Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

First simultaneous measurement of single and pair production of top quarks in association with a Z boson at the LHC

Dissertation

zur Erlangung des Doktorgrades an der Fakultät
für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Federica Cecilia Colombina

Hamburg
2024

Declaration on oath

I hereby declare in lieu of oath that I have written this dissertation myself and that I have not used any auxiliary materials or sources other than those indicated.

Hamburg, May 28, 2024

Federica Cecilia Colombina

Federica Cecilia Colombina

Gutachter der Dissertation:	PD Dr. Andreas Meyer
	Prof. Dr. Johannes Haller
Zusammensetzung der Prüfungskommission:	Prof. Dr. Johannes Haller
	Prof. Dr. Dieter Horns
	Prof. Dr. Geraldine Servant
	PD Dr. Andreas Meyer
	Dr. Maria Aldaya Martin
Vorsitzender der Prüfungskommission:	Prof. Dr. Dieter Horns
Datum der Disputation:	13.05.2024
Vorsitzender des Fach-Promotionsausschusses Physik:	Prof. Dr. Markus Drescher
Leiter des Fachbereichs Physik:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

In the years 2016–2018 the CMS Experiment at CERN's Large Hadron Collider recorded a large amount of proton-proton collision data at a centre-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}.

With this large dataset, the associated production of top quarks with the Z boson has been measured precisely and differentially. In this thesis, for the first time, single and pair production of top quarks in association with a Z boson (tZq, tWZ and $\mathrm{t\bar{t}Z}$) are measured simultaneously and differentially. A multiclass deep neural network is used to separate the signal processes and the backgrounds. The interference between $\mathrm{t\bar{t}Z}$ and tWZ complicates their separation, and therefore their sum is reported in the results. Compared to previous studies, the simultaneous measurement is less dependent on the signal modeling assumptions and improves the sensitivity to new physics scenarios, as it enables to constrain possible deviations from the standard model across different processes.

Studies on muon identification methods with the CMS detector are also presented, with particular focus on novel techniques to correct discrepancies between simulation and the data.

Zusammenfassung

In den Jahren 2016–2018 hat das CMS-Experiment am Large Hadron Collider des CERN eine große Menge an Proton-Proton-Kollisionsdaten bei einer Schwerpunkts-Energie von 13 TeV gesammelt. Die Daten entsprechen einer integrierten Luminosität von 138 fb^{-1}.

Mit diesem großen Datensatz wurde die Z-Boson-assoziierte Top-Quark-Produktion bereits präzise und differenziell gemessen. In dieser Arbeit nun wird zum ersten Mal die Einzel- und Paarproduktion von Top-Quarks in Verbindung mit einem Z-Boson (tZq, tWZ und $\mathrm{t\bar{t}Z}$) gleichzeitig und differenziell gemessen. Ein tiefes neuronales Netzwerk für Multiklassifikation wird verwendet, um die Signalprozesse vom Untergrund zu trennen. Die Interferenz zwischen $\mathrm{t\bar{t}Z}$ und tWZ erschwert ihre Trennung, weshalb in den Ergebnissen ihre Summe angegeben wird. Im Vergleich zu früheren Studien ist die simultane Messung weniger abhängig von den Annahmen der Signalmodellierung und verbessert die Empfindlichkeit gegenüber neuen physikalischen Szenarien, da sie es ermöglicht, mögliche Abweichungen vom Standardmodell über verschiedene Prozesse hinweg einzuschränken.

Es werden auch Studien zur Identifizierung von Myonen mit dem CMS-Detektor vorgestellt, mit besonderem Schwerpunkt auf neuartigen Techniken zur Korrektur von Diskrepanzen zwischen Simulation und Daten.

Contents

Introduction **1**

1 Theoretical foundations **5**

1.1 Particle content of the SM 6

1.2 Quantum Electrodynamics 7

1.3 Quantum Chromodynamics 8

1.4 Weak interactions 10

1.5 The Standard Model of particle physics 13

1.6 Proton-proton collisions 15

1.7 The role of the top quark 18

1.8 Top quark production in association with a Z boson 19

1.8.1 Theoretical modeling 21

1.8.2 New physics scenarios 23

2 The LHC and the CMS experiment **27**

2.1 The accelerator complex at CERN 27

2.2 The CMS detector 29

2.2.1 Silicon tracker 31

2.2.2 Electromagnetic and hadronic calorimeters 33

2.2.3 Muon chambers 35

2.2.4 Triggering and data acquisition 37

2.3 Event simulation 38

2.4 Object reconstruction 39
2.4.1 Tracks and primary vertex 40
2.4.2 Jets and b-tagging 40
2.4.3 Electrons 43
2.4.4 Muons 45
2.4.5 Determination of shape-based scale factors 47

3 Event selection and reconstruction 57
3.1 Simulation and datasets 57
3.2 Object selection 59
3.2.1 Leptons 59
3.2.2 Jets 61
3.3 Signal definition 63
3.3.1 Neutrino reconstruction 66
3.3.2 Top quark reconstruction 70
3.4 Nonprompt background estimation 72

4 Classification of signals and background 77
4.1 Multiclass neural networks 77
4.2 Input variables and hyperparameters 80

5 Signal extraction 91
5.1 Systematics 92
5.2 Simultaneous inclusive measurement 97
5.3 Simultaneous differential measurement 104
5.4 Outlook 112

Conclusion 119

A Response matrices 123

B Covariance matrices 125

Introduction

The standard model of particle physics (SM) currently provides the best understanding of the known elementary particles and their interactions, as it can describe almost all the experimental results in the field of particle physics with unprecedented precision. It includes three of the four fundamental forces in nature: the electromagnetic, the strong and the weak force; only gravity is not part of the theory and the development of a common mathematical framework is one of the major open challenges in particle physics. However, gravity becomes negligible at microscopic scales and the SM remains a robust and successful theory for the description of the subatomic world. The SM is believed to be a component of a wider theoretical framework able to incorporate gravity and address the aspects that are still not explained by the theory, like the nature of dark matter and dark energy.

The CERN Large Hadron Collider (LHC) gives the possibility to test the SM interactions at the highest-ever reached collision energy, while looking for possible indications of physics beyond the SM (BSM). The main results presented in this thesis are obtained analyzing data recorded at the Compact Muon Solenoid (CMS) experiment during Run-2, i.e. in the years 2016–2018 from proton-proton collision at the LHC. With this large data set, many rare SM processes could be investigated for the first time; these include, for example, the associated production of top quarks with a Z boson, which is the main topic of this thesis. After almost thirty years since its discovery, the top quark is still of major interest in the particle physics community, and the LHC provides the means to study the top quark to unprecedented detail and precision. With a mass of about 172 GeV, it is the heaviest known elementary particle and the only quark that decays before hadronization, offering the possibility of measuring the bare quark properties. The top quark pair production via the strong interaction is the dominant production mode at the LHC, but the single top quark production is also possible through electroweak processes. The Z-boson associated top quark production, in pairs ($\mathrm{t\bar{t}Z}$) or with a single top quark (tZq and tWZ), provides direct access to the coupling of the top quark to the Z boson, which could be measured for the first time with the Run-2 data set. Furthermore, various BSM scenarios predict significant modifications to these interactions, which enhances the importance of these studies. This thesis presents the first simultaneous measurement of the tZq, $\mathrm{t\bar{t}Z}$ and tWZ cross sections. This approach reduces the dependency on the modeling of the processes and enhances the sensitivity to possible new physics scenarios compared to previous measurements [1–3], where the signals were measured individually while assuming the SM expectations for the other ones. The measurements are performed on the full Run-2 data set produced from proton-proton collisions at the LHC, with a center-of-mass energy equal to 13 TeV. Final states with exactly three leptons, either electrons or muons, are targeted for all the signal processes, as they provide the highest sensitivity. Most of the background contributions are taken from simulated samples, with the exception of those arising from misidentified leptons, which are estimated with a data-driven method. The information coming from the reconstruction of the event topology and in the kinematics of the final state particles are exploited to build a multiclass neural network to improve the discrimination between signals and backgrounds. The final results are obtained performing binned maximum likelihood fits on the neural network output distributions, including both statistical and systematic uncertainties. The signal cross sections are measured inclusively and differentially as functions of five physics observables that are sensitive to the modeling of the processes as well as to BSM scenarios. These results are currently under review by the CMS Collaboration and have been presented for the first time at the conference ”Moriond/EW2024: 58th Rencontres de Moriond on Electroweak Interactions & Unified Theories” in March 2024.

This thesis is structured as follows: first, an overview of the SM and the notions needed to understand the theory behind the experimental results of this thesis is given in Chapter 1. This chapter also contains an overview about the physics of the top quark and its associated production with a Z-boson, starting from Section 1.8. The experimental setup is presented in Chapter 2, with a description of the CMS detector and the techniques employed to reconstruct the final-state particles from the recorded data. A novel technique to fix discrepancies between simulation and the data in muon identification methods is discussed in Section 2.4.5. The main analysis is presented from Chapter 3, starting with a description of the samples and the requirements applied to select signal events and estimate the backgrounds. Chapter 4 is dedicated to the description of the neural network and its training, while the results are presented and discussed in Chapter 5.

CHAPTER 1

Theoretical foundations

In this chapter, a general introduction of the electromagnetic, strong and weak interactions is given, delineating the motivation for the measurement of top quark production in association with a Z boson. The interactions between elementary particles are described by quantum field theories (QFTs), i.e. mathematical frameworks in which particles emerge as excitations of quantum fields. In each of them, the laws that limit the possibilities of a state transformation in a collision or a decay are formulated by the conservation of the corresponding *charges*: the electric charge, the colour charges and the weak charges. The combination of these theoretical frameworks forms the standard model of particle physics (SM), discussed in Section 1.5. This model provides an accurate description of all the fundamental interactions apart from gravity, which to date is only described macroscopically in general relativity. In order to give an overview of the theory behind the main results of this thesis, a description of proton-proton interactions and the physics of the top quark, together with its interaction with the Z boson, is presented in the second part of this chapter.

In this work, the numerical quantities will be expressed using the natural units system, where the reduced Planck constant $\hbar$, the speed of light c and the electric constant ϵ_0 are chosen to be $\epsilon_0 = \hbar = c = 1$; this allows to express all the quantities in the unit GeV and thus simplifying most mathematical expressions. The Feynman representation, introduced by Richard Feynman in 1948 [4], is also employed: it is a widely used method in high energy particle physics to visualise the different interaction processes intuitively, as well as to facilitate the computation of their cross sections. Most of the information contained in this chapter is taken from References [5, 6].

1.1 Particle content of the SM

A summary of all the elementary particles described by the SM and their properties is reported in Figure 1.1.

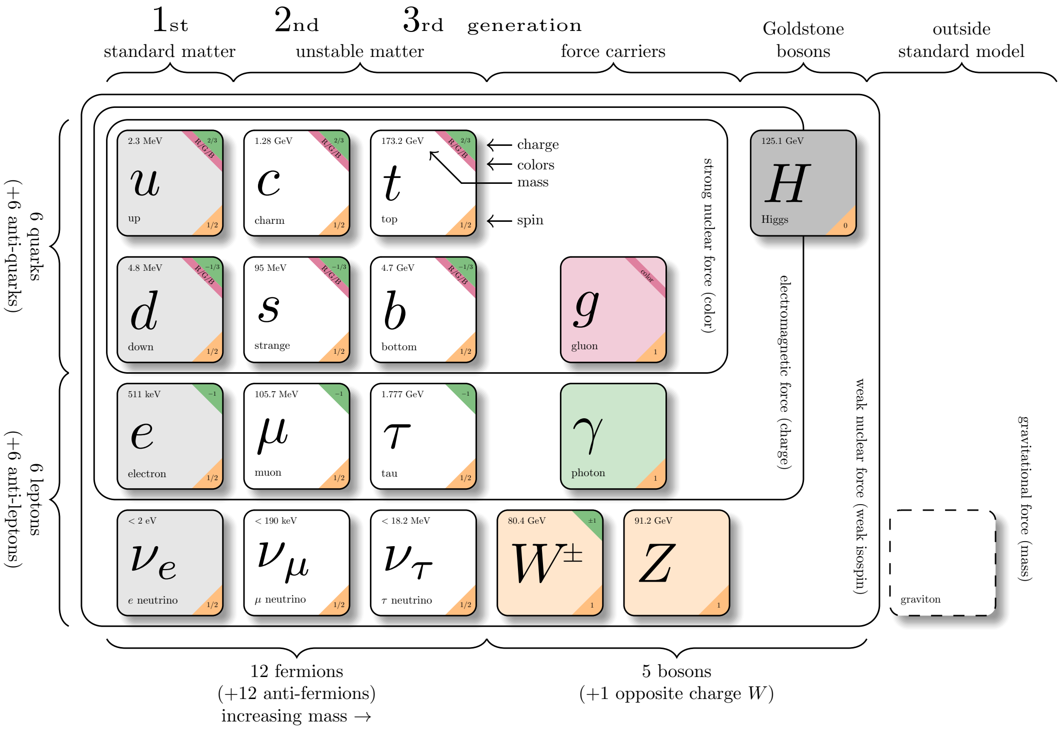


Figure 1.1: Summary of the elementary particles described by the SM, together with their main properties. Taken from Ref. [7].

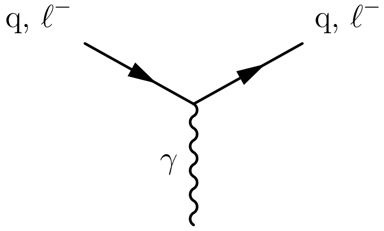
The known elementary fermions, the quarks and the leptons, are grouped into three different generations. Each fermion has a corresponding antiparticle with an opposite sign of all quantum numbers, but the same mass. Each generation has two type of quarks, known as up-type and down-type. The up (u), charge (c) and top

(t) quarks are the up-type quarks of the first to third generation, while the corresponding down-type quarks are the down (d), strange (s) and bottom (b) quarks. The leptons from the three generations are electrons (e), muons (μ) and taus (τ), each of them having a corresponding neutrino: ν_e, ν_μ and ν_τ. The term *flavour* is commonly used to refer to the six different kinds of quarks and leptons. The electromagnetic, strong and weak interactions are all mediated by spin-1 force-carrying particles known as *gauge bosons*. In the case of the electromagnetic and strong force, these are the photon (γ) and the gluon (g), respectively. By contrast, the weak interactions have three distinct gauge bosons, two charged, W^+, W^-, and one neutral, Z. Finally, the Higgs boson (H) is the only known elementary scalar boson in the SM.

To date, the presence of exactly three generations of elementary fermions has not been fundamentally understood, but the existence of a fourth generation is strongly constrained by experimental results [8].

1.2 Quantum Electrodynamics

The electromagnetic interactions of charged particles are mathematically described by the theory of quantum electrodynamics (QED). Electromagnetic processes are mediated by photons, which are massless. The couplings of gauge bosons to elementary *fermions*, i.e. leptons and quarks, are generally described by the SM interaction vertices, each associated to a coupling strength, which is related to the probability of a fermion emitting or absorbing the mediator of the interaction. More accurately, the transition matrix element $\mathcal{M}$ of a process includes a factor α for each interaction vertex and a term associated with the exchanged particle, which is commonly referred to as the *propagator*; physical observables like cross sections and decay rates are proportional to $|\mathcal{M}^2|$. Feynman diagrams such as



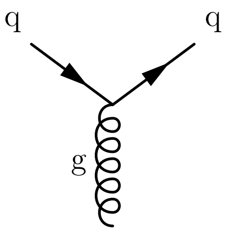
can be used to represent the interaction vertex of QED. The major contribution to a cross section or decay rate in QED is usually given by the Feynman diagram with the minimum number of interaction vertices, called leading-order (LO) diagram. Each vertex contributes to cross section or decay rate with a factor α, which, in the low-energy limit, corresponds to the fine-structure constant $\alpha = e^2/4\pi$. Additionally, there is an infinite series of higher-order-diagrams leading to the same final state. The total amplitude of a given process $\mathcal{M}_{fi}$ can be expressed by externalizing the dependence on α as

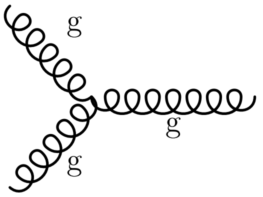
$$\mathcal{M}_{fi} = \alpha \mathcal{M}_{LO} + \alpha^2 \sum_j \mathcal{M}_{1,j} + \dots , \tag{1.1}$$

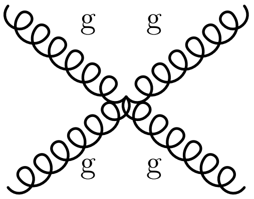
where $\mathcal{M}_{LO}$ refers to the matrix element corresponding to the LO Feynman diagram and $\mathcal{M}_{1,j}$ refers to the next-to-leading-order (NLO) diagrams, characterized by an additional interaction vertex, while the dots indicate the higher-order diagrams. The different contributions can interfere either positively or negatively and the experimentally measured interaction strengths are the result of the sum over all diagrams. The infinite series of corrections to the propagators involves loops of particles that can have arbitrarily large momenta, leading to diverging integrals and unphysical results. This issue is resolved by introducing a cut-off scale Λ, above which the correction terms are ignored. This mathematical procedure, commonly referred to as *renormalisation*, introduces an energy scale dependence. This implies that the experimentally observable quantities need to be expressed at some arbitrary renormalisation scale μ_R, which is usually chosen to be of the same order as the energies associated with the process of interest.

