α in QED. Intuitively, this behavior can be understood by considering that the cloud of gluons surrounding a quark enhances the bare color charge of the quark, in contrast to the screening effect of the virtual pairs in QED. The measured values of α_s at different energy scales are shown in Figure 1.4.

This behaviour has several consequences. At low energy scales, the value of α_s is large, and a perturbative approach is not possible. This causes the value of α_s obtained from Eq. (1.38) to diverge for $q^2 = \Lambda_{QCD} \sim O(300 \text{ MeV})$ [14]. This energy scale is called the QCD scale, and represents the energy scale at which the strong interaction becomes non-perturbative and the quarks and gluons are confined into hadrons. On the other hand, for higher energies, and therefore lower distances, $\alpha_s \approx 0$, allowing the quarks inside the hadrons to be treated as free particles. Hence, when considering the scattering of hadrons at

high energy scales, the perturbative approach is valid, and the interaction can be described as a scattering of free quarks and gluons.

1.9 Proton-proton collisions at the LHC

The analyses described in this thesis work employ data collected by the CMS experiment at the LHC, which collides protons at a center-of-mass energy of $\sqrt{s} = 13$ TeV. Protons are composite particles made up of three valence quarks, two up and one down quarks, and a *sea* of gluons and virtual quark-antiquark pairs, collectively named *partons*.

1.9.1 Factorization

As mentioned in Section 1.8, at the high collision energy of the LHC, the partons can be treated as free particles. This behaviour is summarized by the QCD factorization theorem[32], which states that the cross section $\sigma_{pp\to X}$ of a process $pp \to X$ can be factorized as:

$$\sigma_{pp\to X} = \sum_{i,j} \int\int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij\to X}(x_a p_a, x_b p_b, \mu_F^2, \mu_R^2) \ , \tag{1.40}$$

where $f_a(x_a, \mu_F^2)$ and $f_b(x_b, \mu_F^2)$ are the parton distribution functions (PDFs) of the partons i and j in the protons, which describe the probability of finding that parton with a fraction x of the proton momentum. $\hat{\sigma}_{ij\to X}$ is the partonic cross section, i.e., the cross section related to the scattering of the partons i and j into the final state X, also defined as *hard process*. The PDFs depend on the factorization scale μ_F, which is the energy scale at which the PDFs are evaluated. $\hat{\sigma}_{ij\to X}$ additionally depend on a renormalization scale μ_R, which affects the value of the couplings, as described in Section 1.8, and on the center-of-mass energy of the partonic collision $\sqrt{\hat{s}}$.

Figure 1.5 shows a Feynman diagram representing the process $pp \to t\bar{t}$ proceeding through the hard process $gg \to t\bar{t}$. The remnants of the protons, which are not involved in the hard scattering, are considered through the parton showering and hadronization methods, described in Sections 1.9.2 and 1.9.3.

The PDFs are not yet calculable from first principles. In the analyses described in this thesis work, the PDFs used for the simulation of the proton-proton collisions are the NNPDF3.1 set [33], which are determined from a global fit to a wide range of experimental data. The extrapolation of the PDFs to an arbitrary energy scale is performed using the DGLAP evolution equations [34–36]. Figure 1.6 shows the PDFs of the proton as a function of the momentum fraction x at two different factorization scales.

When considering partonic processes that are initiated by b quarks, two schemes are employable for the PDFs. The 4-flavor scheme (4FS) does not consider the b quark as a parton in the proton, and therefore the PDFs do not include the contribution of the b quark, which can only originate from splitting of a gluon, and can be massive. On the other hand, the 5-flavor scheme (5FS) considers the b quark as a massless parton in the proton with its own PDF.

Figure 1.5: Feynman diagram representing the process $pp \to t\bar{t}$ proceeding through fusion of two gluons coming from the protons.

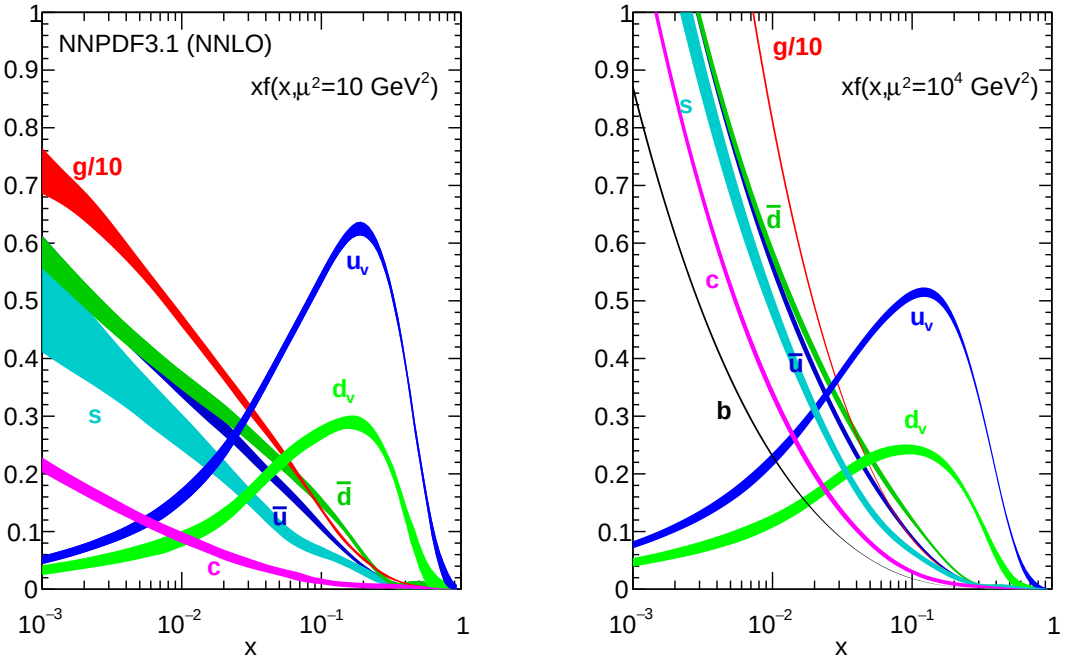


Figure 1.6: PDFs of the proton as a function of the momentum fraction x at the factorization scale $\mu_F^2 = 10$ GeV^2 (left) and $\mu_F^2 = 10^4$ GeV^2 (right), from Ref. [33]. The v subscript indicates the valence quarks. The fraction of momentum carried by partons from the sea increases with the energy scale.

1.9.2 Parton showering

When using the factorization theorem, the hard process can only account for energy scales higher than the factorization scale. The range of phenomena that occur at lower energy scales is usually modeled by factorizing them in other components, so to cover the whole energy phase space of a hadronic collision.

The parton showering model accounts for the radiation of additional particles, e.g., partons and photons, from the initial and final particles of the hard process. These are defined as Initial State Radiation (ISR) and Final State Radiation (FSR), respectively. The model is based on the DGLAP equations, which in this case describe the probability of an initial state parton to be radiated from an additional parton, and the probability of a final state particle to radiate an additional parton or a photon. The same formalism is used to describe the radiation of photons in QED from fermions, and the conversion of photons into fermion-antifermion pairs, as well as the decay of possible resonances created in the hard process.

At the same energy scale as the parton showering, Multiple Particle Interactions (MPI) can occur, which are the result of the interaction of the remnants of the protons that are not involved in the hard scattering, described in Refs. [37–39].

1.9.3 Hadronization

The parton showering model uses a perturbative approach which is valid at energies larger than the QCD scale Λ_{QCD}, below which the colored partons recombine into colorless hadrons. The process of hadronization is non-perturbative; simulation of QCD processes rely on phenomenological models for hadronization. One such model, used in the `PYTHIA8` generator [40], is the Lund String model [41, 42].

This model considers two interacting quarks as connected by a flux tube of gluons, called a *string*. The energy of the string is linearly proportional with the distance between the quarks. When this energy is higher than a certain threshold value, the string can break, and the energy is converted into the mass of new quark-antiquark pairs, which recombine with the original partons to form the colorless hadrons.

Several models exist to describe between which partons the strings are formed, the one used in `PYTHIA8`, and therefore in the simulation of events in this thesis, is the *leading color* model [43].

1.10 The top quark

This thesis work is centered around the study of the top quark, which is the heaviest known elementary particle. It was first predicted, together with the b quark, in 1973 by Kobayashi and Maskawa, to describe the CP violation in the weak interaction [25]. It was discovered in 1995 by the CDF and D0 experiments at the Tevatron collider at Fermilab, through the observation of the production of pairs of top quark and antiquark in proton-antiproton collisions [44, 45].

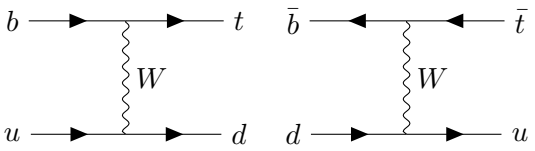


Figure 1.7: Feynman diagrams representing the t-channel single top quark (anti-quark) production, on the left (right).

The measured mass of the top quark is 172.57 ± 0.29 GeV, which makes it the heaviest elementary particle in the SM. Its total decay width is $\Gamma_t = 1.42 + 0.19 - 0.15$ GeV [14], which is one order of magnitude larger than the QCD scale, which is the hadronization scale. Therefore, the top quark decays before being confined into hadrons, a special case in the SM, which allows for the study of a bare quark.

The top quark decays weakly, and almost exclusively into a W boson and a b quark. The decay of the W boson can proceed into a pair of quark and antiquark, in 67% of the cases, or into a charged lepton and a neutrino, for the remaining 33%.

At the LHC, the main production mechanism of top quarks is the $pp \to t\bar{t}$ process, of which an example Feynman diagrams is shown in Figure 1.5. The production cross section measured by the CMS experiment at $\sqrt{s} = 13$ TeV is $\sigma_{t\bar{t}} = 791 \pm 25$ pb [46]. The top quark can also be produced in weak interactions, mainly through t-channel single top quark production, for which a Feynman diagram is shown in Figure 1.7, using the 5FS. The cross section of t-channel single top quark production at $\sqrt{s} = 13$ TeV was measured by the CMS collaboration to be $\sigma_{tq} = 130 \pm 1$ (stat) $\pm$ 19 (syst) pb for the quark, and $\sigma_{\bar{t}q} = 77 \pm 1$ (stat) $\pm$ 12 (syst) pb for the antiquark [47]. The difference in the cross section between the two processes is due to the different PDFs for the up and down-quark, which are present in the initial state of the two processes, as shown in Figure 1.7.

The summary of the production cross sections measured by the CMS experiment for all processes involving the top quark are shown in Figure 1.8.

This thesis work focuses on the production of a single top quark in association with a W boson and an additional photon or Z boson. The details of the production of these processes are described in Chapter 2.

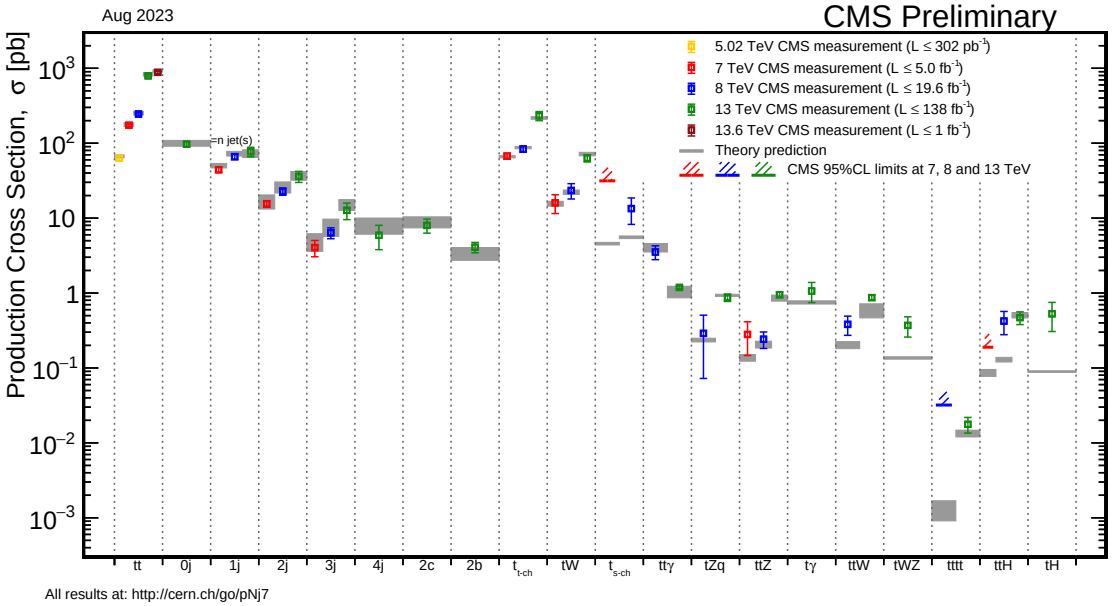


Figure 1.8: Summary of the production cross sections for all processes involving the top quark as measured by the CMS experiment as of Summer 2023, from Ref. [48].

2 Modeling of the tWZ and $tW\gamma$ processes

One of the many points in common between the tWZ and $tW\gamma$ processes is the difficulty in their modeling in event generators caused by the overlap with the $t\bar{t}Z$ and $t\bar{t}\gamma$ processes, respectively. This chapter will describe the methods used to model the two processes and their similarities and differences. Hereafter, the tWZ and $tW\gamma$ processes are described as tWX when considered together, where X represents either a photon or a Z boson.

The tWX production at LHC is well defined at LO in QCD, using the 5FS. A Feynman diagram of the partonic process $gb \to tWX$ is shown in Figure 2.1.

The situation becomes more complicated when the tWX process production is considered at NLO in QCD. In this case, one possible partonic process contributing is $gg\ (b\bar{b}) \to tW^-X\bar{b}$. The same initial and final states, with the same order in the perturbative expansion in the coupling constants α_S and α_W, can also be the result of the QCD LO process $gg\ (b\bar{b}) \to t\bar{t}X$, with the top antiquark decaying into a W^- boson and a bottom quark. The diagrams related to the LO process are defined as resonant diagrams. Illustrative diagrams of the two different processes are shown in Figure 2.2.

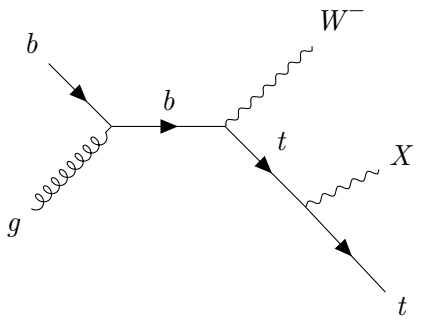


Figure 2.1: Example of leading order Feynman diagram for the tWX process.

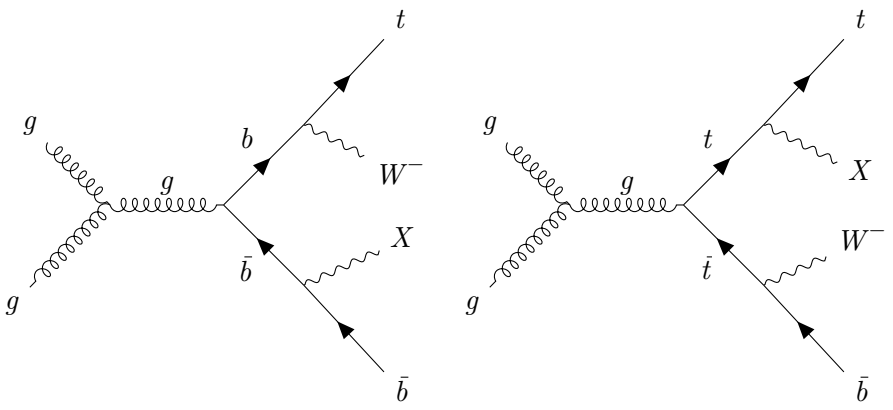


Figure 2.2: Example of Feynman diagrams with the same gg initial state and $tW^-X\bar{b}$ final state. Production of the tWX process at NLO in QCD, with the real emission of a bottom antiquark (left). Production of the $t\bar{t}X$ process at LO in QCD, with the decay of the top antiquark into a W^- boson and a bottom quark (right).

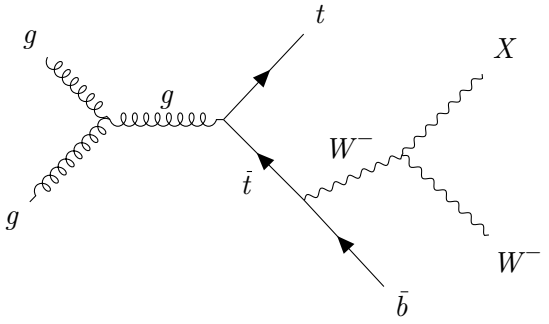


Figure 2.3: Example of leading order Feynman diagram for the $t\bar{t}$ process where the top antiquark decays into a b antiquark, a W^- and an X boson.

Furthermore, diagrams related to the $t\bar{t}X$ process are not the only ones appearing in the perturbative expansion of tWX. Figure 2.3 shows that considering the same powers of the coupling constants α_S and α_W, we also obtain diagrams that would naturally be regarded as pertaining to the $t\bar{t}$ process, where the top antiquark decays into three bodies: the bottom antiquark, the W^- and the X bosons. Contrary to $t\bar{t}X$, these diagrams are treated differently for tWZ and $tW\gamma$.

When the top quark is close to being on-shell, the perturbative expansion of the LO $t\bar{t}X$ ($t\bar{t}$) process is enhanced by a factor proportional to $1/\alpha_S$, coming from the top quark propagator. Because of this, the perturbative expansion of the tWX process itself becomes dominated by the resonant $t\bar{t}X$ ($t\bar{t}$) process, effectively spoiling it by obscuring the contributions from diagrams that genuinely originate from the tWX process.

There is one more problem with the top quark being treated in two different ways in these doubly-resonating diagrams: it appears simultaneously as both a stable asymptotic state (the final-state top quark) and an intermediate state (the decaying top quark). A final state top quark requires its width to be treated as zero in order to cancel infrared divergencies. In contrast, in the case of a decaying top quark, the finite decay width must be used so

as to regularize the propagator; including both treatments at the same time violates the gauge invariance [49]. A more complete review of this issue can be found in [50]. In the next section, we will describe the algorithms used for the overlap removal, often used in literature to face these issues, and how we have applied them to study the modeling of tWZ and $tW\gamma$.

2.1 Overlap Removal

The problem of overlapping diagrams described in the previous section is not exclusive to the tWZ and $tW\gamma$ processes. A similar issue exists for $t\bar{t}$ diagrams overlapping with the tW production or for $t\bar{t}H$ production overlapping with tWH. Moreover, overlapping resonant diagrams are fairly frequent in BSM models that include new particles. To address this issue, Refs. [50] and [51] describe in depth the concepts behind the *Simplified Treatments of Resonances* (STR).

The starting point is considering a process that is well-defined at LO in some interaction:

$$\alpha\beta \to \delta\varphi. \tag{2.1}$$

At NLO, the real emission diagrams will bring about a final state:

$$\alpha\beta \to \delta\varphi\vartheta. \tag{2.2}$$

If there exists a resonance Δ such that the decay

$$\Delta \to \delta\vartheta \tag{2.3}$$

is possible, and

$$\alpha\beta \to \Delta\varphi \tag{2.4}$$

is a possible LO process, then it is also possible to have the same final state as in (2.2) through a double-stage process:

$$\alpha\beta \to \Delta\varphi \to \delta\varphi\vartheta. \tag{2.5}$$

The diagrams relative to this final stage are related to the LO production of $\Delta\varphi$, not the $\delta\varphi$ final state we are interested in. Furthermore, if the production cross section of $\Delta\varphi$ is considerably larger than the one of $\delta\varphi$, it becomes impossible to have a definite modeling for the process under study, which gets extremely watered down by the more prominent background.

In order to have a proper definition of the $\delta\varphi$ final state, the STR schemes have been devised. Let the amplitude corresponding to (2.2) be decomposed into:

$$\mathcal{A}_{\alpha\beta\to\delta\varphi\vartheta} = \mathcal{A}^{\not\Delta} + \mathcal{A}^{\Delta}, \tag{2.6}$$

with $\mathcal{A}^{\Delta}$ being the amplitude corresponding to the process in (2.5), and $\mathcal{A}^{\not\Delta}$ corresponding to the remaining ones. The matrix element is, therefore, proportional to:

$$|\mathcal{A}_{\alpha\beta\to\delta\varphi\vartheta}|^2 = |\mathcal{A}^{\not\Delta}|^2 + 2\mathcal{R}(\mathcal{A}^{\not\Delta\dagger}\mathcal{A}^{\Delta}) + |\mathcal{A}^{\Delta}|^2. \tag{2.7}$$

Clearly, in order to describe the $\delta\varphi$ process, it is desirable to minimize the effect of the second and third terms on the right-hand side of (2.7) as much as possible so that the contributions from the resonant diagrams become negligible compared to the process under study. To do so, Diagram Removal (DR) and Diagram Subtraction (DS) schemes are implemented. The DR scheme straightforwardly removes the terms in (2.7) that contain $\mathcal{A}^{\Delta}$, leaving only the term $|\mathcal{A}^{\not\Delta}|^2$.

While easy to implement and understand, the DR scheme introduces an undesirable dependence on the gauge choice because of the indiscriminate removal of Feynman diagrams from the calculation. Nevertheless, Ref. [51] studied the STR in the context of the tW production at the LHC and found that, for all the simulated observables, every choice of gauge led to the same results within the statistical uncertainties, effectively making Diagram Removal usable for the modeling of the tW process.

On the other hand, DS is devised in such a way as to conserve gauge invariance. The scheme's endpoint is that the removal of resonant diagrams should be local, i.e., only in the part of phase space where the δ particle is on-shell. This is obtained by defining the DS infinitesimal cross section as:

$$d\sigma_{\alpha\beta\to\delta\varphi\vartheta}{}^{DS} = d\sigma_{\alpha\beta\to\delta\varphi\vartheta} - d\sigma_{\alpha\beta\to\delta\varphi\vartheta}{}^{\text{subt}}, \tag{2.8}$$

with $d\sigma_{\alpha\beta\to\delta\varphi\vartheta}$ being calculated from the Eq. (2.7), and $d\sigma_{\alpha\beta\to\delta\varphi\vartheta}{}^{\text{subt}}$ is defined so that:

$$d\sigma_{\alpha\beta\to\delta\varphi\vartheta}{}^{\text{subt}} \propto |\mathcal{A}_{\alpha\beta\to\delta\varphi\vartheta}|^2 d\phi - f(m_{\delta\vartheta})\mathbb{P}|\mathcal{A}^{\Delta}|^2 d\phi, \tag{2.9}$$

where $d\phi$ is the differential final state three-body phase space, f is an arbitrary function which respects the limit

$$\lim_{m_{\delta\vartheta}\to m_{\Delta}} f(m_{\delta\vartheta}) = 1, \tag{2.10}$$

and $\mathbb{P}$ is a kinematic projector operator that maps any configuration of $\delta\varphi\vartheta$ to one that has $m_{\delta\vartheta} = m_{\Delta}$.

The $\mathbb{P}$ operator effectively works by reshuffling the momenta of the particles involved in the process to produce an on-shell Δ particle, while conserving the spin correlations in the process. There are many possible choices for both f and $\mathbb{P}$. Ref. [50] introduced `MadSTR`, a `MadGraph_aMC@NLO` [52] plugin that automatically implements STR, providing two possible reshuffling strategies for projecting to the resonant configuration: either by varying the momenta of the recoil system or the center of mass energy.

Regarding f, the choice typically used is to define it as:

$$f(m_{\delta\vartheta}) = \frac{BW_{\Delta}(m_{\delta\vartheta}, x)}{BW_{\Delta}(m_{\Delta}, x)}, \tag{2.11}$$

where BW_{Δ} is the generalized Breit-Wigner function

$$BW_{\Delta}(m, x) = \frac{1}{\left(m^2 - m_{\Delta}^2\right)^2 + x^2\Gamma_{\Delta}^2}, \tag{2.12}$$

and x can be chosen in `MadSTR` to be either m_{Δ} or $m_{\delta\vartheta}$, respecting Eq. (2.10) in either case.

These conditions on the DS subtraction term are enough to guarantee that the DS scheme conserves the gauge invariance while still minimizing the contribution from the resonant diagrams as much as possible.

Despite the possibility of implementing DS in most cases with `MadGraph_aMC@NLO` and `MadSTR`, this is not the case for processes where one resonance can decay into three particles. This is the case for the tWX process with the diagrams shown in Figure 2.3, where the reshuffling of the momenta is not easily performed. The literature does not currently provide examples of DS being applied in this kind of case.

However, it is important to note that the difference between the results obtained in the DR and DS scheme is a proxy for the impact of the interference between the resonant and non-resonant diagrams:

$$d\sigma_{\alpha\beta\to\delta\varphi\vartheta}{}^{DS} - d\sigma_{\alpha\beta\to\delta\varphi\vartheta}{}^{DR} \propto \mathcal{R}(\mathcal{A}^{\not{\Delta}\dagger}\mathcal{A}^{\Delta}). \tag{2.13}$$

Therefore, Ref. [51] suggests that, in order for the production of $\delta\varphi$ to be a proper physical process, the difference between the DR and DS results should be smaller than the theoretical uncertainties on the calculation themselves. Otherwise, the process cannot be well-defined and should only be considered as a subset of the $\alpha\beta \to \Delta\varphi$ process.

Subsequent studies of the tWH and tWZ processes [53, 54], followed the path of looking for selection criteria that would define a phase space region where the interference term is minimized, and was therefore pure in the single top quark process in study.

Experimentally, the most-used strategy in analyses of the tW process in the CMS collaboration is to consider the DR prediction as the nominal one and the DS prediction as a systematic uncertainty on the modeling, as done in Refs. [55, 56].

Finally, the one last STR scheme to be mentioned is the DR with interference (DRI or DR2), which consists in only removing the rightmost term in Eq. (2.7), i.e., the squared module of the resonant diagrams, while keeping the interference term. Like DR, DRI is not gauge-invariant. However, it is conceptually more similar to DS, as it keeps the interference term but has a less involved implementation.

For the tWZ and $tW\gamma$ analyses presented in Chapters 5 and 6, DR has been used as nominal prediction, and DR2 has been considered for setting an uncertainty on the modeling. Hereafter, the standard DR procedure, without interference, is referred to as DR2.

2.2 *tWZ* modeling

This section presents the theoretical predictions regarding the production of the tWZ process in proton-proton collisions at $\sqrt(s) = 13$ TeV. The first complete study of the process, both at Fixed Order (FO) and NLOPS, was presented in Ref. [54]. While the scope of the paper is to provide a complete description of the process in the context of the Effective Field Theory, this section focuses on the modeling of the process in the SM, with comparisons between DR1 and DR2, and the selection criteria that have been devised in order to define a tWZ-pure phase space. FO refers to the simulation at the matrix element level, i.e., before the decay of resonances, the parton shower, and hadronization. NLOPS, on the other hand, includes the parton shower and hadronization.

2.2.1 Fixed Order predictions

Events are simulated in the final state where the Z boson decays into two leptons using `MadGraph_aMC@NLO`. The PDF set used is the NNPDF3.1, interfaced to `Madgraph` using LHAPDF6 [57]. The DR1 and DR2 methods are applied on the resonant diagrams attributable to both the $t\bar{t}Z$ and $t\bar{t}$ processes.

As discussed in the introduction to this chapter, the overlap between the tWZ and $t\bar{t}Z$ process descends from the fact that at NLO in QCD, the tWZ final state can be accompanied by the real emission of a b quark. These quarks are typically produced with low energy, in the order of a few GeV, and in a direction close to the beam axis, as opposed to the b quarks in the $t\bar{t}Z$ process, which come from the decay of the top quark. In the rest frame of the decaying top quark, the b quark is produced with an energy of ~ 67 GeV and, therefore, a higher p_{T} that makes it distinguishable from real-emission quarks.

The authors of Ref. [53] introduced a threshold on the characteristics of hard b quarks in order to define a phase space region where the interference term is minimized for the tWH process. Similar requirements, named *b-veto*, are used in the tWZ study [54], where events containing a b quark with p_{T} higher than 30 GeV and $|\eta| < 2.5$ (hard b quark) are vetoed.

Figure 2.4 shows how the four different scenarios, i.e., DR1 and DR2, with and without *b-veto*, describe the distribution of the invariant mass of the two leptons and the associated W boson in the simulated tWZ process. The *b-veto* is clearly effective in reducing the interference term contribution, as the two predictions are much closer to each other when it is applied.

In addition, it is worth mentioning that when the *b-veto* is not applied, the DR2 and DR1 prediction drift away from each other rapidly at high invariant masses, showing that the interference between the resonant and non-resonant amplitudes becomes more relevant.

The cause of such behavior might be traced back to the fact that in this high-energy region, the top quarks in the $t\bar{t}$ and $t\bar{t}Z$ diagrams are not likely to be produced on-shell anymore, and the removed diagrams lose their special status of doubly resonant.

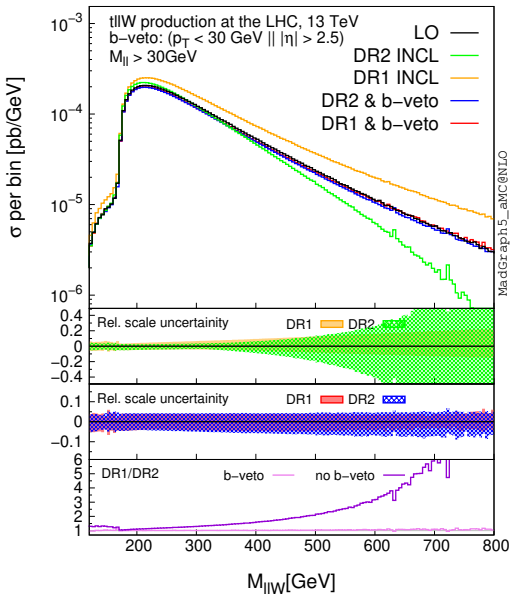


Figure 2.4: Invariant mass of the two leptons and the W boson in the tWZ process, comparing the DR1 and DR2 schemes, with and without the *b-veto* selection [54]. The first and second insets show the relative scale uncertainties of DR1 and DR2 without *b-veto* (upper), and with the *b-veto* (center). The lower inset shows the ratio between DR1 and DR2 with and without *b-veto*. The *b-veto* cut is effective in reducing the interference term contribution and defines a phase space region which is purer in the tWZ process.

Another FO study from Ref. [54], relevant for this thesis, is the comparison of the invariant mass of the two leptons, considering the Z boson in tWZ on-shell and off-shell. The results are shown in Figure 2.5, where the two distributions are compared for the DR1 and DR2 schemes, with the *b-veto* applied in both cases. The plot clearly shows that the two distributions are very similar in the region around the mass of the Z boson, which is the region of interest for the experimental tWZ analysis. Also, it allows us to only consider the modeling of the process with an on-shell Z boson in the NLOPS studies presented in subsection 2.2.2.

2.2.2 NLOPS predictions

The signal events used in the experimental tWZ analysis have been simulated using `MadGraph_aMC@NLO` interfaced with `PYTHIA8` to simulate the parton shower and hadronization, and using `MadSTR` to implement the Diagram Removal schemes. The Z boson is forced to be on-shell. Given the high mass of the final state, an on-shell top quark cannot decay into a b quark, a W boson, and a Z boson. Therefore, the diagrams with the $1 \rightarrow 3$ decay (Figure 2.3) of the top quark cannot feature an on-shell top and need not be removed. As a result, the DR1 and DR2 schemes are only applied to remove the overlap with the

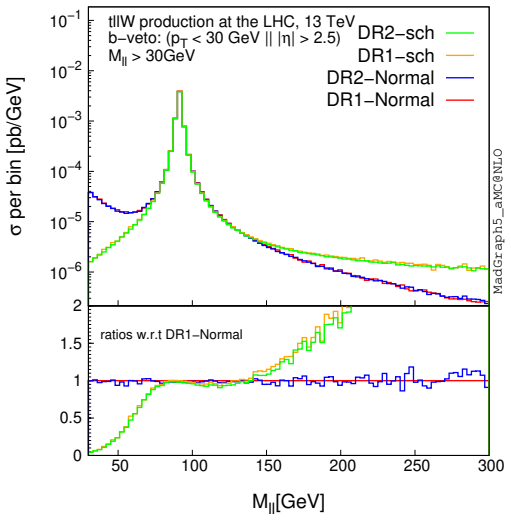


Figure 2.5: Invariant mass of the two leptons coming from the Z boson in the tWZ process, comparing the DR1 and DR2 schemes, with (s-ch) and without (normal) forcing the Z boson to be on-shell before the decay [54]. The lower plot shows the ratio between the 3 other cases and DR1 with off-shell Z boson. It displays how all four distributions are very similar in the region around the mass of the Z boson.

$t\bar{t}Z$ diagrams, analogously to how it was done for the phenomenological tWH study in Ref. [53].

At NLOPS, the definition of the *b-veto* needs to be updated with respect to FO in such a way as to take into consideration the b quark coming from the decay of the top quark. The NLOPS *b-veto* is defined such as to keep events which include at most one hard b quark, defined as having $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$. Additionally, the associated Z and W bosons are required to be central, $|\eta| < 2.4$.

The results of these simulation studies and selection criteria are showcased in Figure 2.6 for the transverse momentum of the top quark (left) and the W boson (right). For both distributions, it is evident how the initial disagreement between DR1 and DR2 is successfully fixed through the application of the *b-veto*, minimizing the contribution from the interference with $t\bar{t}Z$.

The difference between the DR1 and DR2 for the p_{T} of the top quark before the *b-veto* seems to have narrowed down to the different normalizations of the two predictions, $\sigma_{DR1} = 129$ fb and $\sigma_{DR2} = 120$ fb. On the other hand, the interference term presents a quite manifest shape effect on the p_{T} of the W boson, causing the two predictions to have increasing disagreement with the increasing p_{T} of the boson. This is also successfully recovered by the *b-veto*, even if not as well as for the p_{T} of the top quark.

2.3 $tW\gamma$ modeling

Original studies on the $tW\gamma$ process were conducted in the context of this thesis work and, similarly to the tWZ results presented in the previous sections, include predictions

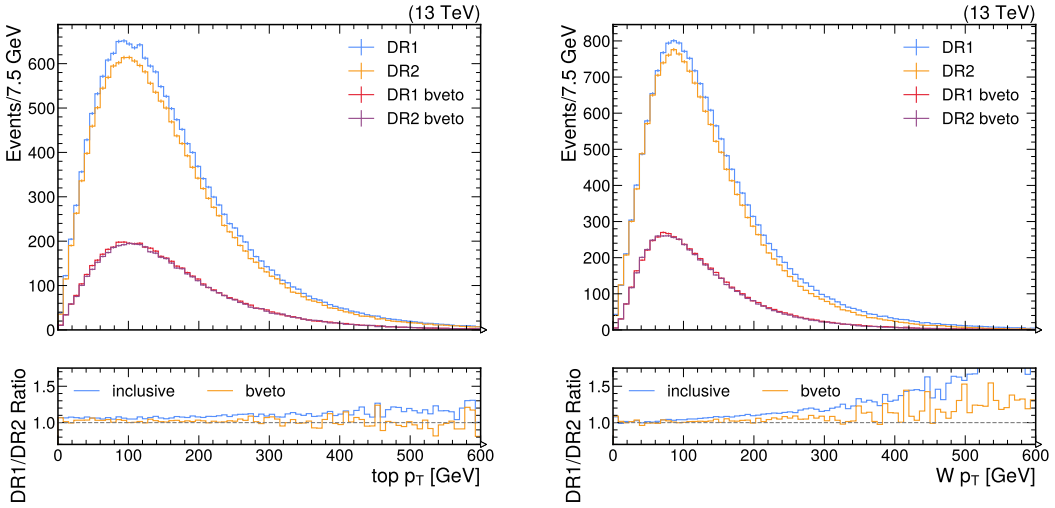


Figure 2.6: Transverse momentum of the top quark (left) and the W boson (right) in the tWZ process at NLOPS, comparing the DR1 and DR2 schemes, with and without the *b-veto* cut. The *b-veto* cut is effective in reducing the interference term contribution, but only in the bulk of the distribution for the W boson, where the disagreement displays a pronounced shape effect. The plots are normalized to an arbitrary integrated luminosity.

both at the FO and NLOPS levels. The specific differences between the $tW\gamma$ and tWZ processes mainly originate from the photon being massless, as opposed to the substantial mass of the Z boson. The most evident consequence is the possibility of the photon being radiated from the decay of either the top quark or the W boson, with Figure 2.7 showing one potential Feynman diagram for such an occurrence in the final state with two charged leptons.

Another point of difference between the two processes is that the overlap between $tW\gamma$ and $t\bar{t}$ must be accounted for when implementing the Diagram Removal schemes. Diagrams of the class shown in Figure 2.3 now allow for the top quark mediator to be on-shell for a wide range of energy of the photon and, therefore, give rise to the same kind of issues existing with $t\bar{t}\gamma$. Hence, the `MadSTR` plugin cannot be employed for the simulation of $tW\gamma$, as it is not able to handle the case where the resonance to be removed decays in a $1 \to 3$ body process. Luckily though, `MadGraph_aMC@NLO` automatically generates all the diagrams of the process and stores the relative information before performing the necessary calculation for event generation, thus allowing the implementation of DR1 and DR2 at the diagram level *by hand*.

Staying in the scope of the dileptonic final state of $tW\gamma$, one feasible way to model the process would be a complete simulation of $pp \to b\ell_1^+\nu_{\ell_1}\ell_2^-\bar{\nu}_{\ell_2}\gamma$, as done in Ref. [58] for the simultaneous simulation of both $tW\gamma$ and $t\bar{t}\gamma$ at LO in QCD. It is not yet possible to simulate the same final state at NLO to achieve a more accurate description of the process, both in terms of the total cross section and of the shape of various observables.

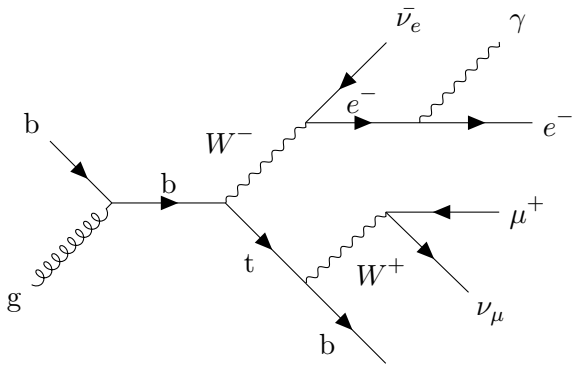


Figure 2.7: Example of possible Feynman diagram for the $tW\gamma$ process where the photon is radiated by any decay product of either the top quark or the W boson. Due to the masslessness of the photon, this class of diagrams needs to be considered in the modeling of the $tW\gamma$ process, differently from what is done for the tWZ process.

One remarkable example of the importance of the NLO accuracy for this kind of processes is the search for the SM tWZ process presented in Chapter 5, where the inclusion of NLO modeling for the signal process granted not only the possibility of a better discrimination from the main $t\bar{t}Z$ background but also a novel understanding of $t\bar{t}Z$ itself, which was also had already been thoroughly studied.

A different approach, which is the one used in the remainder of this section and in Chapter 6, is to separately simulate cases when the photon is emitted by any lines of the tW process (from now on referred to as *photon from hard process*) as illustrated in Figures 2.1 and 2.2, and cases in which the photon is radiated by any of the decay products of the top quark and the W boson (*photon from decay*), demonstrated in Figure 2.7. The hard process is simulated in `MadGraph_aMC@NLO` by requiring $pp \to t\,W^-\gamma$ (and $\bar{t}\,W^+\gamma$) and the DR schemes are implemented *by hand* before the event generation; the top quark and W boson are then decayed into leptons through the use of `MadSpin` [59], before interfacing the results to `PYTHIA8` for parton showering and hadronization.

The events with photons from decay are also generated in `MadGraph`, as $pp \to t\,W^-$ (and $\bar{t}\,W^+$). At the ME level, the only overlap is between tW and $t\bar{t}$. Therefore, we can use `MadSTR` to apply DR1 and DR2. After simulating the decays into leptons with `MadSpin`, which does not support three-body decays, the events are interfaced to `PYTHIA8`. Finally, only the events with photons coming from the decay products of the top quark and the W boson are kept. The emission of the photons is simulated through the Parton Showering (PS) algorithm. This choice allows us to retain the NLO accuracy of the simulation at production. A drawback of using PS to generate photons from decay is that it is not possible to have photons from decay at FO level. As a consequence FO studies can only be performed using photons from the hard process.

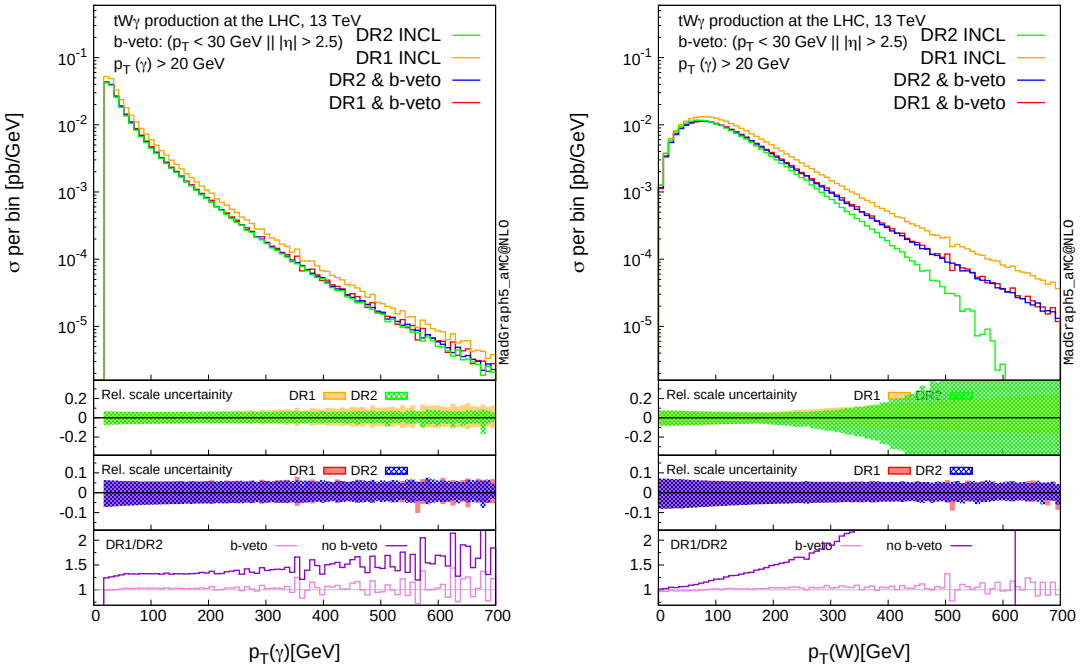


Figure 2.8: FO distribution of p_{T} for the photon (left) and W boson (right) in the $tW\gamma$ process, in the scenarios with DR1 and DR2, with and without the *b-veto*. The first two insets show the relative scale uncertainties of DR1 and DR2 without *b-veto* (upper), and with the *b-veto* (center). The lower inset shows the ratio between DR1 and DR2 with and without *b-veto*. In both cases, the *b-veto* cuts allow the definition of a phase space where the interference from $t\bar{t}$ and $t\bar{t}\gamma$ is minimized.

2.3.1 Fixed Order predictions

At Fixed Order, the $tW\gamma$ events are generated with requirements on the photon to have $p_{\mathrm{T}} > 20$ GeV and to respect the conditions of the Frixione isolation [60] with

$$\delta_0 = 0.4,\ \epsilon_\gamma = 1,\ n = 1. \tag{2.14}$$

After applying DR1 or DR2 by removing both $t\bar{t}$ and $t\bar{t}\gamma$ diagrams, it is found that the calculated cross sections are $\sigma_{DR1} = 258$ fb and $\sigma_{DR2} = 200$ fb. The effect of the *b-veto* is studied, following the same definition from subsection 2.2.1.

Figure 2.8 shows the distributions of the p_{T} of the photon and the W boson in the four scenarios: DR1 and DR2, with and without *b-veto*. The *b-veto* succeeds in defining a phase space where the contribution of the interference is minimized: the large difference between DR1 and DR2 predictions is cured after the application. Additionally, it is interesting to notice the differences between the behaviour of the two predictions for the two different observables. As for the p_{T} of the photon, interestingly, the *b-veto* seems to define the same phase space as DR2, and the discrepancy between DR1 and DR2 before the cut has a much less pronounced shape effect than it is for the W boson, where, furthermore, DR1 and DR2 both considerably drift away from their *b-veto* counterparts in the higher-p_{T} regions.

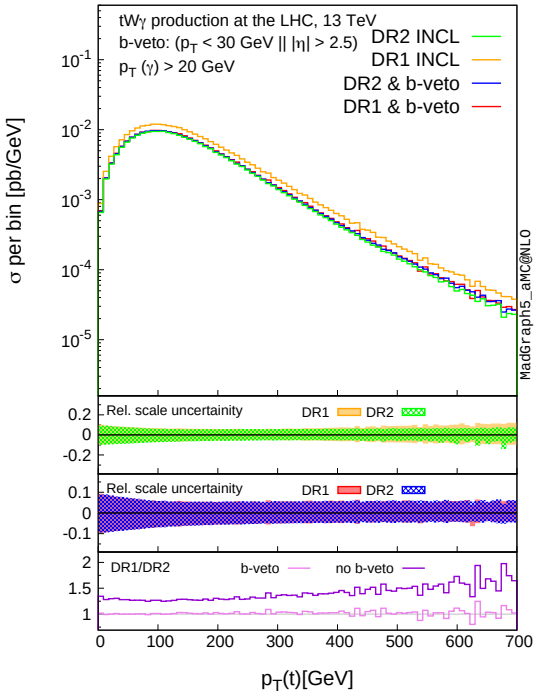


Figure 2.9: Distribution of the transverse momentum of the top quark in the $tW\gamma$ process at FO, simulated with DR1 and DR2, with and without the application of the *b-veto*. The first two insets show the relative scale uncertainties of DR1 and DR2 without *b-veto* (upper), and with the *b-veto* (center). The lower inset shows the ratio between DR1 and DR2 with and without *b-veto*. The difference between DR1 and DR2 is fundamentally just due to a different normalization, and it is successfully removed by the application of the *b-veto*.

As for the p_{T} of the top quark, shown in Figure 2.9, the difference between DR1 and DR2 is also mainly due to the different normalization. A similar feature was observed in the tWZ simulation in Figure 2.6. Similarly to the photon p_{T}, the *b-veto* makes the DR1 and DR2 distributions agree with each other, while the shapes show almost no change before and after the *b-veto*.

2.3.2 NLOPS predictions

The $tW\gamma$ events at NLOPS level have been generated in the final state with two leptons with the recipe described earlier in the current section. Further requirements are the invariant mass $m_{\ell\ell}$ of the two final-state leptons being $m_{\ell\ell} > 30$ GeV, the photon having $p_{\mathrm{T}} > 15$ GeV and $|\eta| < 1.5$, and the Frixione isolation same as for 2.3.1.

The resulting cross sections for this phase space are $\sigma_{DR1} = 23.0$ fb and $\sigma_{DR2} = 21.8$ fb. The NLO predictions are also compared to the ones obtained at LO in QCD, simulated with `MadGraph_aMC@NLO` requiring $pp \to b\ell_1^+\nu_{\ell_1}\ell_2^-\bar{\nu}_{\ell_2}\gamma$, which results in a cross section $\sigma_{LO} = 10.6$ fb.

Figure 2.10 shows the distribution of the p_{T} of the top quark, the W boson, and the photon, normalized to an arbitrary integrated luminosity. The plots display how, in this phase

space, DR1 and DR2 are in very good agreement for all of the observables shown, without the need for any additional selection criteria, as shown in the first inset.

The same figure also shows the comparison between the shapes of DR1 and LO, in the second inset.

It can be seen that $tW\gamma$ at NLO, summing the contributions of photons from the hard process and from decay, is compatible with LO, which includes both photon origins, when the inclusive cross section is factored out. It is also shown that DR1 and DR2 differ only in the overall cross section estimation.

In conclusion, the NLO modeling of the $tW\gamma$ process is found to be solid and it is chosen for this thesis work.

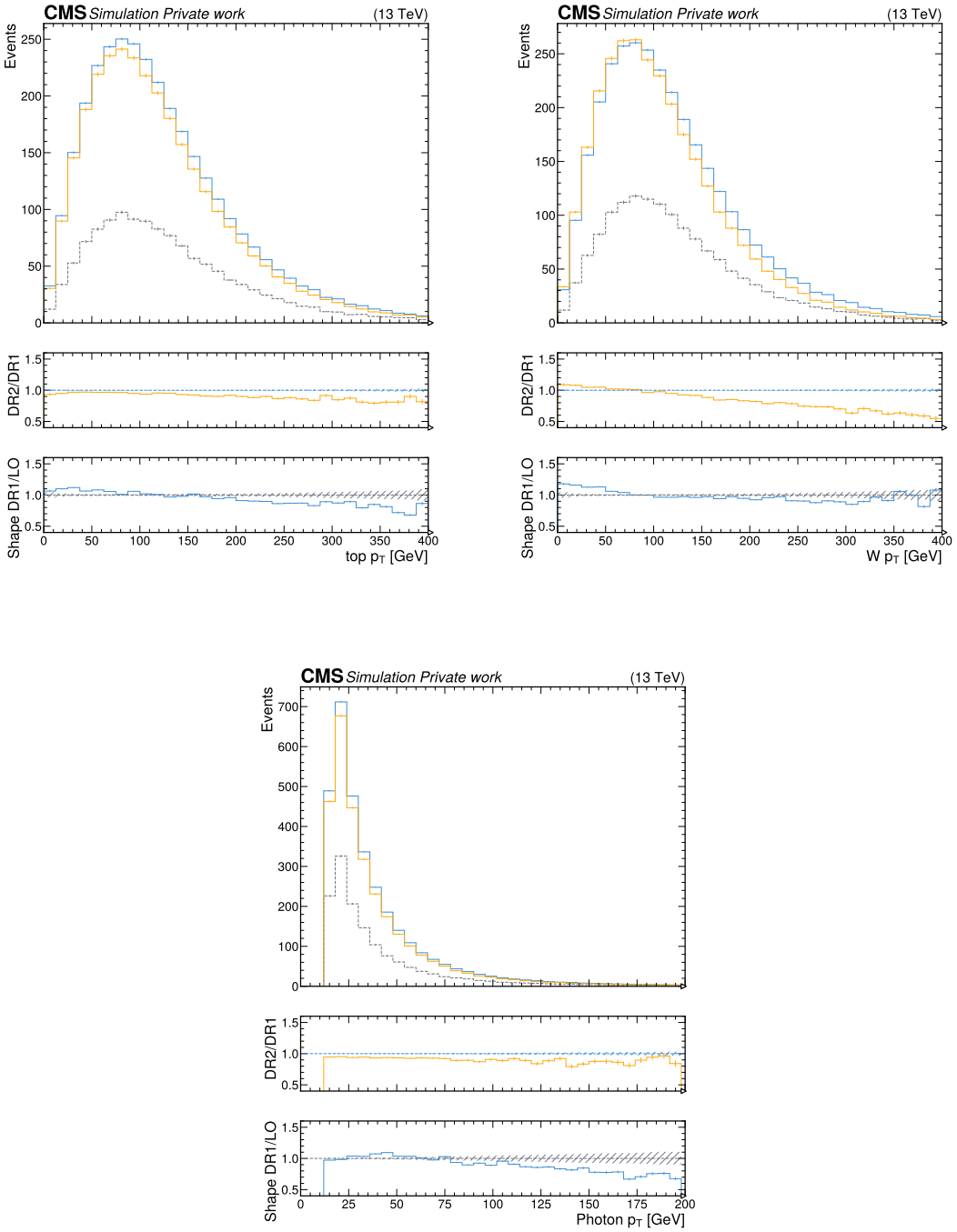


Figure 2.10: NLOPS and LOPS distribution of the p_T of the top quark (left), W boson (right), and photon (lower) for the $tW\gamma$ process, normalized to arbitrary luminosity. For all observables, the central inset shows the comparison between DR1 and DR2. The lower inset displays the ratio between DR1 and LO, with both normalized to unity.

3 Statistical methods and multivariate techniques

The evolution of particle physics has always gone hand in hand with the evolution of the machinery used to study it. New accelerators and new detectors allowed new discoveries, which in turn showed the direction for the development of new machinery. As technology advanced and ushered us into the LHC era of high-energy physics, innovations did not stop at the hardware only. The degree of complexity of the proton-proton collision events at LHC forced particle physicists to develop robust statistical methods and tools in order to derive sound results from the vast amount of collected data.

Some statistical methods described in this chapter include profile likelihood ratio [61] and the CL_s method for defining upper limits in searches, introduced at LEP [62] and used at Tevaton as well [63]. In this thesis work, the statistical methods are implemented using the CMS Combine package [64], which has been devised to discover the Higgs boson in 2012 [3] and is still further developed, including in the analyses presented in this thesis.

A different and complementary approach that has been undertaken is the employment of machine learning methods as means to better reconstruct physics objects and also classify whole collision events, among other tasks.

This chapter will describe the statistical and machine learning techniques used in the searches for the $tW\gamma$ and tWZ processes.

3.1 The statistical model

The statistical model used in both the tWZ and $tW\gamma$ analyses is implemented in Combine, and it is based on the construction of a likelihood function from simulated and observed data.

The events relevant to the analysis are selected among the data collected by the CMS experiment through a combination of requirements on the physics objects and then summarized in histograms, which show the distribution of the chosen observables, e.g., the output

of a multivariate classifier. The same workflow is applied to the simulated events, divided by physics process, and the resulting histograms are used as templates to be fitted to the data.

From the simulated data, it is possible to write the expected number of events in bin i of a histogram as a sum over the P processes:

$$\lambda_i(\vec{\mu}, \vec{\nu}) = \sum_p^P M_p(\vec{\mu}, \vec{\nu}) \omega_{ip}(\vec{\nu}) + E_i(\vec{\mu}, \vec{\nu}), \tag{3.1}$$

where $\vec{\mu}$ are the parameters of interest of the fit, $\vec{\nu}$ are the nuisance parameters, $\omega_{ip}(\vec{\nu})$ is the expected number of events from process p in bin i, $M_p(\vec{\mu}, \vec{\nu})$ is an overall multiplicative factor to the total rate of the process that considers the effects of the parameters of the model, and E_i is a term that takes into account the finite number of simulated events according to the Barlow-Beeston method [65].

In the physics models used in this thesis, the parameter of interest is the signal strength μ, defined for any signal process q as

$$\mu_q = \frac{\sigma_q^{\text{obs}}}{\sigma_q^{\text{exp}}}, \tag{3.2}$$

with σ_q^{exp} being the cross section expected from the theory model, and σ_q^{obs} the measured one. The $M_p(\vec{\mu}, \vec{\nu})$ can therefore be factorized as

$$M_p(\vec{\mu}, \vec{\nu}) = \begin{cases} \mu_p \bar{M}_p(\vec{\nu}), & \text{if } p \text{ is a signal process;} \\ \bar{M}_p(\vec{\nu}), & \text{otherwise.} \end{cases} \tag{3.3}$$

The functions $\bar{M}_p(\vec{\nu})$ only depend on the effect of the nuisance parameters.

To model the systematic uncertainties affecting the shape of the histograms of every process, for each uncertainty two additional histograms are provided through the $\pm 1\sigma$ variation of the nuisance parameter ν. The $\omega_{ip}(\vec{\nu})$ are then produced by interpolating these histograms, normalized to the rate of the nominal histogram, in such a way as to be continuous with their first and second derivatives. The renormalization of the varied histograms is necessary so only to consider the effect that the nuisance parameters have on the shape of the distributions; possible differences in the total event rate are considered in the $M_p(\vec{\mu}, \vec{\nu})$ function through multiplicative factors, together with the effect of systematics which purely affect normalization, e.g., the uncertainty on the measurement of the luminosity, as shown in Ref. [64].

The λ_i functions are then used in the definition of the likelihood function, which consists of the product of a Poisson distribution for every bin count of each histogram and a normal distribution for each nuisance parameter.

$$\mathcal{L}(\vec{\mu}, \vec{\nu} | n_i) = \prod_i \text{Pois}(n_i; \vec{\mu}, \vec{\nu}) \prod_j^S \mathcal{N}(\nu_j; \nu_{0,j}, \sigma_j), \tag{3.4}$$

and therefore

$$\mathcal{L}(\vec{\mu},\vec{\nu}|n_i) \propto \prod_i \frac{\lambda_i^{n_i}(\vec{\mu},\vec{\nu})\mathrm{e}^{-\lambda_i(\vec{\mu},\vec{\nu})}}{n_i!} \prod_j^S \mathrm{e}^{-\frac{(\nu_j-\nu_{0,j})^2}{2\sigma_j^2}}, \tag{3.5}$$

with i going through all the bins of all the histograms and j looping through all the S systematic uncertainties in the model. The observed number of events in bin i is denoted by n_i, whereas $\nu_{0,j}$ and σ_j are the initial value and the prior uncertainty on the nuisance parameter j, respectively.

The best-fit values $\hat{\vec{\mu}}$ for the parameters of interest are then those maximizing $\mathcal{L}$. Equivalently, $\hat{\vec{\mu}}$ is usually obtained by minimizing the negative log likelihood (NLL), which numerically is a less complex problem.

3.1.1 Likelihood scan for uncertainties

The step immediately following the estimation of $\hat{\mu}$ is the estimation of its uncertainty. One prescription for this is given in Ref. [66] and is based on the construction of a profile likelihood ratio test statistic

$$\lambda = \frac{\mathcal{L}(\vec{\mu},\hat{\hat{\vec{\nu}}}(\vec{\mu}))}{\mathcal{L}(\hat{\vec{\mu}},\hat{\vec{\nu}})}, \tag{3.6}$$

where $\hat{\vec{\nu}}$ is the value of the nuisance parameters that maximizes $\mathcal{L}$, and $\hat{\hat{\vec{\nu}}}(\vec{\mu})$ is the value that maximizes $\mathcal{L}$ for a fixed value of $\vec{\mu}$.

According to Wilks' theorem [67], for a large enough sample, $-2\ln\lambda$ follows a χ^2 distribution with r degrees of freedom, with r being the number of parameters of interest. As a result, it is possible to calculate the (1-α)% confidence interval on the parameters of interest by performing a *likelihood scan*: scanning $-2\ln\lambda$ and finding the values of $\vec{\mu}$ for which the value of $-2\ln\lambda$ increases by the α percentile of χ^2 with r degrees of freedom.

Figure 3.1 shows the result of this method for $r = 1$. In this case, given the form of the test statistic with one degree of freedom,

$$-2\ln\lambda \sim \frac{(\mu-\hat{\mu})^2}{\sigma^2}, \tag{3.7}$$

it is sufficient to find the values of μ for which $-2\ln\lambda = n^2$ to obtain the $n\sigma$ confidence level for the parameter μ, with σ being the standard deviation of the distribution of $\hat{\mu}$.

The same procedure can be applied to obtain the uncertainties on the j-th nuisance parameter, by scanning the likelihood as

$$\lambda = \frac{\mathcal{L}(\hat{\nu_j},\hat{\hat{\vec{\theta}}}(\nu_j))}{\mathcal{L}(\hat{\nu_j},\hat{\vec{\theta}})}, \tag{3.8}$$

where $\vec{\theta}$ includes all the nuisance parameters except ν_j, and $\vec{\mu}$.

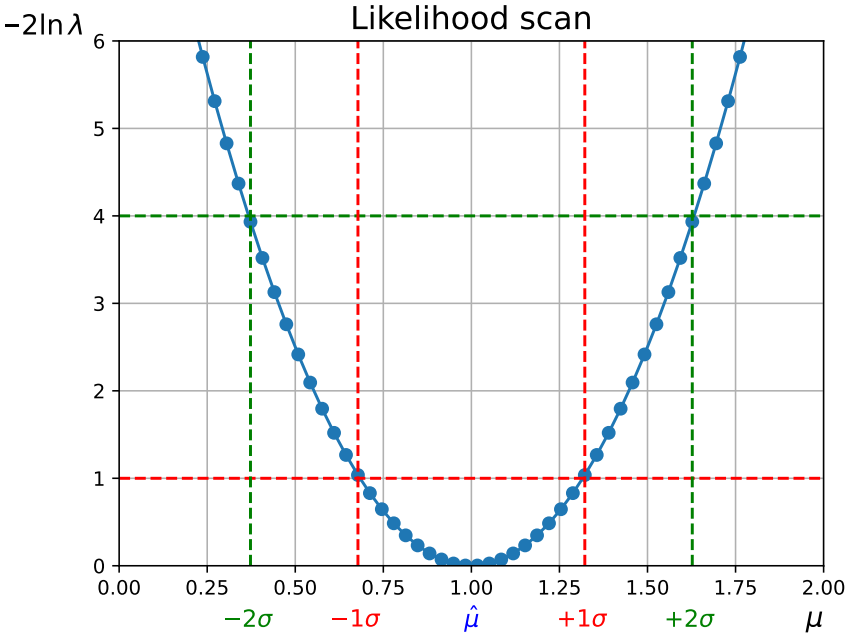


Figure 3.1: Example of a likelihood scan for one parameter of interest. It is possible to find the 1σ (2σ) confidence level for μ by finding the values of μ for which $-2\ln\lambda = 1(4)$.

3.1.2 Significance

This thesis's main topic is searches for rare, still unobserved, processes. Such analyses are said to provide observation (evidence) for the existence of the processes when the statistical significance of the search is 5σ (3σ).

The statistical significance Z in this case is the probability, given as the number of standard deviations of a Gaussian distribution (the higher the number, the lower the probability), that the fitted value of the signal strength $\hat{\mu}$, or a larger value, is obtained under the assumption that the real value is $\mu = 0$. Paraphrasing, the significance represents the probability of rejecting the hypothesis that the process has signal strength $\mu = 0$ (the process does not exist) if this hypothesis is actually true.

The test statistic used to calculate the significance q_0 is also based on the likelihood ratio λ defined in Eq. (3.6):

$$q_0 = \begin{cases} -2\ln\lambda(0), & \text{if} \hat{\mu} \geq 0; \\ 0, & \text{if} \hat{\mu} < 0. \end{cases} \tag{3.9}$$

Considering only the non-trivial case, for a large enough sample,

$$q_0 = \frac{\hat{\mu}^2}{\sigma^2}. \tag{3.10}$$

Ref [61] shows that, under these assumptions, the significance is

$$Z = \sqrt{q_0}. \tag{3.11}$$

3.1.3 Saturated goodness of fit test

Calculating the significance and checking it against a threshold is a way to disprove the null hypothesis ($\mu = 0$) in favor of a well-defined alternative hypothesis ($\mu > 0$). In some occasions, a well-defined alternative hypothesis is not available, for instance, when it is necessary to prove the compatibility between the statistical model and the observed data. These cases are usually resolved through goodness of fit (GoF) tests. They are used in Chapter 6 to study the compatibility between simulated and observed input features for a multivariate classifier.

The GoF test considered in this thesis employs the saturated model, as it considers a test statistic

$$t = -2\frac{\mathcal{L}(\vec{\mu}, \vec{\nu})}{\mathcal{L}_S}, \tag{3.12}$$

where $\mathcal{L}_S$ is the likelihood of the saturated model. The saturated model is defined so that it matches the observed data exactly: $\lambda_i = n_i$ for every i in Eq. 3.5.

In this thesis, the observed data is used to calculate the value t_d from Eq. (3.12), and the statistical model is considered compatible with data when the p-value of t_d is greater than 0.05, i.e., when the value of t_d is within the 95 percentile of the distribution. Unfortunately, the distribution of t is not generally known and is inferred from pseudo-data sets.

The pseudo-data sets are simulated through parametric bootstrap [68]: first $\mathcal{L}(\vec{\mu}, \vec{\nu}$ is fitted to data, in order to find the values $\hat{\vec{\nu}}$ that maximize it, and the respective posterior uncertainties $\hat{\vec{\sigma}}$. Then, the value $\bar{\nu}_j$ of the j-th nuisance parameter is generated according to the probability distribution

$$p(\nu_j; \hat{\nu}_j) = \exp\left(-\frac{(\nu_j - \hat{\nu}_j)^2}{2\hat{\sigma}_j^2}\right), \tag{3.13}$$

and the number of events $\bar{n}_i$ in bin i for the pseudo-data set is defined from Eq. 3.1:

$$\bar{n}_i = \lambda_i(\hat{\vec{\mu}}, \vec{\bar{\nu}}). \tag{3.14}$$

After fitting the statistical model to the pseudo-data sets, the values of t are collected for each instance to evaluate the distribution of the test statistic. Once the distribution is known, the p-value for the observation of t_d can be calculated.

Examples of this procedure and relative distributions are shown in Section 6.4.1.

3.1.4 Upper limits with the CL_S technique

When the accuracy of a search is not enough to have a meaningful measurement of a parameter of interest μ, it is still possible to calculate upper limits; at LHC it is usually done with the CL_S technique [62, 69].

The starting point is the definition of a test statistic:

$$q = \begin{cases} -2\ln\lambda, & \text{if } 0 \leq \hat{\mu} \leq \mu, \\ -2\ln\frac{\mathcal{L}(\mu,\hat{\hat{\vec{\nu}}}(\mu))}{\mathcal{L}(0,\hat{\hat{\vec{\nu}}}(0))}, & \text{if } \hat{\mu} < 0, \\ 0, & \text{if } \hat{\mu} > \mu. \end{cases} \tag{3.15}$$

Being μ a signal strength, once again, the discussion can be restricted to the case where $\mu \geq 0$. The third line in Eq. (3.15) ensures that upward variations of μ do not go against the signal hypothesis.

For a fixed value of μ, the value $q^{\text{obs}}(\mu)$ is calculated from data, together with the two p-values

$$p_\mu = \int_{q^{\text{obs}}(\mu)}^{\text{inf}} f(q(\mu)|\mu)dq, \tag{3.16}$$

$$p_b = \int_0^{q^{\text{obs}}(\mu)} f(q(\mu)|0)dq, \tag{3.17}$$

where the probability distributions $f(q(\mu)|\mu)$ and $f(q(\mu)|0)$ are determined through pseudo-data sets, which can be generated through parametric bootstrap, similarly as how described in Subsection 3.1.3, with the difference that $\hat{\mu}$ in Eq. 3.14 is replaced by either μ or 0. p_μ and $1 - p_b$ represent the compatibility of q^{obs} with the hypothesis that the signal strength has, indeed, value μ or it is 0, respectively. Hence, the CL_S criterion is calculated as:

$$\text{CL}_S = \frac{p_\mu}{1 - p_b}. \tag{3.18}$$

The upper limit at (100-α)% Confidence Level is then calculated by finding the value of μ for which $\text{CL}_S = \alpha$.

3.2 Machine Learning

Machine Learning (ML) is extensively used at different levels in particle physics and executes several key functions in the CMS experiment; in this thesis, ML is used for the classification of physics objects (boosted-top tagging) or events (signal discrimination). The remainder of this chapter is devoted to the description of the ML algorithms used to devise these solutions.

3.2.1 Deep Neural Networks

The basis of the ML algorithms that have been devised in this thesis is the Deep Neural Network (DNN), which consists of a stack of neuron layers that can be trained to achieve a specific task. In particular, in this subsection, the discussion is restricted to fully connected feed-forward DNNs, in which every neuron in a layer is connected to all the neurons in the following layer. The elementary unit is an artificial neuron similar to Rosenblatt's

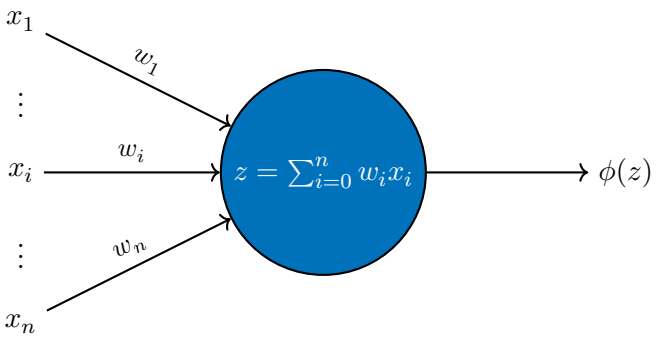


Figure 3.2: The neuron is a map from the input $\vec{x}$ to the output $\phi(\vec{w} \cdot \vec{x})$.

perceptron [70] and, in essence, it is a map from an input vector $\vec{x}$ to an output

$$y = \phi(\vec{w} \cdot \vec{x} - b), \tag{3.19}$$

where ϕ is an arbitrary non-linear function called the activation function, $\vec{w}$ is a vector of weights that are adjusted through the training of the neural network, and b is an additional bias. Usually $\vec{w}$ and $\vec{x}$ are redefined in such a way that they have an extra term which is $w_0 = -b$ and $x_0 = 1$, respectively, so that

$$y = \phi(\vec{w} \cdot \vec{x}). \tag{3.20}$$

A visual representation of this process is given by Figure 3.2.

Considering all neurons in a layer, Eq. (3.20) can be expanded to take the whole output of a layer as:

$$\vec{y} = \vec{\phi}(\mathbf{W} \cdot \vec{x}), \tag{3.21}$$

with $\mathbf{W}$ being the matrix whose rows are the weights of each neuron.

In the case of a DNN these layers are called hidden layers and are placed between two additional layers which take care of the input and output. The input layer injects the DNN's input features into the first layer of neurons, while the output layer returns the DNN's outputs, one for each neuron of the layer.

DNNs can also employ residual connections [71]: this is the case when the input of a layer is summed to its output. In this thesis, *pre-activation* residual connections [72] have been used: the result of the activation function of a layer is summed with the input of the layer itself. Therefore, the output of a layer would be:

$$\vec{y_{\text{res}}} = \vec{\phi}(\mathbf{W} \cdot \vec{x}) + \vec{x}. \tag{3.22}$$

If the input and output of the layer have different dimensions, it is possible to project the input through a learnable linear transformation before the summation:

$$\vec{y_{\text{res}}} = \vec{\phi}(\mathbf{W} \cdot \vec{x}) + \mathbf{W}_{\text{proj}} \cdot \vec{x}. \tag{3.23}$$

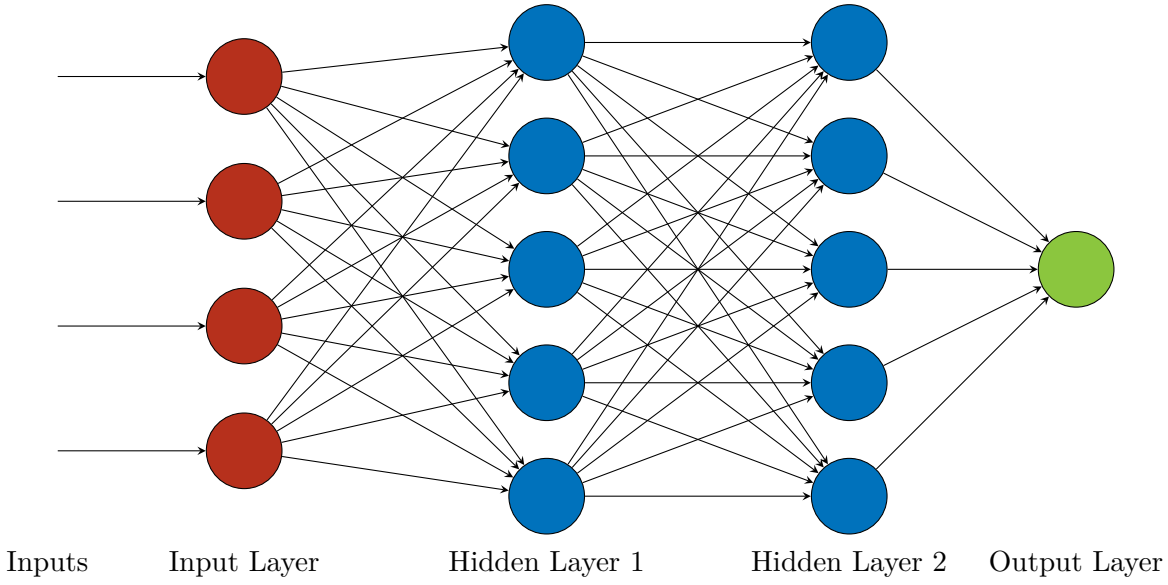


Figure 3.3: Example of a DNN with two hidden layers and one output node. The input features are distributed to the first hidden layer, whose output is the input to the following layer. After the hidden layers, the output layer finally returns the neural network outputs. In this case, this DNN returns one output, as there is one output node.

Several choices are possible regarding the activation function: a common possibility is the Rectified Linear Unit (ReLU) [73]

$$\mathrm{ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{3.24}$$

There are also several variants of the ReLU; in this thesis, the Leaky ReLU [74] and the Sigmoid Linear Unit (SiLU), also called *swish* [75] have been used:

$$\text{Leaky ReLU}(x) = \begin{cases} x, & \text{if } x \geq 0, \\ 0.01x, & \text{if } x < 0 \end{cases}, \quad \mathrm{SiLU}(x) = \frac{x}{1 + \mathrm{e}^{-x}}. \tag{3.25}$$

One more example of activation function is the Scaled Exponential Linear Unit (SELU), which is shown to have self-normalizing properties, i.e., the distribution of the activated outputs of a layer is a normal distribution with mean 0 and variance 1 [76]. It is defined as:

$$\mathrm{SELU}(x) = \lambda \begin{cases} x, & \text{if } x \geq 0, \\ \alpha \mathrm{e}^{x} - \alpha, & \text{if } x < 0 \end{cases}, \tag{3.26}$$

with $\lambda \approx 1.0507$ and $\alpha \approx 1.67326$.

Figure 3.3 shows the design of a DNN with two hidden layers and one output node.

Fully connected feed-forward DNNs are well-suited for problems where there is a fixed number of physics objects with different features considered as input variables, as is the case for the DNN shown in Section 5.4.

Training of a neural network

The training of a DNN is the process through which the network's weights are adjusted so that the DNN can carry out its designed task. In the case of a classification task, the training is performed by learning from a set of data points whose classification is already known, the training sample. In particle physics, the training samples are often derived by MC simulations.

Let the task be a binary classification: in this case, the network needs to learn how to discriminate between points in category $\bar{y} = 1$ (signal) and $\bar{y} = 0$ (background). The output y of the neural network is a composition of the activation functions $\vec{\phi}_i$ of the H hidden layers and the output layer in the form:

$$y = \phi_O(\mathbf{W}_O \cdot \vec{\phi}_H(\mathbf{W}_H \cdot ...\phi_1(\mathbf{W}_1 \ \cdot \vec{x})...)) = \vec{f}(\vec{x}; \mathbf{W}_1, ..., \mathbf{W}_H, \mathbf{W}_O). \tag{3.27}$$

In this task, a sigmoid function is often used for ϕ_O:

$$\phi_O(Z) = \frac{1}{1 + \mathrm{e}^{-z}}, \tag{3.28}$$

whose minimum and maximum are, respectively, 0 and 1, and is therefore suitable to represent the probability that the input data point belongs to the signal category. Hence, the training objective is to update the weights of the DNN so that the output y is as close as possible to the category $\bar{y}$ for each of the N data points x_i in the training sample. The usual way to proceed consists of minimizing a loss function L, which is a measure of how distant the network's predictions y_i are from the true categories $\bar{y}_i$. A common loss for a binary classifier is the binary cross-entropy:

$$L = \frac{1}{N} \sum_{i=1}^{N} \ln(|\bar{y}_i - y_i|). \tag{3.29}$$

The minimization of L as a function of the weights is performed through gradient descent, with different possible implementations. Intuitively, the gradient of a function represents the direction in which the function grows the quickest; therefore, moving against the gradient in the space of weights takes the function to lower values, i.e., where the differences between the predictions of the DNN $\bar{y}$ and the categories of the events y are closer to each other.

The gradient descent algorithm used in this thesis is Adam [77]: the data is divided into fractions, called batches, and the weights are adjusted one batch at a time. The peculiarity of the Adam algorithm is that the adjustments depend not only on the gradient calculated for the current batch but it is combined with information from the previous batches.

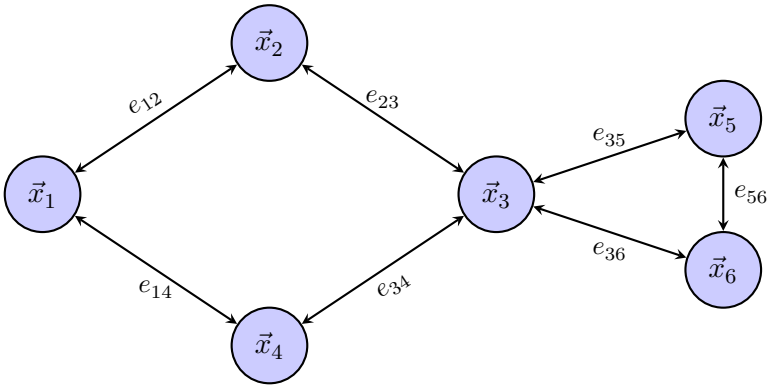


Figure 3.4: An example of a graph. The edges represent connections between the nodes. $\vec{x}_i$ and $\vec{x}_j$ represent the input features of nodes i and j, while $\vec{e}_{i,j}$ are the input features of the edge connecting them.

After the whole training sample has been used, new batches are composed again from the total sample, and the process is restarted. One iteration over the entire training sample is defined as *epoch*.

The gradient itself is calculated through backpropagation [78], which exploits the chain rule of derivatives applied to the activation functions of the layers in the DNN. The use of gradient descent is one of the reasons why the activation functions of the neurons are non-linear: the gradient of a linear function is constant and does not provide any information valuable to the training.

3.2.2 Graph Neural Networks

Graph Neural Networks (GNN) are a class of Neural Networks with the characteristic of accepting a mathematical graph as input [79]. A graph is a collection of nodes and edges connecting some of the nodes. Both can be associated with some input features, as shown in Figure 3.4. Typically, the nodes represent some object, and the edges represent the relationships between pairs of such objects. An example from particle physics is a graph consisting of reconstructed particles as nodes, with transverse momentum and pseudorapidity as features, and the edges representing features of the connected pair of nodes, such as the invariant mass or the angular separation. It is advantageous to use graphs and GNNs to represent the final state of collision events where the number of some components, for example, jets, is not fixed, but the features of all the observed jets are meaningful for the task performed, as it applies to the GNN designed in Section 6.4.

The main element of a GNN is a layer that takes the features of the nodes and, possibly, edges as inputs and returns new features for the nodes. Let all the nodes in the graph be of the same type, e.g., they have the same input features. Let also $\vec{x}_i$ be the input features of node i, $\vec{e}_{i,j}$ the input features of the edge connecting nodes i and j, $\mathcal{N}(i)$ the set of nodes connected to node i by an edge, and $\vec{o}_i$ the output features of node i. Then, one possible

choice for the output of the i-th node is

$$\vec{o}_i = \phi\left(\mathbf{W}_1 \cdot \vec{x}_i + \sum_{j \in \mathcal{N}(i)} (\mathbf{W}_2 \cdot \vec{x}_j + \mathbf{W}_3 \cdot \vec{e}_{i,j})\right), \tag{3.30}$$

where $\mathbf{W}_{1,2,3}$ are the trainable weights of the layer, and ϕ is the activation function.

Compared to the layers discussed in the previous section, a GNN treats each node and its vicinity as a separate instance of the layer, but the weights are shared between the instances. The layer's output is then a new graph with a new set of features for each node, which can be used as input for a subsequent layer.

Generally, three types of prediction tasks can be performed with GNNs: node-level, edge-level, and graph-level. Node-level predictions require an output for each node under study, on top of the graph layers. This output can either be a single output node or a DNN that takes the node output as input. Edge-level predictions demand the extra step of embedding a feature representation for the required edge. One way of performing the task is, for instance, concatenating the features of the two nodes connected by the edge in the last graph layer. The following steps are analogous to the node-level predictions. Graph-level predictions, on the other hand, require an output for the whole graph and need an embedding of the graph itself into a feature representation of fixed size. This can be achieved, for example, by averaging each feature across all the nodes in the last graph layer of the GNN, a procedure called global mean pooling. So, if the output graph of the last graph layer of a GNN has N nodes with feature vectors $\vec{x}'_i$ the graph-level embedding would be:

$$\vec{o} = \frac{1}{N} \sum_{i=1}^{N} \vec{x}'_i. \tag{3.31}$$

Figure 3.5 shows the embedding procedure for the case where each node has M features.

Hence, a graph-level prediction can be obtained by applying an output layer, or a DNN, to the graph-level embedding, as in the previous cases.

Transformer Convolutional layers

The main graph layer used for the design of the GNN in Chapter 6 is the Transformer Convolutional Layer [80], which applies the attention mechanism typical of transformer neural networks [81] to a graph layer. The output $\vec{o}_i$ of the i-th node, using a Transformer Convolutional Layer, is

$$\vec{o}_i = \phi\left(\mathbf{W}_1 \cdot \vec{x}_i + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j}(\mathbf{W}_2 \cdot \vec{x}_j + \mathbf{W}_3 \cdot \vec{e}_{i,j})\right), \tag{3.32}$$

with the attention coefficient $\alpha_{i,j}$

$$\alpha_{i,j} = \text{softmax}\left(\frac{(\mathbf{W}_4 \cdot \vec{x}_i)^\top (\mathbf{W}_5 \cdot \vec{x}_j + \mathbf{W}_6 \cdot \vec{e}_{i,j})}{\sqrt{d}}\right), \tag{3.33}$$

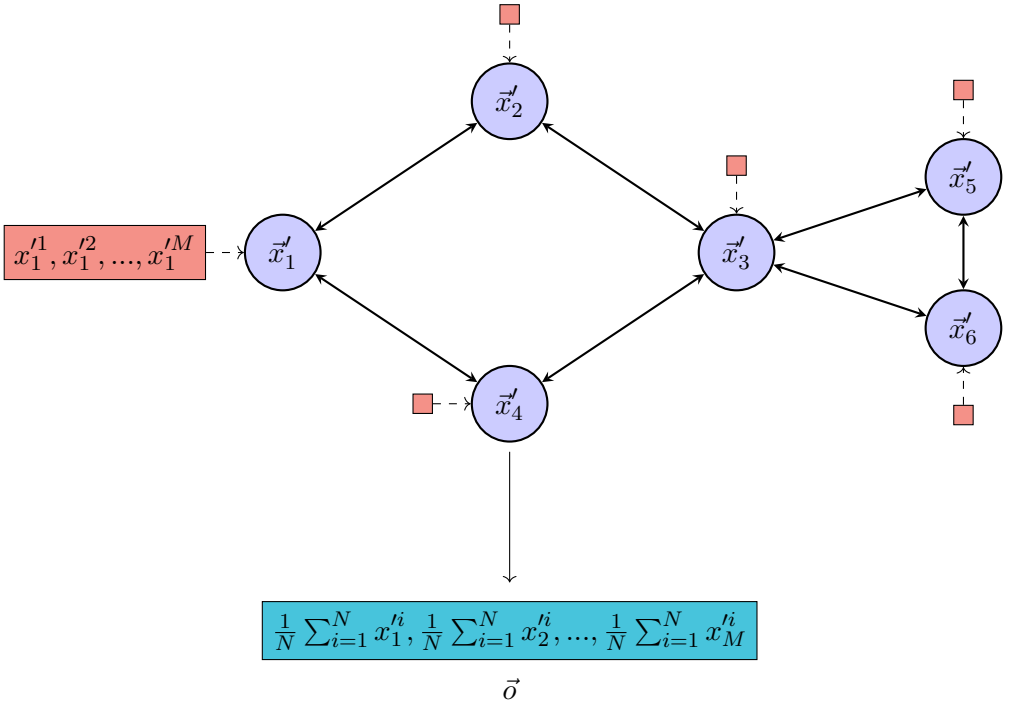


Figure 3.5: Example of embedding of a graph into a fixed-size feature representation through global mean pooling. The output of the last layer of the GNN is averaged across all the N nodes.

where, as in Eq. (3.30), ϕ is the activation function, $\vec{x}_i$ is the feature vector of the i-th node, $\vec{e}_{i,j}$ is the (possible) feature vector of the edge between the i-th and j-th node, and the $\mathbf{W}_{1,...,6}$ are the trainable weights of the layer. Additionally, d is the sum of the dimensions of $\vec{x}_i$ and $\vec{e}_{i,j}$, and the softmax function applied to a vector $\vec{z}$ of size K is defined as

$$\text{softmax}\,(\vec{z})_i = \frac{\mathrm{e}^{z_i}}{\sum_{j=1}^{K} \mathrm{e}^{z_j}} \tag{3.34}$$

and is treated as an activation function for the attention coefficient.

The attention mechanism allows a node to find the values from the connected notes that are more important, i.e., that need to be *paid more attention*. This kind of layer is used in Section 6.4, where the physics object represented by the node (lepton, jet, photon) is one-hot encoded in the node features. For instance, the attention coefficient between two leptons might be larger than the one between a lepton and a jet, signaling that the lepton node *is paying more attention* to the fact that a connected node is a lepton, compared to a different object. This mechanism allows the GNN to autonomously find which edges are meaningful to the designed task.

One last thing to notice is that the Transformer Convolutional Layer can also be defined with multiple attention coefficients and weight matrices, in a configuration called multi-headed attention. When designed this way, the layer can *pay attention* to multiple patterns in data and retain more information.

The output of H-headed attention Transformer Convolutional Layer $\vec{o}_i$ is defined as

$$\vec{o}_i = \phi\left(\mathbf{W}_1 \cdot \vec{x}_i + \frac{1}{H}\sum_{h=1}^{H}\sum_{j\in\mathcal{N}(i)} \alpha_{i,j}^h(\mathbf{W}_2^h \cdot \vec{x}_j + \mathbf{W}_3^h \cdot \vec{e}_{i,j})\right); \tag{3.35}$$

$$\alpha_{i,j}^h = \text{softmax}\left(\frac{(\mathbf{W}_4^h \cdot \vec{x}_i)^\top(\mathbf{W}_5^h \cdot \vec{x}_j + \mathbf{W}_6^h \cdot \vec{e}_{i,j})}{\sqrt{d}}\right). \tag{3.36}$$

Another implementation of H-headed attention for a Transformer Convolutional Layer consists of simply having H different outputs in the form of Eq. (3.32), with the $\mathbf{W}_{1,2,3}$ in common and separate own $\alpha_{i,j}^h$.

Of course, increasing the number of attention heads is not the only possible way to improve the accuracy of a model, and as with every procedure in ML, can bring to significant drawbacks, such as overfitting and increased computational time and resources.

4 The CMS experiment

The Compact Muon Solenoid (CMS) is one of the four experiments located place at the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) in Geneve, Switzerland. CERN was founded in 1954 and throughout its history it has been the stage of many discoveries in particle physics, such as the discoveries of weak neutral currents with the Gargamelle bubble chamber, and the discovery of the Higgs boson at the LHC, already mentioned in Chapter 1. Furthermore, the W and Z bosons were discovered at CERN, in 1983, by the UA1 and UA2 experiments [82–84].

4.1 The Large Hadron Collider

The Large Hadron Collider [85] is the world's largest and highest-energy particle collider, where two beams of particles travel at about the speed of light and are collided in four points of intersection, where the four main experiments are located: the Compact Muon Solenoid (CMS) [86], ATLAS [87], A Large Ion Collider Experiment (ALICE) [88], the Large Hadron Collider Beauty (LHCb) [89]. The LHC is able to accelerate protons and ions.

LHC is the last stage of a complex of different accelerators: each machine brings the energy of the particle beams to a threshold before injecting them into the next stage. As of Run 2 of LHC, between 2016 and 2018, protons are obtained from a hydrogen tank using an electric field to strip the atoms of their electrons. They are injected into [90]:

- Linac 2, a linear accelerator, which accelerates them up to 50 MeV of energy;
- Proton Synchrotron Booster (PSB), which brings them to 1.4 GeV;
- Proton Synchrotron (PS), which accelerates them to 25 GeV;
- Super Proton Synchrotron (SPS), which raises the energy to 450 GeV before injecting the beam into the LHC.

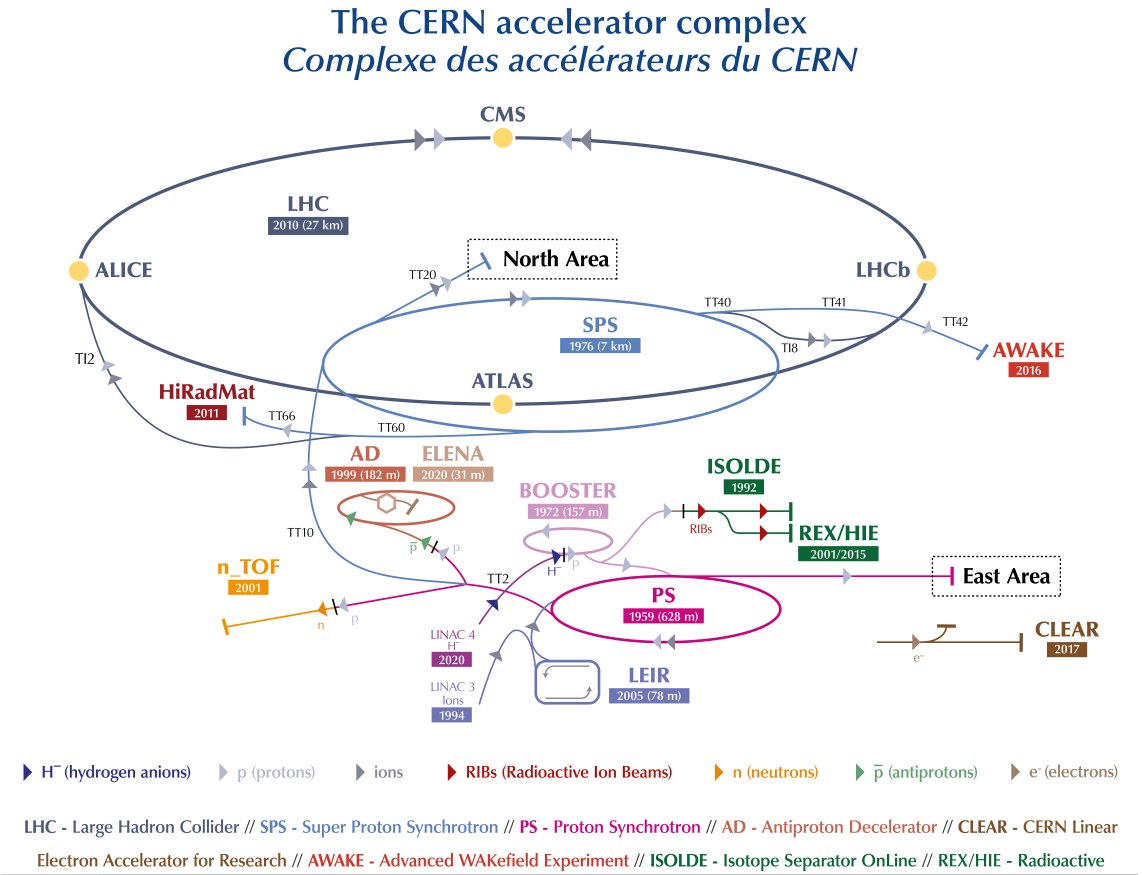


Figure 4.1: A schematic view of the accelerator complex at CERN, including the LHC and its accelerator chain. For each accelerator, the year in which the operations began is reported. From Ref [91].

Since 2020, Linac 2 has been replaced by Linac 4, which accelerates negative hydrogen ions to 160 MeV before feeding them to the PSB. A schematic view of the LHC and its current injector chain is shown in Figure 4.1.

The LHC is housed in a tunnel 27 km long in circumference, which was originally built for the Large Electron-Positron Collider (LEP), which was in operation from 1989 to 2000 [92]. Protons in the LHC are accelerated by sixteen superconducting radiofrequency cavities, which bring the energy of a proton beam to a maximum of 6.8 TeV. The use of radiofrequency cavities forces the protons not as a continuous beam, but in bunches, which are separated in time by 25 ns. During Run 2, the maximum energy was 6.5 TeV per beam, corresponding to a center-of-mass energy of 13 TeV.

A second parameter to describe the performance of a proton-proton collider is the luminosity (or instantaneous luminosity) $\mathcal{L}$, which is related to the rate of proton-proton collisions R in the machine through the total inelastic cross section σ_{pp}:

$$R = \mathcal{L}\sigma_{pp}. \tag{4.1}$$

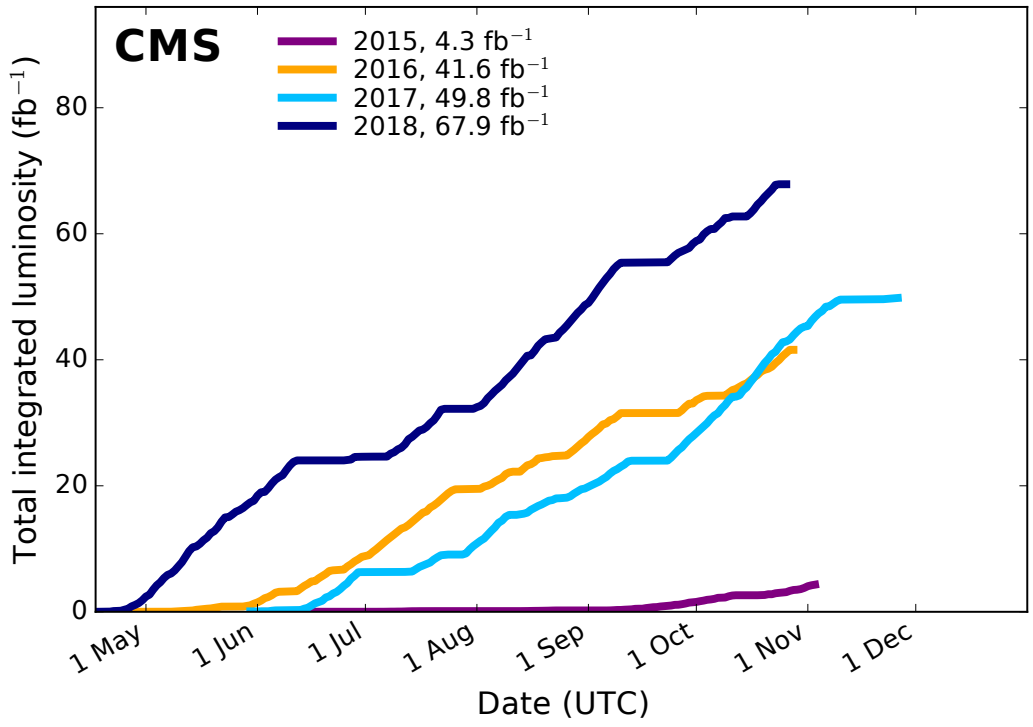


Figure 4.2: LHC integrated luminosity over time for the 2015-2018 period. From Ref. [95]

The luminosity only depends on the characteristics of the collider. If the beams collide with each other without a crossing angle (*head-on*), the luminosity for a single colliding bunch i can be calculated as [93]:

$$\mathcal{L} = N_1 N_2 f \int \rho_1(x,y)\rho_2(x,y)dxdy, \tag{4.2}$$

where N_1 and N_2 are the number of protons in the two colliding bunches, $f = 11246$Hz is the LHC orbit frequency, and ρ_1 and ρ_2 represent the normalized particle density distributions in the two bunches. The values of ρ are obtained using a method known as Van Der Meer scan, from its inventor Simon Van Der Meer, which consists in measuring the rate of interactions as a function of the beam separation at the interaction point [94].