1.3 Quantum Chromodynamics

The theoretical description of the strong force, which binds together the quarks by exchange of gluons, is called quantum chromodynamics (QCD). The corresponding charges are called *colours* and they are responsible for the system of attraction and repulsion between quarks. Leptons, being colour-neutral, do not undergo the strong force. Quarks can exist in three orthogonal colour states, labelled as red, blue and green, and antiquarks carry the opposite colour charges; gluons, in contrast, carry a combination of colour and anticolour. The interaction vertices of QCD can be represented as







where the last two diagrams are showing the gluon self-couplings.

The coupling constant of QCD α_S can be mathematically expressed as a function of a given energy scale q as

$$\alpha_S(q^2) = \frac{g_S^2(q^2)}{4\pi} \simeq \frac{\alpha_{S(\mu_R^2)}}{1 + \beta_0 \alpha_{S(\mu_R^2)} \ln \frac{q^2}{\mu_R^2}} , \tag{1.2}$$

where g_S is the coupling strength and β_0 is a dimensionless constant that depends on the number of quark flavours and colours. The value of α_S hence decreases logarithmically towards high energy scales (Fig. 1.2), while it becomes larger at small energies. This behaviour is usually referred to as the *running* of α_S. The high value of α_S at low energies also leads to a phenomenon known as *colour confinement*, stating that coloured objects are always confined to colour singlet states. This behaviour also implies that the method described in Equation 1.1 in the context of QED, called perturbative approach, cannot be used in this case and other calculation techniques need to be used. The most common one is the lattice QCD [10], which it performs calculations on a discrete lattice of space-time points. Conversely, since large energies correspond to lower values of α_S, quarks and gluons in high energy interactions can be described as free particles (a property called *asymptotic freedom*) and the perturbative approach in terms of α_S is expected to yield reliable results.

Quarks are experimentally detected as jets of colourless particles and the process leading to their formation is known as *hadronization*: whenever quarks are produced in an interaction, they start to separate at high velocities and the energy

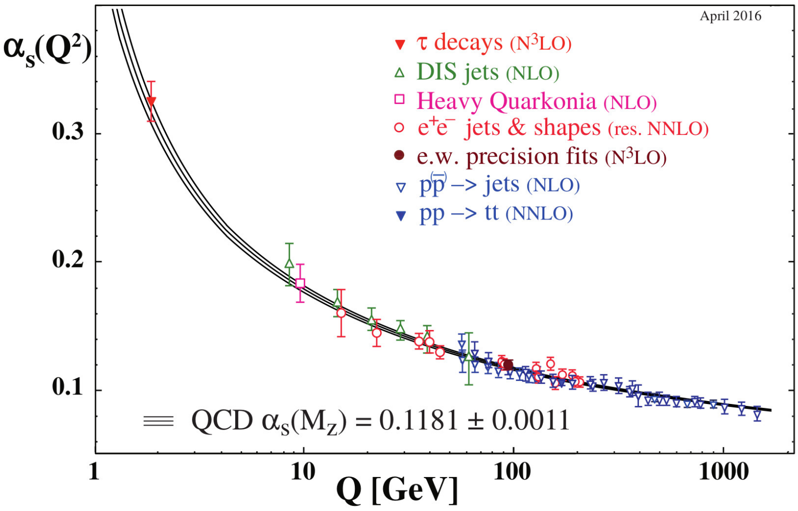


Figure 1.2: Value of α_S as a function of the energy scale Q, as reported in the QCD review of the Particle Data Group [9].

stored in the colour field between them increases until it is sufficient to create a new q$\bar{\text{q}}$ pair, which becomes energetically favourable after a certain distance; the process continues until all the quarks can combine into colourless hadrons. These bounded hadronic states are usually in the form of mesons (q$\bar{\text{q}}$), baryons (qqq) and antibaryons ($\bar{\text{q}}\bar{\text{q}}\bar{\text{q}}$). More rarely, quarks can also combine into four- and five-quark particles, known as tetraquarks and pentaquarks, respectively.

1.4 Weak interactions

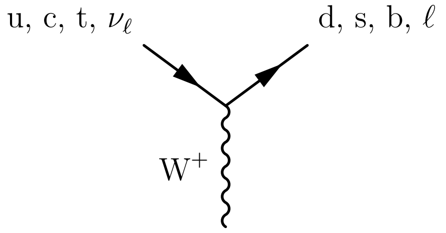
Apart from photons and gluons, all the known particles undergo weak interactions: quarks, leptons, neutrinos and heavy bosons. As mentioned in the introduction of this chapter, these are mediated by three spin-1 bosons, two of which, the W^+ and W^- bosons, are electrically charged, and the Z boson is neutral. Their masses are about 80.4 GeV and 91.2 GeV, respectively [9]. The dimensionless coupling constant can be expressed as

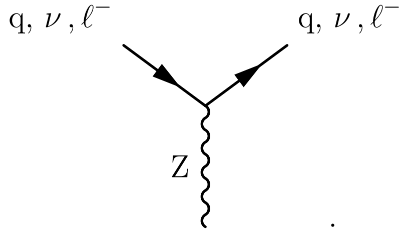
$$\alpha_W = \frac{g_W^2}{4\pi} \simeq \frac{1}{30}. \tag{1.3}$$

This implies that the weak force is intrinsically stronger than the electromagnetic interaction. However, the cross section term associated to each propagator is equal to

$$P \simeq \frac{1}{q^2 - m^2}\,, \tag{1.4}$$

where q and m are the four-momentum and the mass of the propagator, respectively. The large masses associated to the W and Z bosons therefore suppress the weak interactions at low energies, while in the high energy limit where $|q^2| > m^2$ the strength becomes comparable with that of QED. The Z boson is the mediator of the *neutral* weak interaction while the *charged* weak interactions are mediated by the $\mathrm{W}^{\pm}$ bosons and they are the only SM processes that can change the quark flavours. In the latter case, the mediators can couple to a neutrino and a lepton, or to an up-type quark and a down-type quark; the interaction vertices of the weak interactions are drawn as





The couplings in the quark sector can be expressed in terms of the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix, defined as

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} \tag{1.5}$$

where the mixing parameters V appear in the mathematical formulation of the charged weak interaction strengths, such that $|V_{\mathrm{ij}}|^2$ quantifies the transition probability between the up-type quark i and the down-type quark j. In the SM, the CKM matrix is unitary, implying that

$$|V_{\mathrm{ud}}|^2 + |V_{\mathrm{us}}|^2 + |V_{\mathrm{ub}}|^2 = 1\,, \tag{1.6}$$

$$|V_{\mathrm{cd}}|^2 + |V_{\mathrm{cs}}|^2 + |V_{\mathrm{cb}}|^2 = 1\,, \tag{1.7}$$

$$|V_{\mathrm{td}}|^2 + |V_{\mathrm{ts}}|^2 + |V_{\mathrm{tb}}|^2 = 1\,. \tag{1.8}$$

So far, all experimental measurements of the individual CKM matrix elements have been found to be consistent with these three unitarity relations. A violation of unitarity would be a clear sign of new physics.

Weak interactions are the only SM scenario that does not include parity conservation, which is equivalent to spatial inversion through the origin, $\mathbf{x} \to -\mathbf{x}$. Additionally, it has been experimentally demonstrated that the simultaneous transformation of charge conjugation, i.e. replacing particles by their antiparticles, and parity (CP) is not conserved in the weak interactions [11]. CP violation is mathematically implemented as a complex phase in the CKM matrix and, even though not sufficient, it is essential to explain the matter-antimatter asymmetry in the Universe. While CP is violated, CPT, the combined operation of C, P and time reversal T, is believed to be an exact symmetry of the Universe. CP violation implies that the T symmetry also needs to be violated. One consequence of the CPT symmetry is, for example, that particles and antiparticles have identical masses.

A summary of the fundamental interactions and their properties can be found in Table 1.1.

Table 1.1: Summary of the four known fundamental interactions and their properties.

Force	Boson	Spin	Mass/GeV
Strong	Gluon	1	0
Electromagnetism	Photon	1	0
Weak	W and Z bosons	1	80.4 and 91.2 [9]
Gravity	Graviton?	2	0

1.5 The Standard Model of particle physics

Starting from the 1960s, S. Glashow, A. Salam and S. Weinberg developed a unified description of the electromagnetic and weak interactions. The theory predicts the existence of three vector bosons, W^+ , W^- and Z, which gain masses through the Brout-Englert-Higgs (BEH) mechanism as they interact with a massless scalar field, i.e. a field associated with spin-0 particles: the Higgs field.

In a QFT, the dynamics of the quantum mechanical fields describing spin-0, spin-half and spin-1 particles are determined by the corresponding Lagrangian density, which is usually referred to simply as the Lagrangian. For a scalar field ϕ, the kinetic term can be written as

$$\mathcal{L}_{\text{kin}} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) \tag{1.9}$$

where $\partial_\mu \phi$ indicates the partial derivatives of the field with respect to each of the four space-time coordinates. A more detailed description of the Lagrangian formalism can be found in References [12, 13]. Considering now a scalar field ϕ with the potential

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4 \tag{1.10}$$

where the term proportional to ϕ^4 refers to the self-interactions of the scalar field, the corresponding Lagrangian will be given by

$$\mathcal{L} = \mathcal{L}_{\text{kin}} - V(\phi) = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4 \,. \tag{1.11}$$

The lowest energy state of the field ϕ, i.e. the vacuum state, is the minimum of the potential $V(\phi)$. It can be shown that in order to have a finite minimum, λ has to be greater than zero. However, there are no constraints on μ^2, which is also the reason why this term cannot be associated with the mass of the particle; in the case where $\mu^2 < 0$, the potential has two minima at

$$\phi = \left| \pm\sqrt{\frac{-\mu^2}{\lambda}} \right| \,, \tag{1.12}$$

implying the presence of two possible vacuum states. The choice of one of these states is a process known as *spontaneous symmetry breaking*, as it breaks the symmetry of the Lagrangian. The excitations of the fields describing the particle states are obtained from perturbation of the field around the chosen vacuum state v:

$$\phi(x) = v + \eta(x)\,. \tag{1.13}$$

By rewriting Equation 1.11 in terms of the field η, it can be shown that the term proportional to η^2 can be interpreted as a mass:

$$m_\eta = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}\,. \tag{1.14}$$

This procedure can also be applied to a complex scalar field

$$\phi = \frac{1}{2}(\phi_1 + i\phi_2) \tag{1.15}$$

and when $\mu^2 < 0$, the potential will have an infinite set of minima defined by

$$\phi_1 + \phi_2 = \frac{-\mu^2}{\lambda} \tag{1.16}$$

which all lie on a circle in the complex ϕ plane, as illustrated in Figure 1.3. By expanding ϕ around the vacuum state, a non-diagonal mass matrix for the vector boson fields is introduced in the Lagrangian. This means that the observed boson fields are a linear combination of the original fields written in the Lagrangian before the symmetry breaking; in the case of the Z boson and the photon, this mixing is described by the matrix

$$\begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \tag{1.17}$$

where θ_W is the so called *weak mixing angle*, the only free parameter needed to predict the transition probabilities of electroweak processes. The mass terms for the W and Z bosons arise from adding these new combinations in the Lagrangian, while the photon remains massless. This is the aforementioned BEH mechanism, which can also be exploited to generate the fermion masses. The coupling terms be-

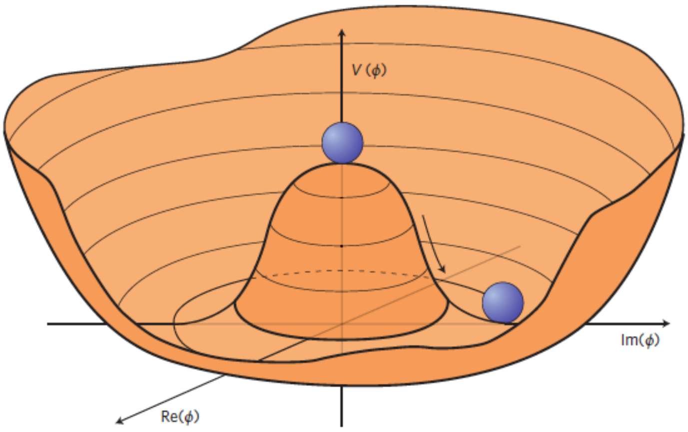


Figure 1.3: Shape of the potential $V(\phi)$ in the case of a complex scalar field. The minima are an infinite set of sates corresponding to a circle in the complex ϕ plane and the choice of one of these points spontaneously breaks the symmetry of the model. Plot taken from [14].

tween the Higgs field and the fermions are mathematically described by the Yukawa interaction [15], which causes the fermion to acquire masses after symmetry breaking. Their exact values are not predicted by the BEH mechanism and they are free parameters of the SM.

1.6 Proton-proton collisions

Many practical applications of the theoretical principles of the SM are based on proton-proton collisions, a process that nowadays is studied at hadron colliders. These tools are designed to accelerate hadrons to extremely high energies and then make them collide; they allow to achieve higher center-of-mass energies compared to circular electron-positron colliders, as the radiation loss through bremsstrahlung is inversely proportional to the particle mass to the power of four. Hadron colliders also play a critical role in the search of new particles at high energy scales: the W, Z and Higgs bosons as well as the top quark were all discovered with hadron colliders [16–23]. The underlying process in proton-proton collisions involves the interaction of two partons, which can be quarks, antiquarks or gluons. This is one of the consequences of asymptotic freedom, as it implies that in the high energy limit partons can be considered as free particles inside the proton.

The probability that a given parton participates in the hard interaction is described by the *Parton Distribution Functions* (PDFs); these are reflecting the structure of the proton, which does not only include three valence quarks but also a *sea* of strongly interacting gluons, quarks and antiquarks. The evaluation of PDFs is essential to make cross section predictions for hadron colliders, as protons are not elementary particles and each parton interaction occurs at a different energy. While this means that the center-of-mass energy of the hard partonic interaction is different in each collision, it also allows to explore a wide energy spectrum of the particles in the final state.