Related to the luminosity is the integrated luminosity L, which is the time integral:

$$L = \int \mathcal{L}dt, \tag{4.3}$$

integrated over the time in which the collider is in operation. Figure 4.2 shows the integrated luminosity collected by the CMS experiment during the Run 2 period as a function of time. While Run 2 extended from 2015 to 2018, the analyses presented in this thesis work are based on data collected from 2016 to 2018, corresponding to an integrated luminosity of 138 fb^{-1}.

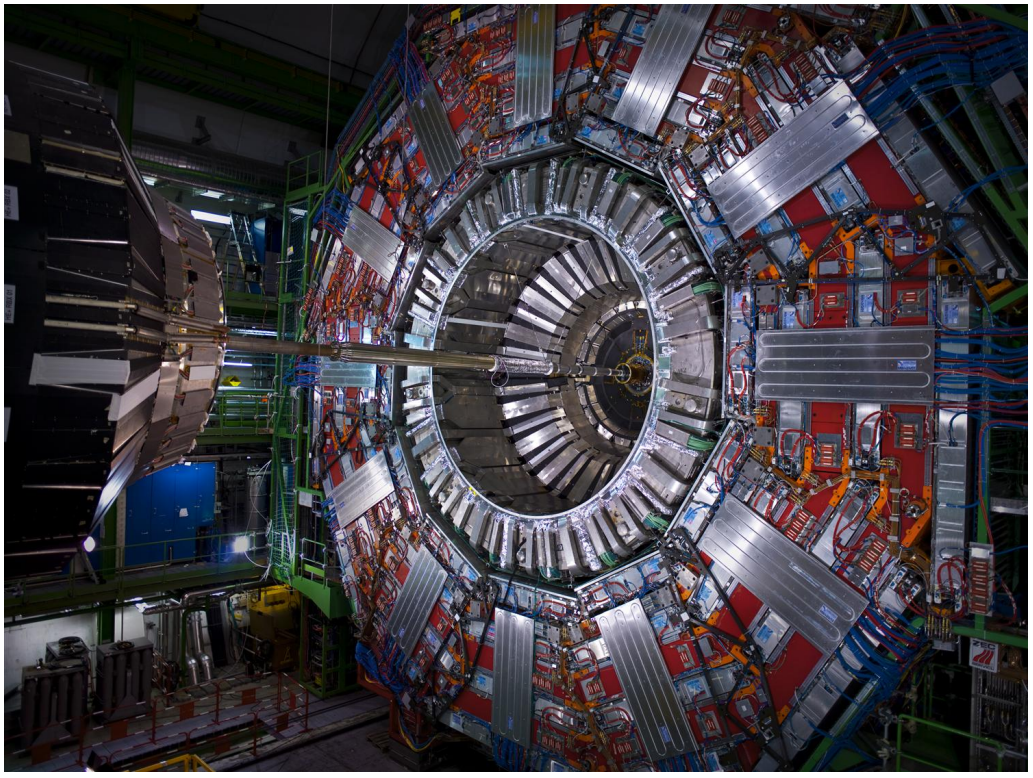


Figure 4.3: A picture of the CMS detector. From Ref. [96].

4.2 The Compact Muon Solenoid

CMS is a general-purpose particle detector, i.e., it has a broad physics program which ranges from the study of the Standard Model to the search for new physics beyond it. The detector is built around a superconducting electromagnet, the *Solenoid* in the CMS acronym, which provides the apparatus with a magnetic field of 3.8 T, bending the trajectory of charged particles and therefore allowing to precisely measure their momentum. The detector is in the shape of a 28.7-meter-long horizontal cylinder with a 15-meter-long diameter, centered around the collision point to exploit the cylindrical symmetry of the interaction. Figure 4.3 shows a picture of a layer of CMS during maintenance in Run 2. The CMS detector is divided in different subdetector systems, each designed for specific measurements and with precise requirements.

The experiment uses a right-handed Cartesian system of coordinates to describe the collision events:

- the x-axis points to the center of the LHC ring;
- the y-axis points upwards, orthogonal to the ring;
- the z-axis is along the beam pipe in the counterclockwise direction, observing the ring from above.

Furthermore, a cylindrical coordinate system is used to describe the quantities related to the particles arising from the collision:

- the radial distance r from the z-axis;

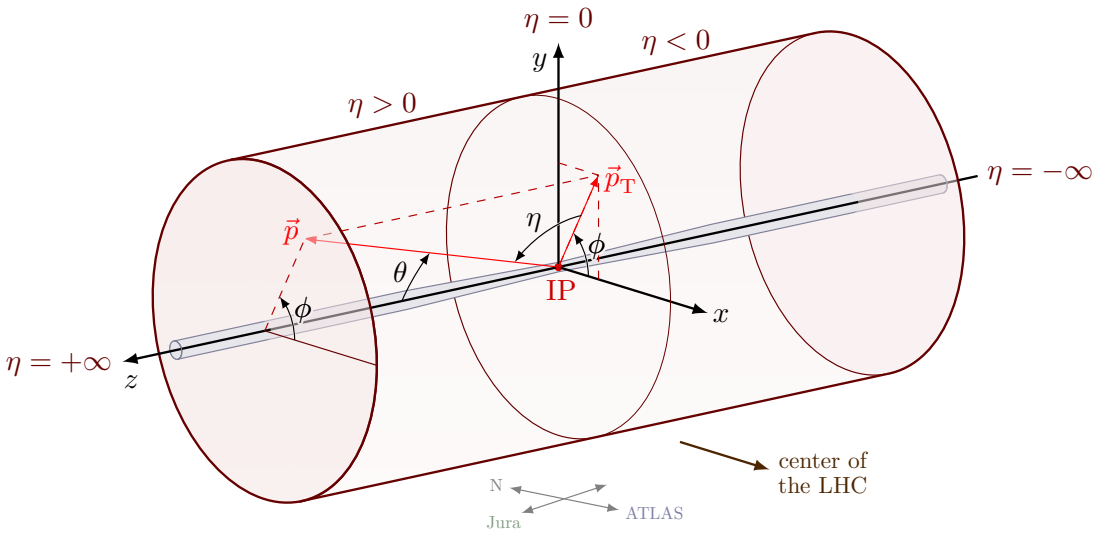


Figure 4.4: The coordinate systems in CMS. Image produced according to instructions in Ref. [97].

- the azimuthal angle ϕ in the xy plane, increasing from the x-axis;
- the polar angle θ in radial direction, increasing from the positive z-axis.

Most commonly, an alternative variable is used instead of θ: the pseudorapidity

$$\eta = -ln\left[tan\left(\frac{\theta}{2}\right)\right], \tag{4.4}$$

which transforms linearly under the effect of a Lorentz transformation along the z-axis, if the energy of the particle is much larger than its mass. Because of this property, the difference in pseudorapidity between two particles is, in this approximation, invariant under this kind of boosts. Figure 4.4 shows the coordinate systems used in CMS.

Other quantities which are invariant under a Lorentz transformation along the z-axis are:

- the angular distance:
$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}; \tag{4.5}$$
- the transverse momentum p_T:
$$p_T = \sqrt{{p_x}^2 + {p_y}^2}. \tag{4.6}$$

4.2.1 The subdetector systems of the CMS experiment

The CMS detector is composed of several subdetector systems and components:

- Beam Radiation, Instrumentation, and Luminosity (BRIL);
- tracking system;
- electromagnetic calorimeter (ECAL);
- hadronic calorimeter (HCAL);

- superconducting solenoid;
- muon system.

Many of the systems are divided in two subcomponents: the barrel, which is a cylindrical layer coaxial with the beam pipe, and the endcaps, which are plane layers placed at the two ends of the cylinder to ensure the detection of particles which travel close to the z-axis.

The tracker

The tracker is designed to precisely detect the passage of charged particles arising from the collision, allowing for a good reconstruction of their trajectory. It is the subdetector closest to the interaction point, as it starts only 30 mm away from the beam pipe. The short distance allows for the best-possible accuracy in the reconstruction of the secondary vertices of interaction, e.g., the decay vertices of mesons containing b quarks. The tracker has a diameter of 2.5 m and a length of 5.8 m; it covers the pseudorapidity range $-2.4 < \eta < 2.4$. It is composed of the inner tracker, made of silicon pixel detectors with pixels of size 100×150 µm, and the outer tracker, made of silicon strip detectors with strip pitches ranging from 10 cm $\times$ 80 µm to 25 cm $\times$ 205 µm.

When a charged particle passes through the tracker, it knocks out electrons from the silicon atoms, creating electron-hole pairs. The pairs are directed by an electric field to electrodes applied to the sensors, where they are collected. The movement of the electrons and holes generates a current which is picked up and amplified by a readout system.

The inner detector has an accuracy of 10 µm and of 20 µm for the radial and transverse positions, respectively [98]. It was originally designed to include 60 millions pixels distributed in three barrel layers and the two endcaps. After an upgrade performed between the years 2016 and 2017, it features 124 million pixel detectors distributed in four barrel layers and three endcaps [99]. Additionally, the outer tracker has an accuracy between 35 µm and 52 µm for the radial position and 530 µm in the transverse position.

The Electromagnetic Calorimeter

The ECAL is a hermetic homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals [100, 101]. Its main purpose is the measurement of the total energy of high-energy electrons, or positrons, and photons which enter the detector. Electrons interact in the material of the ECAL by undergoing bremsstrahlung, which is the radiation of a photon. Photons, instead, interact in the material by converting into electron-positron pairs. The combination of these phenomena causes the energy of an incoming particle to be quickly dissipated into the production of a multitude of lower-energy particles; a process called electromagnetic shower. The energy of the shower is deposited into the crystals of the ECAL, causing its atoms to reach excited energy states. The return to the ground energy state of the atoms of the $PbWO_4$ lattice is mediated by the emission of photons in the visible spectrum, which are collected by different photodetectors, generating an electric signal. Avalanche photodiodes are used in the barrel, while the endcaps are equipped with vacuum phototriodes. The ECAL is placed at a radial distance between 1.25 m and 1.8 m from the beam pipe.

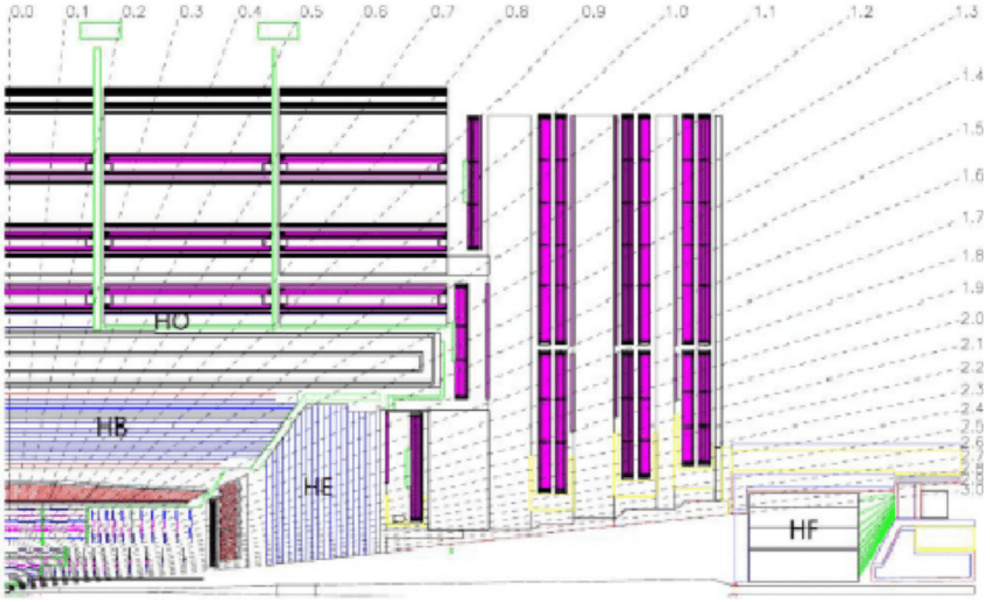


Figure 4.5: Longitudinal view of one quarter the CMS detector. The locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters are shown. From Ref. [86].

The Hadronic Calorimeter

The HCAL is a sampling calorimeter, i.e., it is made of layers of active material and absorber. The active material is made of a fluorescent scintillator, the absorber is cartridge brass. The main purpose of the hadronic calorimeter is the measurement of the energy of hadrons produced in the collision event. The hadrons interact with the nuclei of the atoms of the absorber through the strong interaction, which causes the production of hadronic showers. The showers present an electrically charged component, e.g., charged mesons or leptons from their (semi-)leptonic decays, which create a signal in their passage through the fluorescent scintillator, with a similar mechanism as in the ECAL. The presence of the absorber is made necessary by the fact that the mean free path for a hadron in a material is one order of magnitude larger than for an electron, or a photon, in the same material [14]. The mean free path is inversely proportional to the density of the material, as the presence of more nuclei translates to a higher probability of a hadron interacting with one of them. Hence, alternating layers of active material and higher-density absorber allows the calorimeter to contain the entire development of the hadronic shower in the thickness of the detector.

The HCAL is divided in Barrel (HB) and Endcap (HE) sections, with a range $|\eta| < 3.0$. Additionally, an Outer Barrel (HO) and a Forward (HF) sections are present, realized with different materials. A view of all four HCAL section is displayed in Figure 4.5.

The HF is 11.2 m away from the interaction point along the z-axis, covering the pseudorapidity range $3 < |\eta| < 5.2$. It is made of quartz fibers embedded within a 165 cm long steel absorber and exploits a technology based on the Cherenkov effect. The HO is composed of

scintillators that are placed outside the solenoid to ensure adequate sampling depth and to measure late shower development.

Solenoid

The solenoid is a superconducting coil surrounding the tracker and calorimeters, apart for a small portion of HCAL. It provides the detector with a uniform 3.8 T magnetic field. An iron return yoke is located around the magnet to avoid border effects, and to bend the lines of force in a way to provide an almost constant magnetic field of 1.8 T magnitude outside the solenoid bore, as well.

Muon system

The muon system is the part of the detector which is used for muon identification, momentum measurement and triggering. Muons are the final products in many processes, and can pass through multiple layers of material while losing only a small fraction of their energy. Therefore, the muon system of CMS is the outermost layer of the detector, placed in between layers of the iron return yoke, where only muons are expected to survive. It covers an active area of about 25000 m^2. Three types of gas detectors are used: drift tubes (DTs), cathode strip chambers (CSCs) and resistive plate chambers (RPCs). Their functioning principle is based on the ionization of the gas by the passage of the muon. The electrons and ions are then collected by electrodes (wires, strips, or plates), generating a signal which is read out by the dedicated electronics.

The DTs are placed in the barrel and cover a pseudorapidity range $|\eta| < 1.2$. They are arranged in four stations which form concentric cylinders around the beam line. The DTs are used are the main tracking system for muons in the barrel.

The CSCs are placed in the endcaps, covering a range $0.9 < |\eta| < 2.4$. The cathode strips and anode wires in the CSCs are arranged in a cross pattern, allowing to measure two position coordinates for each passage of a muon. In a CSC, the electrons knocked out of the gas atoms create electronic avalanches next to the anode wires, creating a clear signal.

The RPCs are placed both in the barrel and the endcaps and have very fast response, in the order of nanoseconds, and very good time resolution. These features make them especially valuable for triggering purposes.

Trigger

The LHC provides proton-proton bunch crossings at a maximum rate of 40 MHz, with an average of 35 collisions per crossing. The maximum rate of collision events collected by CMS that can be stored was, for Run 2, around 1 kHz. This drastic reduction in the rate of stored events is achieved by the trigger, which is composed of Level-1 trigger (L1 trigger) and High Level Trigger (HLT).

The main principle behind the trigger selection is to select all possible *interesting events*, which are the ones that can be used for new physics analyses and measurements.

The L1 trigger, built with custom-designed electronics, has a latency of 4 µs, reducing the incoming rate to 100 kHz [102]. It is divided in a muon component and a calorimeter component.

The former is based on information from the muon systems, whose signals are combined together into trigger primitives (TPs). The TPs are used in the reconstruction of muon tracks, which are ranked in quality and sent to the Global Muon Trigger, which removes duplicate tracks coming from different TPs and in turn sends the information to the Global Trigger.

The L1 calorimeter trigger is divided in two layers. The first layer reconstructs TPs from the energy deposits in the ECAL and HCAL. The second layer combines this information into the reconstruction of the physics objects, such as electrons, photons, jets, and energy sums. The second layer communicates with the Global Trigger, whose task is to take the final decision to reject the event or to pass it to the HLT for further evaluation.

The HLT, is an offline software trigger which runs a streamlined version of the reconstruction algorithms, described in the next section, and finally select events for further analysis and storage [103, 104]. The HLT functioning revolves around the concept of *trigger paths*, which are sets of algorithmic processing steps. These include steps to reconstruct the particles produced in the collision events, and to select the events if the reconstructed objects respects predetermined thresholds, e.g., in transverse energy, isolation, reconstruction quality.

BRIL

The BRIL subsystem is responsible for the monitoring of the luminosity collected by the CMS detector. It uses both detectors specifically built to measure luminosity, such as the fast beam conditions monitor BCM1F and the pixel luminosity telescope (PLT), and information coming from other subsystems, such as the inner tracker, the HF, and the DTs. The redundancy results in a reduction in the systematic uncertainties of the luminosity measurements and in the safeguard against the possibility of all detectors not being available in the same period.

The measurement of the luminosity is performed through the measurement of a rate R specific to the detector (hits, tracks, clusters), with the proportionality constant given by the visible cross section σ_{vis}:

$$\mathcal{L} = \frac{R}{\sigma_{\text{vis}}}. \tag{4.7}$$

The value of σ_{vis} is specific to each detector and includes factors due to the conditions of the detector, its geometric acceptance and efficiency. It is usually measured according to the provisions in Ref. [93] in special LHC fills devoted to VdM scans.

4.3 Event reconstruction

The event reconstruction at CMS is based on a Particle-Flow (PF) algorithm [105], i.e., a global reconstruction technique which optimally combines information from all subdetector

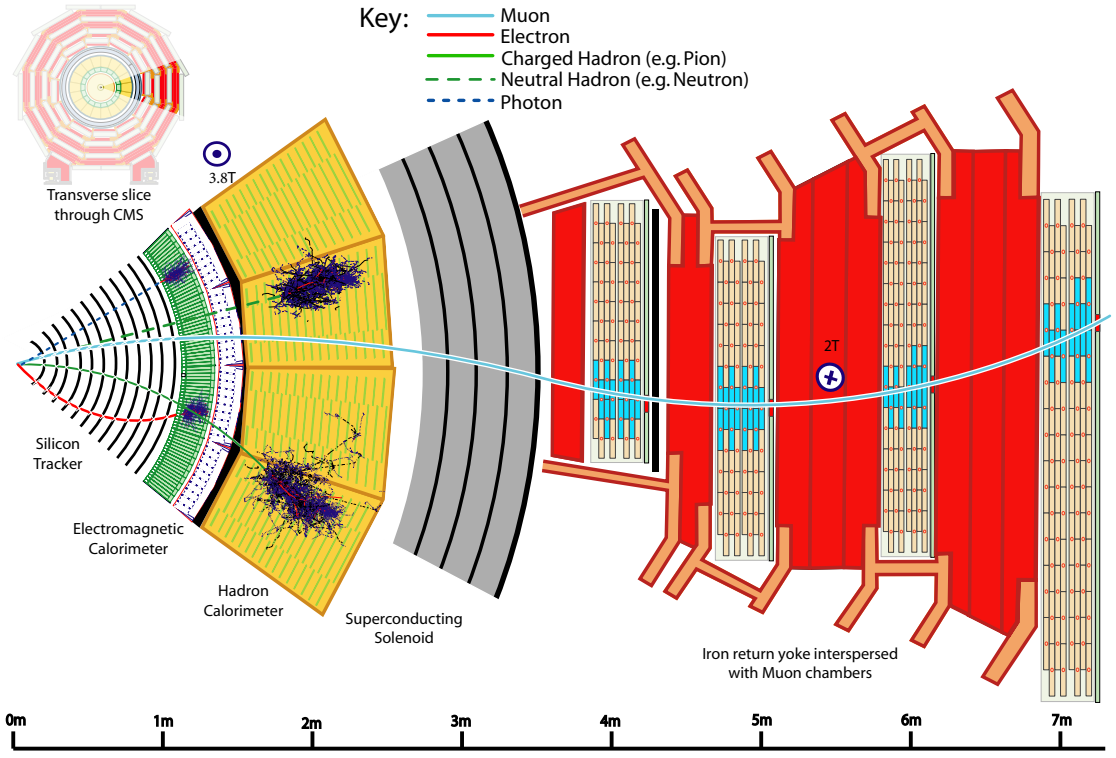


Figure 4.6: A schematic view of a transverse slice of CMS, showing the interactions of the particles with the subdetector systems. From Ref. [105].

systems to reconstruct all final-state particles produced in the collision event. Typically, in CMS, electrons can be reconstructed by the hits they generate in the tracker, and a matching energy deposit in the ECAL. Photon reconstruction is based on the energy deposit in the ECAL. Charged hadrons are also reconstructed by hits in the tracker, followed by a hadronic shower, which can start in the ECAL, but ultimately is absorbed by the HCAL. Neutral hadrons are reconstructed by the energy deposit in the ECAL and HCAL. Finally, muons can be reconstructed by the hits they generate in all detector layers, as they are usually the only particles which pass through the entire detector. Their transverse momentum is reconstructed by the curvature of their trajectory in the magnetic field of the solenoid.

The interactions of the particles in the volume of the detector are graphically summarized in Figure 4.6, where a schematic view of a transverse slice of CMS is shown.

This section describes the methodologies used in the reconstruction of the different particles produced in the collision events.

Tracks and primary vertices

The first reconstruction step is the determination of the tracks of the charged particles produced in the collision. This is done through a combinatorial track finder algorithm based on a Kalman filter [106, 107]. It proceeds in three stages: a few hits compatible with the passage on one particle are taken as seed; a pattern recognition algorithm is used

to collect all hits compatible with the seed track; a global fit is performed to extract the properties of the charged particle.

This algorithm is run iteratively for ten times, with looser quality requirements for the tracks at each step. Hits corresponding to reconstructed tracks are not considered in following iterations of the track finder algorithm. Moreover, the last iteration is designed to consider hits in the muon system as seed, aiming at increasing the reconstruction efficiency for muons. This procedure ensures high efficiency in the reconstruction of the tracks, while maintaining the misreconstructed track rate at the same level as the first iteration.

The reconstructed tracks are used for the primary-vertex reconstruction, i.e., the measurement of the location of all proton-proton interaction vertices in an event, including both the *signal* vertex and vertices from pileup (PU) collisions. The PU collisions are comprised of all other proton-proton collisions that are simultaneous with the signal collision under study. Primary-vertex reconstruction consists of three steps [98]. First, tracks that are compatible with the primary interaction region are selected. This region consists of the volume close to the interaction point. Secondly, the tracks that appear to originate from the same vertex are clustered together, and are then used in a fit to extract the position of each vertex.

Muons

In CMS, isolated muons are reconstructed with an efficiency close to 100%, and the reconstruction of their tracks is different from the one generally used in the PF algorithm. Three non-mutually-exclusive types of muons are reconstructed in CMS. Hits in the DTs and CSCs detectors are clustered to form track segments which act as seeds used by a pattern recognition algorithm to cluster all compatible hits in the muon system. These tracks are defined as *standalone muons*. If the standalone muon is compatible with a track reconstructed in the inner tracker of the detector, the hits from the two tracks are combined to perform a new fit which returns a *global muon*. Lastly, each track reconstructed in the inner tracker with $p_{\mathrm{T}} > 0.5$ GeV and $p > 2.5$ GeV is extrapolated to the muon system. If at least one track segment matches the track, this latter is considered a *tracker muon*.

Electrons and photons

Electrons and photons are expected to deposit all their energy in the ECAL, through an electromagnetic shower. Because of the similarities in the way they interact with the detector, electrons and photons are reconstructed in the same step. Because of the thickness of the tracker, the electromagnetic shower can be initiated before the particle is able to reach the ECAL, generating multiple particles which, in turn, produce scattered energy clusters in the ECAL, which need to ultimately be related to the original incoming electron or photon. Additionally, the emission of a photon from an electron in the tracker causes the electron to bend while in the tracker. In order to accurately reconstruct electrons, their track finding algorithm is based on a Gaussian-Sum Filter (GSF) [108], and not on a Kalman filter.

The reconstruction proceeds in the following steps [100]. ECAL clusters are formed by grouping together crystals which measured energy exceeding a predefined threshold. In any specific region, a seed cluster is defined as the one with the most energy, which needs to be greater than 1 GeV. Clusters within a certain area around a seed cluster are combined together into a supercluster (SC). Seeds in the pixel detector that are compatible with the position of the SC, and the trajectory of an electron, are used as seeds to the GSF algorithm.

At the same time, a dedicated algorithm [109] is used to identify tracks which are compatible with the conversion of a photon into an electron-positron pair in the volume of the tracker, defined as *conversion tracks*. Electrons are identified by SCs linked to a GSF track compatible with the trajectory of an electron. Photons are identified by SCs which do not have a GSF track, or that are compatible with conversion tracks.

Missing transverse energy

The last step of the PF algorithm is the calculation of the missing transverse energy (MET), which is defined as the negative vector sum of the transverse momenta of all the particles reconstructed in the event:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = \sum_{i=particle} \vec{p}_{T,i}. \tag{4.8}$$

The interacting partons in the colliding protons have no transverse momentum, so the total transverse momentum of the particles in the final state should be zero. A deviation from zero, i.e., a non-null MET can indicate the presence of undetected particles, such as neutrinos.

4.3.1 Detector simulation

After parton showering and hadronization, simulated proton-proton collision events need to be processed through a full simulation of the CMS detector, in order to be compared with the observed data. The simulation of the passage of the particles through the detector, and their interactions with the subdetector systems and the magnetic field are performed with the GEANT4 toolkit [110]. The software includes a full model of the CMS detector. The output of the GEANT4 simulation is then processed by the PF algorithm in the same way as the data collected by the detector. Furthermore, the simulated events at CMS include the effect of pileup, whose distribution is based on the one expected in the observed data.

5 First evidence for the production of the tWZ process

The first analysis discussed in this thesis is the search for the SM tWZ process, which resulted in the first evidence for the process and was published in Ref. [111].

The tabulated results of this analysis are provided in the HEPData record [112].

Section 5.1 presents an introduction to the analysis and its motivations. The overview of the strategy and workflow of this search, outlining the trigger selection, the reconstruction and identification of the physics objects, and the processes that represent a background to the tWZ process, are given in Section 5.2. Section 5.3 defines the signal and control regions in which the analysis is performed; it also describes the design, training, performance, and validation of the classifiers used to discriminate between the signal and the background processes in the signal regions. Section 5.4 is dedicated to the detailed description of the DNN algorithm which is used to classify top quarks with high transverse momentum which decay in leptons. Section 5.5 details the systematic uncertainties that are considered in the maximum likelihood fit used to extract the signal strength and statistical significance of the tWZ process. Finally, the results of the search are presented in Section 5.6.

5.1 Introduction

The associate production of a top quark with a W and a Z boson is conceptually similar to the tW production process. The main difference is, of course, the emission of a Z boson from any internal or external line of the Feynman diagrams for the production process. These are shown in Chapter 2, together with the studies performed on the modeling of the physics process.

The theoretical interest in tWZ does not stop at the definition of a reliable and sound modeling; the process has been found to be a very promising channel to uncover effects of physics beyond the Standard Model in the context of the Effective Field Theory (EFT)

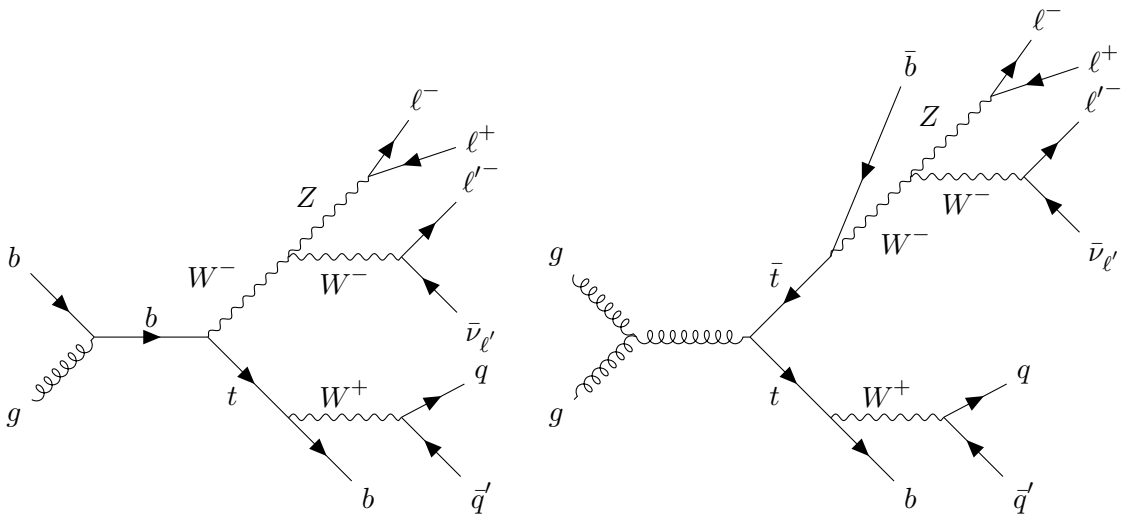


Figure 5.1: Illustrative Feynman diagrams for the tWZ (left) and $t\bar{t}Z$ (right) processes in the final state with three leptons. The main difference between the two is represented by the presence of one hard b jet coming from the decay of the top anti-quark in $t\bar{t}Z$.

approach [113]. Furthermore, recent theoretical studies suggested that these effects are enhanced in the regions of phase space where the final state particles are produced with high energy [54].

The most challenging aspect of the experimental search for the tWZ process is undoubtedly represented by the extreme similarity with the $t\bar{t}Z$ process. This similarity not only affects the modeling of the signal process, but it also makes it complicated to distinguish between collision events traceable to either process. The only difference in the final state signatures consists of one additional b jet arising from the decay of the second top quark in the $t\bar{t}Z$ process, which is not always reconstructed or identified.

The two illustrative Feynman diagrams in Figure 5.1 show the difference between tWZ (left) and $t\bar{t}Z$ (right) in the case where three leptons are present in the final state.

One more hurdle in the identification of tWZ events is represented by the fact that its cross section is significantly smaller than the theoretical cross section for the production of the $t\bar{t}Z$ process. It is calculated to be $\sigma_{t\bar{t}Z} = 859^{+76}_{-84}$ fb at NLO, in QCD and electroweak, and next-to-next-to-leading logarithmic accuracy [114]: almost seven times as large as the theoretical tWZ cross section of $\sigma^{DR1}_{tWZ} = 130$ fb. The latest measurement performed by the CMS experiment resulted in a measurement of $\sigma^{\text{exp}}_{t\bar{t}Z} = 0.95 \pm 0.05\ (\text{stat}) \pm 0.06\ (\text{syst})$ pb, compatibly with the theoretical prediction.

The tWZ analysis is carried out in the three- and four-lepton final states, where the Z boson decays into oppositely charged leptons, and either the W boson or the top quark, or both, decay with a charged lepton in the final state. The final state signature with three leptons is completed by the presence of a b jet coming from the decay of the top quark and two light jets which come from the hadronic decay of the W boson, in addition to a neutrino, which results in high MET. The four-lepton final state is characterized by the

presence of the b jet alone and two neutrinos, which are also responsible for high MET. The charged leptons considered are electrons and muons.

The strategy employed in this search is based on dividing the collected data into multiple signal and control regions, so to constrain the uncertainties coming from all the contributing processes as well as possible.

Additionally, the analysis relies heavily on the use of machine learning algorithms to create classifiers to separate the signal from the background processes in each signal region of the analysis. Most of the classifiers are used as templates to be fitted through the Maximum Likelihood Estimation technique, as described in Chapter 3.

Furthermore, the analysis employs a boosted leptonic top-tagger, described in Section 5.4, which is used to identify combinations of b jets and leptons that are likely to come from the decay of a high-p_{T} top quark. The aim of this high-energy region is to define a phase space where the tWZ process might be more sensitive to the effects of new physics.

The result of the search is presented in Section 5.6, where the statistical significance of the signal is calculated through both a fit where the signal strength of tWZ μ_{tWZ} is the only parameter of interest, and a simultaneous fit of μ_{tWZ} and $\mu_{t\bar{t}Z}$, the signal strength of the $t\bar{t}Z$ process. In either case, the significance of the signal is found to be greater than three standard deviations from the background-only hypothesis, representing first evidence of the tWZ process at the LHC.

5.2 Analysis overview

The search for the SM tWZ process has been carried out using proton-proton collision data collected by the CMS experiment at $\sqrt{s} = 13$ TeV in the years 2016-2018, corresponding to an integrated luminosity $\mathcal{L} = 138$ fb^{-1}. The data-taking was divided in four periods: 2016 pre-VPF, 2016 post-VFP, 2017, and 2018. While the division between different years is justified by upgrades made to the CMS subdetector systems at the end of every year, the division between 2016 pre-VFP and post-VFP is due to an issue in the silicon tracker during the Summer of 2016. It was found that the discharge time of the pre-amplifiers used in the readout of the tracker showed a dependency on the occupancy rate, i.e., to the instantaneous luminosity. This caused a loss of efficiency in the detection of tracker hits. The solution was the application of a voltage bias to the pre-amplifier (VFP) which increased the discharge speed of the pre-amplifiers [115].

This section describes the simulation of the signal samples, the trigger logic used to select the events of interest, the physics objects reconstruction, and the background processes present in the analysis.

5.2.1 Signal simulation

The signal process is simulated according to the prescriptions given in Chapter 2. The simulated samples are produced using the `MadGraph_aMC@NLO` generator at NLO in QCD with the 5FS, using `MadSTR` to implement the DR1 and DR2 schemes. The DR1 modeling

is considered as the nominal prediction for the process, while the DR2 scheme is used to estimate the theoretical uncertainties. The samples produced by `MadGraph_aMC@NLO` are generated to the matrix element level. `Pythia8` is then used to simulate the parton showering of the events and the hadronization.

Furthermore, in order to increase the number of simulated events used to model the high-energy region of the phase space, additional samples have been produced with a requirement on the p_{T} of the top quark, which is set to be greater than 270 GeV. In order to correct the disagreement between the DR1 and DR2 schemes in the high-p_{T} region, the *b-veto* criterion is applied to these samples, as described in Section 2.2.2. The number of events is then renormalized to the pre-veto value.

5.2.2 Trigger selection

Only a part of the events collected by the CMS experiment can be stored for further analysis, due to the high rate of collision events produced by the LHC. The decision on whether to record or discard a collision event is taken by the CMS two-tiered trigger system. As explained in Chapter 4, the final decision on the selection of a collision event is taken by the HLT. For this analysis, the events are collected using an extensive combination of single-, di- and tri-lepton (muon or electrons), plus photon trigger paths. These latter ones are included in the analysis to increase the efficiency for the selection of high-p_{T} electrons, which can be wrongfully reconstructed as photons. This combination has first been used for the measurement of associated production of a top quark and Z boson (tZq) in Ref. [116], where it has been proved to have a selection efficiency close to 100% for the selection of events with three leptons with transverse momenta $p_{\mathrm{T}} > 25, 20, 10$ GeV.

5.2.3 Objects reconstruction and identification

Following the trigger selection, the PF candidates are required to pass a series of offline selection criteria in order to be considered for the subsequent steps of the analysis.

Leptons

The main strategy of this analysis is using a multi-lepton final state in order to reduce the contamination from the background processes, as the SM does not allow many processes that produce three or four leptons in the final state at LHC; the ones that are allowed are described in Section 5.2.5. Nevertheless, this choice comes at the cost of increased probability of selecting nonprompt leptons. Prompt leptons are the ones coming from the decay of the W and Z bosons or from the leptonic decay of a τ lepton, which itself originated from the decay of a boson. Nonprompt leptons are misidentified hadrons, misidentified photons, or genuine leptons, which are produced in semileptonic decays of hadrons, especially b hadrons. They are identifiable by being less isolated, originating from a secondary vertex, or being close to a b jet.

It is fundamental to reduce the rate of nonprompt leptons that are erroneously selected as much as possible. The TOP `Lepton`MVA boosted decision tree algorithm has been

developed to identify prompt leptons with high efficiency and low rates of nonprompt lepton misidentification and has been profitably used in CMS analyses in multileptonic final states [116, 117].

Table 5.1: Preselection criteria for the TOP `LeptonMVA` algorithm for electrons and muons.

Electron	**Muon**
$p_{\mathrm{T}} > 10$ GeV	$p_{\mathrm{T}} > 10$ GeV
$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.4$
$d_{xy} < 0.05$	$d_{xy} < 0.05$
$d_z < 0.1$	$d_z < 0.1$
$\mathrm{SIP}_3 < 8$	$\mathrm{SIP}_3 < 8$
$\mathrm{I}_{\mathrm{mini}} < 0.4$	$\mathrm{I}_{\mathrm{mini}} < 0.4$
$N_{\mathrm{miss}} < 2$	Has track in both tracker and muon system
	Medium ID

The PF lepton candidates are first required to respect a set of preselection criteria, which are reported in Table 5.1, where d_{xy} and d_z are the transverse and longitudinal impact parameters, respectively, and SIP_3 is the significance of the impact parameters in three dimensions, defined as the ratio of the impact parameter itself and its uncertainty. Additionally, N_{miss} is the number of missing hits in the inner tracker, and the muon Medium ID refers to a set of identification criteria described in Ref. [118]. Also, the muons undergo a further correction of the measured p_{T}, as prescribed in Ref. [119].

Furthermore, the relative isolation of the leptons is defined as the scalar sum of the p_{T} of the PF candidates in a cone centering on the lepton direction, excluding the lepton itself. For electrons, the relative isolation is given by:

$$\mathrm{I}_{\mathrm{rel}} = \frac{1}{p_{\mathrm{T}}} \left(\sum_{\text{charged}} p_{\mathrm{T}} + \max \left(0, \sum_{\text{neutral}} p_{\mathrm{T}} + \sum_{\text{photon}} p_{\mathrm{T}} - \rho A \right) \right), \tag{5.1}$$

where ρ is the average p_{T} flow density in the event, and A is the area of the geometric cone used to calculate the isolation. For muons, the term ρA is replaced by $\frac{1}{2} \sum_{\mathrm{PU\ charged}} p_{\mathrm{T}}$, where the p_{T} is summed over the charged PF candidates not associated with the primary vertex [120]. In the case of the mini-isolation, used for the preselection, the radius of the isolation cone R is given by

$$R = \begin{cases} 0.2, & p_{T,\ell} \leq 50 \text{ GeV}, \\ \frac{10 \text{ GeV}}{p_{T,\ell}}, & 50 \text{ GeV} < p_{T,\ell} < 200 \text{ GeV}, \\ 0.05, & p_{T,\ell} \geq 200 \text{ GeV}. \end{cases} \tag{5.2}$$

The idea of correlating the size of the isolation cone with the p_{T} comes from considering the decay of a massive particle with a lepton and a jet in the final state [121], as it is the case for the leptonic decay of a top quark. If the top has high energy, the decay products tend to be collimated, with a typical angular distance between them $\Delta R_{\mathrm{prod}} \approx 2 \frac{m_t}{p_{T,t}}$, with m_t and $p_{T,t}$ being the mass and the p_{T} of the decaying top quark, respectively. Decreasing

R with the increasing p_{T} of the lepton reduces the probability of the isolation cone and the b jet accidentally overlapping.

The TOP `LeptonMVA` employs input features related to the kinematic properties, isolation, and impact parameters of the lepton, and properties of the closest jet.

The leptons in this analysis are selected when their MVA output is above the medium working point, for which the prompt lepton selection efficiency is 93%, and the nonprompt lepton selection efficiency is $\mathcal{O}(1\%)$. Both efficiency values are calculated with respect to the preselection.

Jets

In the final state of the tWZ process, hadronic jets arise either as b jets from the decay of the top quarks or from the hadronic decays of the W boson. Furthermore, the presence of an extra b jet in the selected event is a good indicator of the $t\bar{t}Z$ process, and it is useful in the discrimination between the two processes. In this analysis, the jets are reconstructed using the anti-k_T algorithm with a distance parameter of 0.4 (AK4 jets) [122], with the PF candidates used as input. The charged PF candidates which are associated with a pileup vertex are not considered in the jet reconstruction, following the Charged Hadron Subtraction (CHS) algorithm [123].

Jets are selected in the analysis if they satisfy the requirements outlined in Ref. [123], they have $p_{\mathrm{T}} > 25$ GeV and are central, $|\eta| < 2.5$. In case their p_{T} is smaller than 50 GeV, they are also required to pass the loose working point of an MVA-based PU jet ID, defined as selecting non-pileup jets with 99% efficiency. Furthermore, jets need to be separated from the selected leptons by $\Delta R > 0.4$.

The jets undergo a series of corrections to the energy scale calibration, and the differences in the jet energy calibration (JEC) and resolution (JER) between data and simulation are corrected according to the prescriptions in Ref. [124].

A jet is considered b-tagged if it passes the medium working point of the DeepFlavour algorithm [125, 126]. The DeepFlavour algorithm is a DNN that leverages information from PF candidates and secondary vertices in a jet to provide the probability that a given jet is a b jet, a c jet, or a light jet.

Fat jets

In addition to AK4 jets, the analysis also makes use of anti-k_T jets with a distance parameter of 0.8 (AK8 jets), with the aim of reconstructing the hadronic decays of high-p_{T} top quarks. The decay products from a high-p_{T} top quark are expected to receive a large Lorentz boost, which collimates them and makes them appear as a single large-radius jet. An AK8 jet is selected if it has $p_{\mathrm{T}} > 300$ GeV and $|\eta| < 2.4$, and it is required to be separated from the selected leptons by $\Delta R > 0.8$. Additionally, the AK8 jet is required to be within $\Delta R < 0.8$ from a selected AK4 jet.

Missing transverse energy

The neutrinos produced in the leptonic decays of the W bosons cannot be detected in CMS, as they exit the detector without interacting with it. The MET is used as a proxy to reconstruct its properties. In this analysis, type 1 MET is used, which is calculated as:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\left(\sum_{\text{Unc. PF}} \vec{p}_T + \sum_{\text{Jets}} \vec{p}_T \right), \tag{5.3}$$

where the unclustered PF candidates are the ones that are not associated to any jet. The jets, after the application of energy corrections, are used instead of the PF candidates they were reconstructed from. This increases the precision of the $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement, as the p_{T} of jets is usually measured with higher precision than the p_{T} of its components.

5.2.4 Top quark reconstruction

The properties of the top quarks in the events contribute significantly to the discrimination between the tWZ and $t\bar{t}Z$ processes, and they are employed in the phase space with three leptons in the final state. There are two main complications in the reconstruction of top quarks from their decay products. The first one is the presence of a neutrino in the final state, whose longitudinal momentum cannot be measured. The second one is that there is no unique way to assign the leptons and the jets to either the top quark or the W boson in the process.

Neutrino longitudinal momentum

The momentum of the neutrino is reconstructed according to the recipe given in Ref.[127]. It is based on imposing the conservation of energy in the $W\ell\nu$ vertex, where the lepton is the one not associated with the Z boson, and the neutrino is considered to be the only source of MET. Therefore, the longitudinal component of the neutrino momentum $\vec{p}_\nu$ is calculated as the solution of the equation:

$$(\vec{p}_\nu + \vec{p}_\ell)^2 = m_W^2, \tag{5.4}$$

with

$$\vec{p}_{T,\nu} = p_{\mathrm{T}}^{\mathrm{miss}}. \tag{5.5}$$

A detailed description of the solution to Eq. (5.4) is given in Appendix A.

Assignment of the decay products

Having reconstructed the momentum of the neutrino, it is now possible to assign the lepton and the jets to the top quark and the W boson. Two chi-square quantities are constructed, χ^2_{had} representing the hypothesis where the top quark decays into jets and the W boson into a lepton and a neutrino, and χ^2_{lep} representing the hypothesis where the lepton comes

from the top quark decay and the W boson decays into jets. The two chi-square quantities are:

$$\chi^2_{\text{had}} = \left(\frac{m_{\text{bjj}-m_t}}{\sigma_{t,\text{had}}}\right)^2 + \left(\frac{m_W - m_{\ell\nu}}{\sigma_{W,\text{lep}}}\right)^2 = \left(\frac{m_{\text{bjj}-m_t}}{\sigma_{t,\text{had}}}\right)^2, \tag{5.6}$$

where the second term vanishes because of Eq. (5.4), and

$$\chi^2_{\text{lep}} = \left(\frac{m_{\text{b}\ell\nu} - m_t}{\sigma_{t,lep}}\right)^2 + \left(\frac{m_{\text{jj}-m_W}}{\sigma_{W,had}}\right)^2. \tag{5.7}$$

In these equations, m_{bjj} is the invariant mass of the three jets selected, $m_{\text{b}\ell\nu}$ is the invariant mass of the b jet and the lepton and neutrino, m_t is the top quark mass, m_W is the W boson mass, $m_{\ell\nu}$ is the invariant mass of the lepton and the neutrino, and m_{jj} is the invariant mass of the two jets which are not b-tagged. If the discriminant in Eq. (6.3) is positive, two solutions are found, and the one that yields the smallest value of χ^2_{lep} is selected. The σ terms are the resolution on the reconstructed invariant masses of the top quark and the W boson and are measured in simulation to be:

$$\sigma_{t,\text{had}} = 22 \text{ GeV}, \tag{5.8}$$

$$\sigma_{t,\text{lep}} = 19 \text{ GeV}, \tag{5.9}$$

$$\sigma_{W,\text{had}} = 12 \text{ GeV}. \tag{5.10}$$

5.2.5 Background processes

Because of its extremely similar topology, the $t\bar{t}Z$ process is the main background in this analysis. Its contribution is estimated using simulated data, which is generated with the `MadGraph_aMC@NLO` generator at NLO in QCD with the 5FS and interfaced with `PYTHIA8` for parton shower and hadronization. Nevertheless, tWZ and $t\bar{t}Z$ are not the only processes that can decay into a final state with three (or four) leptons and a b jet in the final state. Other processes that contribute to the same final space signature have cross sections comparable, or larger, than tWZ.

tZq

The tZq process is the production of a top quark in association with a Z boson and a light quark, as shown in Figure 5.2. When the Z boson decays into a pair of leptons and the top quark decays leptonically, the final state signature is similar to the one of the tWZ process. One main feature of the tZq process is that the light jet tends to be collinear with the beam axis [116]; this information is useful in the discrimination between the two processes. The tZq process is simulated using `MadGraph_aMC@NLO` at NLO in QCD with the 4FS and `PYTHIA8`.

WZ and *ZZ* with associated jets

The associate production of two heavy vector bosons, shown in Figure 5.3, is also a significant background in this analysis, as it can decay into a final state with three or four

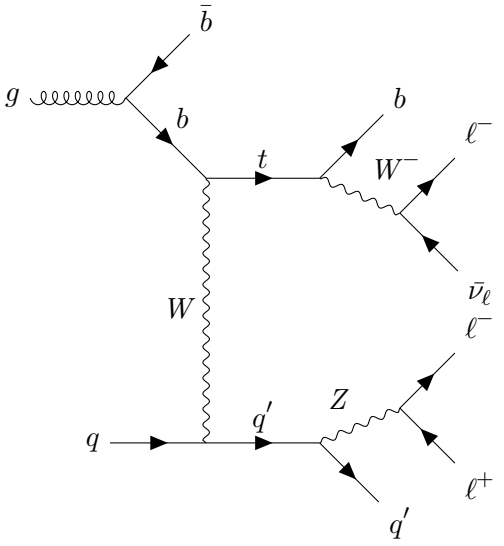


Figure 5.2: Example of a Feynman diagram for the tZq process in the 4FS.

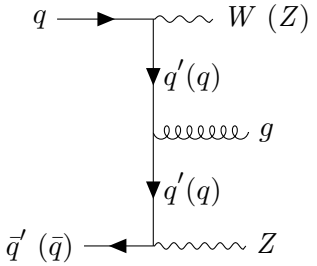


Figure 5.3: Example of a Feynman diagram for the associate production of WZ (ZZ) and hadronic jets.

leptons and be produced in association with hadronic jets via additional QCD radiation. The cross section, being in the order of tens of picobarn [128, 129], makes the contribution from erroneously b-tagged light jets significant. The ZZ production is simulated at LO level using `PYTHIA8`.

A smaller contribution is represented by the production of WZ in association with a genuine b jet. This process is simulated using `MadGraph_aMC@NLO` at NLO in QCD and `PYTHIA8`.

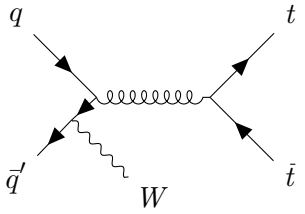
Processes with photons

Processes that produce a pair of leptons and a photon in the final state represent a contribution to the background in the analysis. The photon can convert into a pair of leptons because of the interaction with the detector, and one of the leptons can be produced so that it can pass the required selection criteria. The main processes that contribute to this kind of background are the associate production of a Z boson or of a top quark pair with a photon, simulated with `MadGraph_aMC@NLO` at NLO at LO in QCD, respectively, and interfaced with `PYTHIA8`. The conversion of photons in the detector materials is simulated with GEANT4, which was found to provide a good modeling of the process in Ref. [116].

Other background processes

Other background processes that contribute to the final state signature of the analysis are the associated production of a top quark (pair) with heavy vector bosons, such as $t\bar{t}W$, shown in Figure 5.4, which is generated at NLO in QCD with `MadGraph_aMC@NLO` and `PYTHIA8`. Other examples are the $t\bar{t}VV$ processes, with V being a W or a Z boson, which are all generated at LO in QCD with the same generators. Finally, the VVV processes are also considered, and the used simulation was generated with `MadGraph_aMC@NLO` at NLO in QCD and interfaced with `PYTHIA8`.

Figure 5.4: Example of a Feynman diagram for the associated production of a W boson and a top quark pair.



Processes containing nonprompt leptons

Despite the effectiveness of the TOP `Lepton`MVA in rejecting nonprompt leptons, residual considerable contamination remains. The main sources are represented by the Drell-Yan and $t\bar{t}$ processes, in virtue of their large cross sections. Unlike others, the contribution from this background is not inferred from simulation but is calculated from data, using the fake rate method (also known as tight-to-loose method), largely inspired by Refs. [116, 130].

The method is based on the definition of *fakable* leptons, i.e., leptons which fail the TOP `Lepton`MVA medium working point while still passing the preselection criteria (Table 5.1) and additional requirements defined in Table 5.2. Here, Electron MVA ID is the output of the MVA algorithm developed by the CMS collaboration to identify electrons [131], the associated jet is the one that shares PF candidates with the leptons, if it exists, and $p_{\mathrm{T}}^{\mathrm{ratio}}$ is the ratio between the p_{T} of the lepton and the p_{T} of the associated jet. These requirements are found to be the ones for which the DNN distributions in the Application Regions and Measurement Regions (defined below) are the most similar. For the remainder of the discussion on the fake rate method, the leptons that pass the TOP `Lepton`MVA medium working point are referred to as tight leptons in order to distinguish them from the fakeable leptons.

The following step in the modeling of the nonprompt-lepton background is the construction of Application Regions (AR) which are defined in the same way as the Signal Regions of the analysis, with the key difference that at least one lepton is required to be fakeable. The estimation of the distribution of nonprompt-lepton background in the SR is obtained through the following calculation, which is valid for phase spaces with two leptons; it is then generalized to the cases with three and four leptons, which are relevant to the tWZ analysis. Denoting as f_i the fake rate, i.e., the probability that the i-th lepton in an event

Table 5.2: Additional selection criteria for fakeable leptons.

Variable	**Data-taking era**	**Electron**	**Muon**
Electron MVA ID	All	Loose	-
DeepFlavour score of associated jet	2016 pre-VFP	< 0.15	< 0.60
	2016 post-VFP	< 0.08	< 0.60
	2017	< 0.08	< 0.60
	2018	< 0.06	< 0.50
$\frac{1}{p_{\mathrm{T}}^{\mathrm{ratio}}} - 1$	2016 pre-VFP	< 0.30	-
	2016 post-VFP	< 0.35	
	2017	< 0.30	
	2018	< 0.35	

is nonprompt but tight, and as ϵ_i the probability that the i-th lepton is prompt and tight, it is possible to relate the number of pairs of tight or fakeable lepton to the number of pairs of prompt or nonprompt leptons:

$$\begin{pmatrix} N_{tt} \\ N_{tf} \\ N_{ft} \\ N_{ff} \end{pmatrix} = \begin{pmatrix} \epsilon_1\epsilon_2 & \epsilon_1 f_2 & f_1\epsilon_2 & f_1 f_2 \\ 0 & \epsilon_1(1-f_2) & 0 & f_1(1-f_2) \\ 0 & 0 & (1-f_1)\epsilon_2 & (1-f_1)f_2 \\ 0 & 0 & 0 & (1-f_1)(1-f_2) \end{pmatrix} \begin{pmatrix} N_{pp} \\ N_{pn} \\ N_{np} \\ N_{nn} \end{pmatrix}, \tag{5.11}$$

under the assumption that $1 - \epsilon_i \approx 0$. The subscript t (f) in the i-th position on the left-hand side of Eq. (5.11) indicates that the i-th lepton is tight (fakeable). On the right-hand side, the subscript p (n) indicates that the i-th lepton is prompt (nonprompt).

The factor $\epsilon_1\epsilon_2 N_{pp}$ is taken from simulation and the number of events, in the signal regions, coming from the background of nonprompt leptons is:

$$N_{tt}^{bkg} = \epsilon_1 f_2 N_{pn} + f_1\epsilon_2 N_{np} + f_1 f_2 N_{nn} \tag{5.12}$$

The $3x3$ submatrix of the last three rows and columns of the matrix in Eq. (5.11) can be inverted analytically in order to find:

$$\begin{pmatrix} N_{pn} \\ N_{np} \\ N_{nn} \end{pmatrix} = \begin{pmatrix} \frac{1}{\epsilon_1(1-f_2)} & 0 & -\frac{f_1}{\epsilon_1(1-f_1)(1-f_2)} \\ 0 & \frac{1}{\epsilon_2(1-f_1)} & -\frac{f_2}{\epsilon_2(1-f_1)(1-f_2)} \\ 0 & 0 & \frac{1}{(1-f_1)(1-f_2)} \end{pmatrix} \begin{pmatrix} N_{tf} \\ N_{ft} \\ N_{ff} \end{pmatrix}. \tag{5.13}$$

By replacing into Eq. (5.12), it is found that the distribution of the nonprompt-lepton background in the signal regions is

$$N_{tt}^{bkg} = \frac{f_2}{1-f_2} N_{tf} + \frac{f_1}{1-f_1} N_{ft} - \frac{f_1 f_2}{(1-f_1)(1-f_2)} N_{ff}. \tag{5.14}$$

Repeating the calculation for a higher multiplicity of leptons, the number of events coming from the background of nonprompt leptons in a signal region is given by the data events in

the correspondent AR weighted by:

$$w(l_1, l_2, ..., l_n) = (-1)^n \prod_{i=1}^{n} \frac{f_i}{1 - f_i}, \tag{5.15}$$

where l_i is the i-th fakeable lepton and n is total number of fakeable leptons in the event. Since some of the fakeable leptons can be prompt, the contribution to the AR from prompt leptons is estimated by simulation and subtracted from the data events in the AR before applying the weights.

The fake rates are measured from simulation using QCD multijet samples. Events are required to have exactly one lepton, whether tight or fakeable, and at least one jet within $\Delta R < 0.7$ from the lepton. The fake rates are calculated, separately for electrons and muons in the different years of data-taking, as the ratio of the number of events with a tight lepton to the total number of events, in bins of $p_{\mathrm{T}}^{\mathrm{cone}}$ and $|\eta|$. The quantity $p_{\mathrm{T}}^{\mathrm{cone}}$ is defined as:

$$p_{\mathrm{T}}^{\mathrm{cone}} = \begin{cases} p_{\mathrm{T}}^{\mathrm{lepton}}, & \text{if lepton is tight,} \\ 0.7 \cdot p_{\mathrm{T}}^{\text{closesest jet}}, & \text{if lepton is fakeable and a jet exists within } \Delta R < 0.4, \\ 0.7 \cdot p_{\mathrm{T}}^{\mathrm{lepton}} \cdot (1 + I), & \text{if lepton is fakeable and no jet exists within } \Delta R < 0.4, \end{cases} \tag{5.16}$$

with I being the isolation of the lepton in a cone of radius $R = 0.4$. The use of the $p_{\mathrm{T}}^{\mathrm{cone}}$ variable is motivated by the fact that often nonprompt leptons are produced in the hadronization of a parton. Using the p_{T} of the closest jet instead of the lepton p_{T} improves the calculation of the fake rates by making them more stable across the possible nonprompt lepton sources. The 0.7 pre-factor in the definition of the $p_{\mathrm{T}}^{\mathrm{cone}}$ is chosen so to ensure the $p_{\mathrm{T}}^{\mathrm{cone}}$ is continuous as a function of the TOP `Lepton`MVA output.

5.3 Signal and control regions

The tWZ process is a rare process with a cross section quite smaller than the $t\bar{t}Z$ process, with which it shares an extremely similar final state. As a way to retain as many signal events as possible while maximizing the statistical power, the analysis makes use of several signal regions (SR) and control regions (CR), all of which are orthogonal to each other and fitted simultaneously in order to extract the signal strength μ_{tWZ}.

5.3.1 $\mathrm{SR}_{3\ell,3j}$

The main SR of the analysis is the $\mathrm{SR}_{3\ell,3j}$, which is designed to include all the events in which the tWZ process produces a final state with three leptons. The three jets, of which one is b-tagged, account for the hadronic decay products of either the top quark or the W boson, and the b jet coming from the top quark decay. Figure 5.5 shows an illustrative event display for an event in the $\mathrm{SR}_{3\ell,3j}$. The two green lines represent the two electrons compatible with a Z boson; the red line represents the muon coming from the decay of either the W boson or the top quark, the other branch of the same decay being

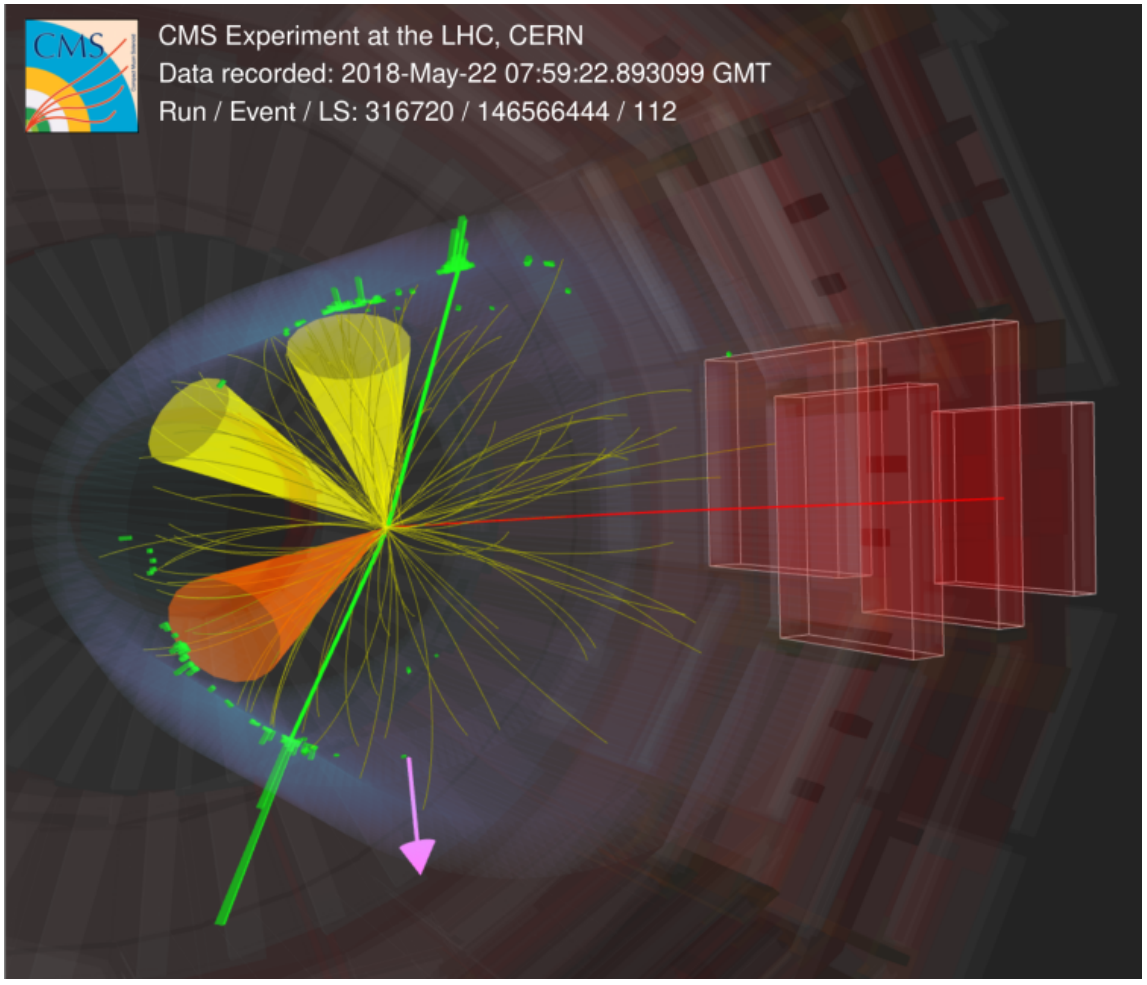


Figure 5.5: Event display for an event in the $SR_{3\ell,3j}$.

the neutrino, which is reconstructed starting from the MET, represented by the purple arrow. Finally, the three cones represent the three jets, of which the orange one is the b-tagged jet.

The $SR_{3\ell,3j}$ is defined by requiring:

- exactly three leptons with $p_T > 25$, 20, 20 GeV;
- at least one pair of leptons compatible with a Z boson:
  - the two leptons have opposite charge and same flavor (OSSF);
  - the invariant mass $m_{\ell\ell}$ of the pair respects $|m_{\ell\ell} - m_Z| < 15$ GeV;
  - if two pairs respect the aforementioned criteria, the pair with the invariant mass closest to m_Z is selected;
- at least three selected AK4 jets;
- at least one AK4 jet is b-tagged.
- the event are not selected in the $SR^{Boosted}$, defined in Section 5.3.1.

In this region, a fully connected feed-forward DNN multiclass classifier (multiclassifier) is used to discriminate between the signal and the background processes. It provides three outputs which represent the probability for the analyzed event to belong to the tWZ process, the $t\bar{t}Z$ process, or any other background process. The event shown in Figure 5.5 has been chosen for display in virtue of its high probability to be a tWZ event, as predicted by the multiclassifier.

The multiclass classifier, together with all the other DNNs designed for this analysis, is realized using `Keras` [132] interfaced with `TensorFlow` [133].

The variables used as input to this DNN are listed in Table 5.3.

Variable	**Description**
$p_{T,2\ell}$	p_{T} of the sub-leading lepton
$p_{T,3\ell}$	p_{T} of the trailing lepton
$\eta_{3\ell}$	η of the trailing lepton
max$\Delta R_{\ell\ell}$	Maximum ΔR separation among lepton pair combinations
maxΔR_{jj}	Maximum ΔR separation among jet pair combinations
njet	Jet multiplicity
nbjet	b jet multiplicity
maxη_j	Maximum η among the jets (useful for the recoiling jet in tZq process)
maxm_{jj}	Maximum invariant mass among jet pair combinations
m_{sys}	Invariant mass of the system composed of leptons, jets, and MET
$p_{\mathrm{T}\ell j}^{\mathrm{max}}$	Maximum p_{T} among any combination of a lepton and a jet
η_{2j}	η of the sub-leading jet
η_{3j}	η of the trailing jet
m_t	Invariant mass of the reconstructed top quark
$m_{\mathrm{had}\ W}$	Invariant mass of the reconstructed W boson decaying into hadrons
$p_{T,t}$	p_{T} of the reconstructed top quark
$p_{T,\mathrm{lep}\ W}$	p_{T} of the reconstructed W boson decaying into leptons
$p_{T,\mathrm{had}\ W}$	p_{T} of the reconstructed W boson decaying into hadrons
$\eta_{\mathrm{had}\ W}$	η of the reconstructed W boson decaying into hadrons
$\eta_{\mathrm{lep}\ W}$	η of the reconstructed W boson decaying into leptons
$\phi_{\mathrm{had}\ W}$	ϕ of the reconstructed W boson decaying into hadrons
$\phi_{\mathrm{lep}\ W}$	ϕ of the reconstructed W boson decaying into leptons

Table 5.3: List of the variables used as input features to the multiclass classifier in the $\mathrm{SR}_{3\ell,3j}$, and their descriptions.

Figure 5.6 shows the distribution of the $p_{\mathrm{T}\ell j}^{\mathrm{max}}$ (left) and m_{sys} (right) input variables in the $\mathrm{SR}_{3\ell,3j}$.

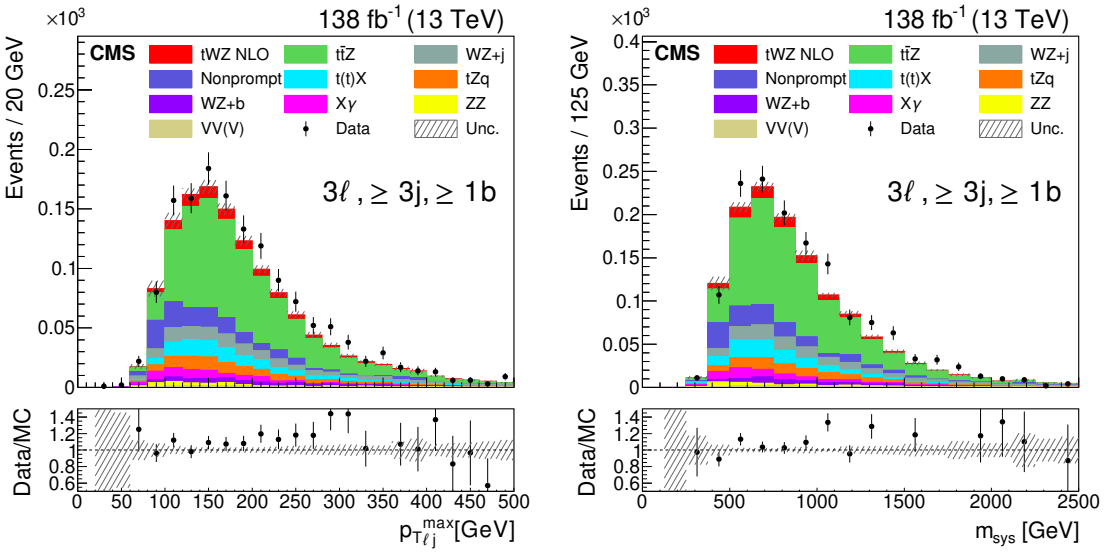


Figure 5.6: Pre-fit distribution of the $p_{\mathrm{T}\ell j}^{\mathrm{max}}$ (left) and m_{sys} (right) observables in the $\mathrm{SR}_{3\ell,3j}$. The lower insets show the ratio between data and simulation, underlining an overall good agreement between the two within uncertainties.

The multiclassifier DNN is composed of two hidden layers of 100 and 64 neurons, respectively, both using a SELU activation function. The output layer is composed of three neurons, each representing the probability for the event to belong to the tWZ, $t\bar{t}Z$, or other background processes, using a softmax activation function. This activation function, described in Section 3.2.2 in the context of the transformer model, is the generalization of the sigmoid function to multiple dimensions, and its outputs are constructed as to sum up to one, therefore representing the probability that the considered data point belongs to each of the classes. The layers are interposed with batch normalization layers and droupout layers. The formers transform their inputs so that they become distributed according to a normal distribution with $\mu = 0$, and $\sigma = 1$ [134]. The dropout layers randomly set 10% of the input units to zero at each update during training, in order to prevent overtraining [135], i.e., fine-tuning the weights of the DNN on the training sample in such a way that it is unable to generalize the learned classification to unseen data

1.5 million simulated events are used for the training and validation of the multiclassifier, composed of 44% tWZ events, 36% $t\bar{t}Z$ events, and 20% other backgrounds. Two equivalent versions of the multiclassifier are trained, each one using one half of this sample, which is furthermore divided into a training sample (80%) and a validation sample (20%).

In the analysis workflow, each multiclassifier acts solely on the half of simulated events that it has not been trained on, so as to avoid any bias. Regarding data, each event is assigned the output of either one of the two multiclassifiers, chosen randomly with equal probability.

As discussed in Section 5.2.5, the identification of two b jets is a strong indicator that the considered event could belong to the $t\bar{t}Z$ process. Therefore, the $\mathrm{SR}_{3\ell,3j}$ is further divided into two regions: the $\mathrm{SR}_{3\ell,3j,1b}$, where the events have exactly one b-tagged jet, and the $\mathrm{SR}_{3\ell,3j,2b}$, where the events have at least two b-tagged jets. The $\mathrm{SR}_{3\ell,3j,1b}$ is purer in tWZ

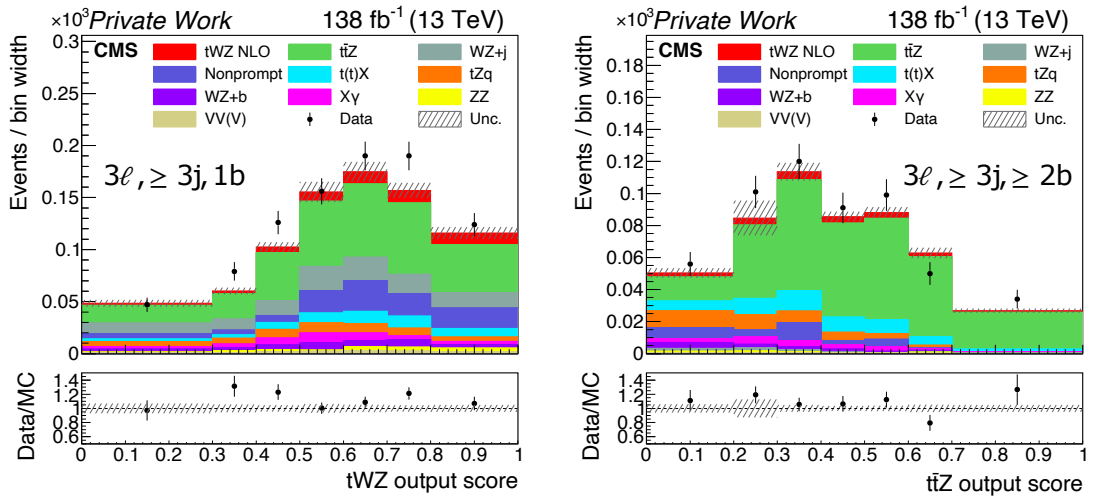


Figure 5.7: Pre-fit distribution of the tWZ output of the multiclass classifier in the $\mathrm{SR}_{3\ell,3j,1b}$ (left) and $t\bar{t}Z$ output in the $\mathrm{SR}_{3\ell,3j,2b}$ (right). The lower insets show the ratio between data and simulation. The low yield of tWZ events in the $\mathrm{SR}_{3\ell,3j,2b}$ makes the region analogous to a control region for the $t\bar{t}Z$ process, thus making it useful to use the $t\bar{t}Z$ output of the multiclassifier as a template in the fit.

events, while the $\mathrm{SR}_{3\ell,3j,2b}$ has a very large contribution from the $t\bar{t}Z$ process and is useful to constrain the $t\bar{t}Z$ background in the fit. Figure 5.7 shows the pre-fit distribution of the tWZ ($t\bar{t}Z$) output of the multiclass classifier in the $\mathrm{SR}_{3\ell,3j,1b}$ ($\mathrm{SR}_{3\ell,3j,2b}$), on the left (right) side. These are the stacks of signal and background histograms which are fitted to data points in these regions.

$\mathrm{SR}^{\mathrm{Boosted}}$

In addition to the division in $\mathrm{SR}_{3\ell,3j,1b}$ and $\mathrm{SR}_{3\ell,3j,2b}$, the $\mathrm{SR}_{3\ell,3j}$ further includes a $\mathrm{SR}^{\mathrm{Boosted}}$, which is defined with the intention of selecting events in which the top quark is produced with high energy. Both decays into jets and leptons are considered, with different requirements.

As mentioned in Section 5.1 interest in the definition of this region is led by the fact that the high-energy spectrum of the tWZ process might be more sensitive to the presence of new physics, compared to lower energies.

Events selected in the $\mathrm{SR}_{3\ell,3j}$ are first checked against the presence of a boosted top quark decaying hadronically, reconstructed as an AK8 jet with the following requirements:

- $p_{\mathrm{T}} > 300$ GeV;
- $105 < m_{SD} < 210$ GeV, where m_{SD} is the soft drop mass of the jet with $\beta = 0$ and $z_{\mathrm{cut}} = 0.1$ [136];
- a b-tagged jet is present within $\Delta R < 0.8$ from the AK8 jet.

The m_{SD} selection is motivated by the property of the soft drop mass to act as a tagger for the top quark when its value is compatible with the mass of the top quark.

Events in which a boosted top quark decaying hadronically is not found are then considered for the presence of a boosted top quark decaying leptonically, identified starting from the lepton which is not compatible with the Z boson. A boosted leptonic top quark is found if:

- the lepton has $p_{\mathrm{T}} > 30$ GeV;
- the lepton is within $\Delta R < 2$ from a b-tagged jet;
- the b-tagged jet has $p_{\mathrm{T}} > 200$ GeV;
- the lepton-jet pair is classified as signal from a purpose-designed DNN, described in Section 5.4.

The selection on $DeltaR$ is justified by the fact that the decay product of a boosted top quark tend to be collimated; $\Delta R > 2$ indicates that either the leptons and b quark do not originate from a boosted top quark, or they are not the decay product of the same particle. The selection threshold for the p_{T} of the b quark acts as a proxy for the p_{T} of the top quark from which it originates. The event yield is used as the input to the fit to data for the $\mathrm{SR}^{\mathrm{Boosted}}$, in view of the low number of selected events: five signal and 81 background events.

5.3.2 $\mathrm{SR}_{3\ell,2j}$

In some events, one of the jets produced in the final state of the tWZ process can fall outside the acceptance of the detector and, therefore, not be reconstructed. With the intention of retaining as many signal events as possible, the $\mathrm{SR}_{3\ell,2j}$ is defined to select events that only feature two jets. The requirements for the $\mathrm{SR}_{3\ell,2j}$ are:

- same selection criteria for the leptons as in the $\mathrm{SR}_{3\ell,3j}$;
- exactly two selected AK4 jets;
- at least one AK4 jet is b-tagged.

The peculiarity of this region is that the $t\bar{t}Z$ process is expected to have a very low yield because of the strict requirement of two jets, two less than expected from $t\bar{t}Z$ events in the three-lepton final state. In fact, in this region, events with nonprompt leptons, the WZ + (b) jets process, and the tZq process are all expected to have comparable yields, greater than the $t\bar{t}Z$ process.

In order to discriminate the contribution of the signal process to this region, a fully connected feed-forward DNN binary classifier is used to discriminate between the tWZ process and all the other backgrounds. The total sample used for the training and validation of this classifier is composed of 950 thousand simulated events, with 40% of tWZ and 60% of background events. The decision to use a binary classifier in place of a multiclass algorithm resides in the multitude of relevant background processes. Not enough events are available separately for each class in order to train two multiclassifiers effectively, according to the same training scheme described in Section 5.3.1.

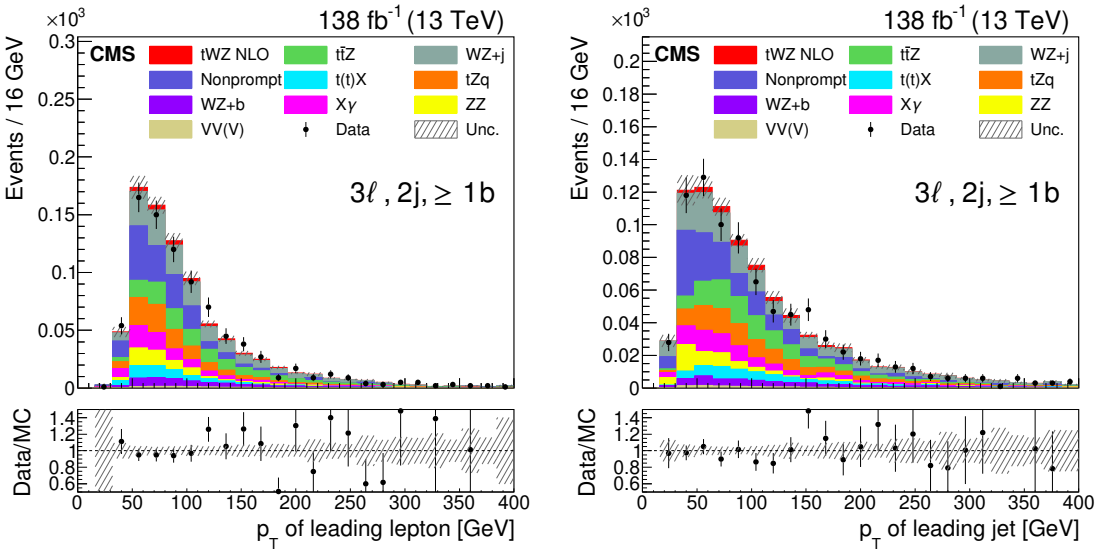
The input features to the binary classifier are listed in Table 5.4.

Variable	Description
$p_{T,1\ell}$	p_{T} of the leading lepton
$p_{T,2\ell}$	p_{T} of the sub-leading lepton
$p_{T,3\ell}$	p_{T} of the trailing lepton
$\eta_{3\ell}$	η of the trailing lepton
nbjet	b jet multiplicity
maxm_{jj}	Maximum invariant mass among jet pair combinations
max$p_{T,jj}$	Maximum transverse momentum among sum of two jets
maxη_j	Maximum η among the jets (useful for the recoiling jet in tZq process)
$p_{T,1j}$	p_{T} of the leading jet
$p_{T,2j}$	p_{T} of the sub-leading jet
η_{1j}	η of the leading jet
η_{2j}	η of the sub-leading jet
m_{sys}	Invariant mass of the system composed of leptons, jets, and MET
$p_{T,\mathrm{sys}}$	Transverse momentum of the system composed of leptons, jets, and MET
$p_{\mathrm{T}\ell j}^{\mathrm{max}}$	Maximum p_{T} among any combination of a lepton and a jet

Table 5.4: List of the variables used as input features to the binary classifier in the SR$_{3\ell,2j}$, and their descriptions.

Figure 5.8 shows the distribution of the transverse momentum of the leading lepton (left) and jet (right), which are two of the input variables to the binary classifier in the SR$_{3\ell,2j}$.

Figure 5.8: Pre-fit distribution of the transverse momentum of the leading lepton (left) and jet (right) in the SR$_{3\ell,2j}$. The lower insets show the ratio between data and simulation, underlining an overall good agreement between the two within uncertainties.



The DNN is composed of two hidden layers of 75 and 25 neurons, respectively, both using a SELU activation function. The two layers are interposed with batch normalization layers and dropout layers with a 10% dropout rate. The output layer is made up of a single neuron, which outputs the probability that the event belongs to the tWZ process after

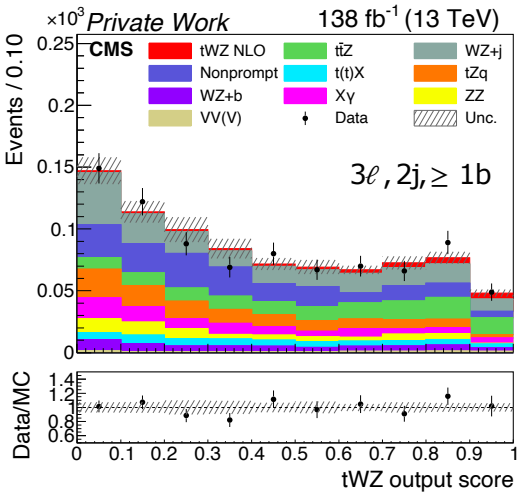


Figure 5.9: Pre-fit distribution of the output of the DNN in the $SR_{3\ell,2j}$. The lower inset shows the ratio between data and simulation.

the application of a sigmoid activation function. The training and validation samples are divided in the same way as for the multiclassifier in the $SR_{3\ell,3j}$, with the same training scheme.

The pre-fit output of the DNN in the $SR_{3\ell,2j}$ is shown in Figure 5.9. This simulated distribution is fitted to data points in the maximum likelihood fit.

5.3.3 $SR_{4\ell}$

The last signal region of the analysis is the $SR_{4\ell}$, whose aim is the selection of events in which the tWZ process produces a final state with four leptons. The total event yields in this region are expected to be very low, as the branching ratio for the four-lepton decay is one fourth of the three-lepton decay, for both the tWZ and $t\bar{t}Z$ processes. An important background to this region is the ZZ process, where both Z bosons decay into leptons.

The selection criteria for the $SR_{4\ell}$ are:

- exactly four leptons, with $p_T > 25$, 20, 20, 10 GeV;
- one pair of leptons compatible with a Z boson, according to the same criteria as in the $SR_{3\ell,3j}$;
- the remaining pair is not compatible with a Z boson;
- at least one b-tagged jet.

The multiplicity distribution of b-tagged jets is chosen to be fitted to data in this region, as it provides a good degree of separation between the tWZ and $t\bar{t}Z$ process in this low-yield region. The pre-fit distribution of the b-tagged jet multiplicity in the $SR_{4\ell}$ is shown in Figure 5.10.

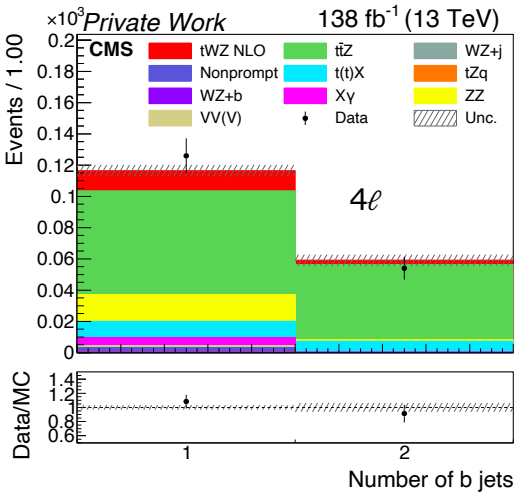


Figure 5.10: Pre-fit distribution of the b-tagged jet multiplicity in the $SR_{4\ell}$. The lower inset shows the ratio between data and simulation.

5.3.4 Control regions

Two control regions, CR_{WZ} and CR_{ZZ}, are defined in the analysis and fitted simultaneously with the signal regions, in order to provide constraints on the normalization of the associated production of WZ and light jets, and of the ZZ process.

The CR_{WZ} is defined by requiring:

- exactly three leptons, with $p_{\mathrm{T}} > 25$, 20, 20 GeV;
- at least one pair of leptons compatible with a Z boson;
- the $p_{\mathrm{T}}^{\mathrm{miss}}$ is greater than 50 GeV;
- no b-tagged jet is present.

In these criteria, the veto on b jets is introduced in order to reduce the contribution from the tWZ and, most importantly, the $t\bar{t}Z$ processes. This causes the CR_{WZ} not to provide a constraint on the associate production of WZ and b jets. Nevertheless, this is a minor background, as discussed in Section 5.2.5.

The CR_{ZZ} is defined by requiring:

- exactly four leptons, with $p_{\mathrm{T}} > 25$, 20, 20, 10 GeV;
- two pairs of leptons compatible with originating from Z bosons.

Both control regions enter the maximum likelihood fit through their event yields.

The expected yield in all the signal (control) regions, for all the processes, together with the observed number of events in data is shown in Table 5.5 (Table 5.6).

5.4 DNN for boosted leptonic top tagging

A *boosted leptonic top-tagger* has been designed as a fully connected feed-forward DNN to recognize high-p_{T} (boosted) top quarks to define the leptonic boosted region.

Table 5.5: Expected yields for all simulated processes and observed number of data events in the signal regions. The uncertainties are related to the limited size of the simulated samples, except for the nonprompt-lepton contribution, which is extracted from data and assigned a 30% uncertainty.

	$SR_{3\ell,3j}$	$SR_{3\ell,2j}$	$SR_{4\ell}$	$SR^{Boosted}$
tWZ	77.47 ± 0.12	28.19 ± 0.07	15.98 ± 0.06	5.44 ± 0.02
$t\bar{t}Z$	657.9 ± 1.6	122.76 ± 0.61	113.86 ± 0.64	59.03 ± 0.50
Nonprompt leptons	139 ± 42	170 ± 51	1.02 ± 0.31	1.94 ± 0.58
tZq	86.45 ± 0.78	108.69 ± 0.71	0.29 ± 0.04	4.37 ± 0.17
ZZ	22.7 ± 2.4	60.6 ± 4.1	20.0 ± 2.3	0.30 ± 0.29
WZ	166.4 ± 3.3	227.8 ± 4.0	0.59 ± 0.19	6.84 ± 0.66
$VV(V)$	15.51 ± 0.11	10.55 ± 0.09	1.35 ± 0.03	0.64 ± 0.02
$t(\bar{t})X$	108.30 ± 0.99	49.4 ± 1.2	17.32 ± 0.34	6.26 ± 0.19
$X\gamma$	54.1 ± 2.6	78.3 ± 3.7	6.92 ± 0.95	1.08 ± 0.31
Total backgrounds	1249 ± 42	822 ± 51	159.9 ± 2.6	80.8 ± 1.1
Data	1463	849	180	77

The input to the neural network is a top quark candidate, consisting of the pair of a b-tagged jet and a charged lepton within an angular distance $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 2$ from each other. Furthermore, so as to consider boosted top quarks, the b jet is required to have $p_{\mathrm{T}} > 200$ GeV, while the lepton needs to have $p_{\mathrm{T}} > 30$ GeV. The output of the network is a value between 0 and 1, which represents the likelihood that the top quark candidate represents a genuine boosted top quark.

The 14 input features to the DNN are:

- p_{T}, η, and ϕ of the b jet;
- p_{T}, η, ϕ, and invariant mass of the top quark candidate;
- DeepFlavour b-tagging score of the b jet;
- mini-isolation $\mathrm{I_{mini}}$;
- $p_{T,rel} = |\vec{p}_\ell \times \vec{p}_b|/|\vec{p}_b|$;
- p_{T} and η of the b jet in the top quark candidate rest frame;
- p_{T} and η of the lepton in the top quark candidate rest frame,

with the top quark candidate consisting of the sum of the four-momenta of the b jet ($\vec{p}_b$) and the lepton ($\vec{p}_\ell$).

The DNN consists of:

- **Batch normalization layer**: normalizes the input of the DNN to be distributed according to a normal distribution centered in zero and with unitary standard deviation;
- **Two dense layers**: two hidden layers of 15 fully connected neurons with SiLU as activation function;

Table 5.6: Expected yields for all simulated processes and observed number of data events in the control regions. The uncertainties are related to the limited size of the simulated samples, except for the nonprompt-lepton contribution, which is extracted from data and assigned a 30% uncertainty.

	CR_{WZ}	CR_{ZZ}
tWZ	31.96 ± 0.08	2.39 ± 0.02
$t\bar{t}Z$	112.41 ± 0.73	14.44 ± 0.23
Nonprompt leptons	1450 ± 430	23.0 ± 7.0
tZq	74.84 ± 0.67	0.05 ± 0.01
ZZ	597 ± 12	2202 ± 22
WZ	10610 ± 25	0.68 ± 0.16
$VV(V)$	166 ± 14	16.52 ± 0.07
$t(\bar{t})X$	39.4 ± 1.8	1.08 ± 0.07
$X\gamma$	519 ± 11	2.53 ± 0.60
Total backgrounds	13520 ± 430	2028 ± 23
Data	12743	2352

- **Gaussian Dropout layer** [137]: multiplies every input by a random noise value drawn from a normal distribution with $\mu = 1$ and $\sigma = 0.3$;
- **Two dense layers**: two more hidden layers, also with 15 fully connected neurons and SiLU as activation function;
- **Output layer**: provides the network output after the application of a sigmoid activation function.

The hyperparameters of the DNN, e.g., the number of layers and neurons, and the standard deviation for the Gaussian Dropout layer have been chosen by performing a hyperparameter grid search, i.e., several possible combinations of hyperparameter values have been tested, and the one yielding the best validation loss, defined below, are chosen.

Two different DNNs with the same architecture have been trained: one for which the lepton considered is an electron, and the second one for muons. The training sample for the electron (muon) DNN consists of 1.846 (2.115) M top quark candidates, selected from simulated events of $t\bar{t}$ production in CMS in the ℓ + jets decay channel. In both cases, half of the training sample consists of genuine top quark candidates (signal), while the other half is made of background top quark candidates. These include the cases in which the b jet does not arise from the primary interaction vertex, or the b jet and the lepton do not originate from the decay of the same top quark.

The training has been performed using the Adam gradient descent algorithm to minimize a binary cross-entropy loss function. In order to avoid overtraining the DNNs, the trend of the loss function during training has also been observed when calculated on a statistically independent validation sample (validation loss), consisting of 273 thousand top quark candidates for the electron DNN and 330 thousand top quark candidates for the muon DNN.

The training is stopped when the validation loss has not decreased for 30 epochs, and the selected trained model is the one at the epoch with the lowest validation loss.

Figure 5.11 shows the trend of the training and validation losses for the electron and muon DNNs.

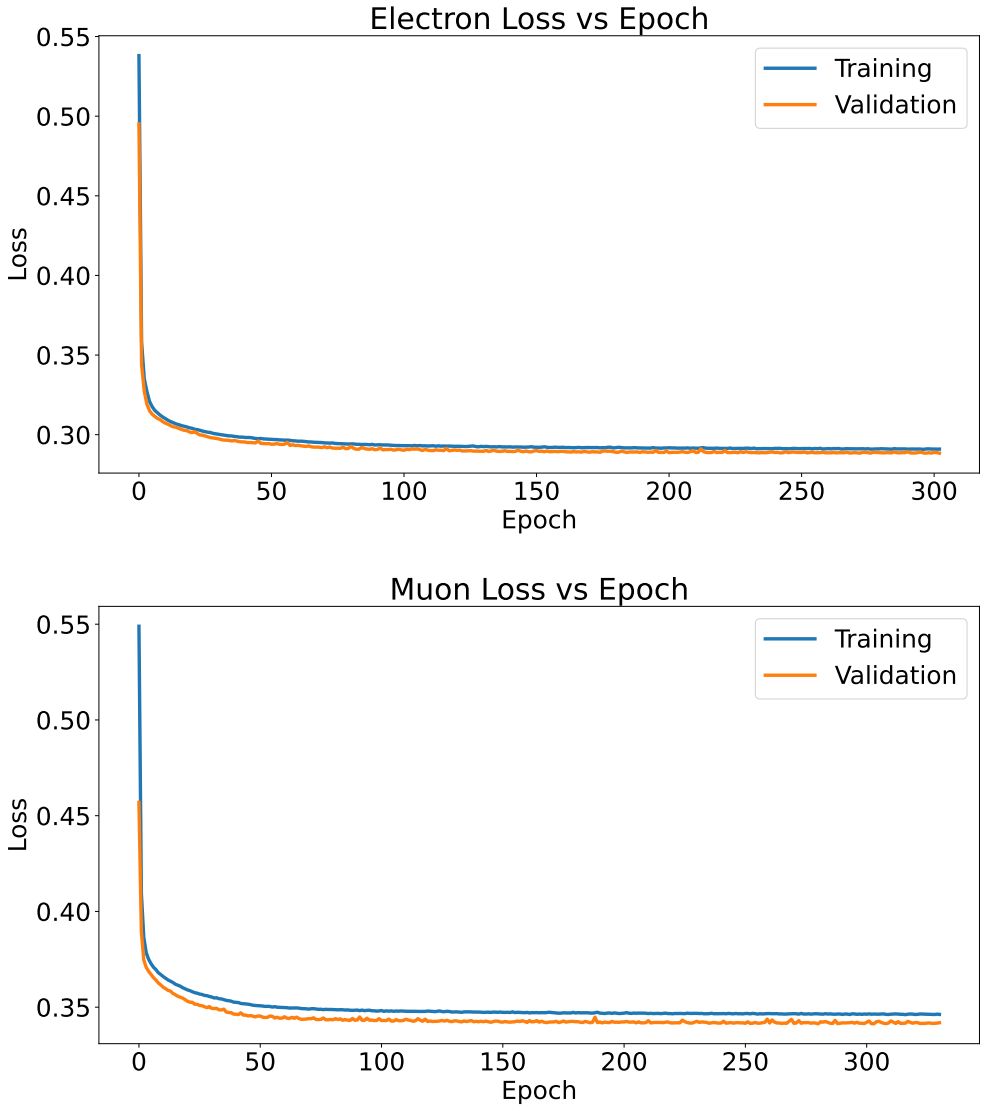


Figure 5.11: Plot of the training and validation losses against epoch of training for electron (upper) and muon (lower) leptonic top-tagger. The training is interrupted after the validation loss has not decreased for 30 epochs, and the selected weights are the ones corresponding to the lowest validation loss.

The output distributions of the DNNs are shown in Figure 5.12. Both cases show the background accumulating towards 0 and the signal towards 1 and a clear distinction between the two distributions. Furthermore, the plot compares the distribution of the outputs for the training and the validation samples, which are sampled from the same distribution. This hypothesis has been tested through Kolmogorov-Smirnov (KS) [138] tests between the distributions. Both the signal and background distributions for muons and

electrons returned a p-value greater than 5% for the compatibility of training and validation distributions. This suggests that the algorithm is able to generalize the classification capabilities to data that have not been used for training it.