PDFs are not calculable, but universal: once determined, they can be used for predictions at other experiments. Their estimation is obtained through deep inelastic scattering events in electron-proton collisions, fixed-target experiments and also hadron colliders [24]. As already mentioned in Section 1.3, it is not possible to use the perturbative approach in QCD at low energies and therefore it is necessary to define a fixed energy scale, the factorization scale μ_F, below which PDFs are used to describe the nonperturbative QCD effects. The cross section of a proton-proton scattering process at a center of mass energy $\sqrt{s}$ is then given by

$$\sigma(\mathrm{pp} \to \mathrm{X})(s) = \sum_{\alpha,\beta} \int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2 f_\alpha(x_1, \mu_\mathrm{F}) f_\beta(x_2, \mu_\mathrm{F}) \cdot \hat{\sigma}(\alpha\beta \to \mathrm{X})(x_1 x_2 s, \mu_\mathrm{F}, \mu_\mathrm{R}) \tag{1.18}$$

where $\hat{\sigma}(\alpha\beta \to \mathrm{X})(x_1 x_2 s, \mu_\mathrm{F}, \mu_\mathrm{R})$ is the cross section of two initial partons α and β to produce the final state X, and $f_{\alpha,\beta}$ are the associated PDFs. Usually, measurements are conducted at different energy scales and the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation [25–27] is then used to evaluate PDFs for a specific μ_F. The results presented in this thesis are based on events simulated with the NNPDF3.1 set [24], which incorporates up-to-date high-precision QCD calculations of hadron collider processes that are directly sensitive to PDFs. These sets are shown in Figure 1.4: it can be seen that the gluon and sea quark distributions rise steeply towards low proton momentum fractions x, while valence quarks tend to carry higher x values. As discussed in Section 1.3, the scattered partons arising from proton-proton interactions are observed as jets as a consequence of hadronization. The residues from each proton can initiate multiple partonic interactions, producing particles that are collectively denoted as *underlying event*, which can interact with those originating from the primary interaction. Furthermore, hadron-hadron colliders accelerate beams of particles, which means that multiple primary interactions take place at the same time.

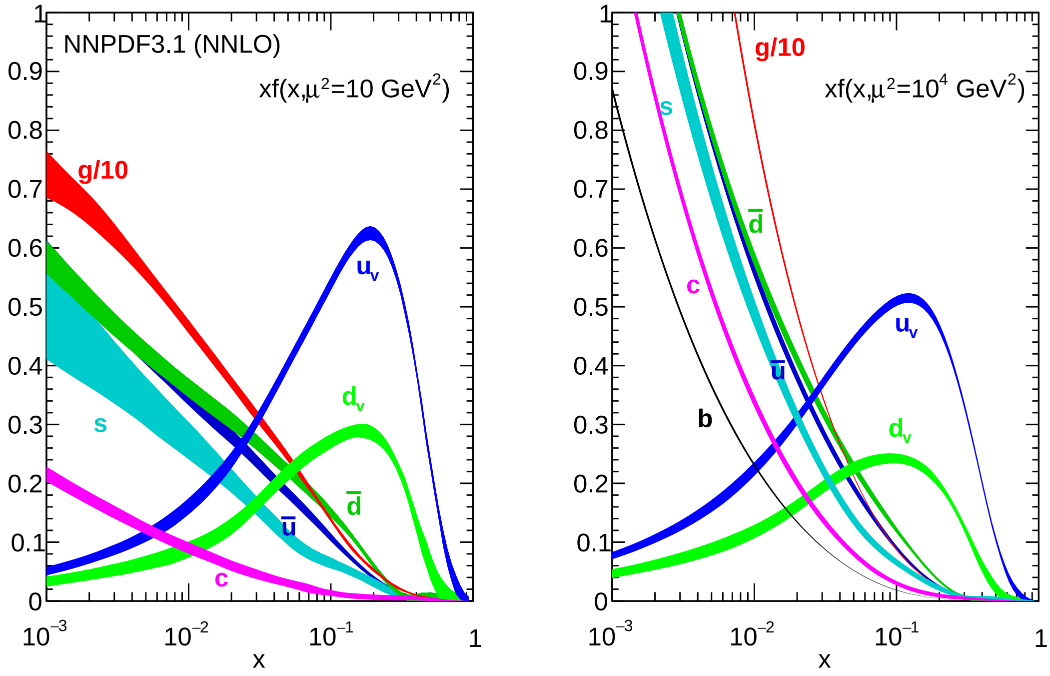


Figure 1.4: NNPDF3.1 sets as a function of the fraction x of the proton momentum, for $\mu_F^2 = 10\,\mathrm{GeV}^2$ (left) and $\mu_F^2 = 10^4\,\mathrm{GeV}^2$ (right). The valence quarks are indexed with v, while the other contributions refer to sea quarks and gluons. The distributions were evaluated with the DGLAP equation at NNLO precision [24].

1.7 The role of the top quark

The results presented in this thesis are primarily focused on the physics of the top quark, the heaviest known elementary particle. Its existence was first predicted by Makoto Kobayashi and Toshihide Maskawa in 1973, to explain the CP violation in the quark sector [28] discussed in Section 1.4, and was only discovered in 1995 in proton-antiproton collisions at the CDF and D0 experiments at the Tevatron [20,21]. Since $|V_{\mathrm{tb}}| \gg |V_{\mathrm{ts}}| > |V_{\mathrm{td}}|$, nearly all top quarks decay into $\mathrm{t}\rightarrow\mathrm{bW}^+$. Due to its large mass of about 172 GeV, its lifetime is really short (about 5×10^{-25} s [9]) and it is the only quark that decays before hadronization, providing the possibility to study the properties of the bare quark. Furthermore, its large mass implies a strong coupling to the Higgs boson: the top quark Yukawa coupling y_{t} has a magnitude of order one in the SM and it brings the most important contributions to loop corrections in electroweak predictions. The high value of its mass also remains unexplained to date and it could be an indication that the top quark is a portal to possible new physics.

The dominant production mode in proton-proton colliders, like the CERN's LHC, is the pair production ($\mathrm{t\bar{t}}$) via the strong interaction, with a cross section of about 830 pb at a center-of-mass energy of 13 TeV [29, 30]. It includes both gluon and quark initial states (Figure 1.5). In contrast, single top quarks (Figure 1.6) are produced through the electroweak interaction, either with an intermediate W boson decaying into a top and an antibottom quark (s-channel) or by a bottom quark exchanging a W boson with an up or down quark (t-channel); additionally, it can

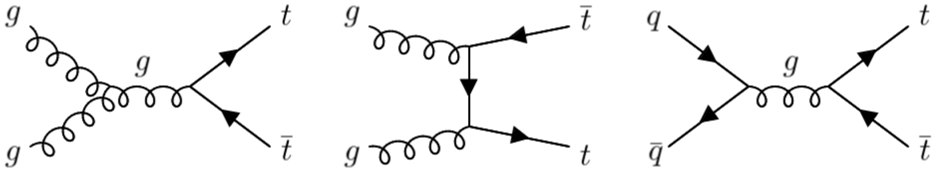


Figure 1.5: LO Feynman diagrams showing the production of top quark pairs in different channels. The gluon fusion (left and center) represents the main production mode at the LHC, while the one having a quark-antiquark pair in the initial state (right) was the dominant channel at the Tevatron in propton-antiproton collisions.

also be produced in association with a W boson. Among the single top production modes, the t-channel is the most common at the LHC, with a cross section of about 220 pb at 13 TeV [31,32].

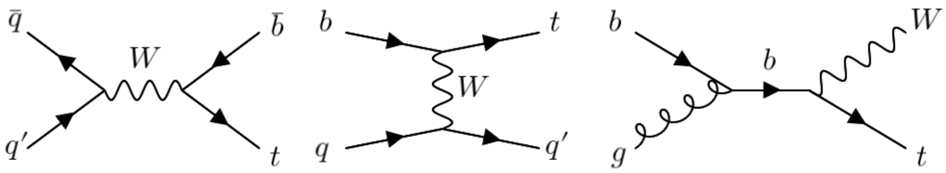


Figure 1.6: LO Feynman diagrams of the single top quark production in the t-channel (left), s-channel (center) and in association with a W boson (right).

The studies in the top quark sector offer unique opportunities to test the SM and look for possible deviations from it. Thanks to the large data sets recorded at the LHC and the high center-of-mass energy, precise measurements are also possible of rarer processes involving the top quark, such as its production in association with bosons and fermions, which allow to study the interactions and couplings of the top quark with other SM particles as well as to look for possible signs of new physics. An overview of the top quark cross section measurements for the different production mechanisms is shown in Figure 1.7; results were produced by the CMS Collaboration at different center-of-mass energies.

1.8 Top quark production in association with a Z boson

In this thesis, a measurement is presented of the Z-boson associated top quark production, which can happen in pair ($\mathrm{t\bar{t}Z}$) and as a single top quark in the t-channel or in association with a W boson (tZq and tWZ, respectively); these production modes are all represented in Figure 1.8. The s-channel contribution is considered to be negligible and it is not taken into account. A primary motivation to examine these processes is their direct sensitivity to the coupling of the top quark to the Z boson, which occurs at LO, as visualized in the Feynman diagrams in Figure 1.8. Sizeable modifications to this interaction are predicted in several Physics-Beyond-

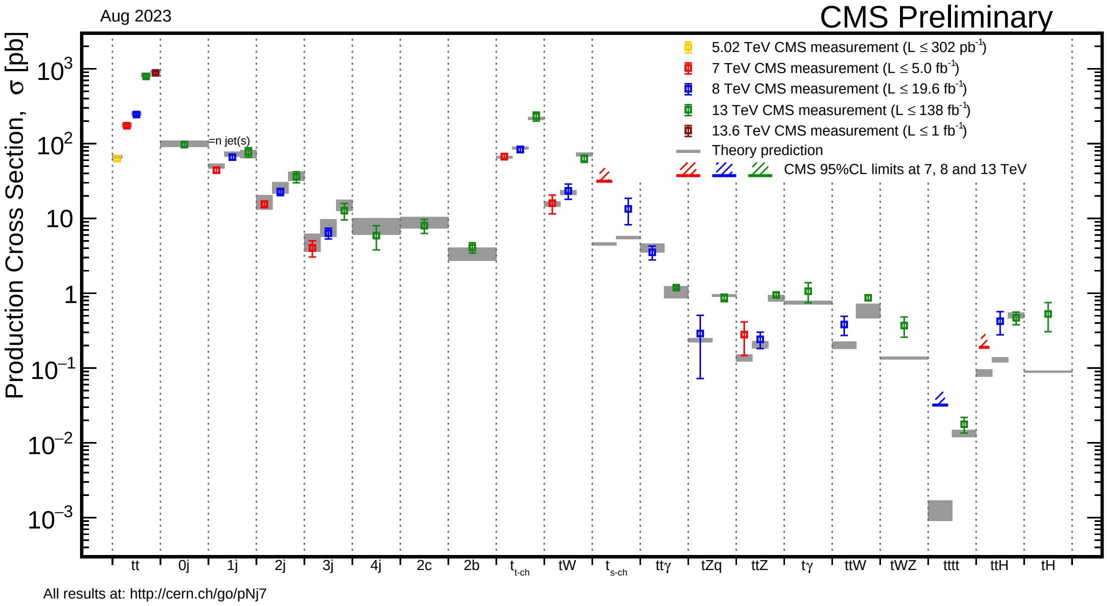


Figure 1.7: Overview of the top quark cross section measurements produced by CMS for the different production modes. Results are shown for different center-of-mass energies. Taken from Ref. [33].

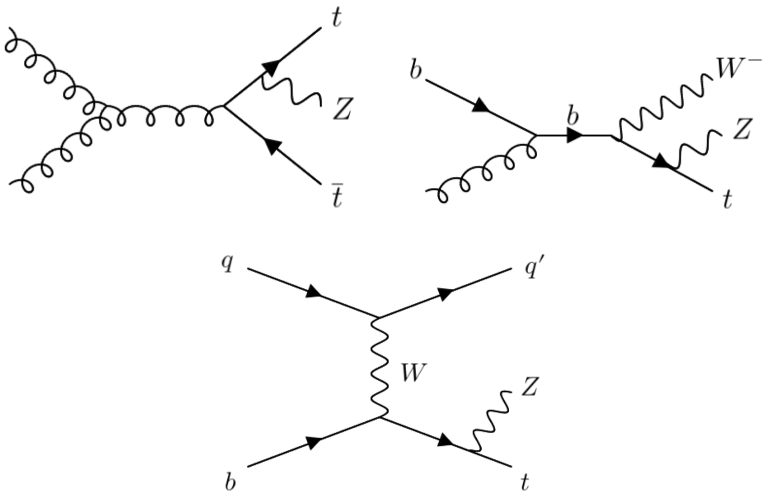


Figure 1.8: Examples of LO Feynman diagrams for $\mathrm{t\bar{t}Z}$ (top left), tWZ (top right) and tZq (bottom) production.

the-SM (BSM) scenarios, which provides an additional point of major interest for these studies. Furthermore, many measurements and searches of rare processes (e.g. tHq, $\mathrm{t\bar{t}}\gamma$ or $\mathrm{t\bar{t}H}$) have large background contributions from tZq and $\mathrm{t\bar{t}Z}$, which therefore need to be deeply understood. As it can be seen from Figure 1.7, the tZq and $\mathrm{t\bar{t}Z}$ production processes have similar inclusive cross sections of about 870 and 740 fb, respectively; precise and differential measurements of these processes have already been performed with data collected at a center-of-mass energy of 13 TeV at the LHC [1, 2, 34, 35]. The tWZ process has a much smaller cross section and differential measurements of this process have not yet been performed. However, evidence for this process was recently reached with an analysis performed by the CMS Collaboration [3]. Despite its small cross section of about 130 fb, the tWZ production has unique features that make it suitable to provide enhanced sensitivity to new physics scenarios [36, 37]: besides being a rare electroweak production process that can probe the weak couplings of the top quark, it is also sensitive to possible energy growth in the $\mathrm{bW}\rightarrow\mathrm{tZ}$ scattering [36].

Experimental results for tZq were found to be in good agreement with SM expectations, while a small excess was observed for both $\mathrm{t\bar{t}Z}$ and tWZ; all the measurements are, however, within 2σ from the prediction in perturbative QCD.

1.8.1 Theoretical modeling

The dominant LO contributions resulting in the $\mathrm{t\bar{t}Z}$ final state in proton-proton collisions come from gluon-gluon and quark-antiquark initial states. The production cross section has been evaluated including both QCD and electroweak corrections at NLO [38]. As shown in Figure 1.9, the tWZ simulation at NLO features some contributions where an intermediate (anti-)top quark becomes resonant and create an overlap with the $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}}$ processes. Such overlap contributions have to be subtracted when simulating the tWZ production; the overlap removal procedures are divided into two main categories: the diagram removal and the diagram subtraction methods (DR and DS respectively) [37]. In the former, the resonant diagrams are set to zero, while in the latter case a subtraction term is exploited to suppress the resonant contributions. The general amplitude $\mathcal{M}$ of the tWZ process can be written as

$$\mathcal{M} = \mathcal{M}_{\text{non-res}} + \sum_i \mathcal{M}_{\text{res},\, i} \tag{1.19}$$

where the index i refers to the resonant contributions from $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}Z}$. The work presented in this thesis relies on the DR scheme, which has two variants:

- DR1, in which the resonant terms are discarded before squaring the amplitudes.
- DR2, where only the squared resonant matrix elements in the squared amplitude are discarded, while their interference with the non-resonant part is kept.

The DR1 method has been employed to simulate the tWZ events, while an alternative sample generated with DR2 is used to assign a modeling uncertainty. The shapes obtained with the DS scheme were found to lie between the ones obtained from DR1 and DR2 [37], and hence they are not considered in this work.

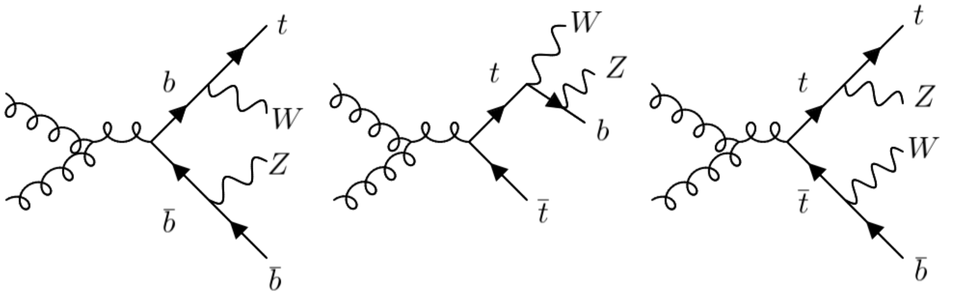


Figure 1.9: Feynman diagrams for the tWZ (left), $\mathrm{t\bar{t}}$ (middle) and $\mathrm{t\bar{t}Z}$ (right) production arising in the simulation of the tWZ process at NLO.

The presence of a b quark in the initial state is expected in both the tZq and tWZ processes. There are two different approaches to calculate the distribution of the initial b quark: it can be integrated into the proton PDFs through the five-flavour scheme (5FS) or simulated through gluon splitting in the hard interaction with the four-flavour scheme (4FS). A comparison between the two schemes is shown for the tZq production in Figure 1.10. The 5FS is generally expected to be more precise in estimating inclusive cross sections. On the other hand, the 4FS is expected to

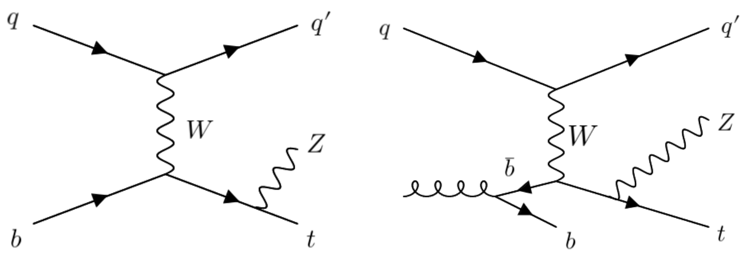


Figure 1.10: Feynman diagrams of the tZq production in the 5FS (left) and 4FS (right). In the former case the b quark is integrated into the proton's PDF, while in the latter it comes from gluon splitting.

provide a better modeling of the kinematic distributions for the particles in the final state [39]. However, in the case of tWZ, NLO computations are currently not available in this scheme and they would be very computationally expensive [37]. For this reason, the results presented in this thesis rely on the 4FS for the simulation of tZq events, while the 5FS is used for both tWZ and $\mathrm{t\bar{t}Z}$.

Examples of predicted differential distributions for the three signal cross sections are shown in Figure 1.11 as a function of the p_{T} of the Z boson [37–39]. Different shapes are used to confront different levels of accuracy in the calculation of $\mathrm{t\bar{t}Z}$, while the comparison in tWZ focuses on the different DR approaches. For tZq, predictions in the 4FS and the 5FS are compared: it can be seen that the prediction in the 5FS is about 5% higher than that in the 4FS. The exact predictions of the expected cross sections used to perform the analysis presented in this thesis will be discussed in Chapter 5.

1.8.2 New physics scenarios

A better understanding of the processes investigated in this thesis is also important for searches involving flavour-changing neutral currents (FCNC): these are hypothetical interactions in which the flavour of a fermion changes without altering its electric charge (Figure 1.12). This includes, for example, the decay of the top quark into a c or u quark under the emission of a Z boson. Even though such processes are heavily suppressed in the SM, several BSM scenarios predict measurably higher

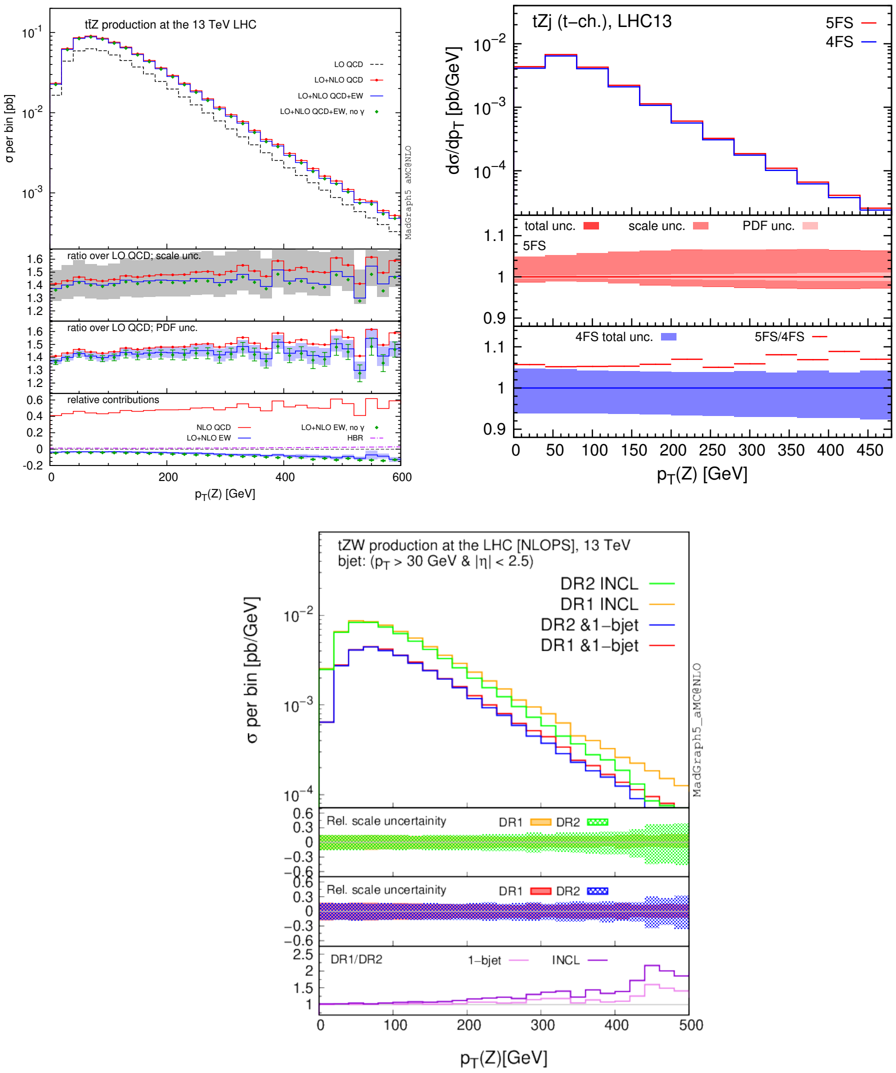


Figure 1.11: Predicted distributions of the differential cross sections for $\mathrm{t\bar{t}Z}$ (top left), tZq (top right) and tWZ (bottom) at a center-of-mass energy of 13 TeV, as a function of the p_T of the Z boson. For $\mathrm{t\bar{t}Z}$, the shapes coming from computations at different levels of accuracy are compared; in the case of tZq, both 4FS and 5FS predictions are shown, while the expected cross section of tWZ is shown with the two different DR methods. The figures are taken from Refs. [37–39].

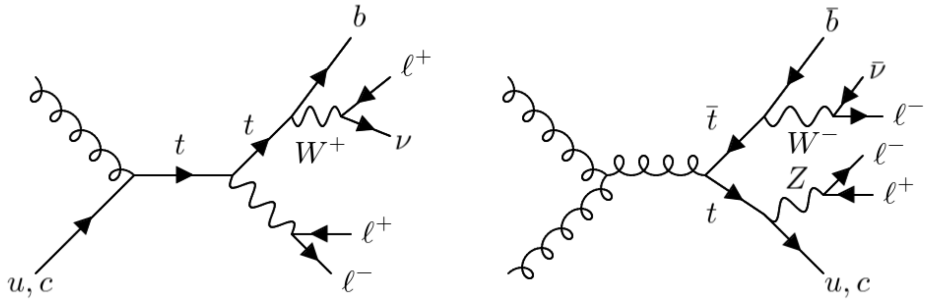


Figure 1.12: FCNC interactions involving the top quark production through the coupling to a gluon (left) and the decay of the top quark into a Z boson and a c or u quark (right).

cross sections. The SM production modes of tZq, $\mathrm{t\bar{t}Z}$ and tWZ are expected to be the main backgrounds for these processes.

A more general and model-independent approach in the search for new physics is realized through the effective field theory (EFT) framework [40,41]: in EFT, the presence of possible new physics scenarios is quantified in a coherent and model-independent way by looking for deviations in well-established processes and results from several measurements can be used in combination. Even if direct observations of new particles might not be possible at the current collider operational energy regimes, new physics scenarios might already be accessible in the existing measurements and EFT aims to find a low energy approximation of the theory behind these possible new interactions. This is done by extending the dimension-4 SM Lagrangian with higher-order operators $\mathcal{O}$ of dimension d, having an interaction strength proportional to $1/\Lambda^{d-4}$ and characterized by a dimensionless Wilson coefficient C_i:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i}{\Lambda^{d-4}} \mathcal{O} \,. \tag{1.20}$$

The factor $1/\Lambda^{d-4}$ suppresses the contributions from higher-order operators, implying that EFT interpretations can be performed considering a finite set of operators at relatively low order in which the Wilson coefficients can be constrained from experimental data. Only one operator can be constructed at dimension-5 after allowing for right-handed neutrinos [42]. However, EFT interpretations in high energy physics usually target dimension-6, which has 59 independent operators, or dimension-8, as they might bring the dominant contribution in cases where the interference between SM and EFT amplitudes at dimension-6 is suppressed [42].

The previous experimental analyses measured the $\mathrm{t\bar{t}Z}$, tZq and tWZ cross sections individually, assuming SM expectations for the other processes, which were then treated as backgrounds. In this thesis, the three processes are measured simultaneously for the first time. This helps to constrain and identify new physics scenarios, as the impact of possible deviations from the SM is considered across different processes, reducing the dependency on specific theory assumptions. The overlap between the $\mathrm{t\bar{t}Z}$ and tWZ processes is handled in the simulation as described in Section 1.8.1; however, the separation of the two processes still leads to challenges in both the theoretical and experimental workflow, and therefore the sum of the two processes is reported in the final results. The differential cross sections are determined for both the combined sum of $\mathrm{t\bar{t}Z}$ and tWZ, along with tZq, and providing full information for the use of the data, e.g. for EFT interpretations.

CHAPTER 2

The LHC and the CMS experiment

This chapter provides an overview of the CERN's accelerator system and the CMS detector, which has been used to collect the data analysed in this thesis. In particle colliders, particles are accelerated to velocities approaching the speed of light through the application of electromagnetic fields. In circular colliders, particles move in opposite directions and produce head-on collisions in which a substantial amount of energy is released. The LHC [43] is the most powerful collider ever built, located in a tunnel 27 kilometers in circumference on the border between Switzerland and France. The following information refers to the configuration used during Run-2, i.e. for data recorded in the years 2016–2018.

2.1 The accelerator complex at CERN

The CERN accelerator complex (Figure 2.1) consists of a succession of particle accelerators that brings the proton beams to increasingly higher energies before injecting them into the LHC for collisions. In order to create proton beams, an electric field is applied to hydrogen gas to remove the electrons from the hydrogen atoms and produce protons, which are then injected into a chain of accelerators that brings the beams up to the energy of 450 GeV. At this point, the protons are transferred into the two beam pipes of the LHC, where they reach the maximum energy of 6.5 TeV[1]. The particle beams circulate in opposite directions while being held in cyclic orbits by a magnetic field. Collisions take place at four interaction points around the LHC ring, with a total centre-of-mass energy of 13 TeV. Under normal operating conditions, a proton beam has 2808 bunches, each of them containing around 10^{11} protons. The bunches get squeezed to about 20 μm as they approach the collision points, in order to maximize the probability of proton-proton interaction. Collisions occur during a time interval of about ten hours until the intensity is diluted; after that, the beam remnants are dumped in a water-cooled graphite block, requiring the LHC to be refilled. During Run-2, collisions took place once every 25 nanoseconds, and up to 40 simultaneous interactions occurred in each crossing; this is commonly referred to as *pile-up*. The average number of events N produced in a given time interval can be expressed as

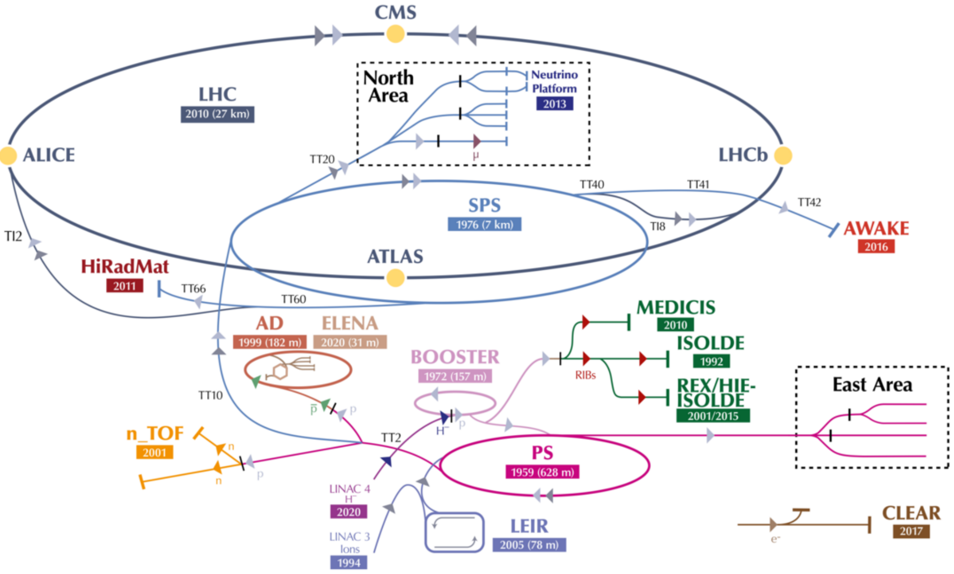


Figure 2.1: Illustration of the CERN accelerator complex. The LHC (dark blue line) is the last element of the particle accelerator chain. Taken from Ref. [44]

$$N = \sigma\mathcal{L} = \sigma\int\mathcal{L}(t)\mathrm{d}t\,, \tag{2.1}$$

---

[1] The number refers to Run-2. In the ongoing Run-3, the maximum energy is 6.8 TeV.

where σ is the cross section of a given process, and $\mathcal{L}(t)$ and $\mathcal{L}$ are the instantaneous and integrated luminosities, respectively. The instantaneous luminosity is widely used in the context of particle colliders to express the frequency of event production, while the integrated luminosity measures the total number of events during a period of data-taking. The stable decay products of the particles produced in the interactions are detected with four different experimental apparatuses built around the interaction points:

- the ALICE detector [45], mainly designed to investigate the QCD sector and the quark-gluon plasma in heavy-ion collisions, in which extreme values of energy density and temperature are reached.
- The LHCb experiment [46], which focuses on precision measurements of heavy flavour physics, mainly CP violation and rare decays of beauty and charm hadrons.
- The ATLAS [47] and CMS [48] detectors, two general-purpose experiments designed to investigate various aspect of particle physics. Cross checks and combination of the results can be done with data recorded by the two experiments, as they target the same type of measurements.

2.2 The CMS detector

The CMS experiment [48] is designed to detect all the stable particles produced in the collisions and measure their properties. The overall layout of the detector is shown in Figure 2.2: it is shaped in an onion-like structure, with its components arranged in several concentric cylinders. The main feature of the detector is a solenoid magnet formed by a cylindrical coil of superconducting fibres. It produces a 3.8 Tesla magnetic field inside the inner tracker, providing a large bending power for charged particles trajectories: this allows to measure the particle charges and their momenta. The solenoid is surrounded by an iron yoke that guides the magnetic field through the outer parts of the detector, where the field strength is still as large as 2 Tesla. The innermost layer of the detector is the CMS tracker, which surrounds the interaction point. The two CMS calorimeters, ECAL (Electromagnetic

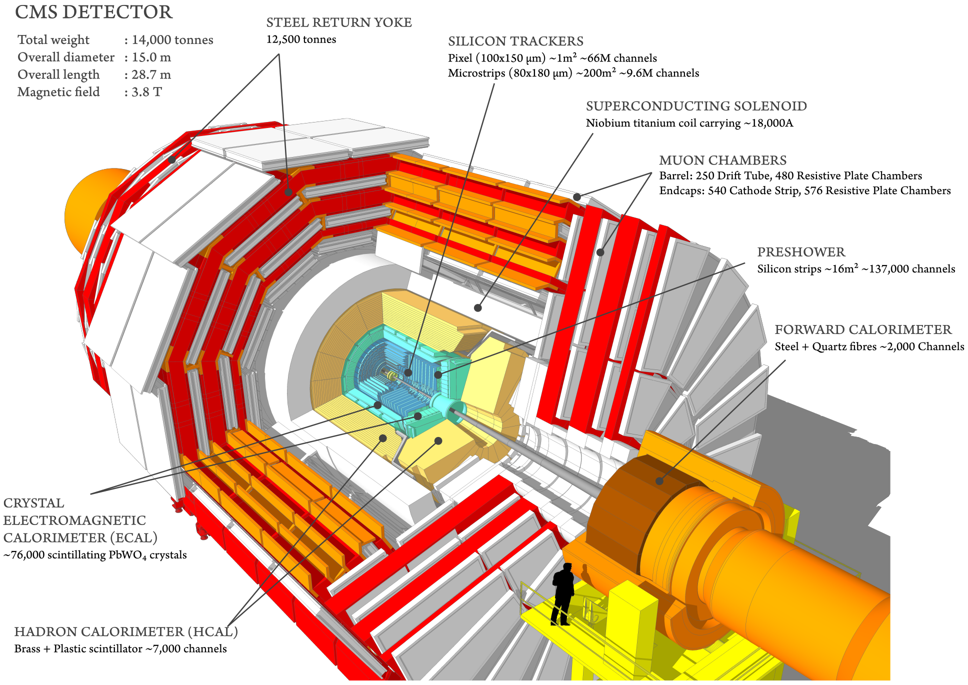


Figure 2.2: Illustration of the CMS detector and its components, taken from Ref. [49].