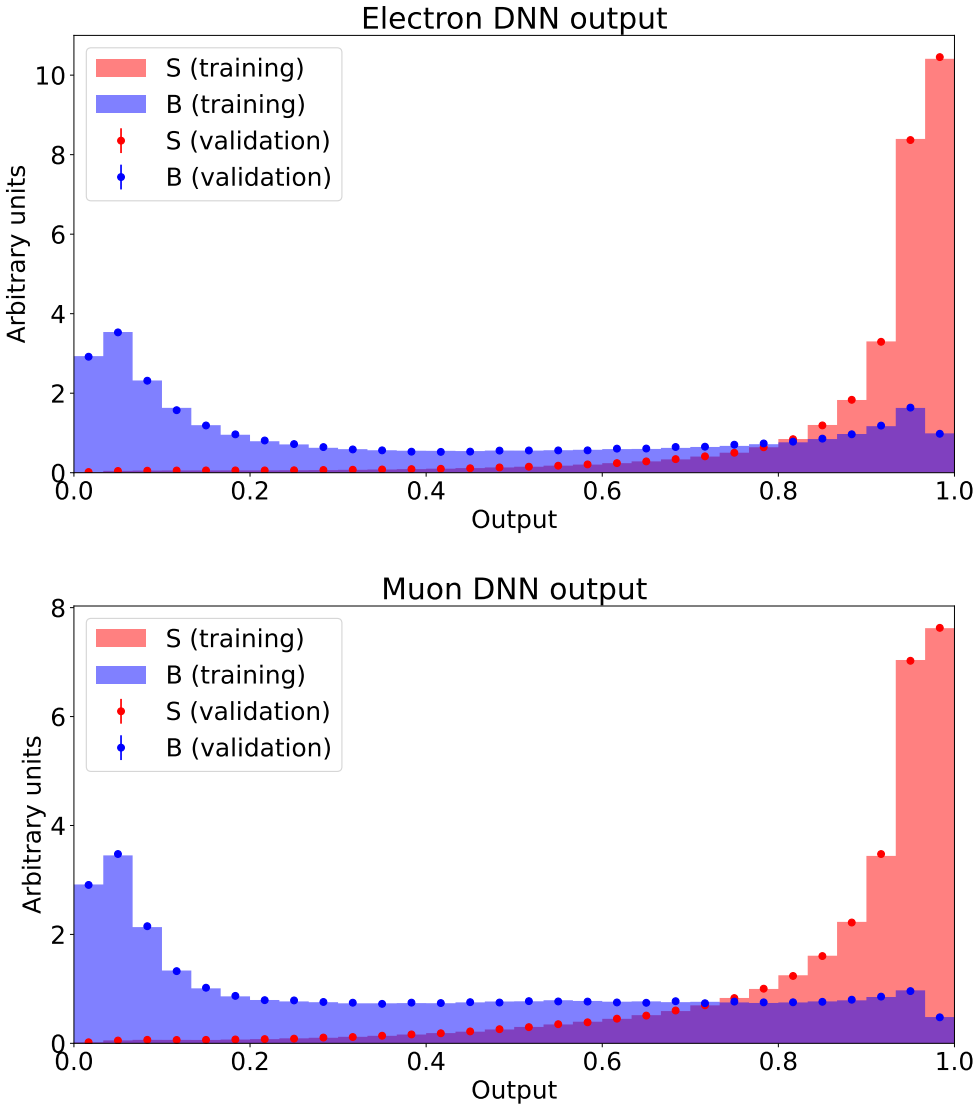


Figure 5.12: Output of the top-tagger for electrons (upper) and muons (lower). In both cases, the background and the signal are clearly discriminated, and the outputs of the validation and training samples follow the same distribution, i.e., the p-value of the KS test between the two is greater than 5%.

In order to quantify how well the classification task is carried out, it is common to use a Receiver Operating characteristic (ROC) curve [139], which shows how the true positive rate (TPR) changes as a function of the false positive rate (FPR). Considering a top quark candidate selected when the output of the classifier is greater than a certain threshold, or working point, the TPR (FPR) is the ratio of the selected signal (background) top quark candidates out of the total signal (background) candidates. Both TPR and FPR are calculated using the validation sample.

The ROC curves for the electron and muon top-taggers are displayed in Figure 5.13, where they are also compared with the curve that would be obtained by an algorithm that selects top quark candidates randomly. The area under the ROC curve (auc) is often used to summarize the classification power of an ML algorithm, with values that range from 0.5 for a random choice to 1 for a perfect classifier. The results of almost 0.9 obtained for the boosted leptonic top-taggers show excellent discriminating power for both leptons.

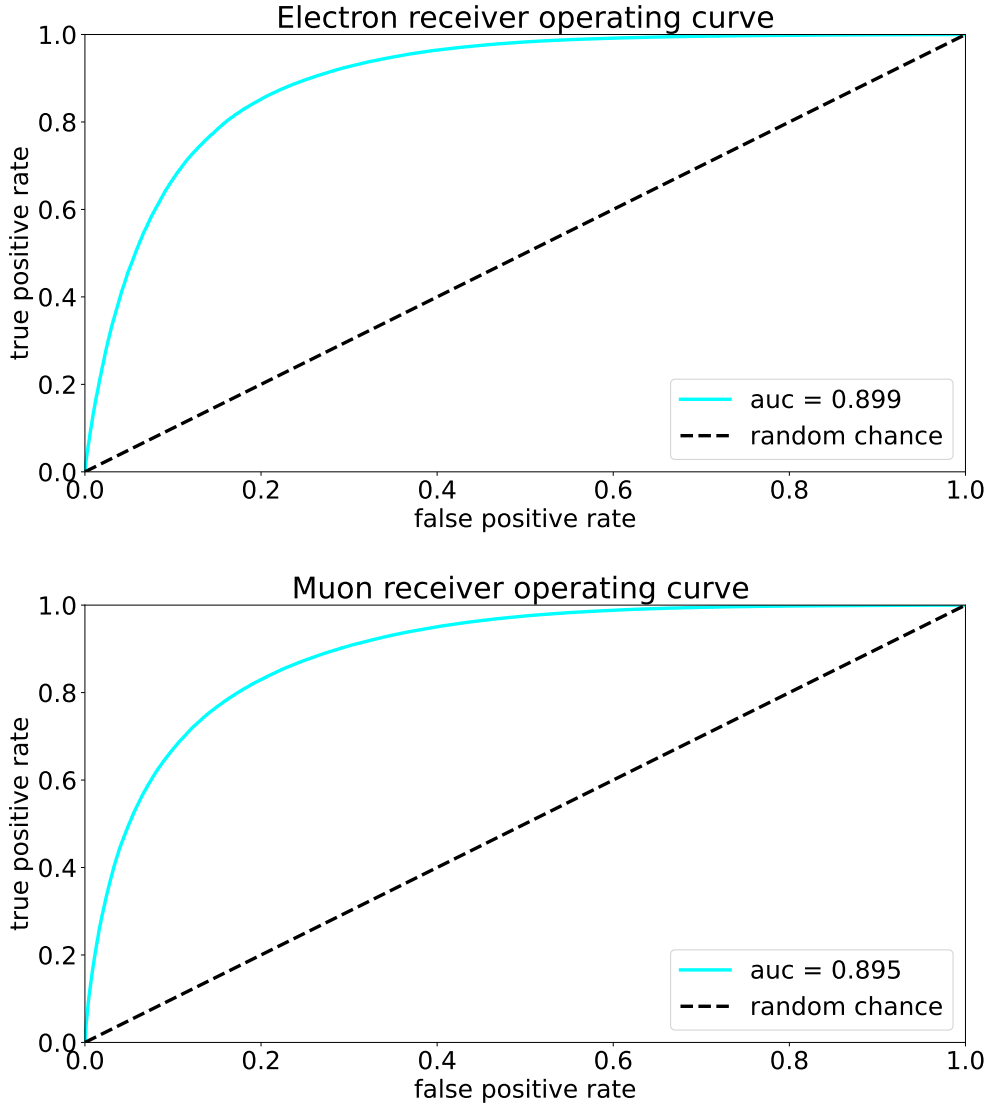


Figure 5.13: ROC curves for the electron (upper) and muon (lower) top-taggers. The diagonal dashed curve represents the ROC curve for an algorithm randomly selecting data. The area under the ROC curve is close to 90% for both classifiers, showing excellent discriminating power.

The following necessary step before being able to use the DNN for the classification task is the choice of a working point (WP), which has been performed by maximizing the product of the TPR and the precision. Precision is defined as the fraction of the selected top quark candidates that are signal, determined from simulation data.

The distribution of this metric with respect to the selection threshold is shown in Figure 5.14. The chosen WP, which maximizes the metric in both cases, is 0.4.

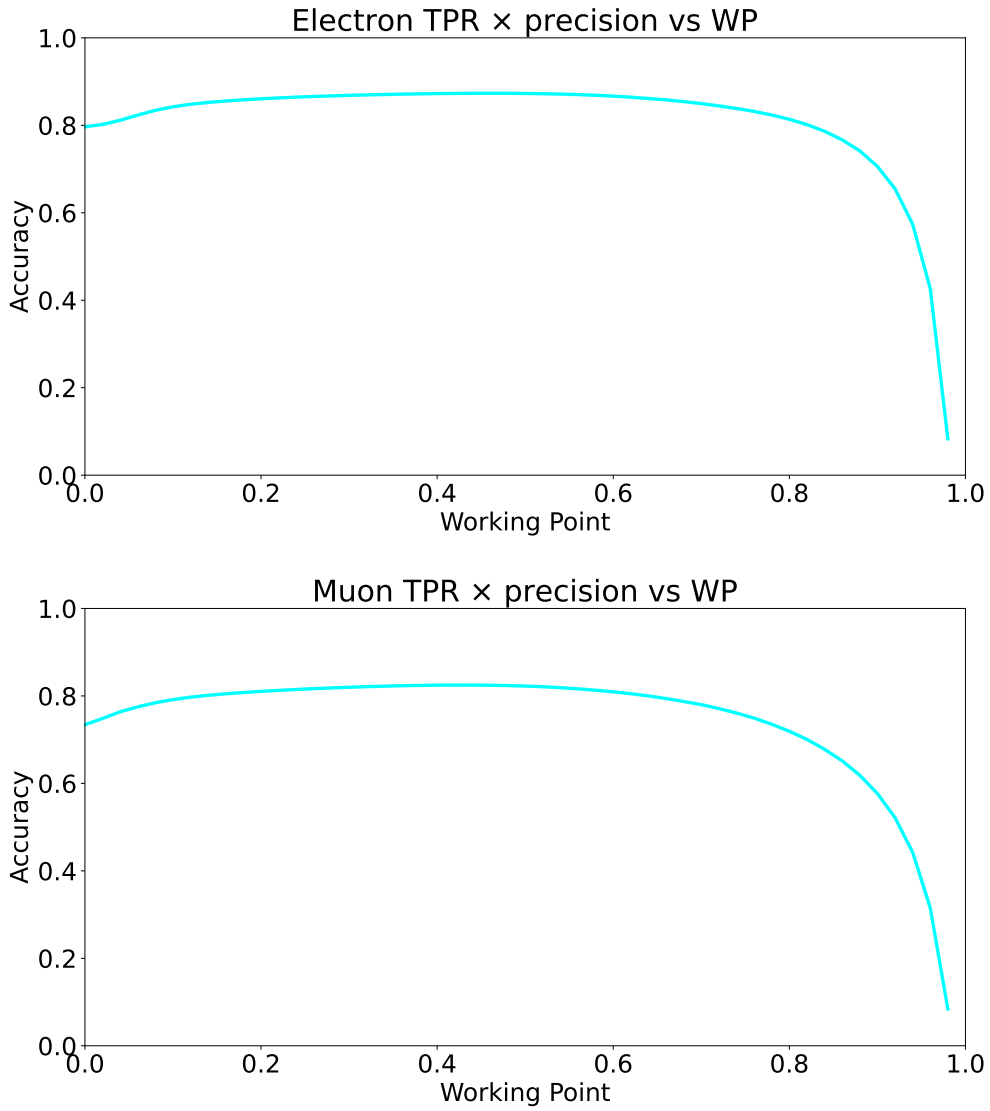


Figure 5.14: Distribution of the product of TPR and precision with respect to the selection threshold for the electron (upper) and muon (lower) DNNs. The threshold that maximizes the metric is 0.4 in both cases, and it is chosen as WP for the classifiers.

5.4.1 Validation of the taggers

Before using the classifiers in the workflow of the search for the tWZ process, the performance of the algorithms needs to be evaluated on real data insensitive to tWZ in order to verify possible discrepancies between the behavior of the simulation used for training and the observed data.

A validation region has been defined by targeting the selection of a phase space rich in events of production of $t\bar{t}$, in the decay channel with two charged leptons. The selection criteria for the validation region are:

- exactly two muons, or two electrons, with opposite charge and $p_{\mathrm{T}} > 25$ GeV;
- invariant mass of the leptons $m_{\ell\ell}$ outside a 15 GeV window around the Z boson mass: $|m_{\ell\ell} - m_Z| > 15$ GeV;
- at least 2 hadronic jets with $p_{\mathrm{T}} > 30$ GeV;
- exactly 1 boosted top quark candidate.

Figure 5.15 shows how, in the validation region, the output of the DNN compares for data and simulation for electrons and muons; the simulation reproduces the data particularly well, especially for the muon DNN. The same does not hold for the electron DNN, which presents disagreements between data and simulation that reach more than 20% in the lowest two bins.

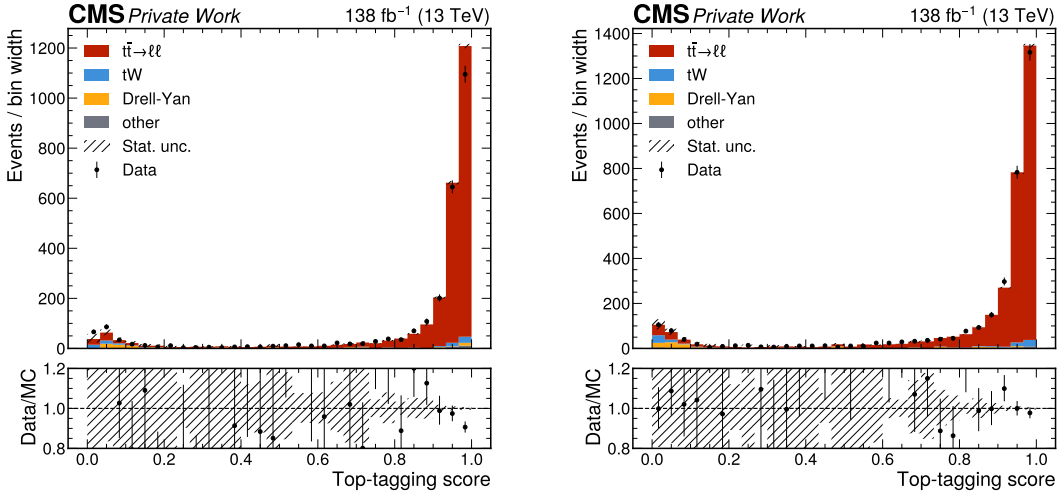


Figure 5.15: Output of the boosted leptonic top-tagger for electrons (left) and muons (right) in a region enriched in dileptonic $t\bar{t}$ events. The lower insets of both plots show the ratio between the observed data and the total MC.

The shape of the output is not directly used in the analysis but only the binary classification between signal and background. Therefore, the disagreement between data and simulation in the electron DNN does not translate to high disagreement in the selection efficiency at the selected WP of 0.4. This is shown in Figure 5.16, which displays the selection efficiency as a function of the selection threshold for electrons and muons. The selection efficiency at the WP is fully consistent between data and MC for the muon DNN, and for electrons, the disagreement is almost negligible, as it is less than 2%.

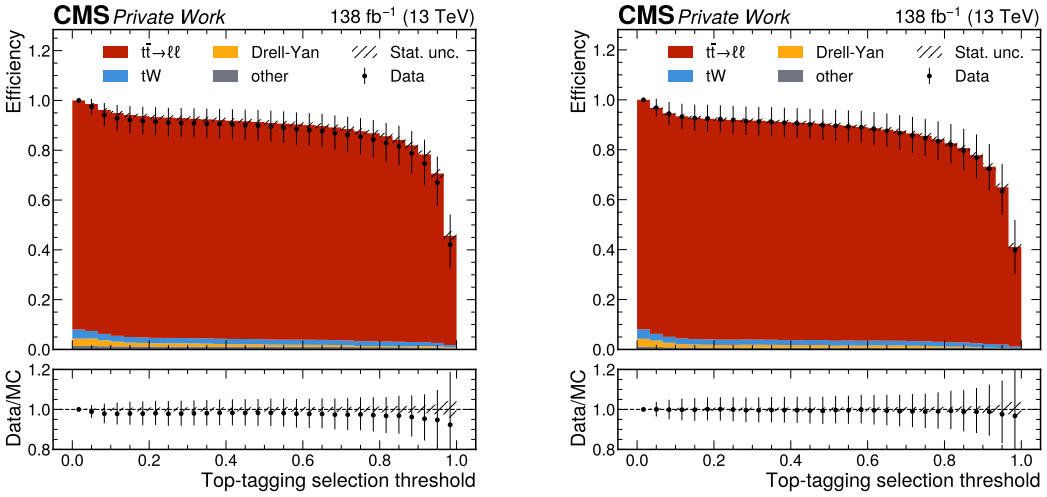


Figure 5.16: Selection efficiency of the boosted leptonic top-tagger for electrons (left) and muons (right) as a function of the selection threshold in a region enriched in dileptonic $t\bar{t}$ events. The lower insets show the ratio between the efficiency in data and MC. At the chosen WP of 0.4, the behavior of data and MC is very consistent for muons, and for electrons, the disagreement is less than 2%.

The benefits arising from this validation region does not end at the selection efficiency, as the $t\bar{t}$ process also allows the validation of the modeling of the input features of the DNNs. Figure 5.17 shows the comparison between data and MC for the $p_{T,\mathrm{rel}}$ variable. Both the electron and muon distributions show that the simulation is able to model the observed data in a fairly consistent manner. The distributions for all the variables are shown in Appendix B.

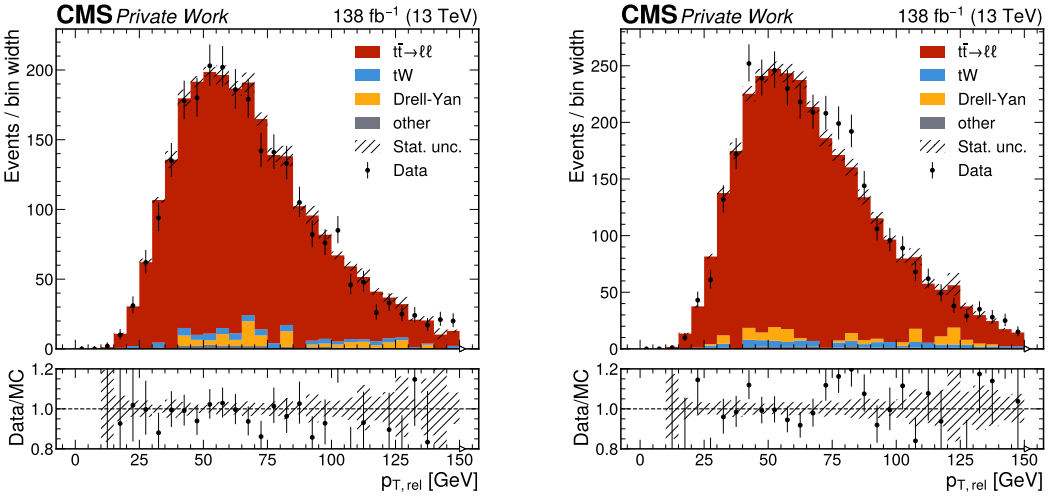


Figure 5.17: Distribution of the $p_{T,\mathrm{rel}}$ variable for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC. For both lepton flavors, the simulation is able to model the data faithfully.

5.5 Corrections and systematic uncertainties

Several corrections are applied to the events and physics objects in the analysis, in order to account for both experimental and theoretical effects, leading to known discrepancies between simulation and data. In the maximum likelihood fit, systematic uncertainties are mapped to nuisance parameters. Correlated systematic effects result in only one nuisance parameter in the likelihood function, while uncorrelated systematic uncertainties are treated as separate nuisance parameters. The definition of the likelihood function and the fit methodology are described in Section 3.1.

5.5.1 Experimental corrections

Pileup

Event scale factors are applied to correct the distribution of pileup interactions in the simulation to match the one observed in data. This number is obtained from the instantaneous luminosity collected by the CMS detector L_i and the total proton-proton inelastic cross section σ_{inel} at $\sqrt{s} = 13$ TeV according to:

$$\mu = \frac{L_i \sigma_{\text{inel}}}{f}, \tag{5.17}$$

where f is the LHC revolution frequency of 11246 Hz. The value used for σ_{inel} is 69.2±3.2 mb, and its 4.6% uncertainty is propagated to the event scale factors, and it is treated as a systematic uncertainty correlated across all the years of data taking.

Luminosity

The uncertainty in the integrated luminosity collected by the CMS experiment is translated into a normalization uncertainty in total event yield in each year of data taking. The value of the uncertainties range between 1.2% and 2.5% and are divided into correlated and uncorrelated components, depending on the different uncertainty sources [93, 140, 141].

Trigger efficiency

As mentioned in Section 5.2.2, the trigger efficiency is measured in data to be compatible with unity. Therefore, no scale factor is applied to the simulated events, but a 2% normalization uncertainty is assigned to the event yields in order to account for the limited size of the data sample used for the efficiency measurement. Since the trigger paths used in the analysis are different in the four periods of data taking, the uncertainty is treated as uncorrelated across the years.

Lepton identification

The lepton selection efficiency is not consistent between data and simulation. Scale factors are applied to the simulated events to correct for this discrepancy. The efficiencies have been measured in data and simulation using the tag-and-probe method [142] in a phase space targeting the selection of Drell-Yan events decaying into two leptons [117]. The

lepton scale factors are obtained as the ratio of the efficiencies in data and simulation in bins of p_{T} and absolute value of the pseudorapidity (pseudorapidity of the supercluster) for muons (electrons). Two sources of uncertainties are considered for the scale factors: the statistical uncertainty arising from the limited size of the data and simulation samples used for the efficiency measurement, and the uncertainty in the modeling of the processes considered.

In this analysis, the simulated events are assigned a weight equal to the product of the scale factors of all the selected leptons. The uncertainties of the scale factors are obtained by varying them within one standard deviation of their uncertainties. The statistical uncertainties are treated as uncorrelated across the years, while the uncertainties in the modeling are treated as correlated.

Jet energy scale and resolution corrections

As mentioned in Section 5.2.3, the jets in the simulation are corrected to match the energy scale and resolution of the jets in data. The relative uncertainties in the corrections are propagated to the jets, and the uncertainty histograms for the maximum likelihood fit are obtained by recomputing the templates with the modified objects. When applying the jet energy scale uncertainties, the MET is also recomputed according to the new values of the energy of the jets.

Missing transverse energy

The MET depends on both the energy scale of all jets and the energy of unclustered objects. While the jet energy scale uncertainties are already considered in the analysis, one additional systematic uncertainty is considered by varying the energy of the unclustered objects by their resolution, and recalculating the MET accordingly.

B tagging efficiency

As for the leptons, the b tagging efficiency is not consistent between data and simulation. Scale factors have been calculated in bins of p_{T}, η, and flavor of the jets (b jet, c jet or light jet) as the ratio of the efficiencies in data and simulation in a region enriched in QCD multi-jet event [143, 144]. A weight is applied on an event-by-event basis to the simulated events:

$$w = \prod_{i=\text{tagged jets}} \mathrm{SF}_i \prod_{j=\text{not tagged jets}} \frac{1 - \mathrm{SF}_j \epsilon_j}{1 - \epsilon_j}, \tag{5.18}$$

where the product is over all the jets in the event, and ϵ is the efficiency of the b tagging algorithm in the phase space used in the analysis, also binned in p_{T}, η, and flavor of the jets. The uncertainties in the scale factors are considered according to their individual sources and the templates for them, to be used in the maximum likelihood fit, are obtained by reanalyizing the events after applying the weight modified by their uncertainties.

Level-1 trigger prefiring

In 2016 and 2017, the Level-1 ECAL trigger system of the CMS detector had an issue that caused some trigger decisions to be applied to the previous bunch crossing [145]. The CMS trigger does not allow two consecutive bunch crossings to fire the same trigger, therefore the events affected by the issue are not recorded. This effect, which is not present in simulation, is corrected by applying a scale factor to the simulated event

$$w = 1 - p(\eta, p_{\mathrm{T}}), \tag{5.19}$$

where p is the probability of the event to be affected by the prefiring issue; η and p_{T} are the pseudorapidity and the transverse momentum of the object that fires the trigger. A similar effect is also present in the L1 muon trigger system, for all the years of data taking, due to the time resolution of the system. The uncertainties in the scale factors are obtained by varying the probability of the event to be affected by the prefiring issue by 20% and are correlated across the years.

HEM15/16 issue

During the 2018 data taking, the CMS detector experienced a failure of the Hadronic Calorimeter in the region with $-3 < \eta < -1.3$ and $-1.57 < \phi < -0.87$. This affected the reconstruction of jets and the measurement of the MET. This effect is not present in the simulation, and it is accounted for by the implementation of an η-dependent systematic uncertainty of between 20 and 35% for the p_{T} of the jets in the affected region.

5.5.2 Theoretical corrections

Normalization uncertainties

Uncertainties in the cross sections of the simulated processes are considered as normalization uncertainties in the templates used in the maximum likelihood fit. A 15% uncertainty is assigned to the $t\bar{t}Z$ process, in order to account for the discrepancy between the measured cross section and the theoretical prediction [146, 147]. Similarly, a 10% uncertainty is assigned to processes with a photon in the final state [148].

Other uncertainties are assigned in order to reflect the experimental precision in the measurement of the inclusive cross sections: 11% is assigned to the tZq process [116], 10% to the di- and tri-boson processes [149–151]. The WZ production in association with b jets is assigned an uncertainty of 20%, coherently with what was observed in Ref. [116].

Finally, the nonprompt lepton background is assigned a conservative 30% bin-by-bin uncertainty, in order to account for the different origins of the nonprompt leptons in the regions used for the measurement of the fake rates and for the analysis.

PDF uncertainties

As discussed in Chapter 1, the PDF set used for the simulation of the samples is the NNPDF3.1. The set provides 103 alternative distributions, obtained with the hessian

method [152]. The uncertainty in the PDF set is translated to the maximum likelihood fit by recomputing the templates for the signal and background processes with the alternative PDF sets. The envelopes of the up- and down-variation, normalized to the same yield as the nominal distribution, are used as templates for the systematic uncertainty.

Separately, the effect of the uncertainty in the value of the strong coupling constant α_S is considered by obtaining alternative distributions by varying the value of α_S by factors of 1/2 and 2, and normalizing the resulting distributions to the same yield as the nominal one.

Parton shower modeling

In order to account for the uncertainty in the parton showering in simulation, alternative distributions are obtained by varying the value of α_S used in the simulation by factors of 1/2 and 2, and normalizing the resulting distributions to the same yield as the nominal one. This is done separately for initial state and final state radiation. In view of the different production mechanisms of EWK-induced (as tWZ and tZq) and QCD-induced ($t\bar{t}Z$) processes, the uncertainties are treated as uncorrelated across the two, for the ISR uncertainty. The FSR uncertainty is treated as correlated across all processes.

Factorization and renormalization scales

Finally, uncertainties in the QCD factorization and renormalization scales, μ_F and μ_R, are considered by varying them by a factor of 2 and 1/2, separately, and recomputing the templates for the signal and background processes. Similarly to is what done for α_S for the ISR PS modeling, the variations are considered uncorrelated between the EWK-induced and QCD-induced processes.

Signal modeling

As mentioned in Section 5.2.1, the signal is modeled using the DR1 prediction. The shape difference between the DR1 and DR2 templates is considered as a systematic uncertainty in the signal modeling in the maximum likelihood fit. Since only one alternative is considered, the difference between DR1 and DR2 is the regarded as variation by one standard deviation in the positive direction for the associated nuisance parameter, while the nominal distribution also acts as variation by one standard deviation in the negative direction.

5.6 Results

The result of the maximum likelihood fit provides a value for the signal strength of the tWZ process

$$\mu_{tWZ} = 2.6 \pm 0.4 \text{ (stat)} \pm 0.7 \text{ (syst)}, \tag{5.20}$$

with an observed statistical significance over the background-only hypothesis of 3.5 standard deviations. This is the first search for the tWZ process and provides the first evidence

of its existence. Furthermore, the result is consistent with the SM prediction within 2 standard deviations.

The observed significance is larger than the one expected through a fit of the simulation to pseudo-data, obtained as the sum of all the simulated processes, which is 1.4 standard deviations. This result is due to the high value measured for the signal strength of the process.

Figure 5.18 shows the post-fit distribution of the templates used in the maximum likelihood fit, for the $\mathrm{SR}_{3\ell,3j,1b}$ (upper left), $\mathrm{SR}_{3\ell,3j,2b}$ (upper right), $\mathrm{SR}_{3\ell,2j}$ (lower left), and $\mathrm{SR}_{4\ell}$ (lower right).

It is interesting to consider the impact of the systematic uncertainties on the signal strength. The impact of the j-th nuisance on the parameter of interest is defined as $|\hat{\mu} - \hat{\mu}(\theta_j \pm \Delta\theta_j)|$, where $\hat{\mu}(\nu_j + \sigma_j)$ is the value of μ that maximizes the likelihood when the j-th nuisance is varied by its uncertainty. The uncertainties in the signal strength, and of the nuisance parameters, are obtained through the procedure described in Section 3.1.1.

Figure 5.19 shows the ten largest impacts of the nuisance parameters.

The most impactful systematic uncertainty is the one related to the normalization of the $t\bar{t}Z$ process, which has an impact of the order of 0.5 on the signal strength. This explains how the systematic uncertainty is the dominant component of the total uncertainty in the signal strength, despite the rather low production cross section of the tWZ process. Due to the extreme similarities between the $t\bar{t}Z$ and tWZ processes, even after using ML discriminants, small differences in the normalization of former process have a large impact on the latter. Furthermore, the deviation of $\sim 0.5\ \Delta\theta_{t\bar{t}Z}$, with $\Delta\theta_{t\bar{t}Z}$ being the pre-fit uncertainty, shows how, inherently, a measurement of the tWZ process also results in a measurement of the $t\bar{t}Z$ process, once again because of the processes are tightly related to each other.

In order to further investigate the correlation between the measurement of the tWZ and $t\bar{t}Z$ processes, and to avoid the bias introduced by the treatment of the normalization of the $t\bar{t}Z$ process as a systematic uncertainty with a prior, a second fit is performed. This time, both the signal strength of the tWZ process μ_{tWZ} and of the $t\bar{t}Z$ process $\mu_{t\bar{t}Z}$ are used as parameters of interest.

The likelihood scan is performed in the two-dimensional space of the two signal strengths, and the result is shown in Figure 5.20. The signal strengths μ_{tWZ} and $\mu_{t\bar{t}Z}$ are found to be strongly anti-correlated as expected, with a correlation coefficient $\rho_{tWZ,t\bar{t}\gamma} = -0.63$, and the significance of the tWZ process is found to be greater than 3 standard deviations in this fit as well.

A previous measurement of the cross section of the $t\bar{t}Z$ process in CMS [146] found a value for the signal strength of the $t\bar{t}Z$ process of $\mu_{t\bar{t}Z} = 1.13 \pm 0.10$. The main difference with Ref [146] is that, in the search for the tWZ process, data samples simulated at NLO in QCD are used for the description of the tWZ process.

As mentioned before, this result represents the first evidence for the tWZ process.

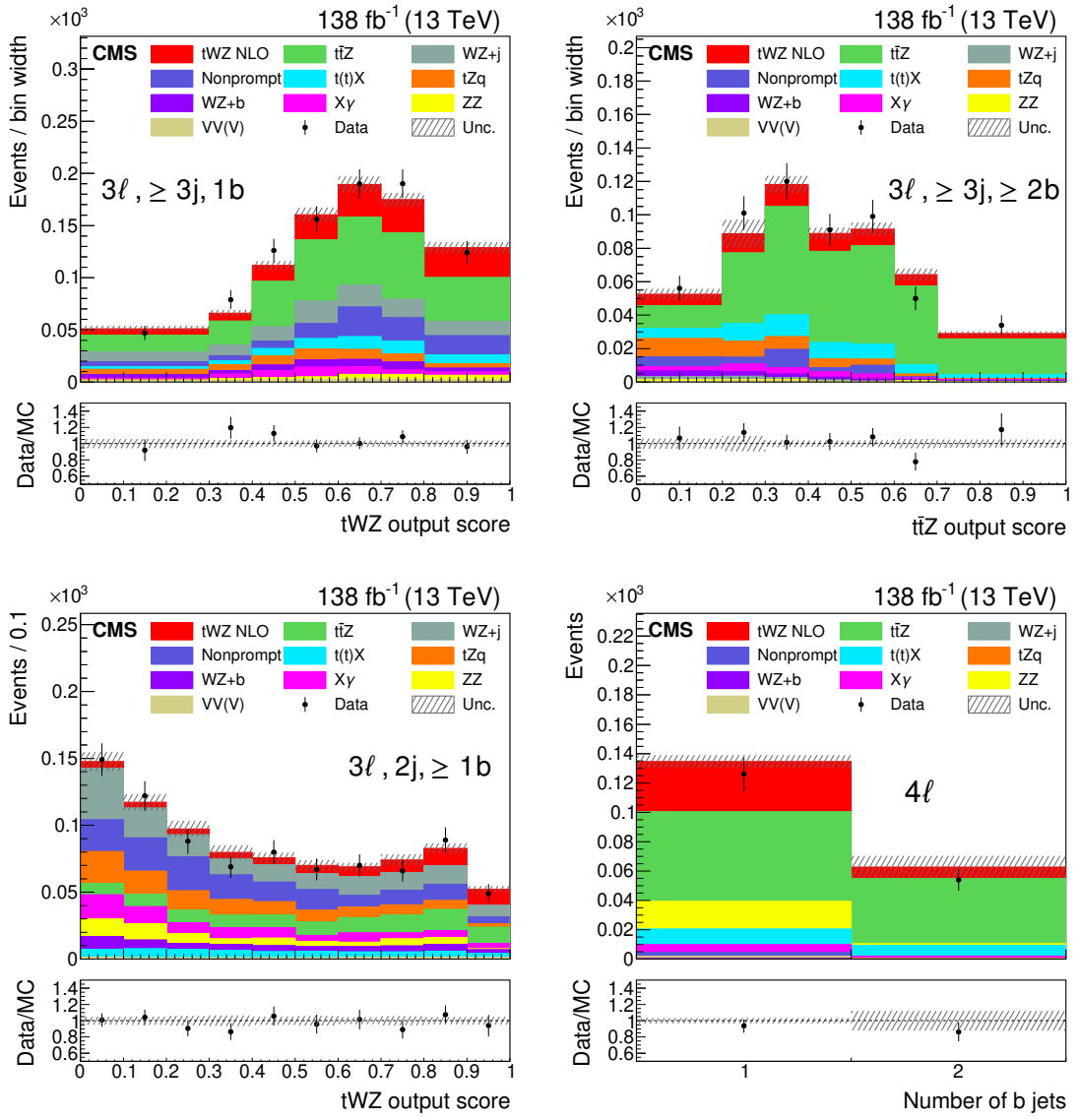


Figure 5.18: Post-fit distribution of the discriminating variables used in the maximum likelihood fit for the $SR_{3\ell,3j,1b}$ (upper left), $SR_{3\ell,3j,2b}$ (upper right), $SR_{3\ell,2j}$ (lower left), and $SR_{4\ell}$ (lower right). The lower insets show the ratio between the observed data and the simulation, which display good agreement between the two, and reduced uncertainties, when compared to the pre-fit distributions (Figures 5.7, 5.9, and 5.10).

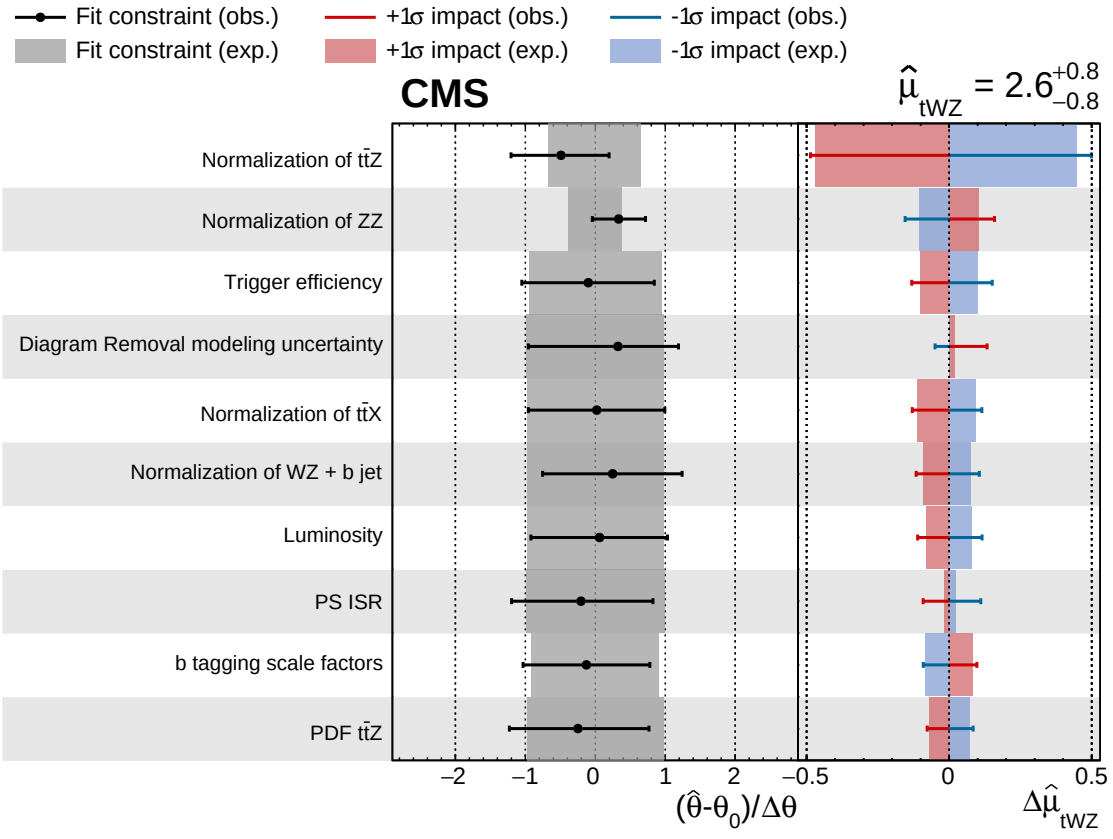


Figure 5.19: Impacts of the ten dominant nuisance parameters on the signal strength of the tWZ process. The central column shows the defined as the difference between the best-fit value and the nominal value, in units of pre-fit uncertainty, for the nuisance parameters. It furthermore shows the observed (expected) post-fit uncertainties as black bars (dark gray band). The right column shows the impact of the nuisance parameters on the signal strength, for both the observed fit (blue and red bars) and the expected fit (blue and red bands). The blue (red) color represents the variation of the nuisance parameter by $+1\sigma$ (-1σ).

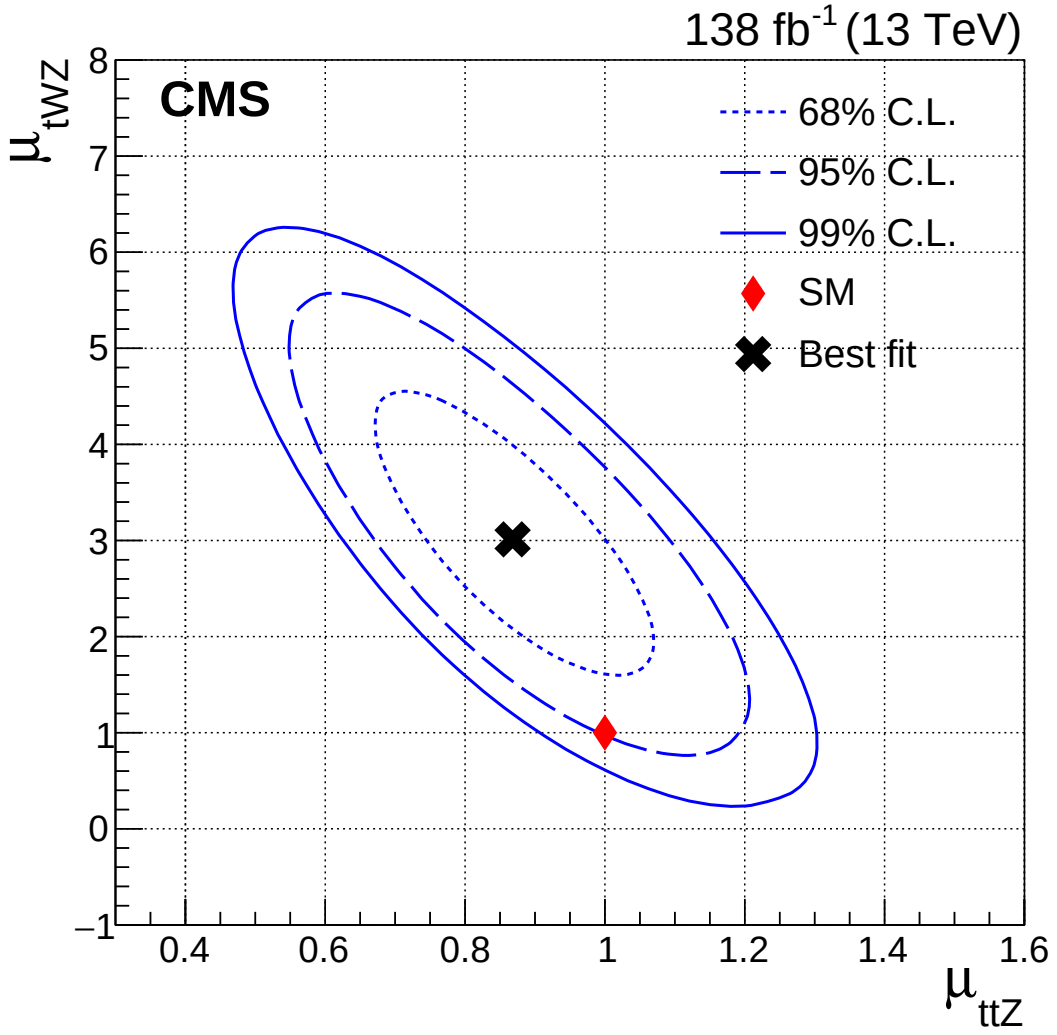


Figure 5.20: Two-dimensional likelihood scan of the signal strength of the tWZ and $t\bar{t}Z$ processes. The plot shows the 68, 95, and 99% confidence level contours, and the best-fit point, which is consistent with the SM prediction within 2σ.

6 Search for the $tW\gamma$ process

This chapter describes the first search, and the first evidence, for the SM production of the $tW\gamma$ process. An introduction to the analysis and its motivations is given in Section 6.1. Section 6.2 provides a full overview of the workflow of this search, outlining the simulation of the signal and background samples, the trigger selection, and the reconstruction and identification of the physics objects. Section 6.3 describes the definition of the signal and control regions in which the analysis is performed. In the signal regions, a GNN-based binary classifier is used to discriminate between the signal and the $t\bar{t}\gamma$ background; the design, training, performance, and validation of this classifier are described in Section 6.4. Section 6.5 details the systematic uncertainties that are considered in the maximum likelihood fit used to extract the signal strength of the $tW\gamma$ process. Finally, the results of the search are presented in Section 6.6.

6.1 Introduction

The associated production of a top quark with a W boson and a photon is a rare single top production process, similar to the tW and tWZ processes. As mentioned in Section 2.3, the photon can be radiated from any internal or external lines of the tW process (*hard process*) (HP), or from the decay products of the top quark or the W boson (*decay*). Figure 6.1 shows illustrative Feynman diagrams for the $tW\gamma$ process in the final state with two charged leptons and a photon coming from the hard process (left) or the decay (right).

As for the case of the tWZ process, the $tW\gamma$ process also has been found to be sensitive to the presence of physics beyond the Standard Model in the context of the EFT framework [113].

The main obstacle to an experimental search for the $tW\gamma$ process is the discrimination of the signal from the $t\bar{t}\gamma$ background, which has a significantly higher cross section and a very similar final state topology. The major difference between the two is the presence of an extra b jet in the final state of the $t\bar{t}\gamma$ process, arising from the decay of the second top quark in the event. For most events, this jet is not reconstructed and/or identified, and the two processes overlap significantly in phase space.

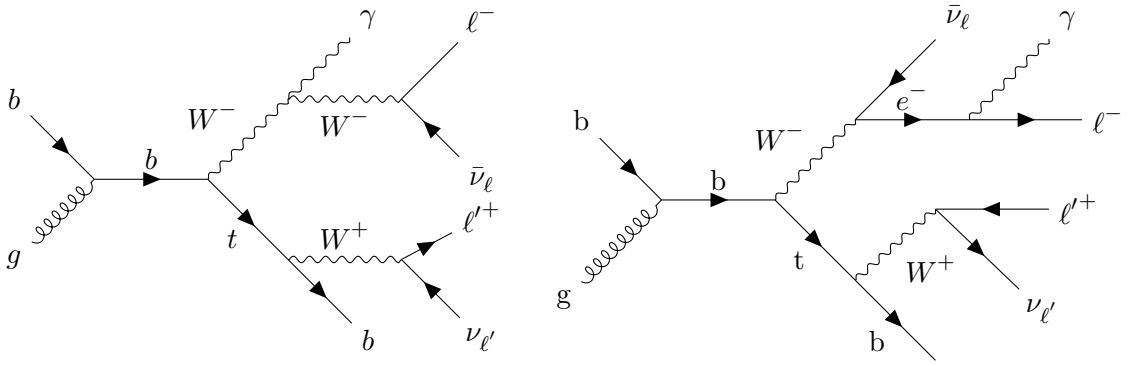


Figure 6.1: Illustrative Feynman diagrams for the $tW\gamma$ process in the final state with two charged leptons and a photon coming from the hard process (left) and decay (right).

The CMS collaboration has published an inclusive and differential measurement of the $t\bar{t}\gamma$ process [148], in which the cross section was found to be 175.2 ± 2.5 (stat) ± 6.3 (syst) fb in the fiducial phase space defined by the presence of two charged leptons, a photon, and at least one b-tagged jet. The measurement was performed using data collected at a center-of-mass energy of 13 TeVin the years 2016, 2017, and 2018, corresponding to an integrated luminosity of 138 fb^{-1}.

The same dataset is the one used in the search for the $tW\gamma$ process performed in this thesis. The considered final state is the one in which both the top quark and the W boson decay with a charged lepton in the final state. The analysis makes use of a GNN-based binary classifier to discriminate between the $tW\gamma$ signal and the $t\bar{t}\gamma$ background in the signal regions. Two control regions and two signal regions are defined in the analysis and the simulated GNN classifier distributions, obtained in these regions, are fitted to data through a maximum likelihood fit, in order to extract the signal strength $\mu_{tW\gamma}$ and the statistical significance of the $tW\gamma$ signal.

6.2 Analysis overview

This section provides an overview of the procedures used for the simulation of signal and background samples, the trigger selection and the reconstruction and identification of physics objects.

6.2.1 Simulation of the $tW\gamma$ and $t\bar{t}\gamma$ samples

The signal process is simulated at NLO in QCD following the recipes outlined in Section 2.3. Two samples are produced: one simulating events in which the photon is emitted from the hard process and a second one in which the photon is radiated from any decay product of either top quark or the W boson, both using the DR1 scheme for overlap removal. Alternative samples are produced using the DR2 scheme in order to estimate a systematic uncertainty in the modeling of the signal. All the signal samples are produced using `MadGraph5_aMC@NLO`.