Calorimeter) and HCAL (Hadron Calorimeter), are positioned within the magnet as well, surrounding the tracker. Muons are ultimately detected in the outer layer of the experiment through the use of muon chambers. The CMS coordinate system conventionally has its origin at the collision point, with the y-axis pointing vertically upward and the x-axis chosen to point to the center of the LHC ring; it follows that the z-axis points along the beam direction. The azimuthal angle ϕ is defined from the x-axis in the (x, y) plane while the polar angle θ is measured from the z-axis. This can be used to define the pseudorapidity

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] , \tag{2.2}$$

an angular variable widely used as spatial coordinate in the physics colliders. The longitudinal momentum in the initial state of a parton collision is unknown, since partons carry different fractions of the proton momentum (Section 1.6). However, it is known that the initial momentum in the direction transverse to the beam line

p_{T} is zero, meaning that the quantities measured in this direction are associated with the particles produced in the collisions. Following the conservation laws, the imbalance of momentum measured in the transverse plane $p_{\mathrm{T}}^{\mathrm{miss}}$ is equal to

$$p_{\mathrm{T}}^{\mathrm{miss}} = -\sum_i p_{\mathrm{T}}(i) \tag{2.3}$$

where i are all the detectable particles in the event, and it is usually attributed to the neutrinos or any hypothetical particle not interacting with the detector material, e.g. dark matter. Figure 2.3 illustrates a slice of the CMS detector and the typical trajectories for the different types of stable particles.

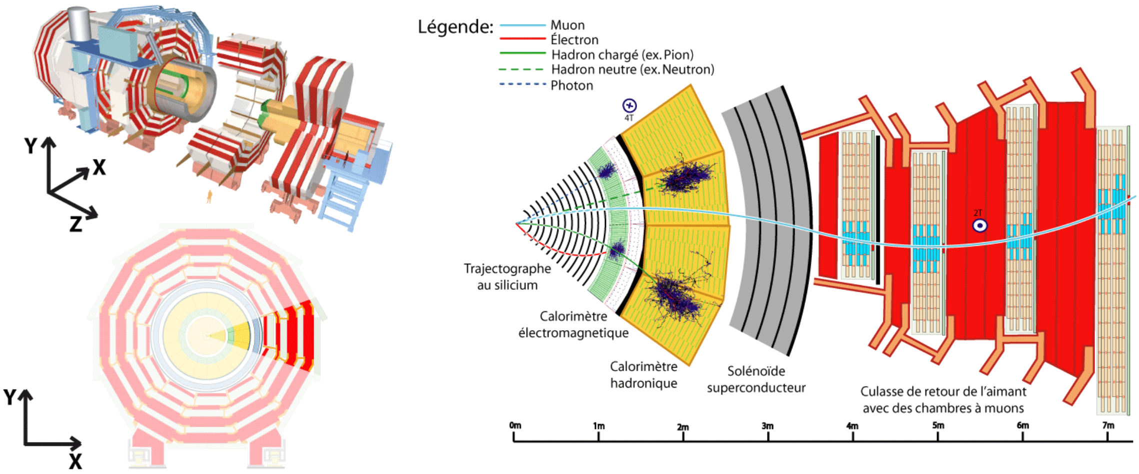


Figure 2.3: Sketch illustrating a transverse slice of the CMS detector and its components. The trajectory of the charged particles is bent due to the presence of a 4 Tesla magnetic field generated by the superconducting solenoid and their path is recorded in the CMS tracking system, the innermost part of the detector. Electrons, positrons and photons are fully absorbed by the ECAL, while hadrons are measured in the HCAL. Finally, muons are travelling towards the outermost part of the experiment and detected in the muon chambers. Taken from Ref. [50].

2.2.1 Silicon tracker

The trajectory of charged particles inside the detector can be used to identify the primary and secondary vertices of the collision; the former is located on the z-axis,

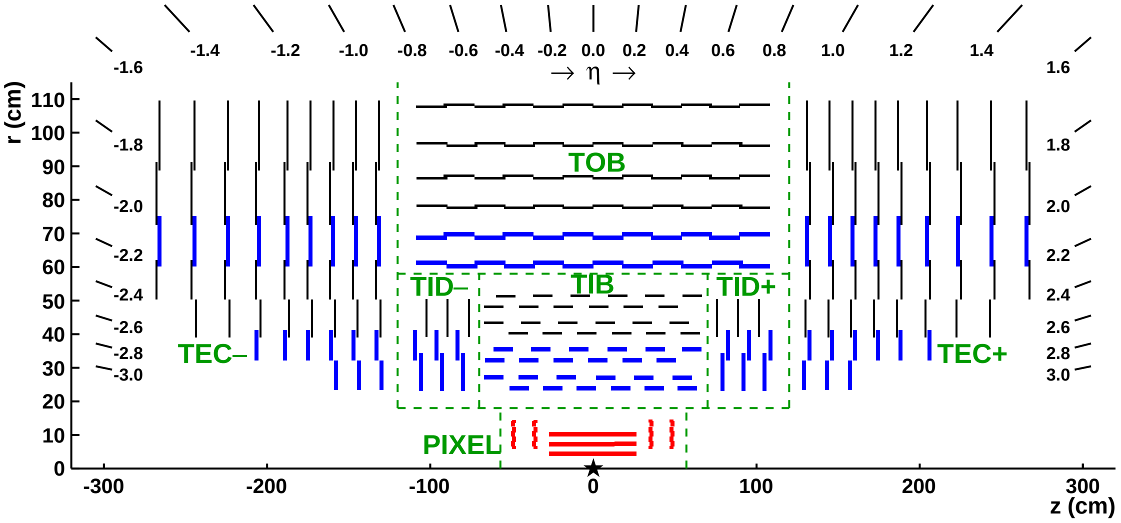


Figure 2.4: Schematic illustration of the CMS tracker. It is composed of two different sub-detectors, the silicon pixel (red) and the silicon strip (blue) tracker. The latter has four and six layers in the inner (TIB) and outer barrels (TOB), respectively. Each endcap has then three inner disk layers (TID) and nine layers in the outer region (TEC). Taken from Ref. [51].

while the latter is displaced in the transverse plane. Furthermore, as mentioned in the beginning of this chapter, tracking the path of charged particles through a magnetic field provides a way to measure their momenta: the more curved the path, the less momentum the particle had. The bending direction of the track can be used to identify the particle charges. For this reason, the CMS tracker has been designed to perform precise and efficient measurements of the particle trajectories by recording their positions in a system of high-granularity silicon detectors. Charged particles travel through these modules producing electron-hole pairs, which in turn generate an electric current that is later amplified and detected by the read-out electronics. Located at a distance of 30 mm from the beam pipe, the tracking system is the innermost part of the detector and it has a length of 5.8 m and a diameter of 2.5 m. It consists of two main sub-detectors: the silicon pixel and the silicon strip tracker, which are set up to cover up to an acceptance range of $|\eta| < 2.5$ (Figure 2.4).

Until the end of 2016, the silicon pixel tracker was composed of three barrel layers and two endcap disks of pixel components mounted around the beam axis in the innermost part of the tracker. The pixel detector was replaced in the beginning of 2017 in order to handle higher instantaneous luminosities: an additional layer (i.e. four barrel layers and three endcap disks) was placed on each side to ensure more accurate measurements. Furthermore, the digital read-out chips were upgraded to have a higher rate capability.

The outer part of the tracker consists of ten layers of silicon micro-strip detectors which are placed up to a radius of 1.1 m; each side is completed by three small and nine large disks in the endcaps. Bigger segments and fewer read-out electronics is needed compared to the pixel detector, as the particle flux is reduced by the increased distance from the collision point. The size of the strip segments in the different layers range from $10\,\mathrm{cm}\times 80\,\mu\mathrm{m}$ to $25\,\mathrm{cm}\times 205\,\mu\mathrm{m}$.

2.2.2 Electromagnetic and hadronic calorimeters

The energies of electrons, positrons and photons are measured with the CMS electromagnetic calorimeter (ECAL); other particles that can interact electromagnetically may also leave tracks in this region of the detector, but they are not fully absorbed. The ECAL (Figure 2.5) is made of lead tungstate ($PbWO_4$) crystals surrounding the silicon tracker and covering the region $|\eta| < 3$. It consists of a barrel and two endcap segments: the barrel part covers the pseudorapidity range $|\eta| < 1.479$ and it features crystals with a transverse size of $2.2\times 2.2\,\mathrm{cm}^2$ or 0.0174×0.0174 in the (η, ϕ) plane; in the endcaps, the covered pseudorapidity range is extended to the range $1.479 < |\eta| < 3$ and crystals are slightly larger, with a front size of $2.9 \times 2.9\,\mathrm{cm}^2$. When electrons interact electromagnetically in the calorimeter, they emit photons through bremsstrahlung, which in turn create electron-positron pairs resulting in an electromagnetic shower of electrons, positrons and photons; the number of particles in the shower is proportional to the energy of the incident particle. As the energy of the electrons falls below a certain level, atoms in the calorimeter are excited and emit photons through a process called scintillation. This signal is then amplified with photomultipliers and finally detected with photodiodes in order to measure the energy that has been deposited in the calorimeter. The scintillation decay time of the lead tungstate crystals is of the same order of magnitude as the LHC bunch

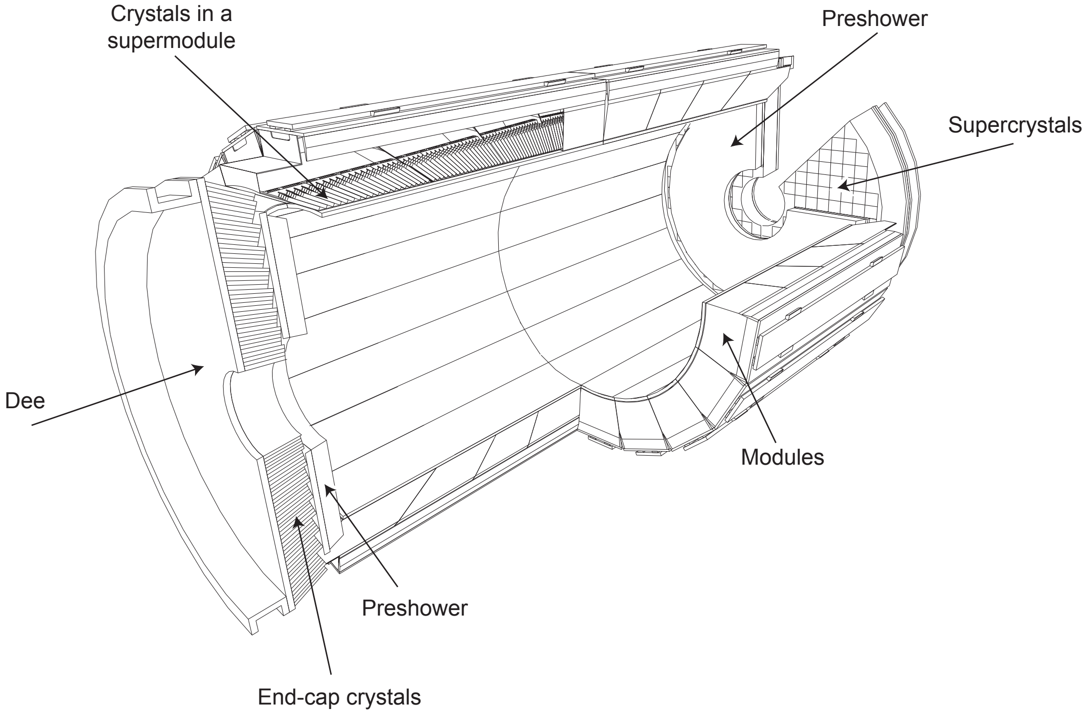


Figure 2.5: Sketch of the CMS ECAL and its components. Taken from Ref. [52].

crossing time and 80% of the light can be collected within 25 ns. Furthermore, a preshower detector is installed before the main ECAL calorimeter in the pseudirapidity range $1.653 < |\eta| < 2.6$; it consists of two layers of lead radiators to initiate the electromagnetic showers and silicon strip sensors to measure the deposited energy, for a total thickness of 20 cm. Its main purpose is to improve the identification of neutral pions in the endcape as well as to help in the discrimination of electrons and positrons against minimum ionizing particles. The CMS hadron calorimeter (HCAL) is essential for accurately measuring hadron jets and the missing transverse energy that can come from neutrinos or exotic particles. The HCAL is made of sampling calorimeters, i.e. calorimeters consisting of alternating layers of absorber and scintillator material, allowing to measure forward jets up to $|\eta| \simeq 5.2$. Interactions occur as hadronic particles pass through the absorber materials and many secondary particles are produced; these result in a particle shower, which is then detected as particles pass through the scintillating layers. The barrel and endcap components (Figure 2.6) are placed inside the solenoid and they are covering a rapidity range of

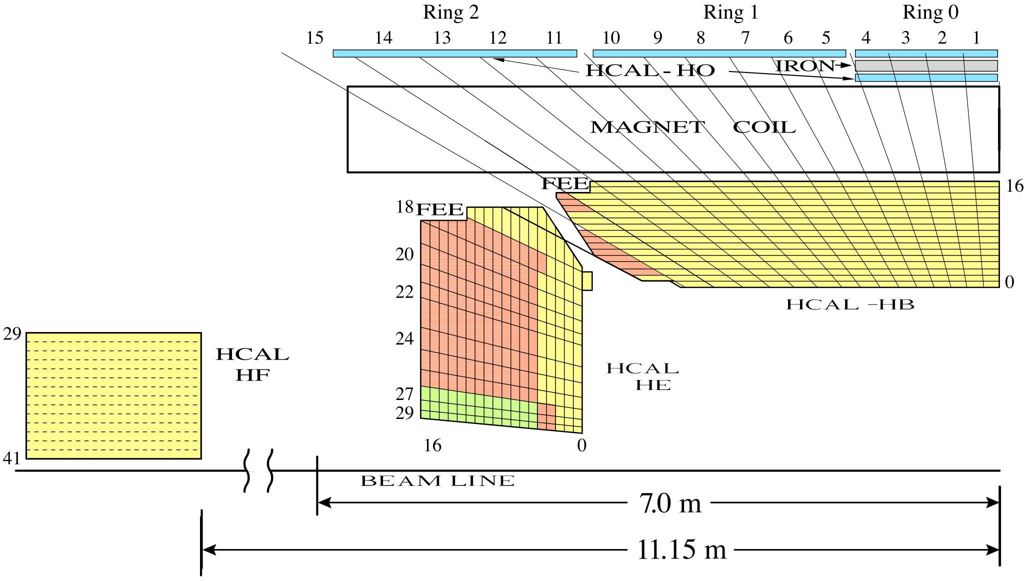


Figure 2.6: Illustration of the HCAL structure. The barrel calorimeter (HB) covers up to $|\eta| < 1.3$, while the two endcap calorimeters extend the range to $1.3 < |\eta| < 3.0$. An outer barrel calorimeter (HO) is placed outside the magnetic coil. The two forward calorimeters (HF) allow to detect jets up to $|\eta| < 5.2$. Taken from Ref. [53].

$|\eta| < 1.3$ and $1.3 < |\eta| < 3$, respectively. The transverse size of the read-out towers is 0.087×0.087 in η and ϕ for $|\eta| < 1.6$, and 0.17×0.17 for $|\eta| > 1.6$. The HCAL is further extended outside the solenoid with an outer barrel calorimeter in order to be able to identify late starting showers and to measure the energy deposited after the barrels. Furthermore, two forward calorimeters are installed to detect jets in the forward region, ensuring a coverage up to $|\eta| < 5.2$. Since the level of radiation in this region is much higher, quartz fibres were chosen as active material, as they can handle such levels of radiations with limited deterioration. The signal is generated when the showers of charged particles hit the quartz fibres and generate Cherenkov light, which is then transported to the read-out system.

2.2.3 Muon chambers

High energy muons are not stopped by any of the CMS calorimeters and tracker system, as they can penetrate several meters of absorber materials without interacting. They are only detected in the outermost part of the experiment, the muon chambers: these detectors are located outside the solenoid magnet and they cover a pseudorapidity region of $|\eta| < 2.4$, without any gaps. Only muons and neutrinos, which are not directly detected by CMS, reach this part of the detector. The CMS muon system employs three different types of gaseous detectors to identify muons and reconstruct their charge and momentum over the entire kinematic range. The full path of the particle is then reconstructed by combining these measurements with the ones from the tracker.

Four stations of drift tubes are installed in the barrel covering a pseudorapidity range up to $|\eta| < 1.2$, as the background is small in this region and the 2 Tesla magnetic field is uniform and mostly contained in the steel yoke. Each tube contains a stretched wire in a gas volume (Figure 2.7): muons ionize the atoms of the gas and produce an electric current which is then amplified to produce a measurable signal. In the endcap regions, where the muon rates and background levels are high and the magnetic field is not uniform, cathode strip chambers are used, covering $1.2 < |\eta| < 2.4$; this choice is motivated by the good radiation hardness and fast response of these detectors. Here arrays of positively-charged anode wires are crossed with negatively-charged copper cathode strips within a gas volume. As muons pass through the wires, they ionize the gas atoms and the resulting electrons converge to the anode wires creating an avalanche of electrons. The positive ions

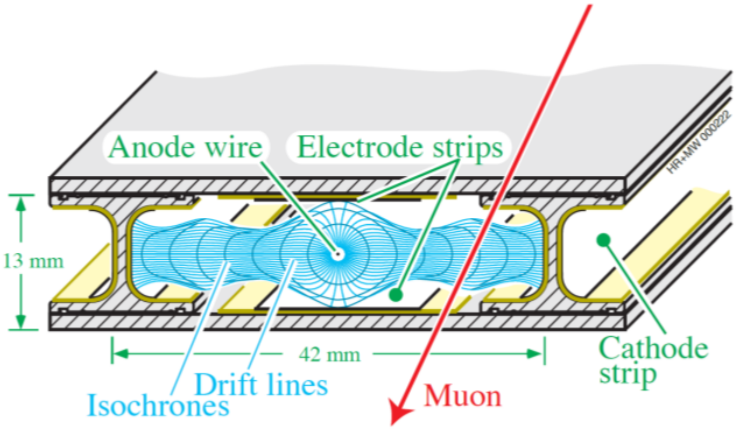


Figure 2.7: Layout of a drift tube cell. The blue lines represent the electric field in the gas volume. Taken from Ref. [54]

move towards the copper cathode and create a signal at the strips. Since the strips and the wires are perpendicular, two position coordinates can be measured for each passing particle.

Resistive plate chambers are installed as complementary detectors in the region $|\eta| < 1.9$. They feature two parallel plates, an anode and a cathode, separated by a thin gas volume. These detectors have a good spatial and time resolution and they can provide a fast measurement of muon's momentum; this information is then used in the trigger system, which will be described in the next section.

2.2.4 Triggering and data acquisition

During Run-2, collisions were occurring with a rate of approximately 40 MHz, an amount of information that is too large to be stored for offline analysis with existing technologies, as it exceeds the data-taking bandwidth limits. As a consequence, CMS developed a two-level trigger system to reduce the acquisition rate and at the same time keep events of possible physics interest. First, the Level-1 (L1) trigger processes the measurements from calorimeters and muon chambers and identifies the possible particle candidates; based on this information, decisions are taken within $3.8\,\mu$s after a collision and thus the event rate is reduced down to 100 kHz. The selected information is then passed to the high-level-trigger (HLT), where the rate is further reduced down to 1 kHz for offline storage. HLT algorithms use data from the whole CMS detector, using full granularity and resolution, and they perform a more sophisticated object reconstruction similar to the techniques employed in offline analyses, including track reconstruction. Both the L1 and HLT output rates can be adjusted by applying a factor ρ called *prescale*: when applied, only a fraction $1/\rho$ of the events selected by a specific trigger path is actually saved for offline analyses. The instantaneous luminosity $\mathcal{L}$ reached peaks of $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ during Run-2; these high values are achieved at the start of an LHC fill and they subsequently decreases over time as a fraction of protons is lost in the repeated bunch crossings and particle distributions are more spread as a consequence of the beam-to-beam interactions. As the two beams deteriorate, prescale and threshold values may be adjusted to keep a constant output rate.

2.3 Event simulation

The accurate simulation of proton-proton collisions and the interactions of the particles in the CMS detector is a crucial ingredient to physics analyses, as simulated events are used to extract physical quantities from data. With LHC, hadronic physics has turned into a high-precision domain and as of now, the lack of clear signals of BSM physics increases the need of precise theoretical predictions for known phenomena.

Various tools can be employed in the distinct stages of the simulation process. In this thesis, signal events are generated following the Monte Carlo (MC) method [55] with the MadGraph5 aMC@NLO [56] package, which makes use of tree-level and one-loop amplitude computations to predict physical observables and final-states at different perturbative accuracy. One of its main features is the possibility to automatically generate Feynman diagrams, calculate cross sections and simulate events from a given initial and final state. This process requires only a small number of input instructions, such as the Feynman rules corresponding to the Lagrangian of a given theory to generate the diagrams. The resulting simulated events are then passed to a separate program to simulate parton showers. Most of the samples employed in this work were produced using matrix element generators interfaced with the pythia [57, 58] software for the modeling of initial and final state radiation (ISR and FSR, respectively) as well as for the hadronization. If events are simulated at NLO precision in α_S, this procedure can lead to double counting of contributions that were already covered in the matrix element simulation. In MadGraph5 aMC@NLO, this is handled by applying negative weights following the FxFx merging scheme [59]. A graphic illustration of event modeling in proton-proton collisions is shown in Figure 2.8.

The final step consists in simulating the interactions of particles arising from the collisions inside the CMS detector. The Geant4 toolkit [60] is employed to model the components of the experiment. The details of the assembly, e.g. cables, electronics and support structures, are included in the simulation as well. This stage involves the simulation of the particle propagation through the detector components and the consequent energy deposits, as well as effects arising from the interactions with the magnetic field and bremsstrahlung. The particles directly emerging from simulated events are usually referred to as *generator-level* particles, while *detector-level* refers to the objects obtained from the event reconstruction, after particles have been detected.

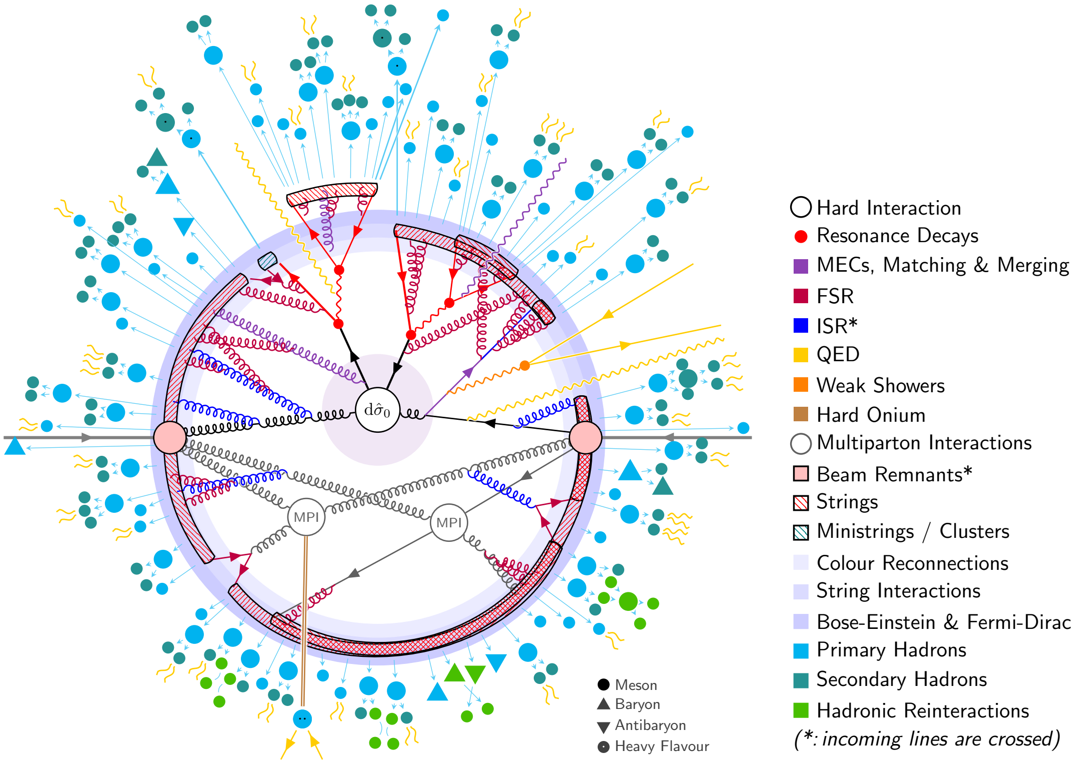


Figure 2.8: Schematic representation of event modeling with PYTHIA. Taken from Ref. [58].

2.4 Object reconstruction

In this section, the reconstruction methods used to convert the CMS raw data into a format containing meaningful physics quantities for the particle candidates are described. This is a crucial first step in every physics analysis, as it allows to extract meaningful physics quantities from the collected data. The description here only focuses on the methods employed for proton-proton collisions, as they were used to obtain the results presented in this thesis.

2.4.1 Tracks and primary vertex

Around one thousand charged particles are expected to pass through the CMS tracking system at each bunch crossing, making the track reconstruction extremely challenging. This procedure also needs to be sufficiently fast to be employed in the HLT. The track reconstruction is performed using the hits, i.e. the signals from the pixel and strip sensors, to estimate the momentum and position parameters of the charged particles and to combine them with the ones coming from adjacent layers; a global fit is then performed to obtain the final track. If the fraction of shared hits between two track candidates is higher than 19%, the track with the fewest hits is removed; if the number of hits is the same, the track that yields the best fit results is kept.

The interaction vertex of a specific event, usually denoted as *primary vertex*, is reconstructed by clustering the tracks that appear to be originating from the same collision and then extracting the vertex position from a fit on its associated tracks. The tracking algorithms can reconstruct the particle paths over the full pseudorapidity range of the CMS tracker ($|\eta| < 2.5$) and find charged particles with a p_{T} down to 0.1 GeV. Identifying the primary vertex with high precision is essential to estimate pile-up contributions, introduced in Section 2.1. The resolution of the reconstructed primary vertex strongly depends on the number of tracks on which the fit is performed, approaching about $10\,\mu$m for vertexes with at least 50 tracks [61]. The vertex reconstruction efficiency has been estimated to be close to 100% when more than two tracks are used. The reconstructed tracks are given as input to the Particle Flow (PF) algorithm [62] to identify the different particle candidates: electrons, muons, photons and jets.

2.4.2 Jets and b-tagging

The PF candidates are clustered into jets using the anti-k_{T} algorithm [63]. First, the distances $d_{i,j}$ between two objects (particles or pseudojets) i and j are defined, together with the ones from the beam $d_{i,\mathrm{B}}$:

$$
\begin{aligned}
d_{i,j} &= \min(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2})\frac{\Delta R_{ij}^2}{R^2}\,, \\
d_{i,\mathrm{B}} &= p_{\mathrm{T},i}^{-2}
\end{aligned}
\tag{2.4}
$$

where $\Delta R_{ij} = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$, with η and ϕ being the rapidity and the azimuthal angle of the particle candidates, respectively. R is the radius parameter and it is commonly set to 0.4. The clustering procedure starts from the smallest of the distances and, if it is a $d_{i,j}$, the two objects are combined together, while i is declared as jet if it is a $d_{i,\mathrm{B}}$, and then removed from the list of objects; this is repeated until no candidates are left. With this algorithm, hard particles (i.e. those having high p_{T}) will accumulate all the soft particles (i.e. with low-p_{T}) in a radius R, forming a conical jet. Soft particles also tend to combine with hard ones before clustering among themselves. These properties makes the algorithm *infrared safe* [64], as illustrated in Figure 2.9. If two jets share a fraction of particles, this is fixed by assigning the overlapping fraction to the jet with the highest momentum; if the momentum of the two jet candidates is the same, the overlapping part is simply equally divided between the two. The algorithm involves a combination of energy and angle in its distance definition, meaning that collinear emissions are clustered at the beginning of the iteration, which makes the procedure *collinear safe* [64]. To

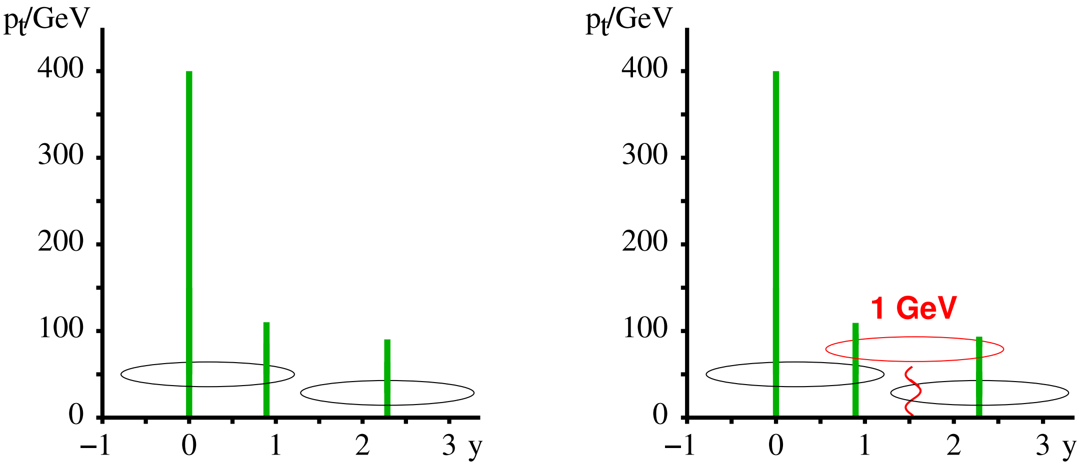


Figure 2.9: Example of infrared unsafety: the vertical lines corresponds to different partons with different p_{T}. With three hard candidates in the same neighborhood, a stable cone can arise from a new soft emission. Taken from Ref. [64].

better understand this concept, an example of collinear unsafety is illustrated in Figure 2.10.