A similar spirit is followed for the $t\bar{t}\gamma$ process. A dedicated $t\bar{t}\gamma$ sample is produced at NLO in QCD with `MadGraph5_aMC@NLO` for the case in which the photon is radiated from the hard process. The contribution from photons from the decay products of the top quarks is obtained from an inclusive $t\bar{t}$ sample. Events from the inclusive sample that do not include a generated photon are used for the modeling of nonprompt photons. The $t\bar{t}$ sample is produced using the `POWHEG` generator [153–156].

All the $tW\gamma$ and $t\bar{t}\gamma$ samples are interfaced with `PYTHIA8` for parton showering and hadronization.

The phase space considered is defined by requiring:

- exactly two charged leptons with opposite charge and $p_{\mathrm{T}} > 15\ GeV$;
- the invariant mass of the two leptons to be $m_{\ell\ell} > 30$ GeV;
- at least one photon with $p_{\mathrm{T}} > 15$ GeV, $|\eta| < 1.5$, and no particle with $p_{\mathrm{T}} > 5$ GeV in a cone of $\Delta R = 0.4$ around the photon direction.

These choices are dictated by the intention of defining a phase space with a slightly larger acceptance than the phase space at the reconstruction level, defined in the next paragraphs. In this way, it is possible to properly consider events whose simulated properties are lower than the selection thresholds, but that pass the reconstruction selection in virtue of the imperfect resolution of the detector.

The simulated cross section for the $tW\gamma$ process is $\sigma_{tW\gamma} = 23.0$ fb, considering both possible photon origins. Additionally, the simulated cross sections for the $t\bar{t}\gamma$ process are $\sigma_{t\bar{t}\gamma,HP} = 67.5$ fb for photons from the hard process, and $\sigma_{t\bar{t}\gamma,D} = 194.0$ fb for photons from the decay products of the top quarks.

6.2.2 Trigger selection

The events in this analysis are selected using a combination of single- and double-lepton trigger paths (muons or electrons). The single-lepton trigger paths employed allow for the selection of events with leptons with p_{T} as low as 24 GeV, for muons, and 23 GeV, for electrons. The selection thresholds are lower for di-lepton trigger paths. As mentioned in Section 5.2.2, the different selection thresholds are a result of the changes to the trigger menu during Run 2 of the LHC.

In order to account for the different trigger efficiencies in data and simulation, scale factors are applied to the simulated events.

6.2.3 Object reconstruction and identification

The analysis makes use of electrons, muons, AK4 jets and loose jets, and photons. The following paragraphs describe the reconstruction and identification criteria for these physics objects.

Table 6.1: Identification criteria for photons.

Variable	**Criterion**
H/E	< 0.02197
$\sigma_{i\eta i\eta}$	< 0.01015
PF charged hadron isolation	< 0.02197
PF neutral hadron isolation	$< 1.189 + 0.01512 \cdot p_{\mathrm{T}}(\gamma) + 2.259 \cdot p_{\mathrm{T}}(\gamma)^2 \cdot 10^{-5}$
PF photon isolation	$< 2.08 + 0.004017 \cdot p_{\mathrm{T}}(\gamma)$

Photons

The photons employed in this analysis are prompt photons, produced by radiation from any internal or external line in the hard process, or from the immediate decay products of the top quarks or W bosons.

In proton-proton collision events, photons can also be produced during the hadronization of partons, for instance in the decays of neutral pions. Another source of reconstructed photons are electrons that are misidentified as photons. These photons are referred to as nonprompt photons and tend to be surrounded by hadronic PF candidates. Additionally, photons produced in the interaction of particles with the detector material tend to deposit a large part of their energy in the HCAL, as opposed to prompt photons which deposit most of their energy in the ECAL.

In order to reduce their contribution to the selected events, the photon candidates are required to pass the identification criteria defined by the CMS collaboration, described in Table 6.1. Here, in the first line, E is the energy of the supercluster associated with the photon, and H is the energy deposited in the HCAL tower which is behind the seed crystal of the supercluster, in the η direction. The quantity $\sigma_{i\eta i\eta}$ is the width of the energy distribution of the supercluster in the η direction. The isolation variables are defined as:

$$\mathrm{PF}_X = \frac{1}{p_{\mathrm{T}}} \max\left(0, \sum_{i \in X} p_{\mathrm{T}} - \rho A\right), \tag{6.1}$$

where $X =$ charged hadrons, neutral hadrons, photons and i runs over the PF candidates present in a cone centering on the photon direction, excluding the photon itself. The width of the isolation cone is $\Delta R = 0.3$. Furthermore, ρ is the average p_{T} flow density in the event, and A is an effective area which depends on the η of the photon.

The (non)prompt photon selection efficiency is estimated to be 80.29% (10.64%) for these selection criteria, in data. In order to account for the difference in the efficiency between data and simulation, scale factors are applied to the simulated events.

Additionally, a reduction in the number of electrons misidentified as photons is achieved by requiring that there is no more than one hit in the pixel detector in the direction of the reconstructed photon (*pixel seed veto*). Scale factors are used to account for the differences in efficiencies between data and simulation in this case as well.

The identified photons are then selected for the analysis if their p_{T} is greater than 20 GeV, and the absolute value of their pseudorapidity is less than 1.44, which corresponds

to coverage of the barrel detector region. The last requirement is that no selected lepton should be within a cone of $\Delta R = 0.4$ around the photon direction.

Leptons

The lepton selection criteria are the same as the ones defined for the tWZ analysis in Section 5.2.3.

During 2024, the CMS collaboration recognized an issue in the calibration of electrons in data: the ones whose associated supercluster had $1.5 < |\eta| < 2$ and $R9 > 0.98$ had their p_{T} erroneously calibrated to a value of 0. The value $R9$ is defined as the ratio between the energy deposited in the 3×3 cristal array centered around the most energetic crystal in the supercluster and the energy of the supercluster. These electrons have therefore been vetoed in the simulated samples for this analysis.

Jets and MET

The AK4 jets are reconstructed and identified according to the definitions in Section 5.2.3, with the addition of the requirement that no selected photons should be found within a cone of $\Delta R = 0.1$ around the jet direction. This analysis also makes use of *loose jets*, which follow the same requirement as the regular jets, but have 20 GeV $< p_{\mathrm{T}} <$ 30 GeV. The missing transverse energy is also calculated as in Section 5.2.3.

6.2.4 Background processes

Analogously to the case of tWZ, the main background to the $tW\gamma$ process is the $t\bar{t}\gamma$ process, with which it shares an extremely similar topology.

$Z\gamma$

Another process that contributes significantly to the phase space with two leptons and a photon is the $Z\gamma$ process, where the Z boson decays into two charged leptons. The photon can be radiated from either the initial state quarks ($Z\gamma_{\mathrm{ISR}}$) or the final state leptons ($Z\gamma_{\mathrm{FSR}}$), as shown in Figure 6.2. The contribution from the $Z\gamma$ process is estimated using a simulation produced at NLO in QCD with `MadGraph5_aMC@NLO` and `PYTHIA8`.

Nonprompt photon background

The modeling of the nonprompt photon background is obtained from simulated samples. The main contribution comes from the $t\bar{t}$ process, where the photon is produced in the hadronization of partons. Smaller contributions come from the Drell-Yan process, simulated at NLO in QCD with `MadGraph5_aMC@NLO`, and the tW process, generated with `POWHEG` [157]. All the samples are interfaced with `PYTHIA8`.

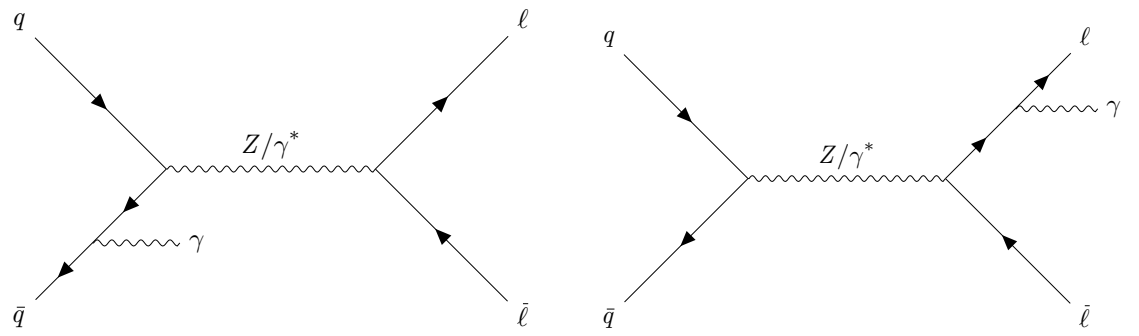


Figure 6.2: Example Feynman diagrams for the $Z\gamma$ process, where the photon is radiated from the initial state quarks ($Z\gamma_{\text{ISR}}$, on the left) or the final state leptons ($Z\gamma_{\text{FSR}}$, on the right).

Other backgrounds

Other backgrounds include processes with two heavy bosons in the final state (WW,WZ,ZZ) and the radiation of a photon. These processes are estimated using simulated samples produced at tree level with `PYTHIA8`. The contribution from the VVV process is also considered and is estimated using a simulation produced at NLO in QCD with `MadGraph5_aMC@NLO` and `PYTHIA8`.

In addition to the tW process, the production of a top quark (pair) in association with a heavy boson (pair) and a photon is also considered, and is estimated using a simulation produced at NLO in QCD with `MadGraph5_aMC@NLO` and `PYTHIA8`.

Also for the minor background processes, the photon can be radiated from either the hard process, or the decay products of the top quarks or the heavy bosons.

6.3 Signal and control regions

The strategy of the search for the $tW\gamma$ process is inspired by the one used for the tWZ analysis. The rare $tW\gamma$ signal has a significantly lower cross section than the $t\bar{t}\gamma$ background while the final state topologies only differ by the presence of an additional b jet. This latter is not reconstructed or b-tagged in more than half of the selected events, leaving a substantial overlap between the topologies of the two processes.

The analysis strategy consists of the definition of two control regions used to estimate the $Z\gamma$ contribution to the phase space, and two signal regions defined to include the $tW\gamma$ signal. To overcome the difficulty of extracting the signal from regions with a very large $t\bar{t}\gamma$ background, a GNN-based binary classifier was developed. The output of the classifier is used as template in a maximum likelihood fit, in order to extract the signal strength $\mu_{tW\gamma}$ and the significance of the tWZ signal.

Additionally, a nonprompt photon control region is defined to assess the uncertainty in the modeling of the nonprompt photon background.

This section describes the selection criteria for the signal and control regions which are fitted simultaneously in the maximum likelihood fit.

6.3.1 Baseline selection

The baseline selection is defined in order to select the events with a final state compatible with the $tW\gamma$ production and decay into leptons. The requirements for an event to be selected are:

- at least one selected photon;
- exactly two selected leptons, with opposite charge;
- the p_{T} of the (sub)leading lepton must respect the conditions $p_{\mathrm{T}} > 25(20)$ GeV;
- the invariant mass of the two leptons must be $m_{\ell\ell} > 30$ GeV;
- at least one selected b-tagged jet.

All the events that pass the baseline selection are then divided into the control and signal regions.

6.3.2 Control regions

Two control regions are defined in the analysis to estimate and constrain the contributions from the $Z\gamma_{\mathrm{ISR}}$ and $Z\gamma_{\mathrm{FSR}}$ processes to the phase space. The purpose of the $\mathrm{CR}_{\mathrm{ISR}}$ is to select the majority of the $Z\gamma_{\mathrm{ISR}}$ events, where the Z boson decays into two leptons. This is achieved by requiring that the two leptons in the event have the same flavor, and that their invariant mass $m_{\ell\ell}$ should respect the condition $|m_{\ell\ell} - m_Z| < 15$ GeV.

The $\mathrm{CR}_{\mathrm{FSR}}$ is instead defined to be pure in $Z\gamma_{\mathrm{FSR}}$ events. In this case, both the leptons and the photon ultimately originate from the decay of the Z boson, as shown in the right diagram in Figure 6.2. Therefore, events are selected in this region if the two leptons have the same flavor, and the dilepton-photon invariant mass $m_{\ell\ell\gamma}$ is such that $|m_{\ell\ell\gamma} - m_Z| < 15$ GeV. Additionally, the invariant mass of the lepton pair must be such that $|m_{\ell\ell} - m_Z| > 15$ GeV, as to suppress the presence of $Z\gamma_{\mathrm{ISR}}$ events and be mutually exclusive to the $\mathrm{CR}_{\mathrm{ISR}}$.

Both control regions feature a prominent disagreement between data and simulation in the distribution of the number of jets in the events, as shown in Figure 6.3.

The modeling of the $Z\gamma$ processes is improved by assigning to it an unconstrained shape uncertainty in the maximum likelihood fit. This procedure works by assigning three systematic uncertainties to the $Z\gamma_{\mathrm{ISR}}$ and $Z\gamma_{\mathrm{FSR}}$ processes: one each to the events with one, two, or more than two jets. These uncertainties cover the range from 10% to 200% of the nominal yields, and are correlated between the $Z\gamma_{\mathrm{ISR}}$ and $Z\gamma_{\mathrm{FSR}}$ processes. In order to account for differences in total normalization between the two processes, an additional normalization uncertainty is assigned to the $Z\gamma_{\mathrm{FSR}}$ process. This decouples the normalization of the two processes while still tying together the shape of the distribution of the number of jets.

To validate this procedure, a fit is performed in the control regions only. The results are shown in Figure 6.4, where the insets show the improved agreement between data and simulation for the distribution of the number of jets in the $\mathrm{CR}_{\mathrm{ISR}}$ and $\mathrm{CR}_{\mathrm{FSR}}$.

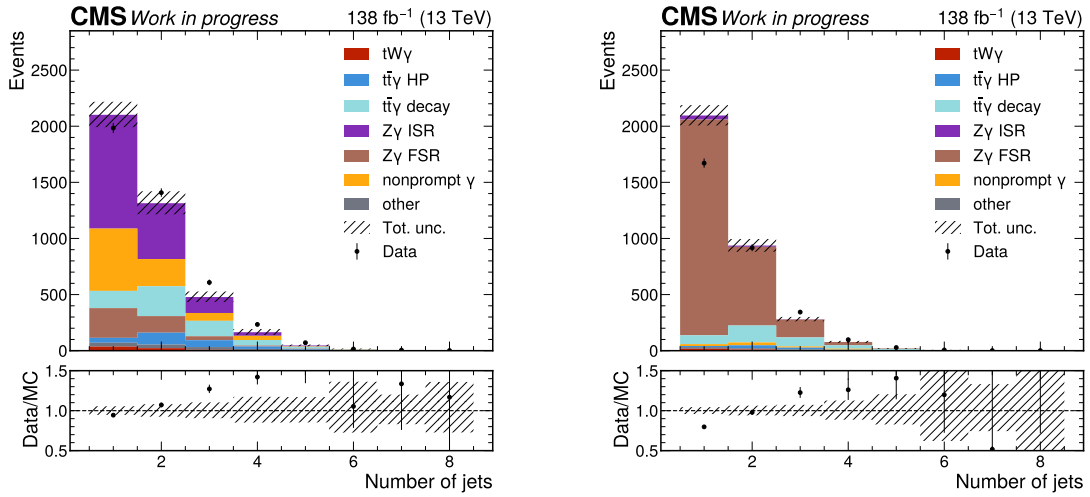


Figure 6.3: Distribution of the number of jets for the events in the CR_{ISR} (left) and CR_{FSR} (right) control regions. The insets show the ratio between data and simulation and highlight a clear disagreement in shape between the two.

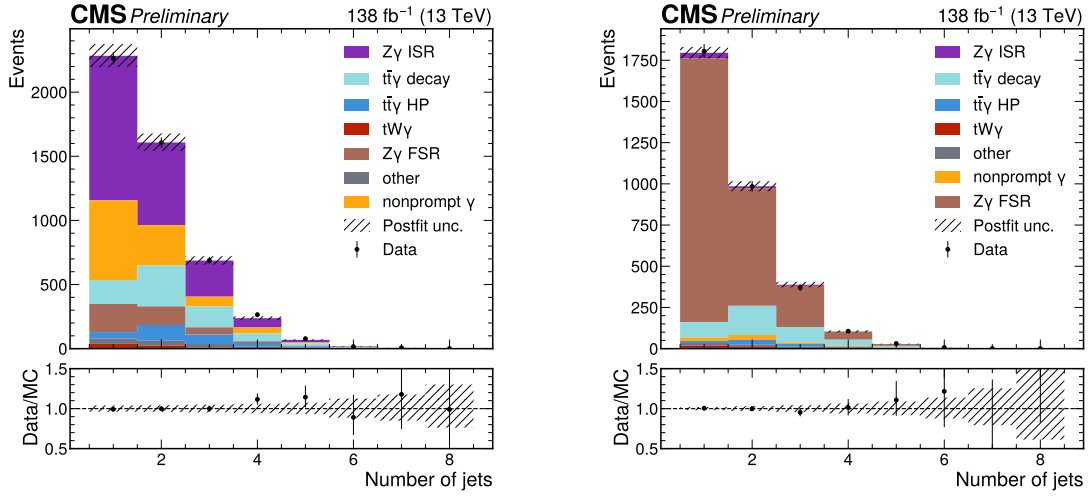


Figure 6.4: Distribution of the number of jets for the events in the CR_{ISR} (left) and CR_{FSR} (right) control regions after a maximum likelihood fit in the background regions only. The insets show the ratio between data and simulation, and highlight the improved agreement between the two after the fit.

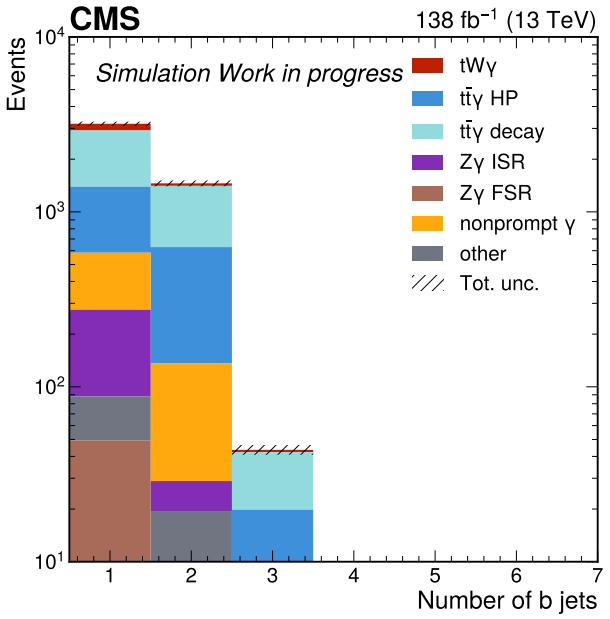


Figure 6.5: Distribution of the number of b-tagged jets for the simulated events in the signal regions. Most of the signal events feature one b-tagged jet, together with more than half of the $t\bar{t}\gamma$ events. The $Z\gamma_{\text{ISR}}$ and nonprompt photon backgrounds are also significantly present in the first bin of the distribution.

6.3.3 Signal regions

Two signal regions are defined in the analysis to select the events with a final state compatible with the $tW\gamma$ process. All events in which the selected leptons have different flavors are included in the signal regions. Additionally, the remaining events are selected if they respect the following conditions: $|m_{\ell\ell} - m_Z| > 15$ GeV and $|m_{\ell\ell\gamma} - m_Z| > 15$ GeV. This is done to suppress the contributions from the $Z\gamma$ processes, which are mostly contained in the orthogonal control regions.

Figure 6.5 shows the distribution of the number of b-tagged jets after this selection for the simulated events. The first bin of the distribution includes the most part of the signal events, as well as more than half of the selected $t\bar{t}\gamma$ events. Additionally, the $Z\gamma_{\text{ISR}}$ and nonprompt photon backgrounds are also significantly present in this bin. This is not the case for the following bins, where there is a smaller fraction of signal events, and the $Z\gamma_{\text{ISR}}$ process is not present at all.

In virtue of the different topologies and fractions of signal and background events in the different bins of the distribution of the number of b-tagged jets, the $\text{SR}_{tW\gamma}$ is defined to include all the events with one b-tagged jet. The remaining events are selected in the $\text{SR}_{t\bar{t}\gamma}$.

In these regions, the output distribution of the GNN-based binary classifier, defined in Section 6.4, is used as the stack of signal and background histograms which is fitted to data points in the maximum likelihood fit.

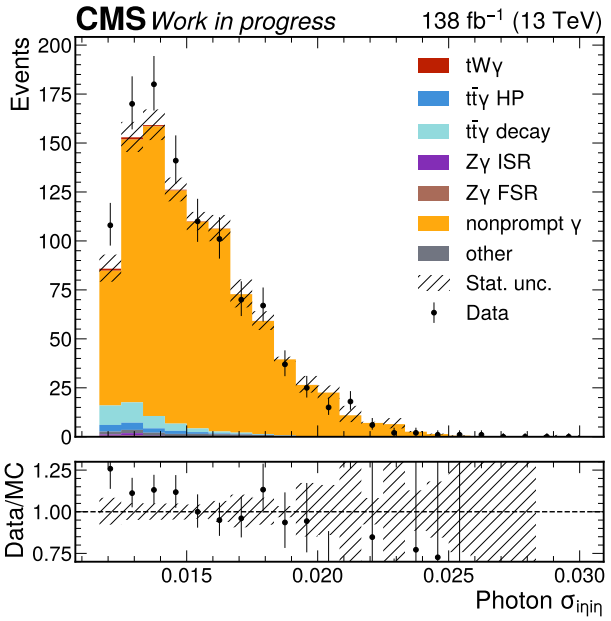


Figure 6.6: Distribution of the $\sigma_{i\eta i\eta}$ variable for the events in the nonprompt photon control region. The lower inset shows the ratio between data and simulation, which have a disagreement of at most 20% in almost all the bins of the distribution.

6.3.4 Nonprompt photon control region

The last region defined in the analysis is the nonprompt photon control region. This latter is not fitted in the maximum likelihood fit, but it is employed to assess the uncertainty in the modeling of the nonprompt photon background.

This control region is established on the definition of a *fakeable* photon. This is a photon candidate that respects all the identification criteria defined in Table 6.1, but fails the condition on the $\sigma_{i\eta i\eta}$ variable. The requisite for a fakeable photon is to have $\sigma_{i\eta i\eta} > 0.012$. The nonprompt photon control region is then defined according to the same selection as the signal regions, but with the additional requirement that there is exactly one photon in the event, and that this photon is a fakeable photon. This region is therefore almost exclusively populated by events with a nonprompt photon.

The distribution of the $\sigma_{i\eta i\eta}$ variable for the events in the nonprompt photon control region is shown in Figure 6.6. The lower inset shows the ratio between data and simulation, which presents an overall good agreement between the two. The same considerations hold true for the distributions of the p_{T} and η of the fakeable photons, shown in Figure 6.7. An overall 20% normalization uncertainty is therefore assigned to the nonprompt photon background in the maximum likelihood fit, in order to account for the maximum discrepancies in the bins that contain the most events.

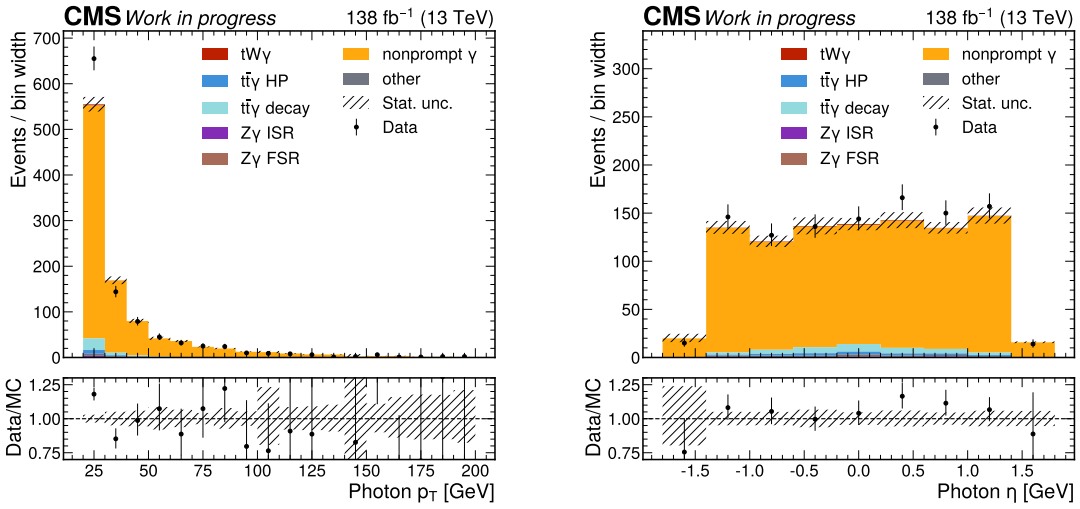


Figure 6.7: Distribution of the p_T (left) and η (right) of the fakeable photons in the nonprompt photon control region. The lower insets show the ratio between data and simulation, which have a disagreement of at most 20% in almost all the bins of the distribution.

6.4 Graph Neural Network for the discrimination of $tW\gamma$ and $t\bar{t}\gamma$

A GNN-based binary classifier has been developed to discriminate between the $tW\gamma$ and $t\bar{t}\gamma$ processes in the signal regions. The choice of a GNN is motivated by the aim of training only one classifier for both signal regions, while still being able to capture the differences in the topologies of the two processes and achieving excellent performance. Indeed, a GNN has the ability of accepting as input the features of a variable number of jets, and of learning the relationships between them. It is particularly useful in this analysis, where the number of jets in an event and their properties carry important information regarding the nature of the event. This is due to the only difference between the $tW\gamma$ and $t\bar{t}\gamma$ processes being the presence of an additional b jet in the former process. If this jet is not reconstructed and/or identified, the GNN is still able to infer the right classification from the other hadronic jets in the event.

6.4.1 Graph structure and input features

The input to the GNN is a fully connected graph, i.e., a graph in which each node is connected to all the other nodes, shown in Figure 6.8.

The nodes of the graph represent the physics objects in each event, i.e., the leading photon, the two leptons, and the hadronic jets. The input features associated with each node are the p_T and η of the object, together with an encoding of the object type. The latter consists of four values which can only be 0 or 1, and represent if the node is associated to a photon, a lepton, a jet, or a b-tagged jet, respectively.

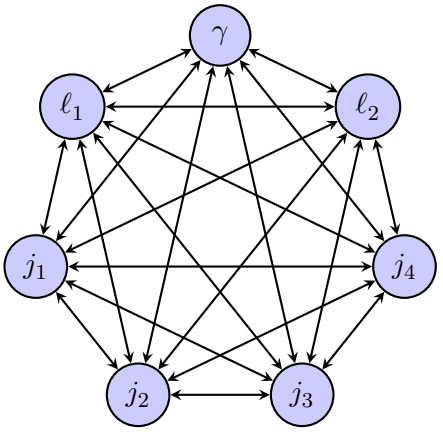


Figure 6.8: Example of a fully connected graph with seven nodes, with each node representing either a photon, a lepton, or a hadronic jet.

The edge features carry information regarding the kinematic relationships between the physics objects. They are: the invariant mass, the transverse momentum, and the pseudo-rapidity of the sum of the four-momenta of the two objects connected by the edge. The last variable is the ΔR between the two objects.

Separately from the graph structure, six *global features* of each event are also included as input to the binary classifier. They are:

- the $p_{\mathrm{T}}^{\mathrm{miss}}$;
- the $m_{\ell\ell\gamma}$;
- the number of loose jets;
- a flag indicating if the leptons in the event have the same flavor;
- the invariant mass of the system of the photon, the lepton pair, and all the jets (m_{sys});
- the ratio between the p_{T} and the energy of the system of the photon, the lepton pair, and all the jets (centrality).

Validation

The modeling of the input features used for the binary classifier is validated with data using the saturated GoF test described in Section 3.1.3. The variables that underwent the test are all the global features, except for the lepton-flavor flag, and the four elements of the momenta of each node, divided by the physics object type. Leptons and jets are additionally ordered in p_{T}. The fits are performed including all the systematic uncertainties described in Section 6.5, considering only the region defined by the sum of the $\mathrm{SR}_{tW\gamma}$ and $\mathrm{SR}_{t\bar{t}\gamma}$. All the fitted variables passed the test, i.e., the observed value of the test statistic t_d obtained from the fit to data has a p-value greater than 0.05.

The distributions of the employed test variables compared to the respective fitted t_d are collected in Appendix C. Figure 6.9 shows the example of the centrality variable, which has a p-value of 0.851.

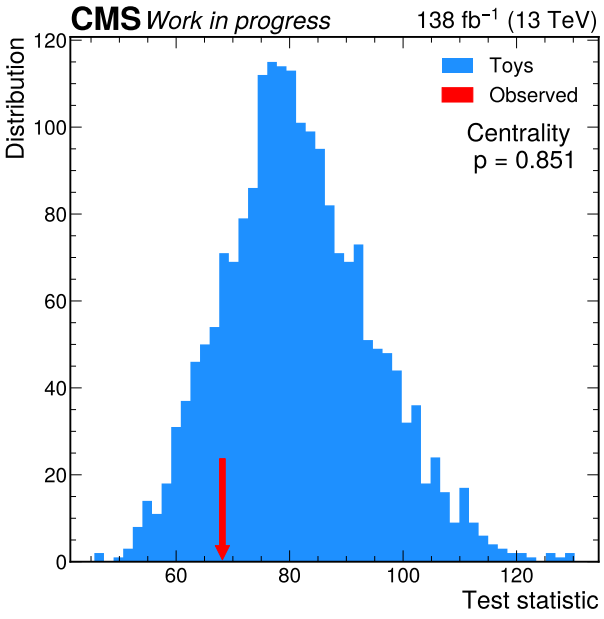


Figure 6.9: Distribution of the test statistic used in the saturated GoF test for the centrality variable, compared to the value t_d obtained from the fit to data. The p-value of the test is 0.851.

6.4.2 GNN architecture

The GNN used in this analysis is based on the use of Transformer Convolutional Layers that act on the graph structure. The graph is then embedded into a fixed-size vector, which is concatenated with the embedding of the global features and passed through a fully connected neural network. The node features, edge features, and global features are all normalized through batch normalization layers before being fed to the GNN.

The GNN architecture can be divided in a graph branch, a global branch, and a final branch. The graph branch is based on the repetition of a module composed of:

- a Transformer Convolutional Layer with three attention heads, which returns a 96-dimensional output for each node and employs a Leaky ReLU activation function;
- a batch normalization layer, which normalizes the output of the Transformer Convolutional Layer;
- a residual connection based on a trainable projector, which increases the dimensionality of the input node variables to 96 and sums the result to the output of the layer.

The residual connection used in this ML architecture always employs the input node variables of the GNN, as this was found to result in better performance than using the output of the previous layer, for this task. The graph branch therefore consists of three of these modules, followed by a global average pooling layer which returns a flat 96-dimensional vector.

The global branch is composed of only one linear layer which takes the global features as input and returns a 6-dimensional vector with a Leaky ReLU activation function. The output is then fed into a batch normalization layer.

The global and graph branches come together into the final branch, where they first undergo a concatenation operation.

The concatenated values are passed to an intermediate output node whose output represents the likelihood of the photon in the event to originate from the hard process or the decay products. The activation function used for this node is a sigmoid function. The intermediate output is not used in the analysis, but it was found to slightly improve the performance of the GNN.

This output is concatenated with its own input and the resulting values are passed to a fully connected neural network with five hidden layers of 64 nodes each, activated by a Leaky ReLU function. Each layer is intertwined with batch normalization layers. The output of this network is a single node, which uses a sigmoid activation function and represents the likelihood of the event to belong to the $tW\gamma$ process.

The described architecture was chosen after a hyperparameter optimization process on simulated data, which found this model to have the best performance in terms of the area under the ROC curve. The optimization employed a Bayesian approach [158], and was performed using the Ax library [159].

6.4.3 Training and testing

The total dataset is made up of 672 batches of 4096 events each. All the batches are composed of:

- 26% of hard process $tW\gamma$ events with the DR1 scheme;
- 12% of hard process $tW\gamma$ events with the DR2 scheme;
- 11% of decay $tW\gamma$ events with the DR1 scheme;
- 5% of decay $tW\gamma$ events with the DR2 scheme;
- 41% of hard process $t\bar{t}\gamma$ events;
- 5% of decay $t\bar{t}\gamma$ events.

The hard process and decay labels refer to the origin of the photon in the event.

The training is performed according to the procedure described in Section 5.3. The dataset is divided into two halves, so that two GNN models can be trained, one for each half. These sub-datasets are further divided into a training and a testing set, with a ratio of 80% and 20%, respectively. In the analysis workflow, the two models then act on the half of the dataset they were not trained on, and the results are combined to obtain the final output of the GNN. The observed data and the simulation samples which are used to train the GNNs are also randomly acted upon by either of the two models.

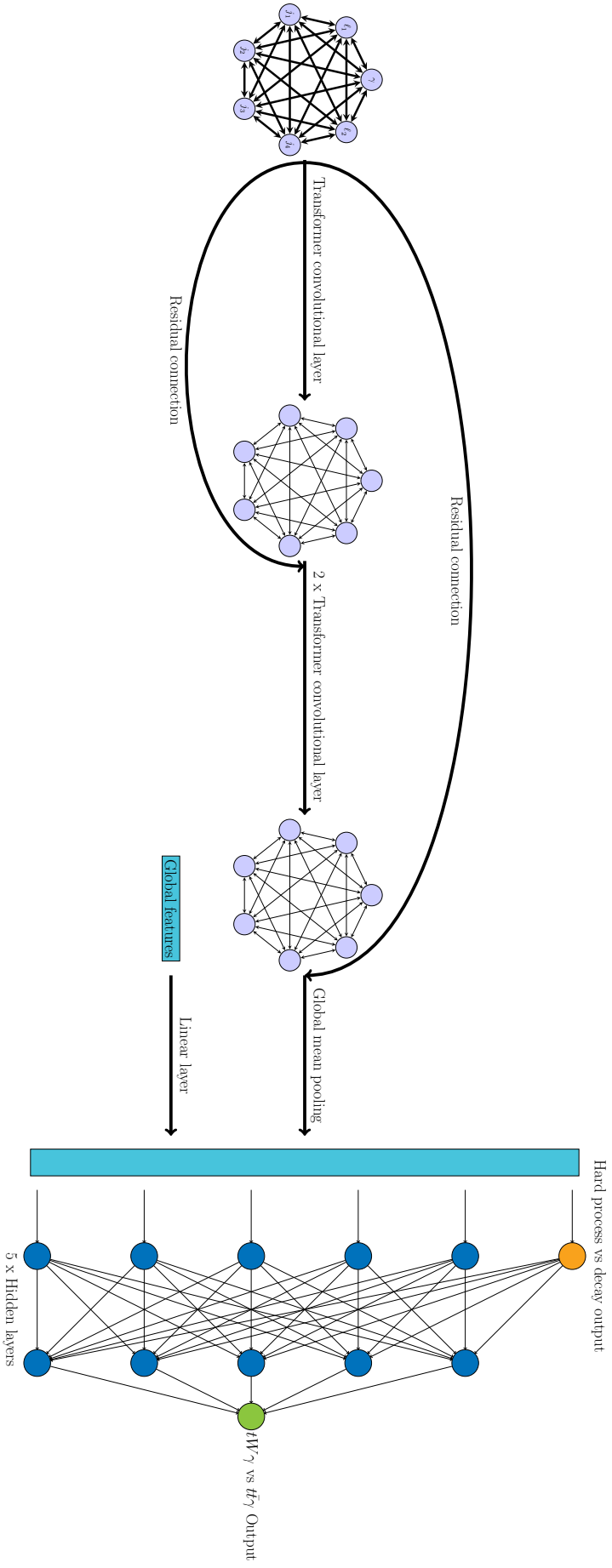


Figure 6.10: Architecture of the GNN developed for the $tW\gamma$ analysis. A complete description is reported in Section 6.4.2.

The training is performed using the Adam optimizer using a binary cross-entropy loss function, both for the $tW\gamma$ vs $t\bar{t}\gamma$ output and the hard process vs decay output. The total loss for a batch of the training sample is defined as:

$$L = \frac{1}{N}\sum_{i=1}^{N} w_i \left(\ln(|\bar{y}_i - y_i|) + 0.4\ln(|\bar{z}_i - z_i|)\right), \tag{6.2}$$

where $N = 4096$ is the number of events in the batch, w_i is the weight of the event, y_i is the output of the GNN for the $tW\gamma$ vs $t\bar{t}\gamma$ classification, z_i is the output of the GNN for the hard process vs decay classification, $\bar{y}_i$ is a label whose value is 1 (0) if the event belongs to the $tW\gamma$ ($t\bar{t}\gamma$) process, and $\bar{z}_i$ is a label whose value is 1 (0) if the photon in the event originates from the hard process (decay). The factor of 0.4 applied to the photon origin part of the loss reflects the fact that the main task of the GNN is to discriminate between the $tW\gamma$ and $t\bar{t}\gamma$ processes. Finally, the weights w_i for the hard process $t\bar{t}\gamma$ and decay $t\bar{t}\gamma$ events are set to the ratio between the number of events in the batch and the number of selected events of same type in the signal regions. The same procedure is used for the hard process and decay $tW\gamma$ events, with the difference that the weight for both DR1 and DR2 events has the number of selected DR1 events in the signal regions as denominator. The weights for the $tW\gamma$ signal are furthermore divided by a factor of two to account for the two DR schemes.

The two models are trained for 3000 epochs, or until the loss of the testing set has not decreased for 100 epochs. In both cases, the final weights of the model are the ones associated with the epoch that has the lowest loss of the testing set. The first model trained for the entire 3000 epochs, while the second one stopped at epoch 2095. Regardless of this, the two models have similar performance.

Figure 6.11 shows the output of the GNN for the $tW\gamma$ vs $t\bar{t}\gamma$ classification for both models for events from both signal regions. The GNN proves very good discrimination abilities, as also shown by the area under the ROC curve values of 0.791 and 0.790, as shown in Figure 6.12. Additionally, Figure 6.13 shows that the GNN performs better in discriminating the signal against the $t\bar{t}\gamma$ background when the photon originates from the hard process, rather than the decay products.

Figures 6.14 and 6.15 show the output of the GNN divided by events in the signal regions $\mathrm{SR}_{tW\gamma}$ and $\mathrm{SR}_{t\bar{t}\gamma}$, respectively. The models exhibit very good discrimination power in both, while returning significantly different distributions for the two signal regions. This is a further reason for splitting the phase space in these two signal regions, as the GNN is able to capture the differences in the topologies of the two processes.

Finally, the GNN is added to the analysis workflow to provide the output for the $tW\gamma$ vs $t\bar{t}\gamma$ classification in the signal regions. This is shown in Figure 6.16, with $\mathrm{SR}_{tW\gamma}$ on the left and $\mathrm{SR}_{t\bar{t}\gamma}$ on the right. These are the templates used in the maximum likelihood fit for the two signal regions.

Regarding the $\mathrm{SR}_{tW\gamma}$ region, the GNN distribution has very good discriminating power against the $t\bar{t}\gamma$ background, but it is not able to separate the $tW\gamma$ signal from the $Z\gamma_{\mathrm{ISR}}$

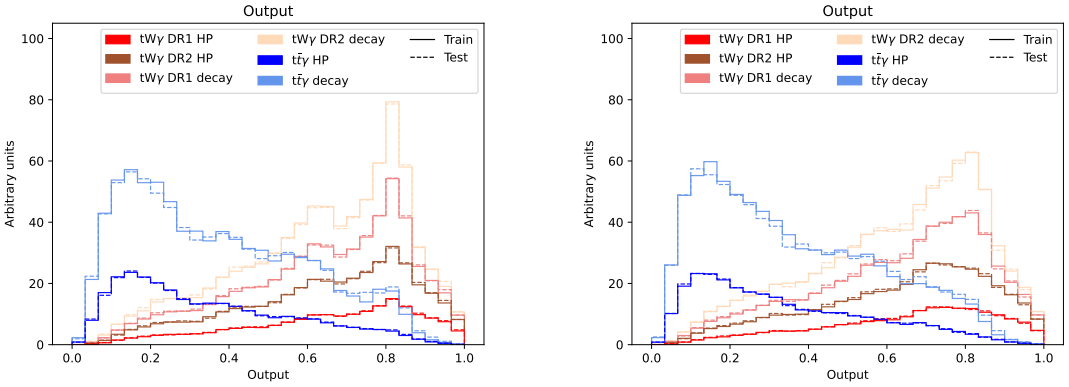


Figure 6.11: Output of the GNN for the $tW\gamma$ vs $t\bar{t}\gamma$ classification for the first (left) and second (right) model for events from both signal regions. The output calculated on the training set is shown in solid colors, while it is as a dashed line for the testing set. In both cases, the GNN shows very good discrimination abilities. The normalization of the outputs is thus that the sum of the events sums to the same arbitrary value for both signal and background, while representing the same ratio of hard process photon to decay photon. For the signal events, the normalization weight is further divided by a factor of two to account for the two DR schemes.

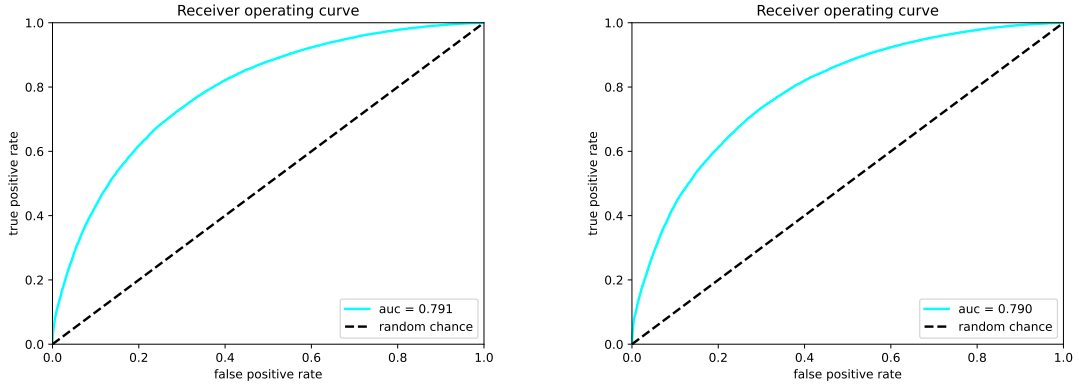


Figure 6.12: ROC curve for the first (left) and second (right) model of the GNN for the $tW\gamma$ vs $t\bar{t}\gamma$ classification. The area under the curve is 0.791 and 0.790, respectively.

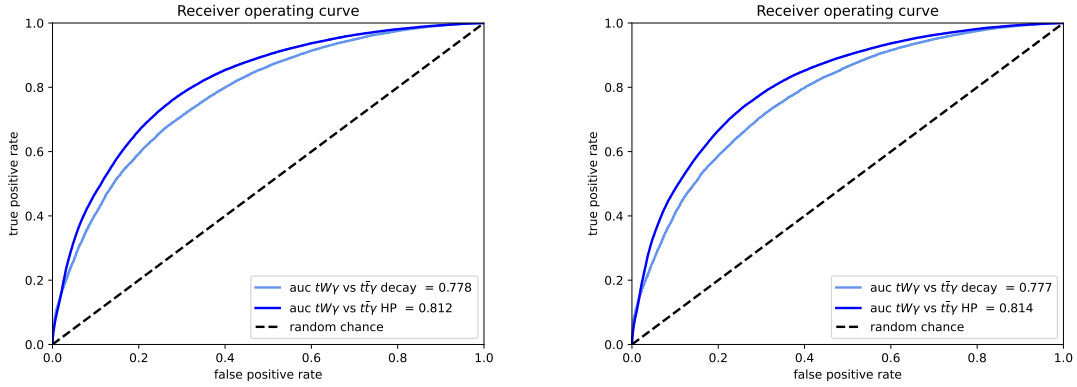


Figure 6.13: ROC curve for the first (left) and second (right) model of the GNN for the discrimination of the signal against the hard process background and the decay background.

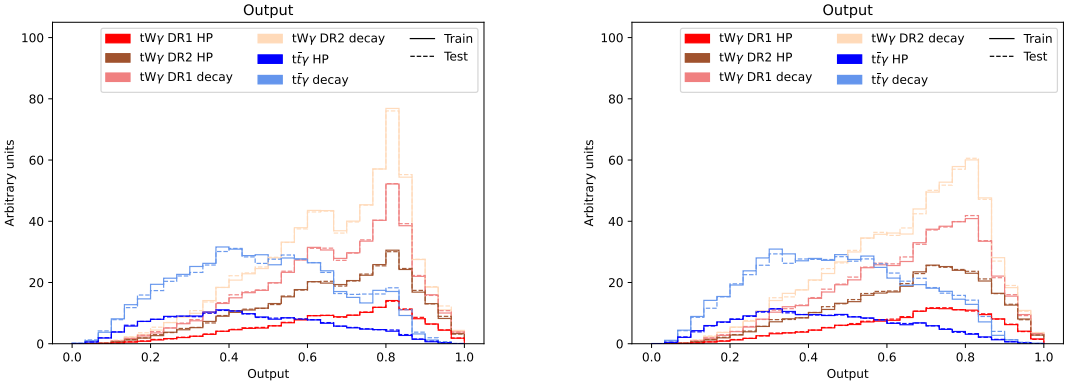


Figure 6.14: Output of the GNN for the $tW\gamma$ vs $t\bar{t}\gamma$ classification for the first (left) and second (right) model for events in the $\mathrm{SR}_{tW\gamma}$, for both the training (solid colors) and testing (dashed lines) sets. The normalization of the outputs is the same as in the caption of Figure 6.11.

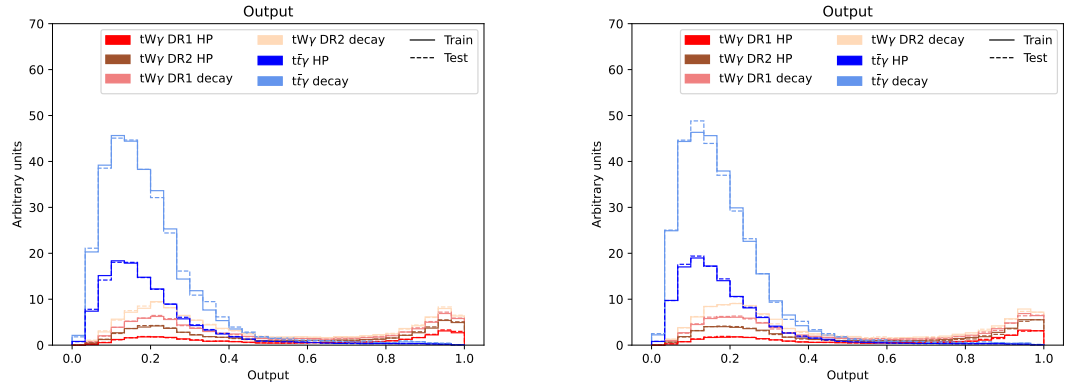


Figure 6.15: Output of the GNN for the $tW\gamma$ vs $t\bar{t}\gamma$ classification for the first (left) and second (right) model for events in the $\mathrm{SR}_{t\bar{t}\gamma}$, for both the training (solid colors) and testing (dashed lines) sets. The normalization of the outputs is the same as in the caption of Figure 6.11.