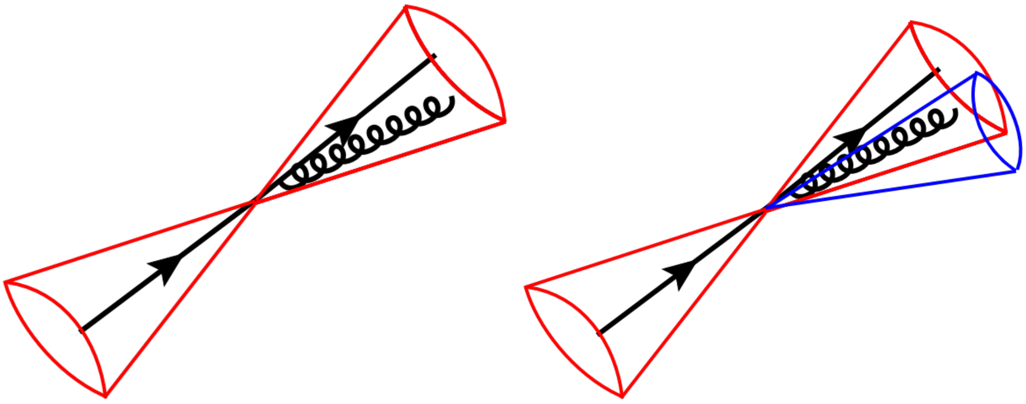


Figure 2.10: Example of collinear safety (left) and collinear unsafety (right). In the case of collinear unsafety, the collinear splitting of a particle leads to the reconstruction of two different jets instead of one. Taken from Ref. [65].

Jets originating from b and c quarks are usually denoted as *heavy flavour* jets, while *light flavour* refers to those coming from u, d and s quarks or gluons. Hadrons containing b or c quarks have a lifetime of the order of picoseconds and hence, they can travel up to a centimeter inside the detector before decaying. Subsequently, a secondary vertex can be reconstructed from the displaced tracks left by their decay products (Figure 2.11). The large masses of b and c quarks also result in daughter particles with higher momenta and a larger number of tracks associated with the jets. These properties can be exploited in machine learning algorithms to tag b jets, which is crucial to be able to reconstruct the full event in top physics analyses: the output scores of such discriminators can be used to select final states with b jets. In this thesis, the DeepJet algorithm [66] is used, which is based on a neural network trained on information coming from the PF candidates in the jet and secondary vertices with no pre-selection applied. The identification of b jets generally has different working points which are defined based on the b jet efficiency and the misidentification probability at some specific output scores. The looser working point has a misidentification probability for b jets of 10%, which improves to 1%

and 0.1% for the medium and tight working points, respectively; in contrast, the b jet efficiency goes from 95% in the loose case down to 65% for the tight one.

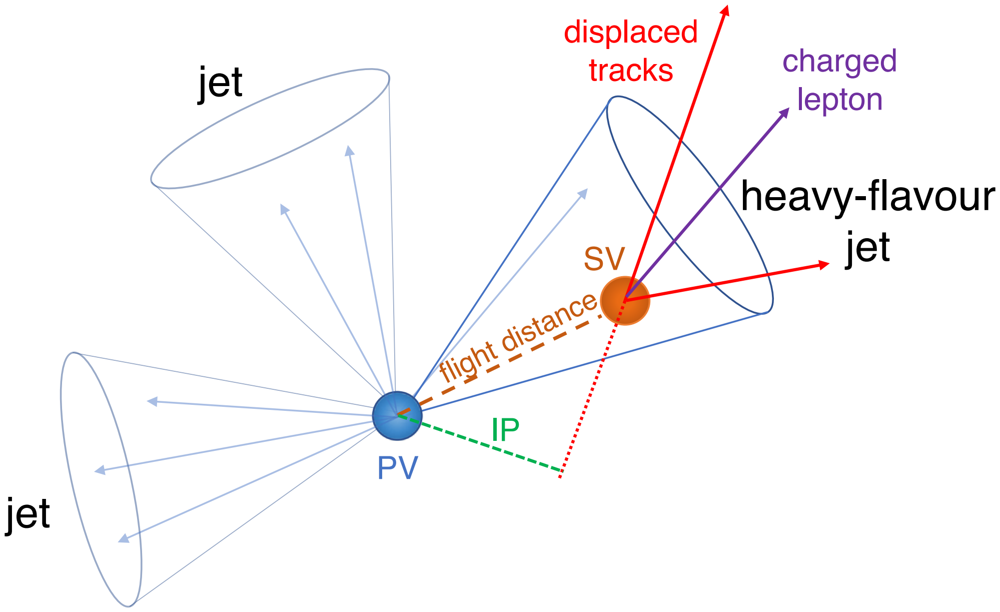


Figure 2.11: Illustration of heavy flavour jet properties. Hadrons have longer lifetimes compared to light flavour jets and they can travel up to a centimeter before decaying. The decay products produce displaced tracks and a secondary vertex (SV) can be reconstructed. The impact parameter is denoted as IP and PV refers to the primary vertex. Taken from Ref. [67].

2.4.3 Electrons

Electrons are measured by combination of the electromagnetic shower in the ECAL and tracks reconstructed from hits in the inner tracking system. In the ECAL, dedicated algorithms are used to combine the clusters emerging from the particle showers into a single object, from which the energy of the primary electron is estimated. The energy deposits are clustered assuming that each local maximum above 1 GeV originates from a single particle. The identification of electrons in CMS can be performed with cut-based techniques or using a multivariate analysis (MVA) approach. The latter is usually the preferred option for precision measurements involving final states with two or more leptons, while the former is widely used in model independent searches featuring unconventional signatures. Three different working points of the electron identification methods are defined corresponding to an efficiency of approximately 70, 80 and 90%, respectively. In all cases, the scale factors are found to be close to unity, as data and simulation efficiencies are usually compatible within 1-5%.

The electron reconstruction and identification efficiencies are estimated with the tag-and-probe method from Z → ee events [68] as a function of the η and p_{T} of the electrons; the efficiency is typically higher than 95% for electrons with a p_{T} above 20 GeV, and compatible between data and simulation within 2% [69]. The tag-and-probe method is summarized in Figure 2.12 and it can generally be used for electrons, muons and photons. In this technique, tight selection criteria are applied to one of the two electrons coming from the resonance decay, the tag; the other electron, the probe, is required to pass the selection criteria for which efficiency is under test.

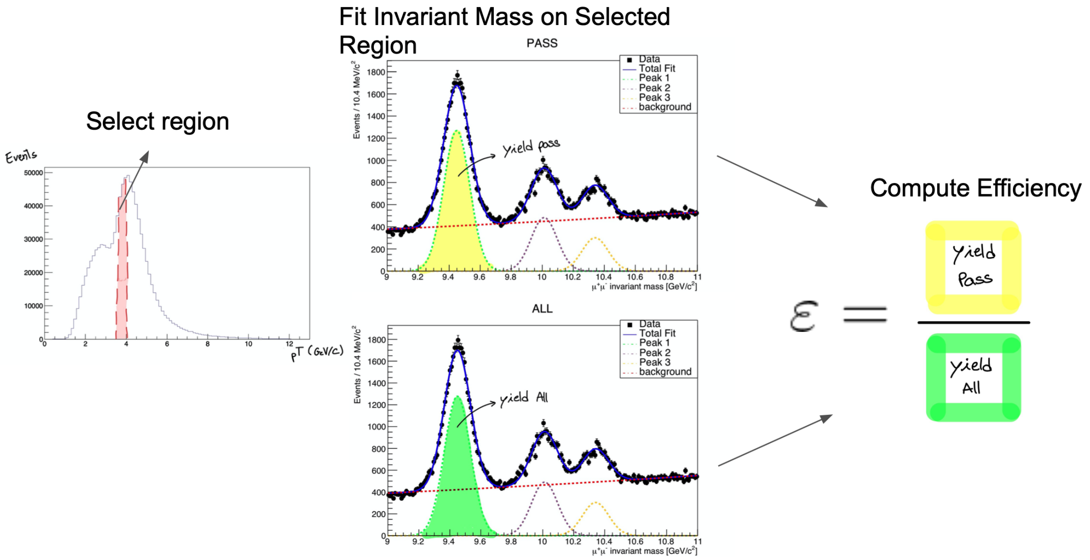


Figure 2.12: Summary of the tag-and-probe method for the efficiency evaluation. First, events from a region with high signal purity are selected (left) and a fit is performed on the invariant mass distributions to extract the signal yields and subtract the background contribution (middle). The efficiency is finally computed with the ratio of the signal yield from the passing probes and the total signal yield (right). Taken from Ref. [70].

The Z $\rightarrow e^+e^-$ events are characterized by a high signal purity and the distribution of the invariant mass of the two electrons is fitted with a signal-plus-background model to extract the signal yield. This is done separately for distributions with probe electrons that pass the selection criteria (*passing* probes) and those that fail the criteria (*failing* probes). The efficiency is then calculated as the ratio between the signal yield from the passing probes and the total signal yield, which includes both passing and failing probes. The ratio of the efficiencies in data and simulation is then usually used to evaluate a scale factor (SF) that is applied as correction to the simulation to ensure a better agreement with data.

2.4.4 Muons

As muons pass through the muon chambers, they ionize the gas and produce electrical signals on the wires and strips. These hits are then employed to reconstruct the particle's path. The tracks in the muon chambers are generally reconstructed independently and then matched with the tracks from the inner tracking system to reconstruct the full path and obtain global muon objects. These are then fed into the PF algorithm to identify the muon candidates.

A set of variables is then used to define identification criteria having different balances between efficiencies and purity. As for electrons, both cut-based and MVA algorithms can be employed for the muon identification [71, 72], depending on the p_{T} range of the particles involved in the analysis as well as on the level of purity required. The muon isolation is usually employed to further distinguish between prompt muons, i.e. muons that directly come from the decay of the particles produced in the collision, and those arising from decays within jets. This variable is evaluated relative to the muon p_{T} by aggregating the energy in geometrical cones around the particle. In this thesis, the so-called *mini-isolation* variable is used, which is based on the scalar sum of the p_{T} of the charged particles, neutral hadrons and photons reconstructed within a narrow cone centered on the lepton (electron or muon) direction:

$$I_\ell = \sum_{\text{charged}} p_\mathrm{T} + \max\left(0, \sum_{\text{neutrals}} p_\mathrm{T} - \rho\mathcal{A}\left(\frac{\mathcal{R}}{0.3}\right)^2\right) , \tag{2.5}$$

where ρ is the energy density of the neutral particles reconstructed within the geometric acceptance of the tracker and $\mathcal{A}$ is the effective area, obtained from the simulation by studying the correlation between I_ℓ and ρ. The size of the cone $\mathcal{R}$ is given by:

$$\mathcal{R} = \begin{cases} 0.05 & \text{if } p_\mathrm{T} > 200\,\mathrm{GeV} \\ 10\,\mathrm{GeV}/p_\mathrm{T} & \text{if } 50 < p_\mathrm{T} < 200\,\mathrm{GeV} \\ 0.20 & \text{if } p_\mathrm{T} < 50\,\mathrm{GeV} \end{cases} . \tag{2.6}$$

The value of the efficiency for muons ϵ_μ is then factorized into the different components:

$$\epsilon_\mu = \epsilon_{\text{reco+ID}} \cdot \epsilon_{\text{iso}} \cdot \epsilon_{\text{trig}} \tag{2.7}$$

where the term $\epsilon_{\text{reco+ID}}$ contains both the reconstruction efficiency in the muon system, including the matching to the inner track, and the identification efficiency. The following terms are evaluated with respect to the previous component, i.e. muons fulfilling the identification and isolation criteria are used as probes in the evaluation of the isolation and trigger efficiencies (ϵ_{iso} and ϵ_{trig}), respectively. Each term in Eq. 2.7 is evaluated independently as a function of the muon p_T and η with the tag-and-probe method, using $\mathrm{Z} \to \mu^+\mu^-$ events. The factorization of the muon efficiencies allows to separate the impact of the different sources and also leads to higher accuracy.

A fit is performed on the invariant mass distribution for passing and failing probes to estimate and subtract the background component. A so-called template distribution, describing the shape of the signal, is built from the simulated $\mathrm{Z} \to \mu^+\mu^-$ events convolved with a Gaussian distribution in order to take into account resolution differences between data and simulation. The background is modeled with the convolution of an exponential decay distribution with an error function; these were found to be suited to describe high- and low-mass regimes, respectively. In

Figure 2.13, the invariant mass distributions for the passing and failing probes are shown, together with the signal and background templates: in both cases, the fit leads to a very good description of the distributions.

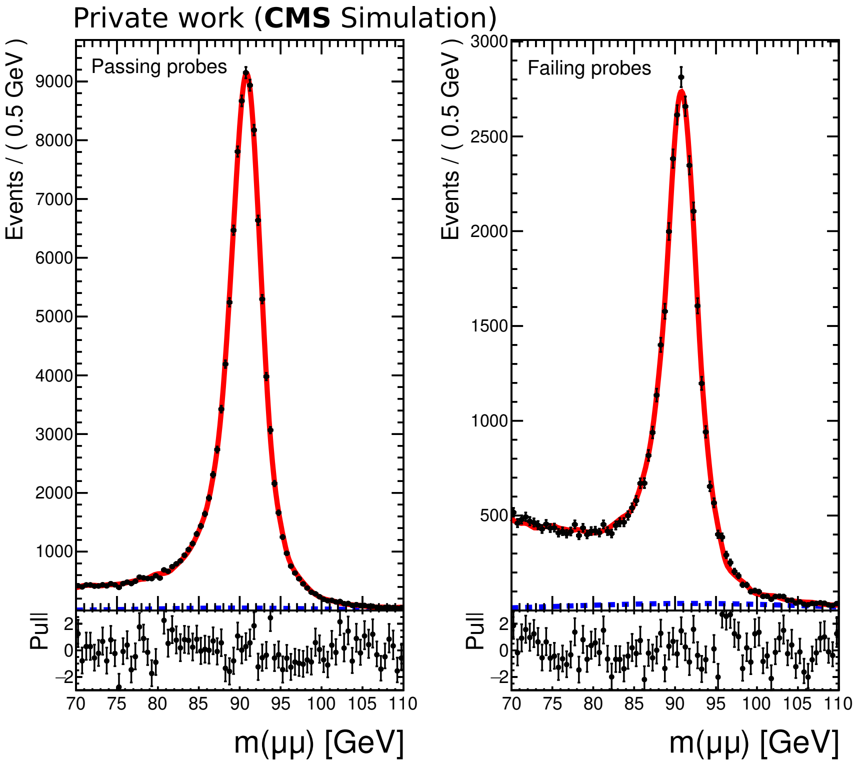


Figure 2.13: Example of tag-and-probe fit perform on the muon invariant mass in Z $\to \mu^+\mu^-$ events for the passing (left) and failing (right) probes. The red line indicates the signal+background model, while the blue dotted line is the background component.

2.4.5 Determination of shape-based scale factors

As already described in Section 2.4.3, MVA algorithms can be used to obtain lepton candidate samples of high efficiency and purity. This can be particularly useful in cases where the performance of the cut-based identification methods is not optimal. Not all the leptons appearing in the detector originate directly from the decay of the particles produced in the collisions, since they can also appear in the decay of hadronized quarks. Additionally, jets might be misidentified and reconstructed as leptons. This background contribution is usually defined as *nonprompt.* In the CMS experiment, one of the most commonly used MVA algorithms is the so-called prompt-muon MVA discriminator [72], which aims to select muons arising from the prompt decays of Z, W and Higgs bosons. The algorithm was developed in the context of multileptonic analyses featuring significant contributions from nonprompt backgrounds. The training was performed with muons passing the loose cut-based identification criteria and having a p_{T} higher than 5 GeV, in order to reduce the contribution from poorly reconstructed muons while keeping an efficiency of almost 100%. The signal samples are obtained from $\mathrm{t\bar{t}H}$ events, while a $\mathrm{t\bar{t}}$ sample is used as background. A comparison between the performance of the prompt-muon MVA and that of the cut-based approach is shown in Figure 2.14, where the efficiency and nonprompt rate are defined as the fraction of muons passing the signal and background selection, respectively. As the figure indicates, at an efficiency of 90%,

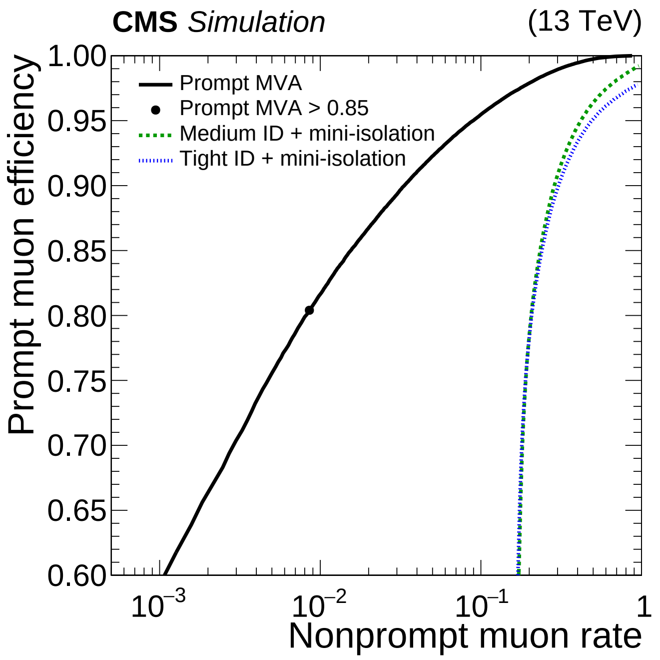


Figure 2.14: Performance of the prompt-muon MVA, compared with that of the tight and medium cut-based identification criteria and mini-isolation requirements. The cut-based selection combines a set of requirements on the mini-isolation variable with the medium and tight working points, i.e. corresponding to a 80% and 90% efficiency, respectively. Taken from Ref. [72].

the background suppression of the MVA algorithm, expressed as nonprompt muon rate is about 3×10^{-2}, a factor 10 better than that of the cut-based algorithm.