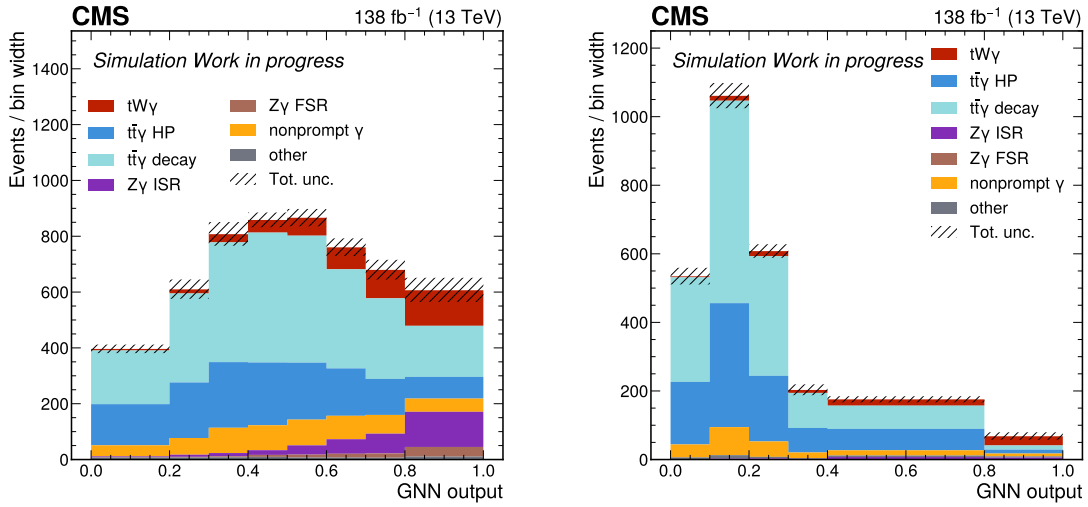


Figure 6.16: Output of the GNN for the $tW\gamma$ vs $t\bar{t}\gamma$ classification for the events in the $\mathrm{SR}_{tW\gamma}$ (left) and $\mathrm{SR}_{t\bar{t}\gamma}$ (right) signal regions. These distributions are used as templates in the maximum likelihood fit.

background, against which it has not been trained. Training the GNN against the $Z\gamma_{\mathrm{ISR}}$ process has not been possible due to lack of simulated events available. On the other hand, the GNN is extremely effective in the $\mathrm{SR}_{t\bar{t}\gamma}$ region, with the last bin of the distribution having a 50% contribution from the signal.

6.5 Corrections and systematic uncertainties

Similarly to the tWZ analysis, several sources of systematic uncertainties are considered in this analysis. The most common sources are related to uncertainties in the corrections applied to the simulated events to account for known differences between data and simulation, or related to uncertainties in the modeling of the processes. The definition of the likelihood function that is fitted, and the fit methodology are described in Section 3.1.

The following uncertainties and corrections are alredy described in Section 5.5:

- pileup;
- luminosity uncertainty;
- lepton identification;
- jet energy scale and resolution;
- missing transverse energy;
- b-tagging efficiency;
- L1 trigger prefiring;
- HEM15/16 issue;
- PDF uncertainties;

- parton shower modeling;
- factorization and renormalization scale;
- signal modeling.

This section describes additional systematic uncertainties which are specific to the $tW\gamma$ analysis.

Trigger efficiency

The trigger selection has different efficiencies between data and simulation. This disagreement is adjusted by applying scale factors to the simulated events. The scale factors are defined as the ratio between the efficiency of the trigger in data and simulation, and are a function of the p_{T} of the leptons in the event.

The efficiencies are calculated according to the following procedure: events are selected using MET triggers and offline selection criteria aiming to define a phase space rich in $t\bar{t}$ events with a final state in two leptons. The selected events are then checked against the dilepton trigger selection used in this analysis and the efficiencies are calculated as the ratio between the number of events that pass both trigger selections and the number of events that pass only the MET trigger selection.

The uncertainties associated to the trigger efficiency correction are assigned to the events by varying the scale factors within their uncertainties before applying them to the simulated events. These uncertainties are considered correlated between the years of data taking.

Photon identification and pixel seed veto

The criteria for the identification of photons show inconsistent selection efficiencies between data and simulation. This disagreement is taken care of by applying scale factors to the simulated events. The scale factors are a function of the p_{T} and η of the photon. A systematic uncertainty is assigned to this correction by varying the scale factors within their uncertainties before applying them to the simulated events. The correction is considered correlated between the years of data taking.

The same reasoning applies for the pixel seed veto. The scale factors are computed as a function of the pseudorapidity of the photon, and are applied to the simulated events. The uncertainty applied to the correction is also considered correlated between the years of data taking.

Normalization uncertainties

As mentioned in Section 6.3.2, the $Z\gamma$ processes are assigned an unconstrained shape uncertainty in the maximum likelihood fit. Furthermore, in order to decouple the normalization of the $Z\gamma_{\mathrm{ISR}}$ and $Z\gamma_{\mathrm{FSR}}$ processes, an additional normalization uncertainty, in the form of a freely floating rate parameter, is assigned to the $Z\gamma_{\mathrm{FSR}}$ process. A rate parameter is a nuisance parameter that is used to scale the yield of a process in the maximum likelihood fit in the same way as a signal strength does.

Furthermore, the nonprompt photon background is assigned a 20% normalization uncertainty in the maximum likelihood fit, as described in Section 6.3.4. Also, a 20% normalization uncertainty is applied to the minor backgrounds, to account for the different processes that are included in this category, and their statistically limited contribution to the phase space.

Finally, two additional parameters are introduced in the maximum likelihood fit to account for uncertainties related to the $t\bar{t}\gamma$ background. These parameters are not assigned a prior uncertainty and are treated as freely floating rate parameters in the fit. They are the signal strength of the total $t\bar{t}\gamma$ process $\mu_{t\bar{t}\gamma}$, and the fraction f of the measured cross section of the hard process $\sigma_{t\bar{t}\gamma,HP}$ in the measured cross section of the total $t\bar{t}\gamma$ process $\sigma_{t\bar{t}\gamma}$. The expected value, taken from simulation, for the latter is:

$$f^{exp} = \frac{\sigma^{\text{exp}}_{t\bar{t}\gamma,HP}}{\sigma^{\text{exp}}_{t\bar{t}\gamma}} = 0.36. \tag{6.3}$$

The parameters $\mu_{t\bar{t}\gamma}$ and f are introduced in the likelihood function, defined in Eq. (3.5), through the substitutions:

$$\mu_{t\bar{t}\gamma,HP} = \mu_{t\bar{t}\gamma}\frac{f}{f_{\text{exp}}}, \quad \mu_{t\bar{t}\gamma,dec} = \mu_{t\bar{t}\gamma}\frac{1-f}{1-f_{\text{exp}}}, \tag{6.4}$$

where $\mu_{t\bar{t}\gamma,HP}$ and $\mu_{t\bar{t}\gamma,dec}$ are the signal strengths of the hard process and decay $t\bar{t}\gamma$ events, respectively.

6.6 Results

The maximum likelihood fit is initially performed using a pseudo-dataset, generated by taking the sum of the contributions of all processes in the template histograms, in order to determine the expected statistical significance of this search. It results in the measurement

$$\mu^{\text{exp}}_{tW\gamma} = 1.00^{+0.23}_{-0.20}, \tag{6.5}$$

corresponding to an expected statistical significance over the background-only hypothesis of 5.71 standard deviations.

The observed data is then fitted using the same likelihood function, and the result is

$$\mu^{\text{obs}}_{tW\gamma} = 0.94^{+0.24}_{-0.22}, \tag{6.6}$$

which is in agreement with the SM prediction well within one standard deviation. This corresponds to a measured value of the cross section of $tW\gamma$ $\sigma_{tW\gamma} = 21.62^{+5.52}_{-5.06}$ fb. It should be noticed that the phase space in which the measurement is performed does not perfectly overlap with the fiducial phase space for which the cross section is defined: 22% of the fiducial phase space is not included in the signal regions and it is extrapolated from simulation.

The measured value translates into an observed statistical significance of 4.97 standard deviations, making up for first evidence of the $tW\gamma$ process.

The reduction in the significance of the signal in the observed data, compared to its expected value, can be understood by observing the differences between the expected and observed impacts of the dominant systematic uncertainties on the signal strength, defined in Section 5.6 and shown in Figure 6.17.

The most prominent difference between the fit to the pseudo-dataset and the observed data is the impact of the DR modeling uncertainty. The DR2 prediction represents the +1-standard-deviation variation of the nuisance parameter for the modeling systematic uncertainty. The DR1 prediction represents both the nominal value and the -1-standard-deviation variation. By construction, the expected fit always returns the nominal value of the nuisance parameters. Therefore, the expected impact on the signal strength of the DR modeling uncertainty is necessarily one-sided; in this case the variation is in the positive direction for $\mu^{obs}_{tW\gamma}$.

On the other hand, the fit to data determines a value of the nuisance parameter associated to the modeling that lies in the middle between the DR1 and DR2 predictions, resulting in a two-sided impact on the signal strength. Hence, the observed uncertainty in the signal strength in the negative direction is larger than the expected one, causing the reduction in the significance of the signal in the observed data. A smaller effect is due to the measured value of $\mu^{obs}_{tW\gamma}$ being smaller than the expected one.

Because of the topological similarities between the $tW\gamma$ and $t\bar{t}\gamma$ processes, the signal strengths $\mu_{tW\gamma}$ and $\mu_{t\bar{t}\gamma}$ are anti-correlated, as shown by the impact of the latter on the former. The value of the correlation coefficient is $\rho_{tW\gamma,t\bar{t}\gamma} = -0.17$, smaller than the analogous value for the tWZ analysis, which is $\rho_{tWZ,t\bar{t}\gamma} = -0.63$. This is due to excellent performance of the GNN in discriminating between the two processes, allowing for a more precise measurement of the signal strength of the $tW\gamma$ process. Additionally, $\mu_{tW\gamma}$ has a positive correlation with the nuisance parameter f, with a value of $\rho_{tW\gamma,f} = 0.25$. This is due to the fact that the GNN performs better in the discrimination against the $t\bar{t}\gamma$ background when the photon originates from the hard process, rather than from the decay products. For increasing values of f, more $t\bar{t}\gamma$ events are of the HP type, which accumulate in the lower bins of the GNN output distribution, leaving room for the $tW\gamma$ signal to increase in the higher bins.

The correlation between $\mu_{tW\gamma}$ and both $\mu_{t\bar{t}\gamma}$ and f can also be observed in the two-dimensional likelihood scans performed around their best-fit values, shown in Figure 6.18. The simultaneous measurement of $\mu_{tW\gamma}$ and f is compatible with the SM prediction within one standard deviation. This result also represents the first measurement of the fraction $\sigma_{t\bar{t}\gamma,HP}/\sigma_{t\bar{t}\gamma}$, which is found to be $f = 0.54 \pm 0.13$.

On the other hand, the simultaneous measurement of signal strength of the $tW\gamma$ and $t\bar{t}\gamma$ is compatible with the SM prediction within two standard deviations. This disagreement is due to measurement of the signal strength of $t\bar{t}\gamma$ being $\mu_{t\bar{t}\gamma} = 0.86^{+0.06}_{-0.05}$. In comparison,

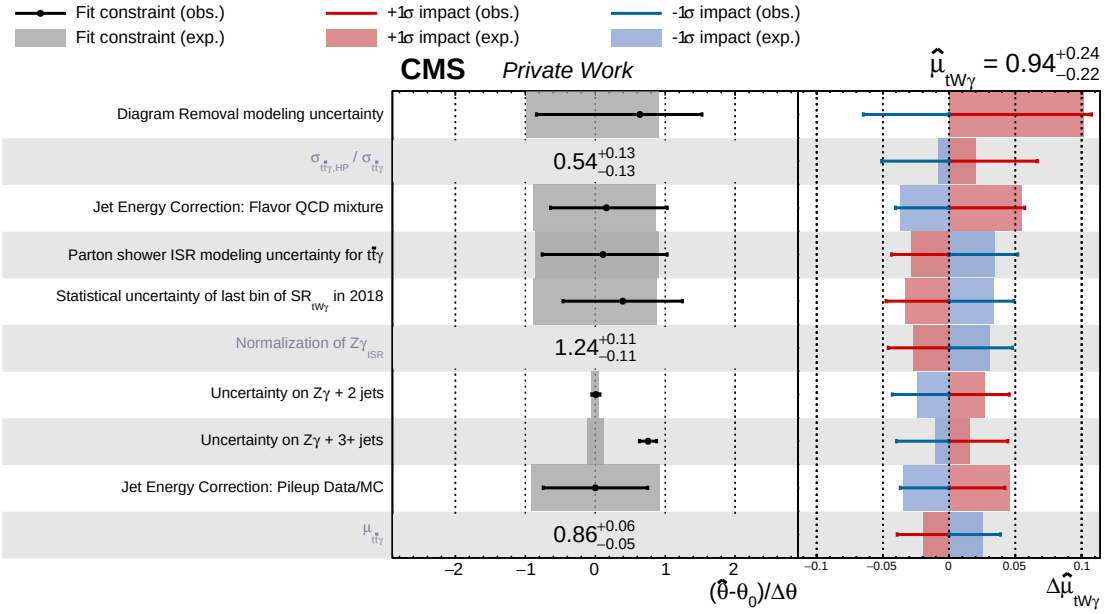


Figure 6.17: Impacts of the ten dominant nuisance parameters on the signal strength of the $tW\gamma$ process. The central column shows the difference between the best-fit value and the nominal value, in units of pre-fit uncertainty, for the nuisance parameters. It furthermore shows the observed (expected) post-fit uncertainties as black bars (dark gray band). For nuisance parameters which are not assigned a prior uncertainty, only the observed value and its uncertainties are shown. The right column shows the impact of the nuisance parameters on the signal strength, for both the observed fit (blue and red bars) and the expected fit (blue and red bands). The blue (red) color represents the variation of the nuisance parameter by $+1\sigma$ (-1σ).

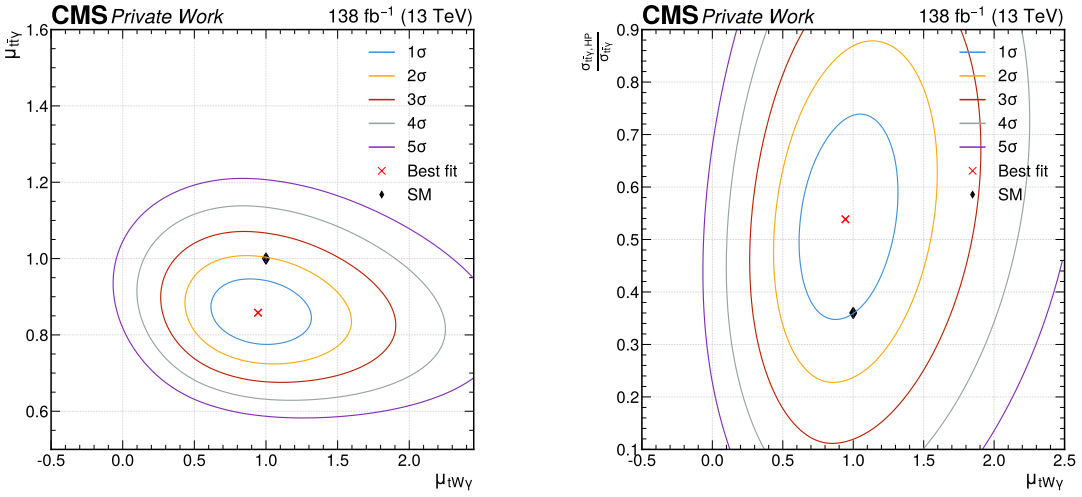


Figure 6.18: Two-dimensional scans of the likelihood function around the best-fit value of $\mu_{tW\gamma}$ and $\mu_{t\bar{t}\gamma}$ (left), and $\mu_{tW\gamma}$ and f (right). The red cross represents the best-fit value, while the black diamond represents the theoretical prediction. The contour lines for one to five standard deviations are shown in different colors.

the most recent measurement of the $t\bar{t}\gamma$ cross section by CMS resulted in a signal strength $\mu_{t\bar{t}\gamma} = 1.13 \pm 0.04$ [148]. The main difference with this previous result is that this analysis employs samples simulated at NLO in QCD for both the $tW\gamma$ and $t\bar{t}\gamma$ processes, for the first time. The discrepancy between the measurements of the signal strength of the $t\bar{t}\gamma$ process has clear similarities with the one observed in the tWZ analysis for the $t\bar{t}Z$ process.

Finally, the post-fit distributions of the GNN output for the regions of the fit are shown in Figure 6.19, displaying the very good agreement between data and the fitted simulation in all regions.

As already mentioned, this analysis consists of the first search for the $tW\gamma$ process, and it results in first evidence of its existence.

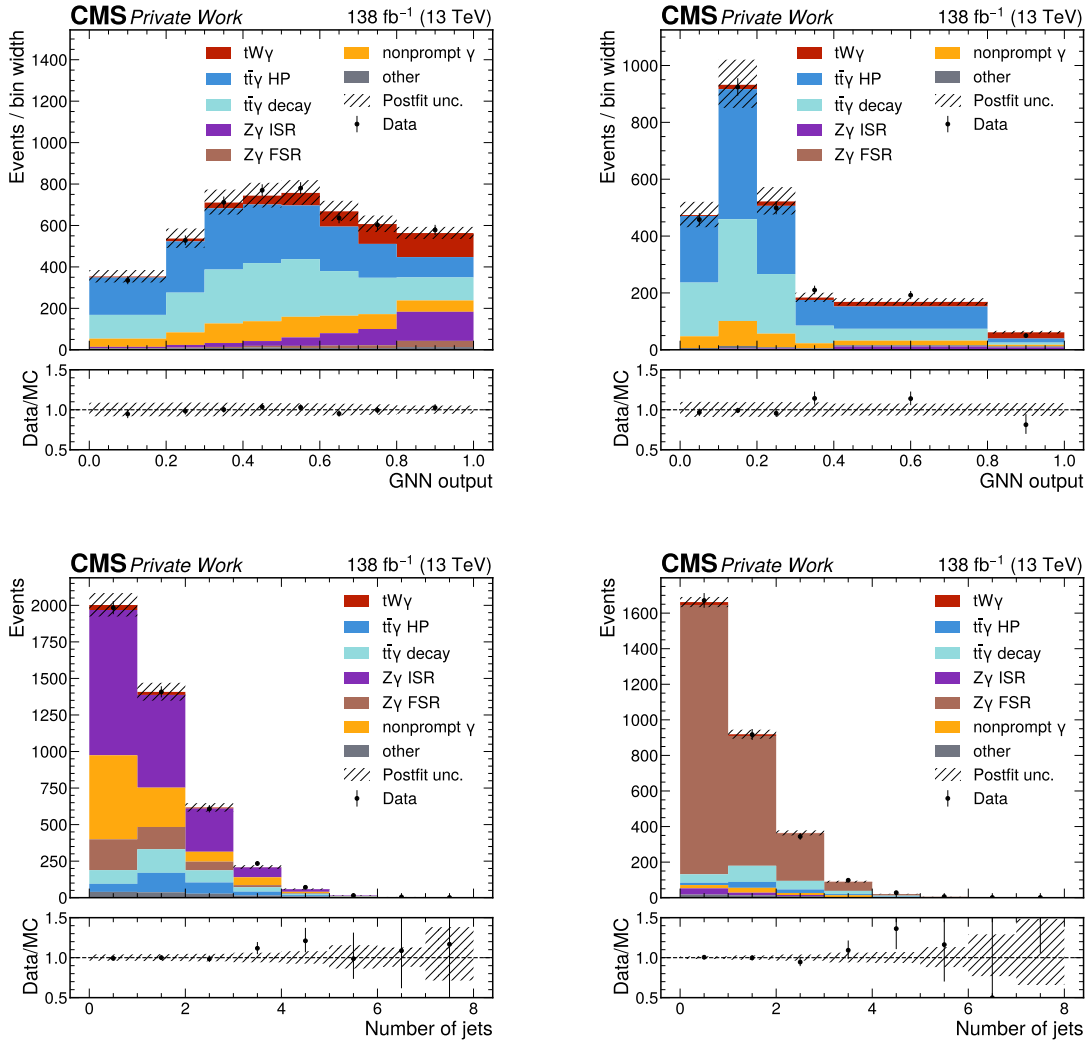


Figure 6.19: Post-fit distributions of the GNN output for the $\mathrm{SR}_{tW\gamma}$ (upper left) and $\mathrm{SR}_{t\bar{t}\gamma}$ (upper right), and of the number of jets for the $\mathrm{CR}_{\mathrm{ISR}}$ (lower left) and $\mathrm{CR}_{\mathrm{FSR}}$ (lower right).

Summary and conclusions

The unprecedented amount of hadron collision data collected by the LHC experiments at the center-of-mass energy of 13 TeV has made possible to observe and study rare processes that were previously inaccessible, because of their low production cross section and high production energy thresholds. The establishment and study of rare processes is an important test of the Standard Model (SM) of particle physics, as deviations observed between theory and observed data can be a sign of new physics beyond the SM.

The first searches for the associated production of a single top quark with a W boson and an additional photon ($tW\gamma$) and Z boson (tWZ) at the LHC are presented in this thesis work. Both analyses have been performed using data collected by the CMS experiment in 2016, 2017, and 2018, corresponding to an integrated luminosity of 138 fb^{-1}.

The results of the tWZ analysis have been published in Ref [111] and represent first evidence for the production of the tWZ process. The analysis is performed in the final state with at least three leptons, where at least one particle, either the top quark or the W boson, decays leptonically, and the Z boson decays into a pair of electrons or muons. This channel has been chosen because of the low background contamination, at the cost of lower branching fractions of the particle decays involved in the process.

In order to select the maximum possible number of signal events, the analyzed phase space is divided into several signal regions, which have different contributions from the signal and background processes. One such region is the *boosted* signal region, where the top quark is produced with a high transverse momentum, defined through the introduction of a boosted leptonic top-tagger: a dedicated machine learning (ML) algorithm that identifies pairs of leptons and b jets compatible with the decay of a top quark with high transverse momentum. The performance of the boosted leptonic top-tagger is validated in a dedicated phase space, which was chosen to be highly pure in events of the $t\bar{t}\gamma$ process, with two charged leptons in the final state.

In order to maximize the sensitivity of the analysis, several fully connected feed-forward neural networks (NN) are trained to discriminate the signal from the background; one for each signal region. Furthermore, control regions are included in the analysis to guarantee a correct estimation of the background processes.

A maximum likelihood fit of the simulated signal and background events to the observed data is performed in bins of the output of the NN classifiers, in order to extract the signal

strength and statistical significance of the tWZ process. The result of the fit shows a signal strength of

$$\mu_{tWZ} = 2.6 \pm 0.4 \text{ (stat)} \pm 0.7 \text{ (syst)}, \tag{6.7}$$

which corresponds to a statistical significance of 3.4 standard deviations: the first evidence for the production of the tWZ process. The measurement has a relative uncertainty of 31%, in large part due to the impact of the systematic uncertainties on the signal strength. The measurement is in agreement with the SM prediction within two standard deviations.

Additional data collected by the CMS experiment in Run 3 of the LHC, which initiated in 2022, will allow to further increase the statistical significance of the tWZ process, and to measure its production cross section with a better precision. This could possibly lead to the observation of the process and an improved understanding of the tension between the measured and predicted values of the cross section.

Additionally, the results of the first search for the $tW\gamma$ process at the LHC are presented in this thesis. The analysis finds its foundations in the modeling studies presented in Chapter 2, which allowed to determine a reliable modeling of the signal process at NLO in QCD. The analysis is performed in the final state where the top quark and the W boson decay leptonically.

The main challenge of the $tW\gamma$ analysis is the separation of the signal from the remarkably similar $t\bar{t}\gamma$ process, which has a much higher production cross section. For this purpose, a dedicated ML binary classifier is developed, which is based on a Graph Neural Network. The graph structure allows the classifier to exploit the correlations between all particles, and all jets in the final state, which carry information fundamental for the discrimination of the signal from the background. The input features of the GNN are validated through saturated goodness-of-fit tests in the analyzed phase space.

Furthermore, the analysis includes dedicated control regions to guarantee a correct estimation of the background $Z\gamma$ process, whose modeling is found to be in consistent disagreement with the observed data in the phase space of the analysis.

The output of the GNN, calculated for the signal and background events, is used as the distribution that is fitted to observed data in order to extract the signal strength and the statistical significance of the $tW\gamma$ process. The signal strength is found to be

$$\mu_{tW\gamma} = 0.94^{+0.24}_{-0.22}, \tag{6.8}$$

well in agreement with the SM prediction within one standard deviation. The statistical significance of the $tW\gamma$ process is found to be 4.97 standard deviations, which represents first evidence for the production of the $tW\gamma$ process at the LHC.

Similarly to what has been stated for the tWZ process, the additional data collected by the CMS experiment in Run 3 of the LHC will allow to further increase the statistical significance of the $tW\gamma$ process, leading to the observation of the process.

Additionally, the analysis would possibly benefit from improvements in modeling the $Z\gamma$ process, and most importantly, the $tW\gamma$ and $t\bar{t}\gamma$ processes, especially concerning the case where the photon in the event is radiated from the decay products of the top quark or the W boson, or the top antiquark for $t\bar{t}\gamma$. One possible advancement in this direction could consist in the simulation of this production mechanism at matrix-element level, as opposed to the current approach, which relies on the parton shower in Monte Carlo event generators.

In conclusion, this thesis work has presented the first searches for the associated production of a single top quark with a W boson and an additional photon ($tW\gamma$) and Z boson (tWZ) at the LHC, establishing the first evidence for the production of both processes. These results have been obtained through the use of dedicated ML algorithms, which have been fundamental in separating the sought signal from the overwhelming $t\bar{t}\gamma$ and $t\bar{t}Z$ backgrounds.

The results of these analyses represent an important starting point for further studies at the LHC, which will allow to establish the observation of the processes, and to shed light on possible tensions between the SM predictions and the measured properties of the associated production modes of a single top quark with two vector bosons.

Bibliography

[1] C. D. Anderson. "The Positive Electron". In: *Phys. Rev.* 43 (6 Mar. 1933), pp. 491–494. DOI: 10.1103/PhysRev.43.491. URL: https://link.aps.org/doi/10.1103/PhysRev.43.491.

[2] P. A. M. Dirac. "The Quantum Theory of the Electron". In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 117.778 (Feb. 1928), pp. 610–624. ISSN: 2053-9150. DOI: 10.1098/rspa.1928.0023. URL: http://dx.doi.org/10.1098/rspa.1928.0023.

[3] The CMS collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[4] The ATLAS collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.020.

[5] N. Aghanim et al. "Planck 2018 results: VI. Cosmological parameters". In: *Astronomy & Astrophysics* 641 (Sept. 2020), A6. ISSN: 1432-0746. DOI: 10.1051/0004-6361/201833910. URL: http://dx.doi.org/10.1051/0004-6361/201833910.

[6] M. Thomson. *Modern particle physics*. New York: Cambridge University Press, Oct. 2013. ISBN: 978-1-107-03426-6, 978-1-139-52536-7. DOI: 10.1017/CBO9781139525367.

[7] A. Bettini. *Introduction to Elementary Particle Physics*. 3rd ed. Cambridge University Press, 2024.

[8] F. Halzen and A. D. Martin. *QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS*. 1984. ISBN: 978-0-471-88741-6.

[9] L. H. Ryder. *Quantum Field Theory*. Cambridge University Press, June 1996. ISBN: 978-0-521-47814-4, 978-1-139-63239-3, 978-0-521-23764-2. DOI: 10.1017/CBO9780511813900.

[10] E. Fermi. "Sulla quantizzazione del gas perfetto monoatomico". In: *Rendiconti Lincei* 3 (1926), pp. 145–149.

[11] P. A. M. Dirac. "On the theory of quantum mechanics". In: *Proc. R. Soc. Lond. A Math. Phys. Sci.* 112.762 (Oct. 1926), pp. 661–677.

[12] S. N. Bose. “Planck’s Law and the Hypothesis of Light Quanta”. In: *Zeitschrift für Physik* 26 (1924), pp. 178–181.

[13] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Physical Review Letters* 81.8 (Aug. 1998), pp. 1562–1567. ISSN: 1079-7114. DOI: 10.1103/physrevlett.81.1562. URL: http://dx.doi.org/10.1103/PhysRevLett.81.1562.

[14] S. Navas et al. “Review of particle physics”. In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[15] 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.

[16] P. 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.

[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] E. Fermi. “Tentativo di una Teoria Dei Raggi β”. In: *Il Nuovo Cimento* 11.1 (Jan. 1934), pp. 1–19. ISSN: 1827-6121. DOI: 10.1007/bf02959820. URL: http://dx.doi.org/10.1007/BF02959820.

[19] J. Schwinger. “A theory of the fundamental interactions”. In: *Annals of Physics* 2.5 (1957), pp. 407–434. ISSN: 0003-4916. DOI: https://doi.org/10.1016/0003-4916(57)90015-5. URL: https://www.sciencedirect.com/science/article/pii/0003491657900155.

[20] 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.

[21] M. Goldhaber, L. Grodzins, and A. W. Sunyar. “Helicity of Neutrinos”. In: *Phys. Rev.* 109 (3 Feb. 1958), pp. 1015–1017. DOI: 10.1103/PhysRev.109.1015. URL: https://link.aps.org/doi/10.1103/PhysRev.109.1015.

[22] S. L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[23] S. 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.

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

[25] M. Kobayashi and T. Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652. eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf. URL: https://doi.org/10.1143/PTP.49.652.

[26] F. Hasert et al. “Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment”. In: *Physics Letters B* 46.1 (1973), pp. 138–140. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(73)90499-1. URL: https://www.sciencedirect.com/science/article/pii/0370269373904991.

[27] H. Yukawa. “On the Interaction of Elementary Particles. I”. In: *Progress of Theoretical Physics Supplement* 1 (Jan. 1955), pp. 1–10. ISSN: 0375-9687. DOI: 10.1143/PTPS.1.1. eprint: https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.1.1/5310694/1-1.pdf. URL: https://doi.org/10.1143/PTPS.1.1.

[28] M. Gell-Mann. “Symmetries of Baryons and Mesons”. In: *Phys. Rev.* 125 (3 Feb. 1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067. URL: https://link.aps.org/doi/10.1103/PhysRev.125.1067.

[29] P. A. M. Dirac. “The Quantum Theory of Emission and Absorption of Radiation”. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 114.767 (Mar. 1927), pp. 243–265. ISSN: 2053-9150. DOI: 10.1098/rspa.1927.0039. URL: http://dx.doi.org/10.1098/rspa.1927.0039.

[30] R. P. Feynman. “Space-Time Approach to Quantum Electrodynamics”. In: *Phys. Rev.* 76 (6 Sept. 1949), pp. 769–789. DOI: 10.1103/PhysRev.76.769. URL: https://link.aps.org/doi/10.1103/PhysRev.76.769.

[31] S. Bethke. “The 2009 world average of α_s”. In: *The European Physical Journal C* 64.4 (Oct. 2009), pp. 689–703. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-009-1173-1. URL: http://dx.doi.org/10.1140/epjc/s10052-009-1173-1.

[32] J. C. Collins, D. E. Soper, and G. F. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[33] R. D. Ball et al. “Parton distributions from high-precision collider data: NNPDF Collaboration”. In: *The European Physical Journal C* 77.10 (Oct. 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5199-5. URL: http://dx.doi.org/10.1140/epjc/s10052-017-5199-5.

[34] V. N. Gribov and L. N. Lipatov. “Deep inelastic e p scattering in perturbation theory”. In: *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[35] Y. L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[36] G. Altarelli and G. Parisi. “Asymptotic Freedom in Parton Language”. In: *Nucl. Phys. B* 126 (1977), pp. 298–318.

[37] V. A. Abramovsky, V. N. Gribov, and O. V. Kancheli. “Character of Inclusive Spectra and Fluctuations Produced in Inelastic Processes by Multi - Pomeron Exchange”. In: *Yad. Fiz.* 18 (1973), pp. 595–616.

[38] T. Sjostrand. “Multiple Parton-Parton Interactions in Hadronic Events”. In: *23rd International Conference on High-Energy Physics*. Aug. 1985.

[39] P. Bartalini et al. *Multi-Parton Interactions at the LHC*. 2011. arXiv: 1111.0469 [hep-ph]. URL: https://arxiv.org/abs/1111.0469.

[40] C. Bierlich et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. 2022. arXiv: 2203.11601 [hep-ph]. URL: https://arxiv.org/abs/2203.11601.

[41] B. Andersson, G. Gustafson, and B. Soderberg. “A General Model for Jet Fragmentation”. In: *Z. Phys. C* 20 (1983), p. 317. DOI: 10.1007/BF01407824.

[42] T. Sjostrand. “Jet Fragmentation of Nearby Partons”. In: *Nucl. Phys. B* 248 (1984), pp. 469–502. DOI: 10.1016/0550-3213(84)90607-2.

[43] G. Hooft. “A planar diagram theory for strong interactions”. In: *Nuclear Physics B* 72.3 (1974), pp. 461–473. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(74)90154-0. URL: https://www.sciencedirect.com/science/article/pii/0550321374901540.

[44] F. Abe et al. “Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab”. In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[45] S. Abachi et al. “Observation of the Top Quark”. In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[46] The CMS collaboration. “Measurement of differential $t\bar{t}$ production cross sections in the full kinematic range using lepton+jets events from proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. D* 104.9 (2021), p. 092013. DOI: 10.1103/PhysRevD.104.092013. arXiv: 2108.02803 [hep-ex].

[47] The CMS collaboration. “Measurement of the single top quark and antiquark production cross sections in the t channel and their ratio in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 800 (2020), p. 135042. DOI: 10.1016/j.physletb.2019.135042. arXiv: 1812.10514 [hep-ex].

[48] The CMS collaboration. *CMS Standard Model summary plots*. Accessed: 11-2024.

[49] A. Denner and S. Dittmaier. “The complex-mass scheme for perturbative calculations with unstable particles”. In: *Nuclear Physics B - Proceedings Supplements* 160 (Oct. 2006), pp. 22–26. ISSN: 0920-5632. DOI: 10.1016/j.nuclphysbps.2006.09.025. URL: http://dx.doi.org/10.1016/j.nuclphysbps.2006.09.025.

[50] S. Frixione et al. “Automated simulations beyond the Standard Model: supersymmetry”. In: *Journal of High Energy Physics* 2019.12 (Dec. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep12(2019)008. URL: http://dx.doi.org/10.1007/JHEP12(2019)008.

[51] S. Frixione et al. “Single-top hadroproduction in association with a W boson”. In: *Journal of High Energy Physics* 2008.07 (July 2008), pp. 029–029. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/07/029. URL: http://dx.doi.org/10.1088/1126-6708/2008/07/029.

[52] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079.

[53] F. Demartin et al. “tWH associated production at the LHC”. In: *The European Physical Journal C* 77.1 (Jan. 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-4601-7. URL: http://dx.doi.org/10.1140/epjc/s10052-017-4601-7.

[54] H. El Faham et al. “Single top production in association with a WZ pair at the LHC in the SMEFT”. In: *Journal of High Energy Physics* 2022.1 (Jan. 2022). ISSN: 1029-8479. DOI: 10.1007/jhep01(2022)100. URL: http://dx.doi.org/10.1007/JHEP01(2022)100.

[55] The CMS collaboration. “Observation of tW production in the single-lepton channel in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* 2021.11 (Nov. 2021), p. 111. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2021)111. URL: https://doi.org/10.1007/JHEP11(2021)111.

[56] The CMS collaboration. “Measurement of inclusive and differential cross sections for single top quark production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* 2023.7 (July 2023), p. 46. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2023)046. URL: https://doi.org/10.1007/JHEP07(2023)046.

[57] A. Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *The European Physical Journal C* 75.3 (Mar. 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3318-8. URL: http://dx.doi.org/10.1140/epjc/s10052-015-3318-8.

[58] The ATLAS collaboration. “Measurements of inclusive and differential cross-sections of combined $t\bar{t}\gamma$ and $tW\gamma$ production in the $e\mu$ channel at 13 TeV with the ATLAS detector”. In: *Journal of High Energy Physics* 2020.9 (Sept. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep09(2020)049. URL: http://dx.doi.org/10.1007/JHEP09(2020)049.

[59] P. Artoisenet et al. “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”. In: *Journal of High Energy Physics* 2013.3 (Mar. 2013). ISSN: 1029-8479. DOI: 10.1007/jhep03(2013)015. URL: http://dx.doi.org/10.1007/JHEP03(2013)015.

[60] S. Frixione. "Isolated photons in perturbative QCD". In: *Physics Letters B* 429.3–4 (June 1998), pp. 369–374. ISSN: 0370-2693. DOI: 10.1016/s0370-2693(98)00454-7. URL: http://dx.doi.org/10.1016/S0370-2693(98)00454-7.

[61] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[62] A. L. Read. "Presentation of search results: the CL_s technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), pp. 2693–2704. ISSN: 0954-3899. DOI: 10.1088/0954-3899/28/10/313. URL: http://dx.doi.org/10.1088/0954-3899/28/10/313.

[63] W. Fisher. *Experiences from Tevatron Searches*. 2008. DOI: 10.5170/CERN-2008-001.61. URL: https://cds.cern.ch/record/1099970.

[64] The CMS collaboration. "The CMS statistical analysis and combination tool: Combine". Submitted to *Comput. Softw. Big Sci.* 2024. arXiv: 2404.06614 [physics.data-an].

[65] R. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. ISSN: 0010-4655. DOI: https://doi.org/10.1016/0010-4655(93)90005-W. URL: https://www.sciencedirect.com/science/article/pii/001046559390005W.

[66] W. A. Rolke, A. M. López, and J. Conrad. "Limits and confidence intervals in the presence of nuisance parameters". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 551.2–3 (Oct. 2005), pp. 493–503. ISSN: 0168-9002. DOI: 10.1016/j.nima.2005.05.068. URL: http://dx.doi.org/10.1016/j.nima.2005.05.068.

[67] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360. URL: https://doi.org/10.1214/aoms/1177732360.

[68] S. M. Lee and G. A. Young. "Parametric bootstrapping with nuisance parameters". In: *Statistics & Probability Letters* 71.2 (2005), pp. 143–153. ISSN: 0167-7152. DOI: https://doi.org/10.1016/j.spl.2004.10.026. URL: https://www.sciencedirect.com/science/article/pii/S0167715204002895.

[69] The ATLAS and CMS collaborations, and LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1379837.

[70] F. Rosenblatt. "The perceptron: A probabilistic model for information storage and organization in the brain". In: *Psychological Review* 65.6 (1958), pp. 386–408. DOI: 10.1037/h0042519.

[71] K. He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV]. URL: https://arxiv.org/abs/1512.03385.

[72] K. He et al. *Identity Mappings in Deep Residual Networks*. 2016. arXiv: 1603.05027 [cs.CV]. URL: https://arxiv.org/abs/1603.05027.

[73] X. Glorot, A. Bordes, and Y. Bengio. "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by G. Gordon, D. Dunson, and M. Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, Apr. 2011, pp. 315–323. URL: https://proceedings.mlr.press/v15/glorot11a.html.

[74] A. L. Maas. "Rectifier Nonlinearities Improve Neural Network Acoustic Models". In: 2013. URL: https://api.semanticscholar.org/CorpusID:16489696.

[75] P. Ramachandran, B. Zoph, and Q. V. Le. *Searching for Activation Functions*. 2017. arXiv: 1710.05941 [cs.NE]. URL: https://arxiv.org/abs/1710.05941.

[76] G. Klambauer et al. *Self-Normalizing Neural Networks*. 2017. arXiv: 1706.02515 [cs.LG]. URL: https://arxiv.org/abs/1706.02515.

[77] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[78] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Learning representations by back-propagating errors". en. In: *Nature* 323.6088 (Oct. 1986), pp. 533–536.

[79] F. Scarselli et al. "The Graph Neural Network Model". In: *IEEE Transactions on Neural Networks* 20.1 (2009), pp. 61–80. DOI: 10.1109/TNN.2008.2005605.

[80] Y. Shi et al. *Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification*. 2021. arXiv: 2009.03509 [cs.LG]. URL: https://arxiv.org/abs/2009.03509.

[81] A. Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL]. URL: https://arxiv.org/abs/1706.03762.

[82] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2. URL: https://www.sciencedirect.com/science/article/pii/0370269383911772.

[83] M. Banner et al. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". In: *Physics Letters B* 122.5 (1983), pp. 476–485. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91605-2. URL: https://www.sciencedirect.com/science/article/pii/0370269383916052.

[84] G. Arnison et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider". In: *Physics Letters B* 126.5 (1983), pp. 398–410. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)90188-0. URL: https://www.sciencedirect.com/science/article/pii/0370269383901880.

[85] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[86] The CMS collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008). Also published by CERN Geneva in 2010, S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[87] The ATLAS collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3 (2008). Also published by CERN Geneva in 2010, S08003. DOI: 10.1088/1748-0221/3/08/S08003. URL: https://cds.cern.ch/record/1129811.

[88] The ALICE collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08002.

[89] The LHCb collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08005.

[90] M. Benedikt et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-3. URL: https://cds.cern.ch/record/823808.

[91] E. Mobs. "The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019". In: (2019). General Photo. Accessed: 11-2024. URL: https://cds.cern.ch/record/2684277.

[92] *LEP design report*. Report. Copies shelved as reports in LEP, PS and SPS libraries. Geneva: CERN, 1984. URL: https://cds.cern.ch/record/102083.

[93] The CMS collaboration. *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13$ TeV. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2676164.

[94] S. van der Meer. *Calibration of the effective beam height in the ISR*. Tech. rep. Geneva: CERN, 1968. URL: https://cds.cern.ch/record/296752.

[95] The CMS collaboratio. *CMS Luminosity - Public Results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 11-2024.

[96] CERN. *CERN CMS images gallery*. https://home.cern/resources/image/experiments/cms-images-gallery. Accessed: 11-2024.

[97] I. Neutelings. *CMS coordinate system*. https://tikz.net/axis3d_cms/. Accessed: 11-2024.

[98] The CMS collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. ISSN: 1748-0221. DOI: 10.1088/1748-0221/9/10/p10009. URL: http://dx.doi.org/10.1088/1748-0221/9/10/P10009.

[99] Tracker Group of the CMS collaboration. "The CMS Phase-1 Pixel Detector Upgrade". In: *Journal of Instrumentation* 16.02 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304. URL: https://cds.cern.ch/record/2748381.

[100] T. C. collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 16 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[101] The CMS collaboration. *ECAL 2016 refined calibration and Run2 summary plots*. CMS Detector Performance Summary CMS-DP-2020-021. 2020. URL: https://cds.cern.ch/record/2717925.

[102] The CMS collaboration. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Journal of Instrumentation* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[103] The CMS collaboration. "The CMS trigger system". In: *Journal of Instrumentation* 12 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[104] The CMS collaboration. "Performance of the CMS high-level trigger during LHC Run 2". In: (Oct. 2024). arXiv: 2410.17038 [physics.ins-det].

[105] The CMS collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[106] W. Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/934067.

[107] R. E. Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1 (Mar. 1960), pp. 35–45. ISSN: 0021-9223. DOI: 10.1115/1.3662552. eprint: https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35_1.pdf. URL: https://doi.org/10.1115/1.3662552.

[108] W. Adam et al. "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC". In: *Journal of Physics G: Nuclear and Particle Physics* 31.9 (July 2005), N9–N20. ISSN: 1361-6471. DOI: 10.1088/0954-3899/31/9/n01. URL: http://dx.doi.org/10.1088/0954-3899/31/9/N01.

[109] "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *Journal of Instrumentation* 10.08 (Aug. 2015), P08010. DOI: 10.1088/1748-0221/10/08/P08010. URL: https://dx.doi.org/10.1088/1748-0221/10/08/P08010.

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

[111] The CMS collaboration. "Evidence for tWZ production in proton-proton collisions at $\sqrt{s}$=13 TeV in multilepton final states". In: *Physics Letters B* 855 (2024), p. 138815. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2024.138815. URL: https://www.sciencedirect.com/science/article/pii/S0370269324003733.

[112] HEPData record for the search for tWZ. 2023. DOI: 10.17182/hepdata.138419.

[113] F. Maltoni, L. Mantani, and K. Mimasu. "Top-quark electroweak interactions at high energy". In: *Journal of High Energy Physics* 2019.10 (Oct. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep10(2019)004. URL: http://dx.doi.org/10.1007/JHEP10(2019)004.

[114] A. Kulesza et al. "Associated top quark pair production with a heavy boson: differential cross sections at NLO + NNLL accuracy". In: *The European Physical Journal C* 80.5 (May 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-7987-6. URL: http://dx.doi.org/10.1140/epjc/s10052-020-7987-6.

[115] "Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data". In: (2020). URL: https://cds.cern.ch/record/2740688.

[116] The CMS collaboration. "Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: *Journal of High Energy Physics* 2022.2 (Feb. 2022). ISSN: 1029-8479. DOI: 10.1007/jhep02(2022)107. URL: http://dx.doi.org/10.1007/JHEP02(2022)107.

[117] The CMS collaboration. "Observation of four top quark production in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: *Phys. Lett. B* 847 (2023). All the figures and tables, including additional supplementary figures and tables, can be found at http://cms-results.web.cern.ch/cms-results/public-results/publications/TOP-22-013 (CMS Public Pages), p. 138290. DOI: 10.1016/j.physletb.2023.138290. arXiv: 2305.13439. URL: https://cds.cern.ch/record/2859484.

[118] The CMS collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV". In: *Journal of Instrumentation* 13.06 (June 2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. URL: https://dx.doi.org/10.1088/1748-0221/13/06/P06015.

[119] A. Bodek et al. "Extracting muon momentum scale corrections for hadron collider experiments". In: *The European Physical Journal C* 72.10 (Oct. 2012). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-2194-8. URL: http://dx.doi.org/10.1140/epjc/s10052-012-2194-8.

[120] The CMS collaboration. "Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s}$=7 TeV". In: *Physics Letters B* 713.2 (2012), pp. 68–90. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.05.028. URL: https://www.sciencedirect.com/science/article/pii/S0370269312005564.

[121] The CMS collaboration. "Search for supersymmetry in pp collisions at $\sqrt{s} = 13$ TeV in the single-lepton final state using the sum of masses of large-radius jets". In: *Journal of High Energy Physics* 2016.8 (Aug. 2016). ISSN: 1029-8479. DOI: 10.1007/jhep08(2016)122. URL: http://dx.doi.org/10.1007/JHEP08(2016)122.

[122] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-ktjet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://dx.doi.org/10.1088/1126-6708/2008/04/063.

[123] The CMS collaboration. "Pileup mitigation at CMS in 13 TeV data". In: *Journal of Instrumentation* 15.09 (Sept. 2020), P09018–P09018. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/09/p09018. URL: http://dx.doi.org/10.1088/1748-0221/15/09/P09018.

[124] The CMS collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/02/p02014. URL: http://dx.doi.org/10.1088/1748-0221/12/02/P02014.

[125] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. URL: https://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[126] "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector". In: (2018). URL: https://cds.cern.ch/record/2646773.

[127] P. Sturm. "Studies for the Measurement of Single-Top-Quark-Events with the CMS-Experiment. Studien zur Messung von Einzel-Top-Quark-Ereignissen mit dem CMS-Experiment". Presented on 17 Mar 2008. Karlsruhe U., 2008. URL: https://cds.cern.ch/record/1311220.

[128] The CMS collaboration. "Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2022.7 (July 2022). ISSN: 1029-8479. DOI: 10.1007/jhep07(2022)032. URL: http://dx.doi.org/10.1007/JHEP07(2022)032.

[129] The CMS collaboration. "Measurements of pp $\rightarrow$ ZZ production cross sections and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13\,\text{TeV}$". In: *The European Physical Journal C* 81.3 (Mar. 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-08817-8. URL: http://dx.doi.org/10.1140/epjc/s10052-020-08817-8.

[130] The CMS collaboration. "Search for the associated production of the Higgs boson with a top-quark pair". In: *Journal of High Energy Physics* 2014.9 (Sept. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep09(2014)087. URL: http://dx.doi.org/10.1007/JHEP09(2014)087.

[131] The CMS collaboration. “Performance of electron and photon reconstruction in Run 2 with the CMS experiment”. In: (2020). URL: https://cds.cern.ch/record/2725004.

[132] F. Chollet et al. *Keras*. https://keras.io. 2015.

[133] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[134] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG]. URL: https://arxiv.org/abs/1502.03167.

[135] G. E. Hinton et al. *Improving neural networks by preventing co-adaptation of feature detectors*. 2012. arXiv: 1207.0580 [cs.NE]. URL: https://arxiv.org/abs/1207.0580.

[136] A. J. Larkoski et al. “Soft drop”. In: *Journal of High Energy Physics* 2014.5 (May 2014). ISSN: 1029-8479. DOI: 10.1007/jhep05(2014)146. URL: http://dx.doi.org/10.1007/JHEP05(2014)146.

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

[138] N. Smirnov. “Table for Estimating the Goodness of Fit of Empirical Distributions”. In: *The Annals of Mathematical Statistics* 19.2 (1948), pp. 279–281. DOI: 10.1214/aoms/1177730256. URL: https://doi.org/10.1214/aoms/1177730256.

[139] T. Fawcett. “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8 (2006). ROC Analysis in Pattern Recognition, pp. 861–874. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2005.10.010. URL: https://www.sciencedirect.com/science/article/pii/S016786550500303X.

[140] The CMS collaboration. “Precision luminosity measurement in proton–proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS”. In: *The European Physical Journal C* 81.9 (Sept. 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09538-2. URL: http://dx.doi.org/10.1140/epjc/s10052-021-09538-2.

[141] The CMS collaboration. *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13$ TeV. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2621960.

[142] The CMS collaboration. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{(s)} = 7$ TeV”. In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. URL: https://dx.doi.org/10.1088/1748-0221/7/10/P10002.

[143] The CMS collaboration. “Performance summary of AK4 jet b tagging with data from proton-proton collisions at 13 TeV with the CMS detector”. In: (2023). URL: https://cds.cern.ch/record/2854609.

[144] The CMS collaboration. "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.

[145] The CMS collaboration. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of Instrumentation* 15.10 (Oct. 2020), P10017–P10017. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/10/p10017. URL: http://dx.doi.org/10.1088/1748-0221/15/10/P10017.

[146] The CMS collaboration. "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). ISSN: 1029-8479. DOI: 10.1007/jhep03(2020)056. URL: http://dx.doi.org/10.1007/JHEP03(2020)056.

[147] O. B. Bylund et al. *Probing top quark neutral couplings in the Standard Model Effective Field Theory at NLO QCD*. 2016. arXiv: 1601.08193 [hep-ph]. URL: https://arxiv.org/abs/1601.08193.

[148] The CMS collaboration. "Measurement of the inclusive and differential $t\bar{t}\gamma$ cross sections in the dilepton channel and effective field theory interpretation in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2022.5 (May 2022). ISSN: 1029-8479. DOI: 10.1007/jhep05(2022)091. URL: http://dx.doi.org/10.1007/JHEP05(2022)091.

[149] The CMS collaboration. "Measurements of the pp $\rightarrow$ WZ inclusive and differential production cross sections and constraints on charged anomalous triple gauge couplings at 13 TeV". In: *Journal of High Energy Physics* 2019.4 (Apr. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep04(2019)122. URL: http://dx.doi.org/10.1007/JHEP04(2019)122.

[150] The CMS collaboration. "Measurements of the pp $\rightarrow$ ZZ production cross section and the Z$\rightarrow 4\ell$ branching fraction, and constraints on anomalous triple gauge couplings at 13 TeV". In: *The European Physical Journal C* 78.2 (Feb. 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-5567-9. URL: http://dx.doi.org/10.1140/epjc/s10052-018-5567-9.

[151] The CMS collaboration. "Observation of the Production of Three Massive Gauge Bosons at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett.* 125 (15 Oct. 2020), p. 151802. DOI: 10.1103/PhysRevLett.125.151802. URL: https://link.aps.org/doi/10.1103/PhysRevLett.125.151802.

[152] H. Paukkunen and P. Zurita. "PDF reweighting in the Hessian matrix approach". In: *Journal of High Energy Physics* 2014.12 (Dec. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep12(2014)100. URL: http://dx.doi.org/10.1007/JHEP12(2014)100.

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

[154] S. Frixione, P. Nason, and C. Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

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

[156] J. M. Campbell et al. “Top-Pair Production and Decay at NLO Matched with Parton Showers”. In: *JHEP* 04 (2015), p. 114. DOI: 10.1007/JHEP04(2015)114. arXiv: 1412.1828 [hep-ph].

[157] E. Re. “Single-top Wt-channel production matched with parton showers using the POWHEG method”. In: *Eur. Phys. J. C* 71 (2011), p. 1547. DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450 [hep-ph].

[158] J. Snoek, H. Larochelle, and R. P. Adams. “Practical Bayesian Optimization of Machine Learning Algorithms”. In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc., 2012. URL: https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf.

[159] M. Balandat et al. “BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization”. In: *Advances in Neural Information Processing Systems 33*. 2020. URL: http://arxiv.org/abs/1910.06403.

Appendix

A Reconstruction of the longitudinal momentum of neutrinos arising in the decays of top quarks

The reconstruction of the neutrino in the decay of top quarks is based on imposing the conservation of energy in the $W\ell\nu$ vertex, where the W boson originates from the decay of a top quark. The neutrino is considered as the only source of MET in the event, hence it is assumed

$$\vec{p}_{T,\nu} = \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}. \tag{6.1}$$

Therefore, the longitudinal component of the neutrino momentum $\vec{p}_{\nu}$ is calculated as the solution of the equation:

$$(\vec{p}_{\nu} + \vec{p}_{\ell})^2 = m_W^2. \tag{6.2}$$

The solutions are in the form:

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

where

$$\mu = \frac{m_W^2}{2} + p_{T,\ell} \cdot p_{T,\nu} \cdot \cos(\Phi) = \frac{m_W^2 - m_{T,W}^2}{2} + p_{T,\ell} \cdot p_{T,\nu}, \tag{6.4}$$

with m_W being the mass of the W boson, Φ the azimuthal angle between the lepton and the neutrino, and $m_{T,W}$ the transverse mass of the W boson. When the discriminant in Eq. (6.3) is positive, two real solutions are found, and both are taken into account for the reconstruction of the top quark. The discriminant is negative when $m_{T,W} > m_W$, which is possible because of the imperfect resolution of the MET. Naively taking the real part of the right-hand side of Eq. (6.3) as the solution would result in Eq.(6.1) not being satisfied. In order to find a $p_{z,\nu}$ for which the energy conservation constraint is respected, it is imposed

that $M_{T,W} = m_W$, which is equivalent to a quadratic relationship between $p_{x,\nu}$ and $p_{y,\nu}$, with two possible solutions for each value of $p_{x,\nu}$: $p_{y1/2,\nu}(p_{x,\nu})$.

The measurement of the MET is taken as reliable, hence the choice of the value of $\bar{p}_{x,\nu}$ and the solution $\bar{p}_{y,\nu}$ (between $p_{y1,\nu}$ or $P_{y2,\nu}$) is done by minimizing the difference between the reconstructed p_{T} of the neutrino and MET in the event:

$$\delta_{1/2}(p_{x,\nu}) = \left| \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} - \vec{p}_{T1/2,\nu}(p_{x,\nu}) \right|, \tag{6.5}$$

where $\vec{p}_{T1(2),\nu}$ is calculated using the solution $p_{y1(2),\nu}$.

Hence, the reconstructed neutrino momentum is:

$$\bar{p}_{x,\nu} = \underset{p_{x,\nu}}{\mathrm{argmin}} \min_{i=1,2} \delta_i(p_{x,\nu}), \tag{6.6}$$

$$\bar{p}_{y,\nu} = p_{yk,\nu}(\bar{p}_{x,\nu}), \tag{6.7}$$

with

$$k = \underset{i=1,2}{\mathrm{argmin}}\, \delta_i(\bar{p}_{x,\nu}). \tag{6.8}$$

B Validation of the input features of the leptonic boosted top-tagger

In this appendix, the comparison between data and MC is shown for all the input features of the leptonic boosted top-tagger presented in Section 5.4, with the plot for electrons on the left and muons on the right. The *boosted* label in some of the variables shown means that the variable is considered in the rest frame of the top candidate.

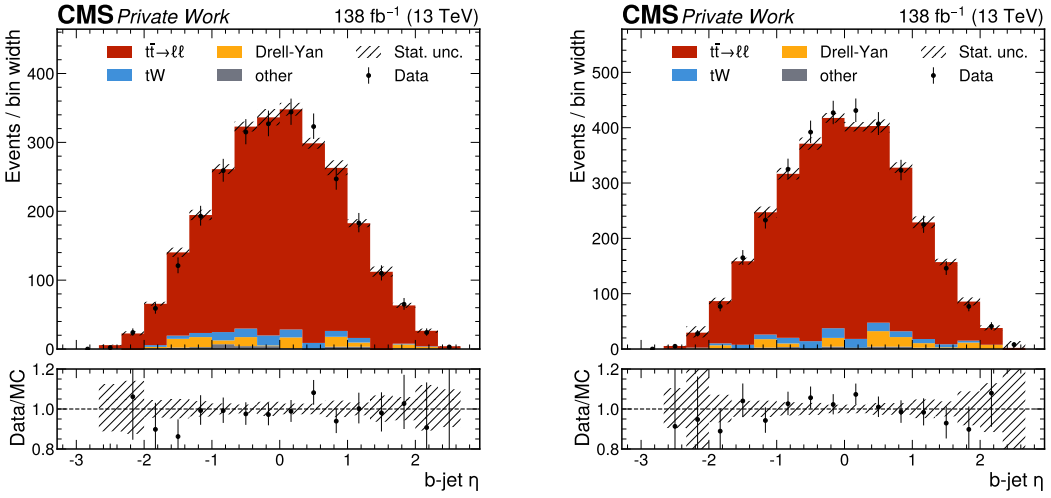


Figure B.1: Distribution of the η of the b-jet for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

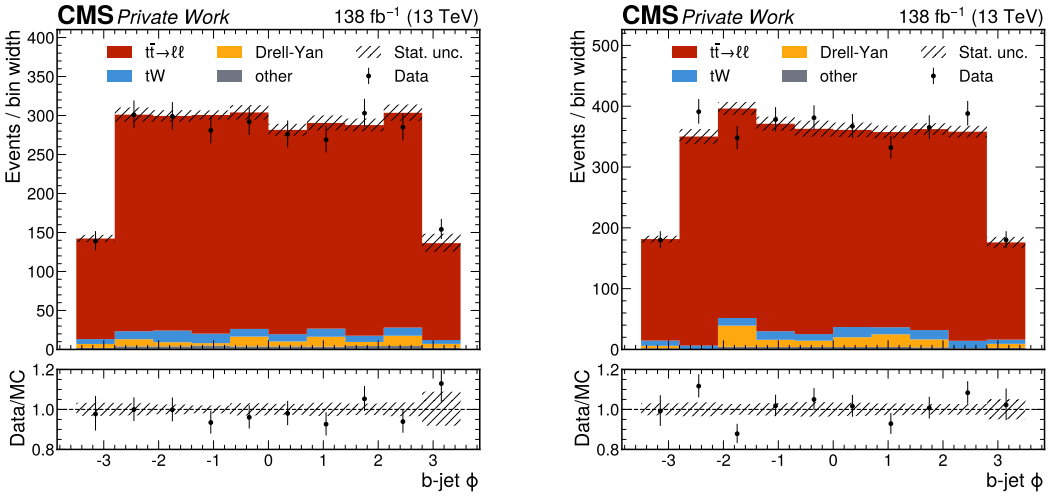


Figure B.2: Distribution of the ϕ of the b-jet for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

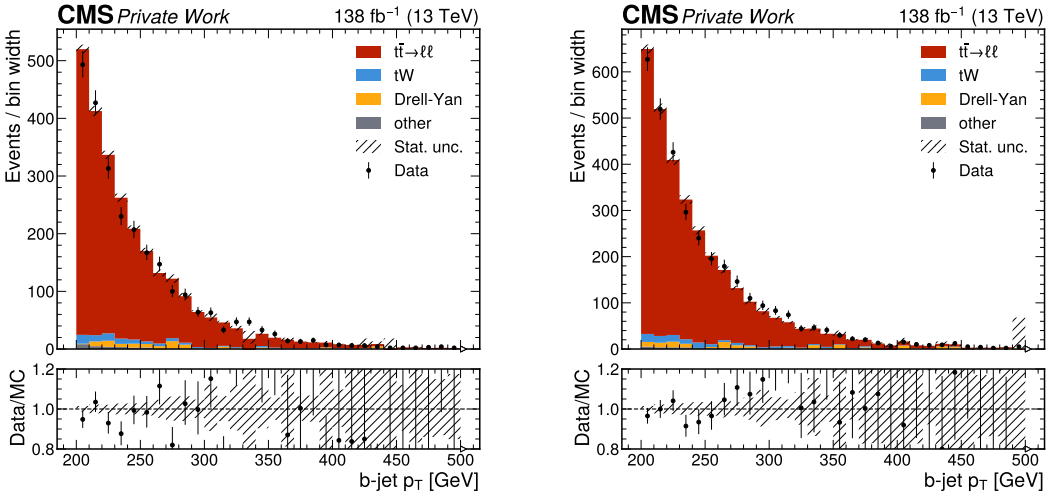


Figure B.3: Distribution of the p_T of the b-jet for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

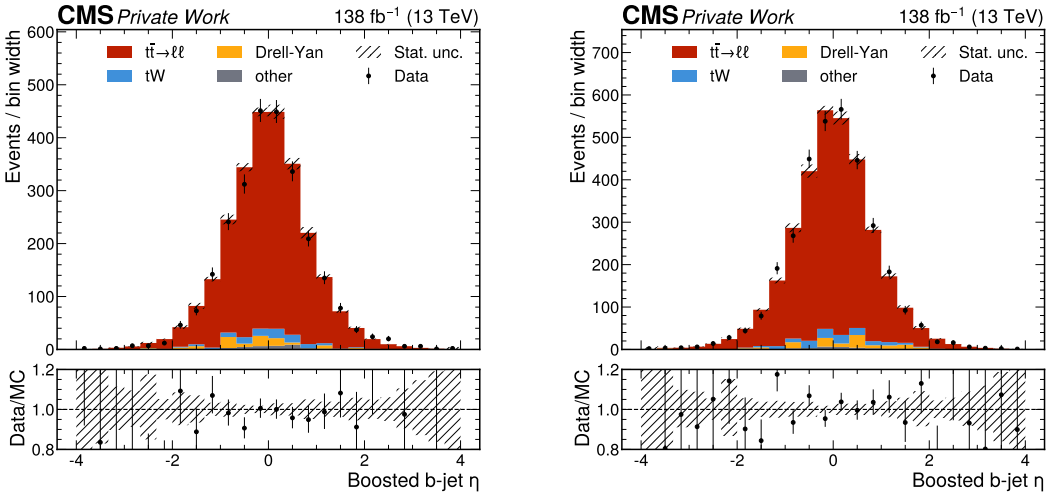


Figure B.4: Distribution of the η of the b-jet in the rest frame of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

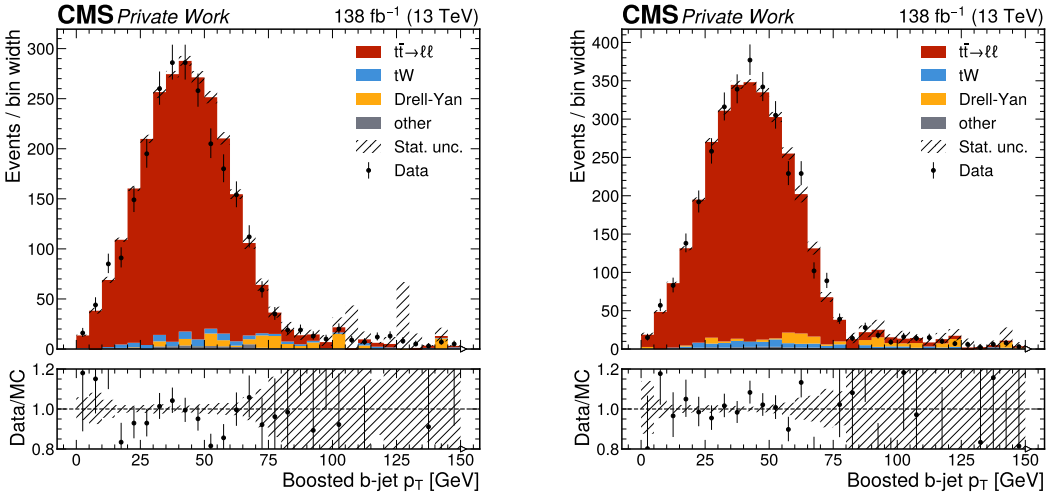


Figure B.5: Distribution of the p_T of the b-jet in the rest frame of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

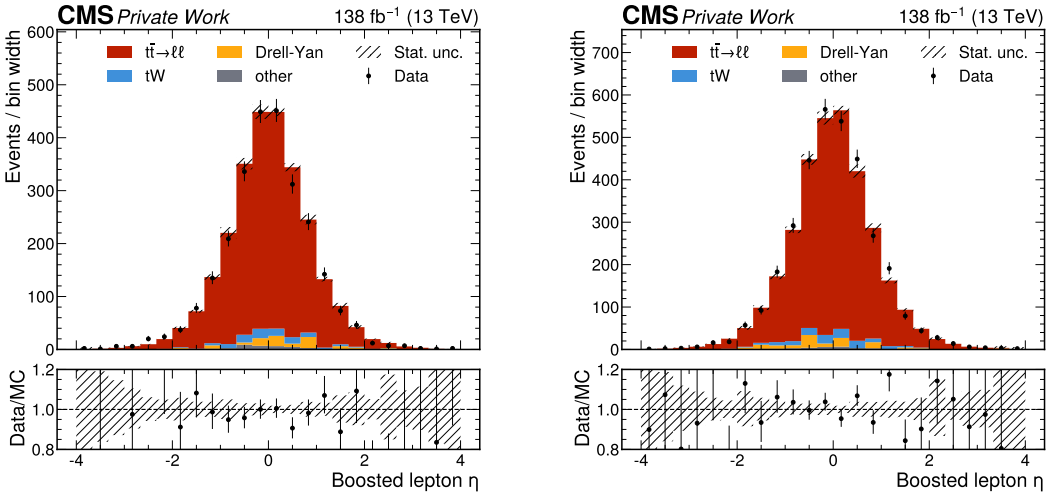


Figure B.6: Distribution of the η of the lepton in the rest frame of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

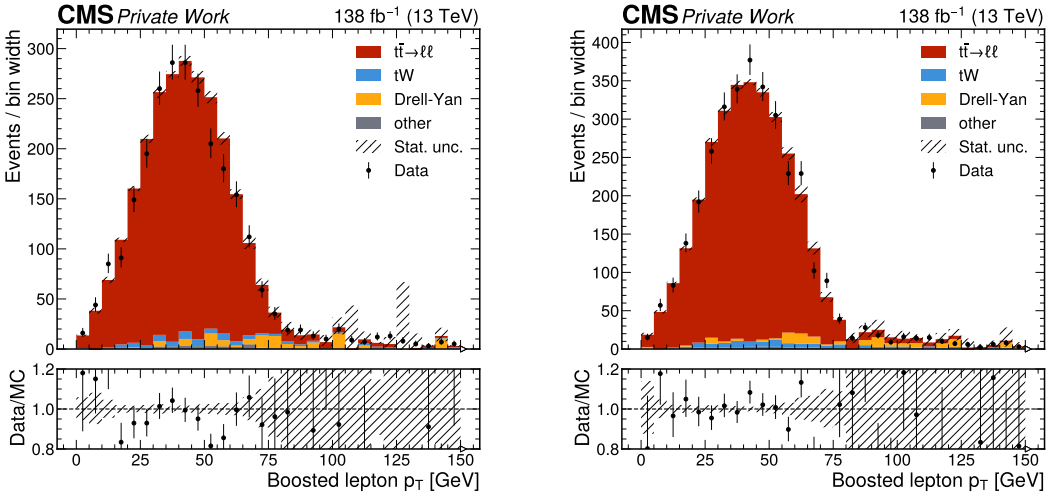


Figure B.7: Distribution of the p_T of the lepton in the rest frame of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

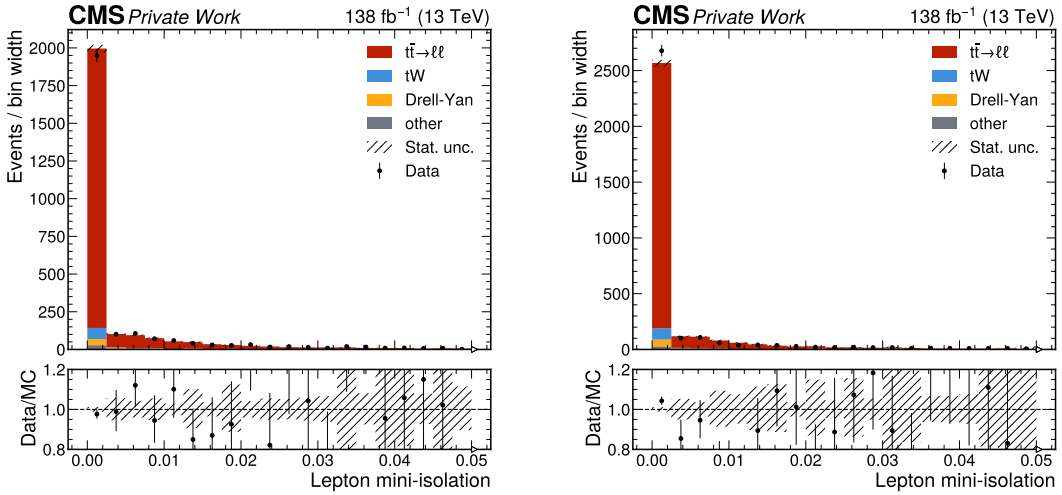


Figure B.8: Distribution of the mini-isolation of the lepton for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

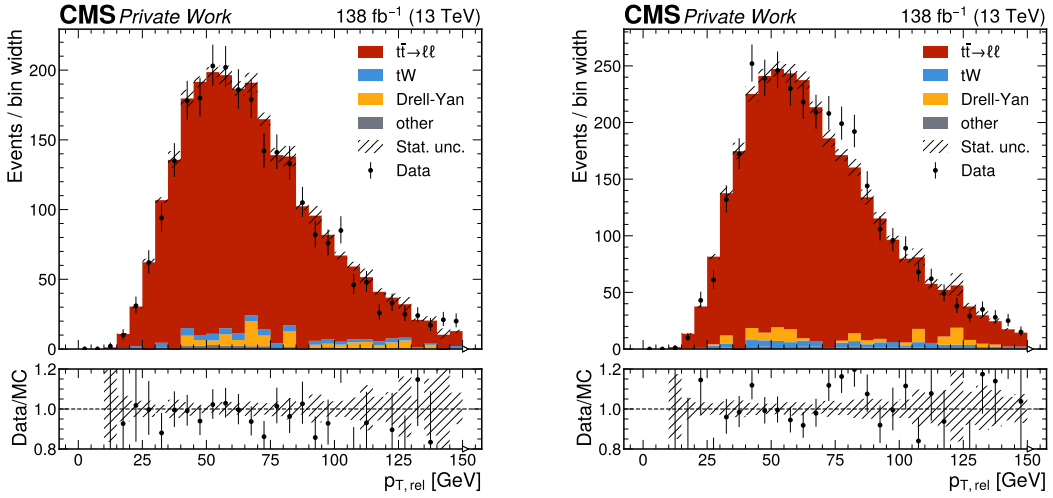


Figure B.9: Distribution of the $p_{T,rel}$ variable defined in Section 5.4 for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

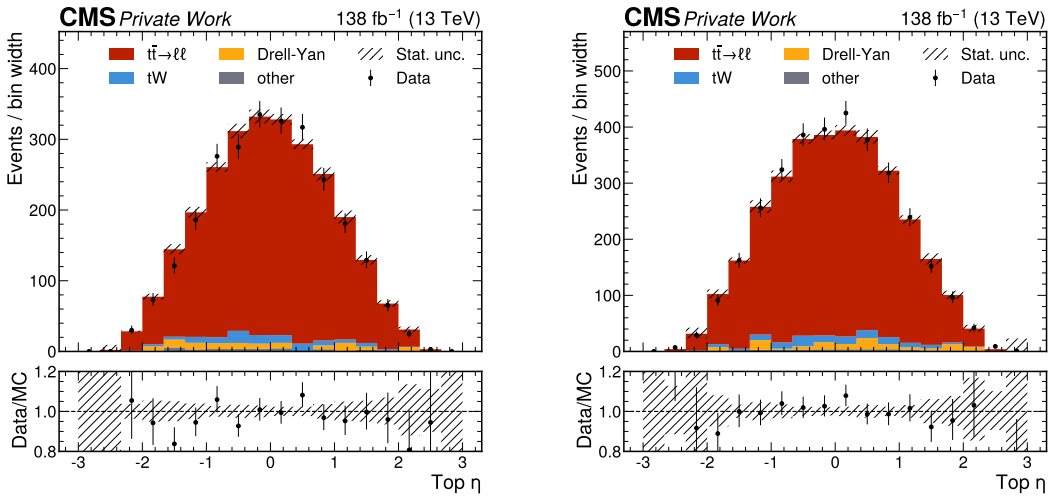


Figure B.10: Distribution of the η of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

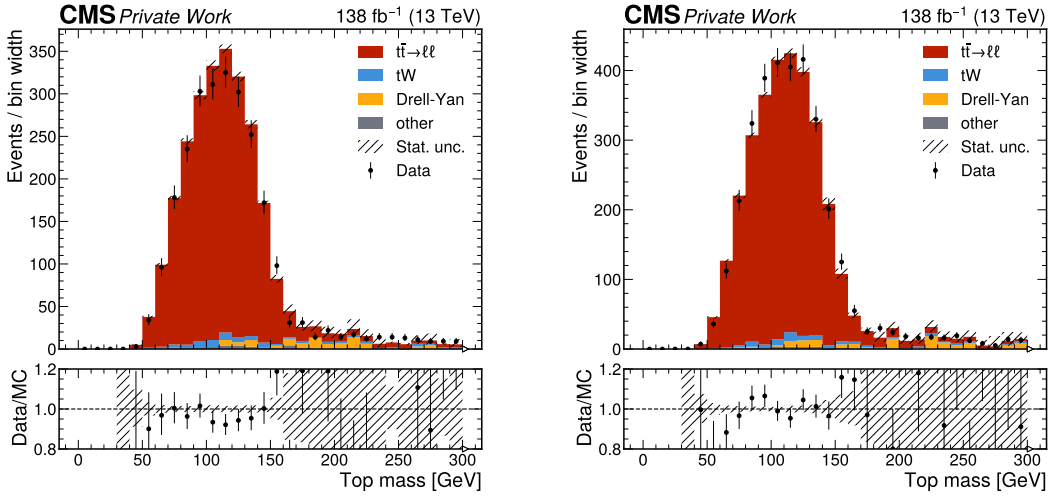


Figure B.11: Distribution of the invariant mass of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

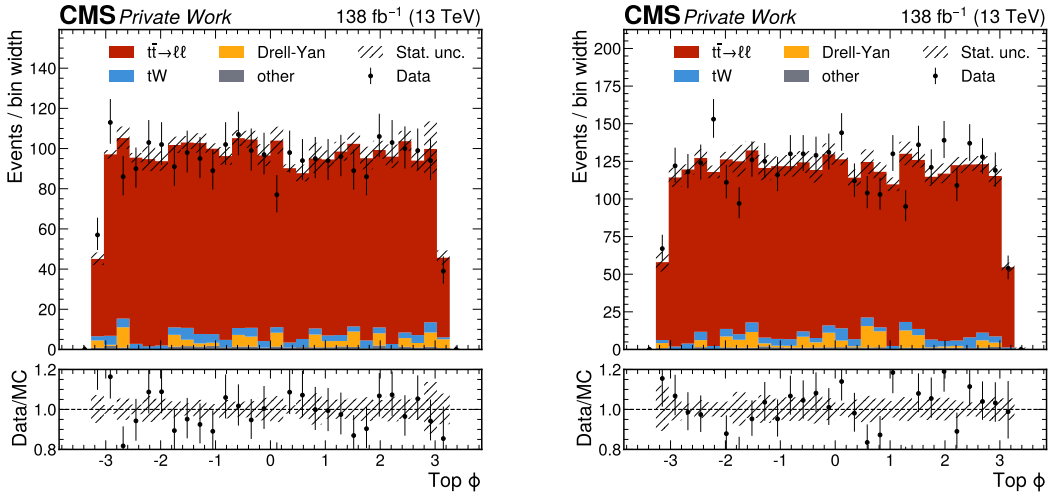


Figure B.12: Distribution of the ϕ of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

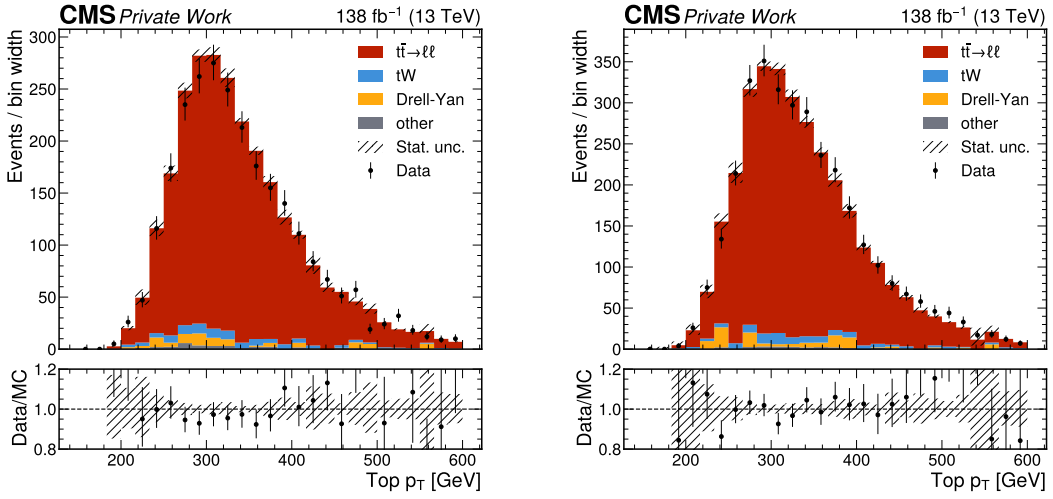


Figure B.13: Distribution of the p_T of the top candidate for electrons (left) and muons (right) in the validation region. The lower insets show the bin-by-bin ratio between data and MC.

C Saturated GoF results for the validation of the input variables to the GNN

This appendix reports the results of the saturated GoF tests employed for the validation of the input features to the GNN presented in Section 6.4.

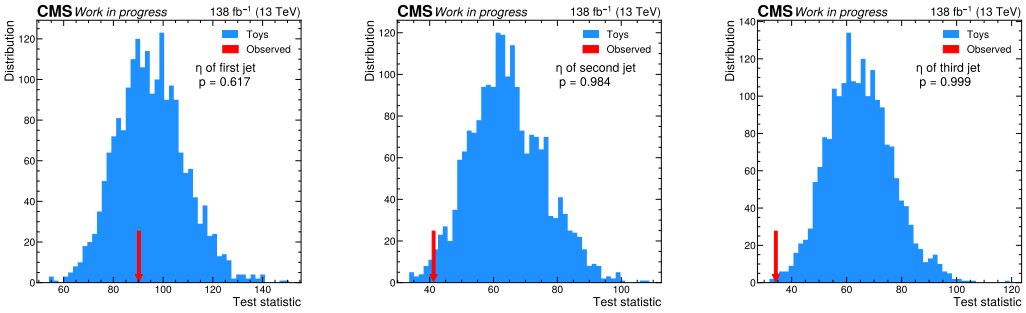


Figure C.14: Distribution of the test statistic used in the saturated GoF test for the η of first jet(left) and the η of second jet(center) and the η of third jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

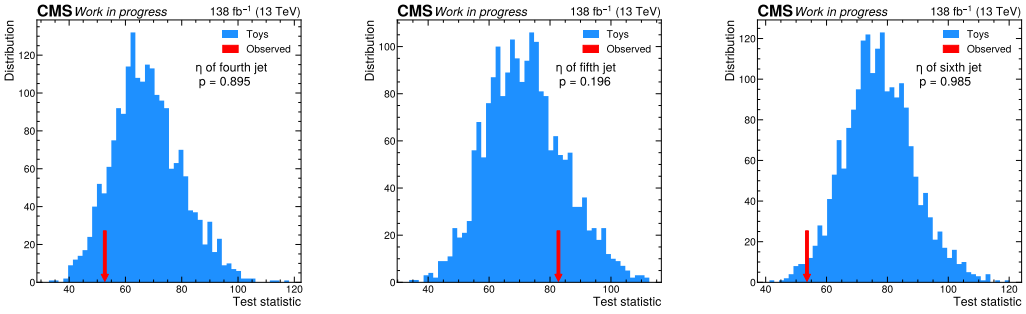


Figure C.15: Distribution of the test statistic used in the saturated GoF test for the η of fourth jet(left) and the η of fifth jet(center) and the η of sixth jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

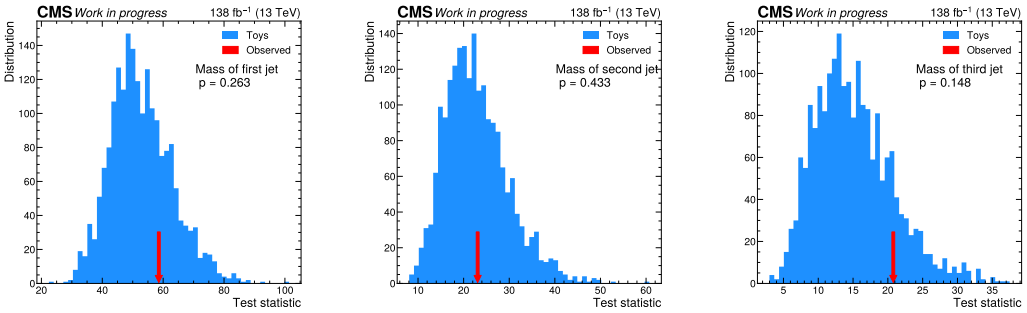


Figure C.16: Distribution of the test statistic used in the saturated GoF test for the mass of first jet(left) and the mass of second jet(center) and the mass of third jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

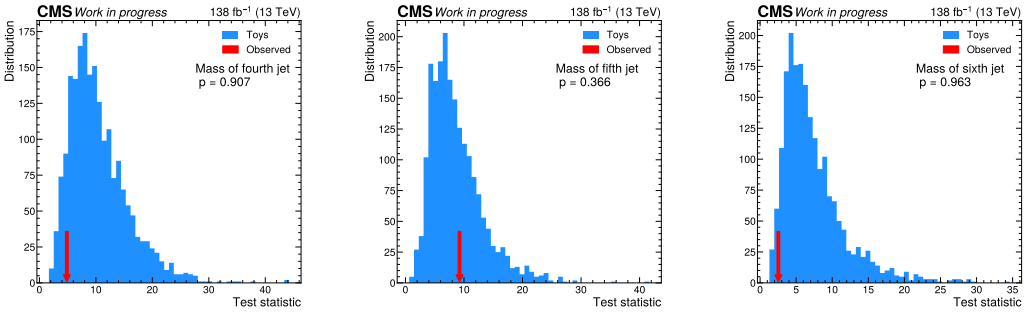


Figure C.17: Distribution of the test statistic used in the saturated GoF test for the mass of fourth jet(left) and the mass of fifth jet(center) and the mass of sixth jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

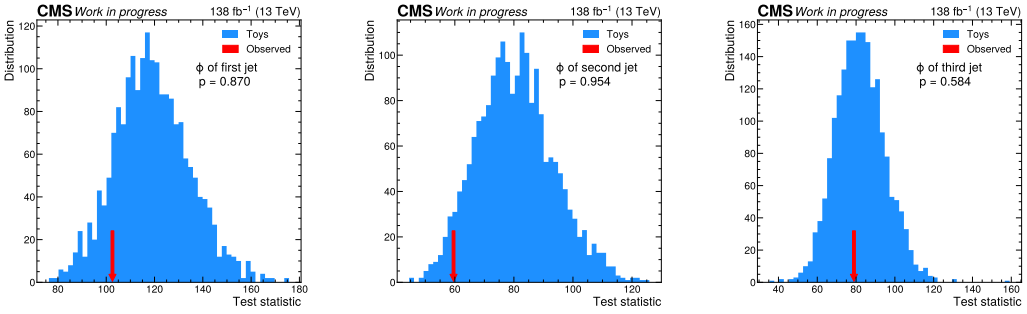


Figure C.18: Distribution of the test statistic used in the saturated GoF test for the ϕ of first jet(left) and the ϕ of second jet(center) and the ϕ of third jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

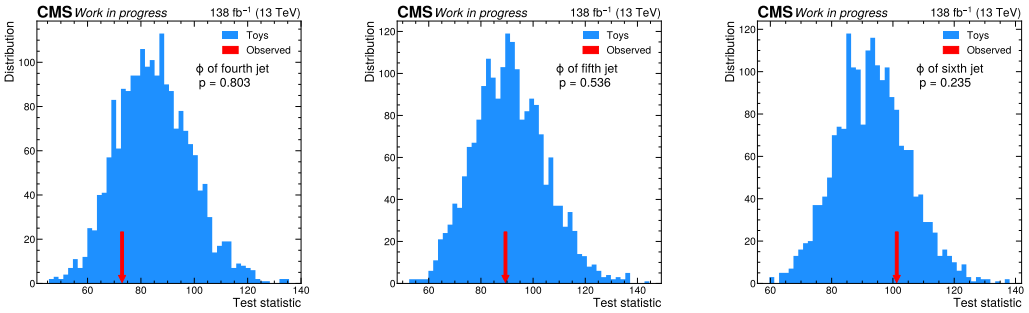


Figure C.19: Distribution of the test statistic used in the saturated GoF test for the ϕ of fourth jet(left) and the ϕ of fifth jet(center) and the ϕ of sixth jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

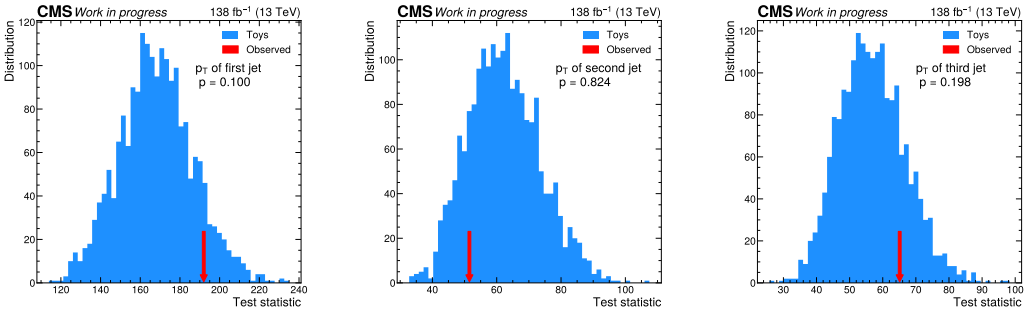


Figure C.20: Distribution of the test statistic used in the saturated GoF test for the p_T of first jet(left) and the p_T of second jet(center) and the p_T of third jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

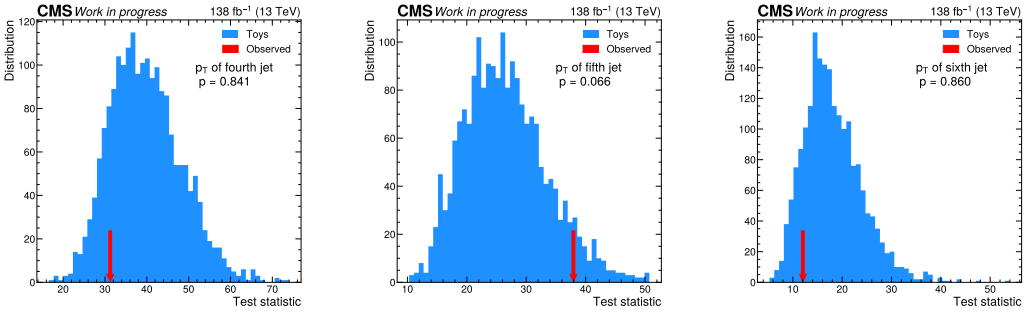


Figure C.21: Distribution of the test statistic used in the saturated GoF test for the p_T of fourth jet(left) and the p_T of fifth jet(center) and the p_T of sixth jetright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

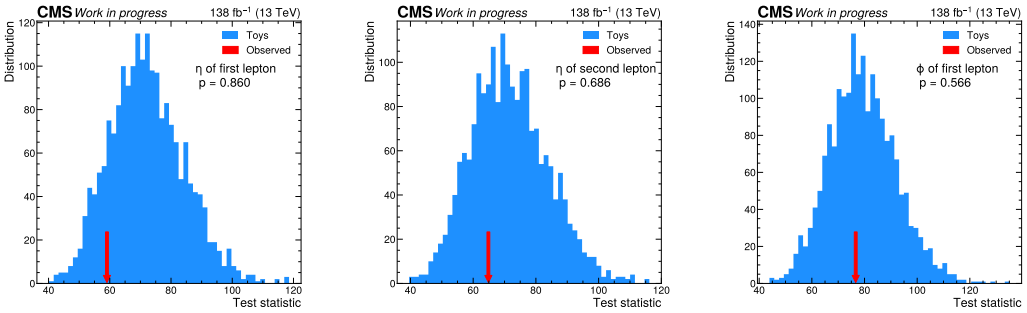


Figure C.22: Distribution of the test statistic used in the saturated GoF test for the η of first lepton(left) and the η of second lepton(center) and the ϕ of first leptonright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

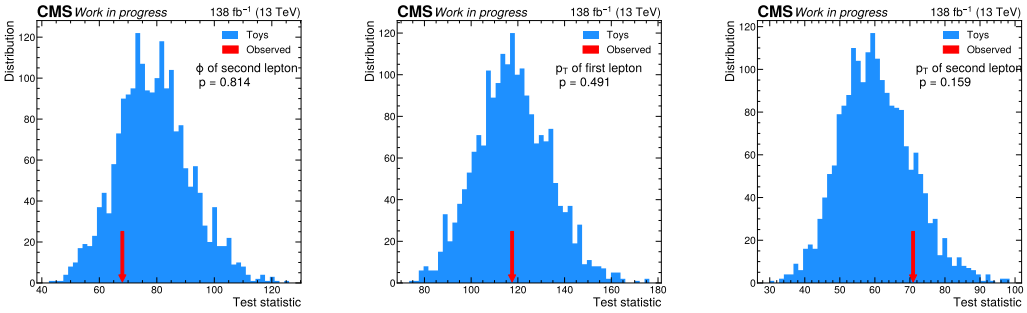


Figure C.23: Distribution of the test statistic used in the saturated GoF test for the ϕ of second lepton(left) and the p_T of first lepton(center) and the p_T of second leptonright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

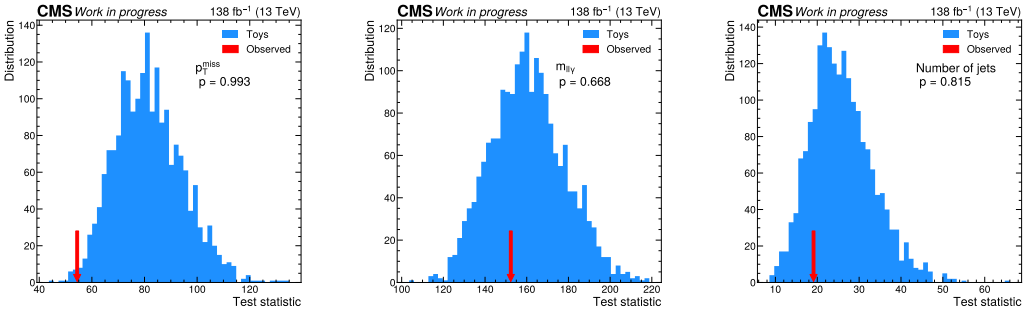


Figure C.24: Distribution of the test statistic used in the saturated GoF test for the p_T^{miss}(left) and the $m_{ll\gamma}$(center) and the number of jetsright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

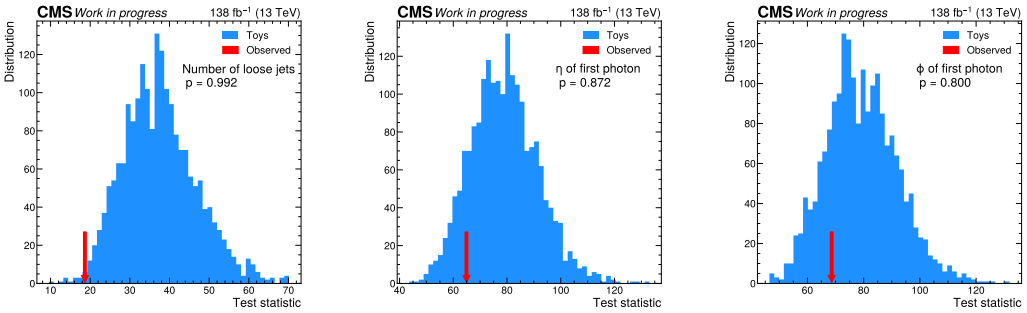


Figure C.25: Distribution of the test statistic used in the saturated GoF test for the number of loose jets(left) and the η of first photon(center) and the ϕ of first photonright compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.

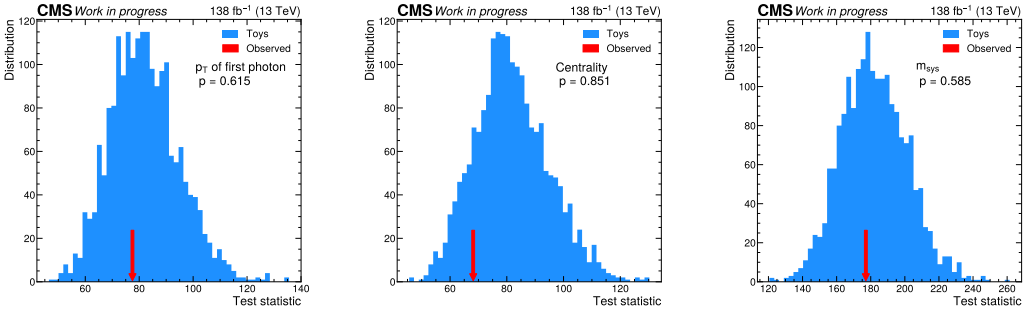


Figure C.26: Distribution of the test statistic used in the saturated GoF test for the p_T of first photon(left) and the Centrality(center) and the m_{sys} right compared to the values t_d obtained from the fit to data. The p-values are reported in the plots.