The discriminator output is a continuous variable which can be used to select the most suitable working point for each analysis. This information has to be included in the definition of tag muons when evaluating the identification efficiencies and SFs. Physics analyses targeting multilepton final states might benefit from exploiting the information contained in the full shape of the discriminator output distribution, e.g. by using it to construct input variables for MVA algorithms that are specifically trained for the analysis. To do so, the SFs are determined in the continuum between the different working points, in order to correct the full shape of the simulated discriminator distribution to that of the data. This allows for finding an optimal working point for a given physics analysis, rather than having to choose among a few predefined working points for which SF have been determined. In this thesis, such shape-based SFs were determined for the first time. The software is now implemented for any MVA-based identification method in the `spark_tnp` Framework [73], a package commonly used in CMS to evaluate efficiencies with the tag-and-probe method, and is now available for the use with future data sets.

The procedure is implemented as follows:

- first, the SFs are evaluated with the tag-and-probe method for different working points of the discriminator output by requiring both a lower and an upper threshold (e.g. an output score between 0.85 and 0.9);
- subsequently, the SFs are interpolated in order to determine a continuous set of SF values as a function of the discriminator value.

Requiring an upper threshold of the discriminator output in the first step is essential to avoid overlaps among the different SFs. With the upper limit, the distinct ranges ensure a non-overlapping segmentation of the discriminator output and avoid ambiguities in the interpolation process.

The efficiencies and SFs can still be determined as a function of the muon's kinematic properties, and the interpolated SFs will also depend on those. As an example, the values of the extracted SFs as a function of the prompt-muon MVA output are shown in Figure 2.15 for different ranges of muon p_T and η; the values are obtained with data from proton-proton collisions recorded with the CMS detector in 2018.

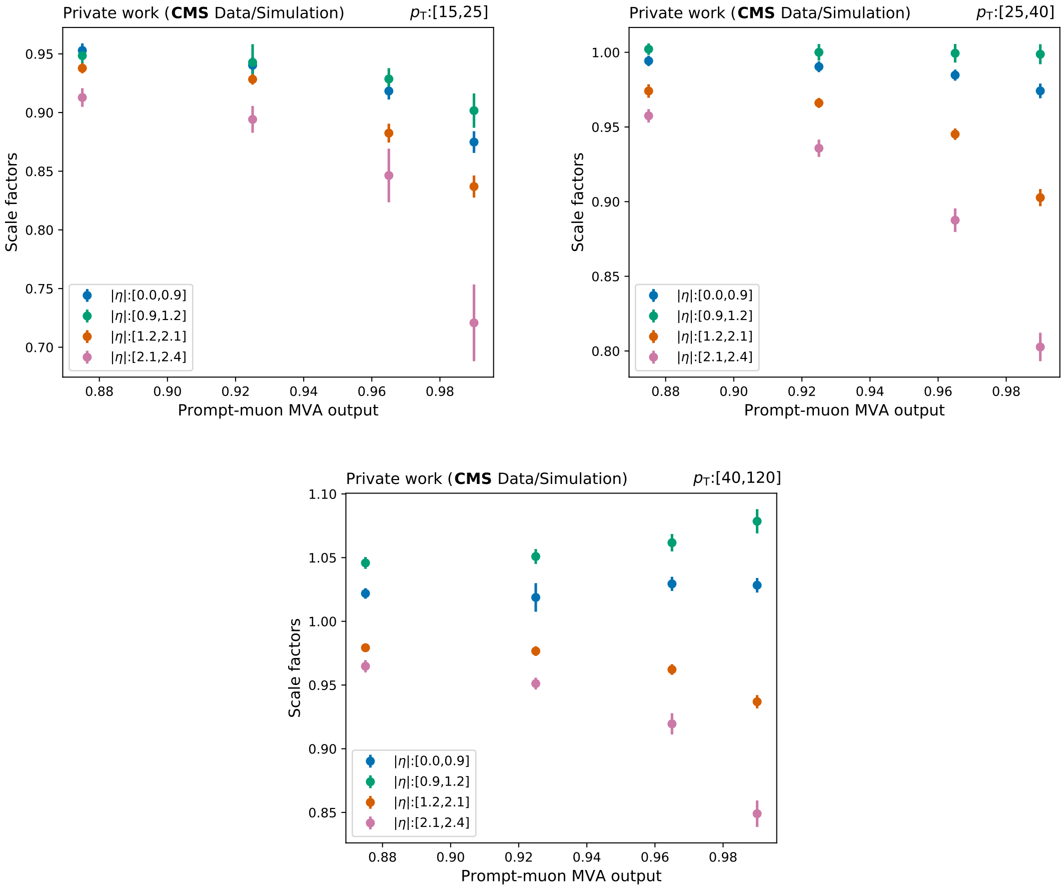


Figure 2.15: Muon SFs for the prompt-muon MVA ID, extracted with the tag-and-probe method in different p_T and η bins. The error bars include both systematic and statistical uncertainties, which are found to be higher for muons with low p_T and $|\eta| > 2.1$. However, this is not expected to have a significant impact on physics measurements, as the prompt-muon MVA ID is meant to be employed in analyses featuring muons in the medium p_T range.

The SFs are generally close to unity, and tend to reach lower values at the end of the discriminator output distribution, in the high pseudorapidity region. This trend was observed across various MVA algorithms and it is due to a slightly worse performance of the simulation in describing the data in that region. These SFs are evaluated with the tag-and-probe method and will be used as input for the interpolation procedure. The systematic uncertainties on the extracted SFs are shown in Figure 2.16; they include an alternative template for the signal modeling, a shift in the isolation cuts applied on the tag muon, and variations of the binning and range of the muon invariant mass distribution on which the fit is performed. Their impact is found to be generally small compared to the statistical uncertainties.

The interpolation procedure is performed with SciPy [74], a Python library widely employed in scientific and technical computing. First, the input data are triangulated with Qhull [75], a set of command-line tools developed in the context of computational geometry. Subsequently, the linear barycentric interpolation is performed on each triangle to construct the interpolant. The resulting interpolated object can be employed in physics analyses to estimate the muon SFs as a function of its MVA output score and other kinematic properties. The stability of this method can be assessed by comparing the interpolated SFs with the ones evaluated with the tag-and-probe method, for a chosen working point. This check is shown in Figure 2.17 for the case where the interpolation is performed from the SFs shown in Figure 2.15; it can be seen that the two procedures give very similar quantitative results in the centers of the bins where the SF are determined. The interpolation also makes the trend of the SFs smoother as a function of the muon p_{T} and η.

The shape uncertainties are estimated by varying the SF data points coherently up and down within their statistical and systematic variations, and then repeating the interpolation. With this method, an envelope around all possible variations is determined. Technically, this slightly overestimates the uncertainties, which are largely bin-to-bin uncorrelated, because the statistical uncertainty is dominant. On the other hand, this effect is expected to be negligible for most physics analyses due to the small size of the uncertainties and it also covers for eventual systematic uncertainties arising from the choice of the interpolation method.

Following the example of the prompt-muon MVA ID, a comparison between the shape of the prompt-muon MVA discriminator before and after the application of

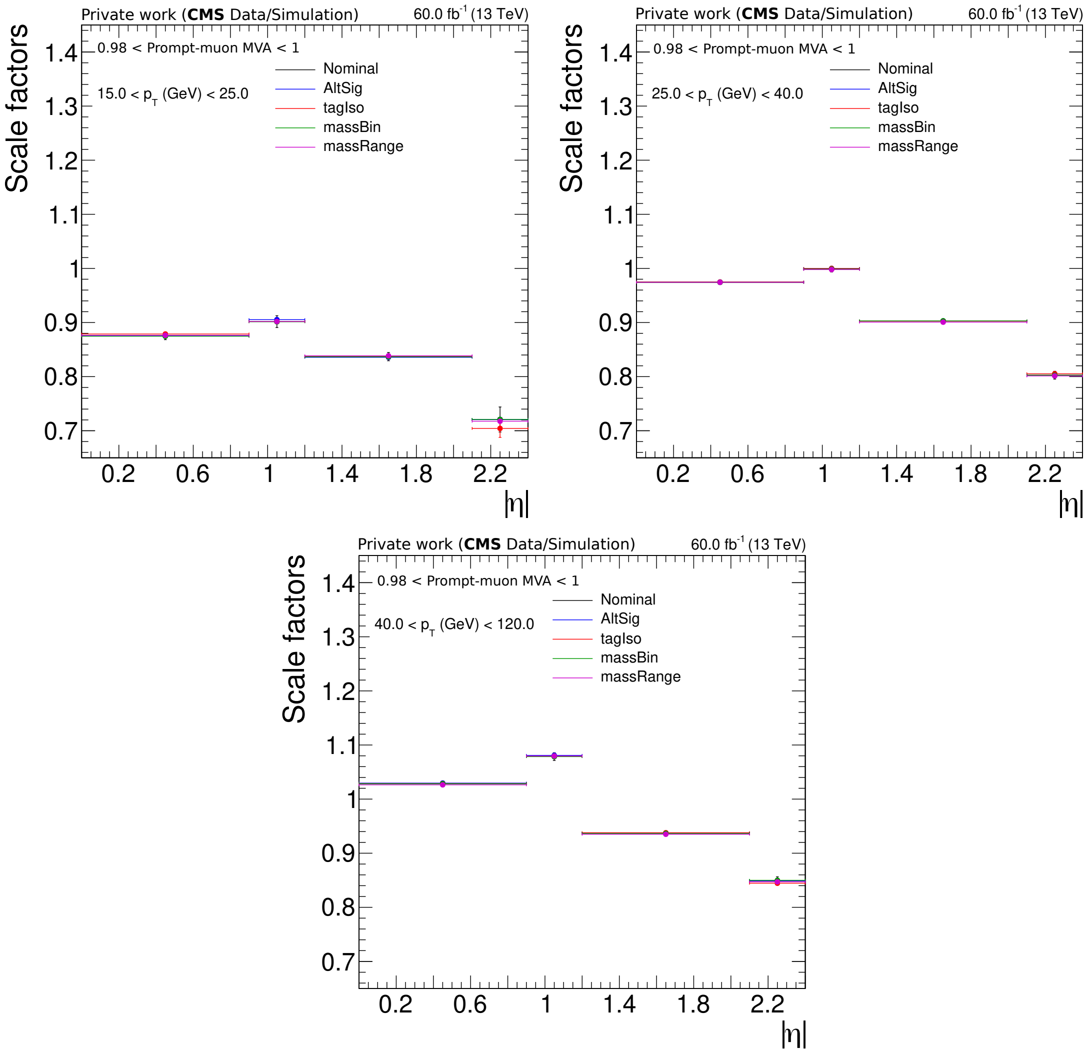


Figure 2.16: Muon SFs for the prompt-muon MVA ID, extracted requiring an output score between 0.98 and 1, in different p_{T} and η bins. The procedure is repeated for different ranges of the discriminator before performing the interpolation. The systematic variations are plotted together with the nominal values and they include alternative signal template (AltSig), different isolation cuts on the tag muon (tagIso), and a shift in the binning and range of the muon invariant mass distribution (massBin and massRange, respectively).

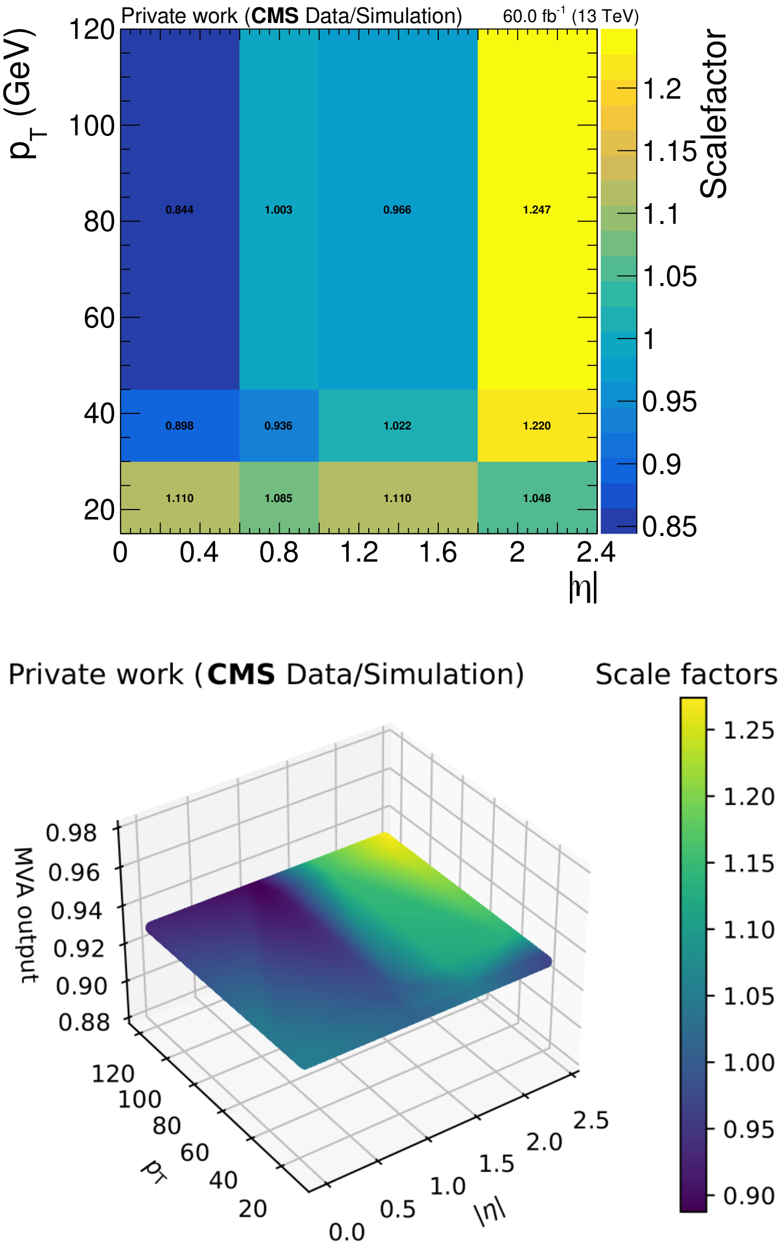


Figure 2.17: Top: muon SFs for an output score of the prompt-muon MVA ID between 0.92 and 0.94, evaluated with the tag-and-probe method in bins of p_{T} and η. Bottom: projection of the results from the three-dimensional interpolation of the muon SFs for a discriminator output value of 0.93. The interpolation is performed as a function of the prompt-muon MVA output score, and the muon p_{T} and η.

the continuous SFs is shown in Figure 2.18, together with the corresponding systematic uncertainties. Even though the data-to-MC agreement is already good, it can be seen that the application of the SFs brings an overall improvement. Remaining discrepancies in the distributions of data and MC may also be attributes to possible differences in the physics model of the MC generator.

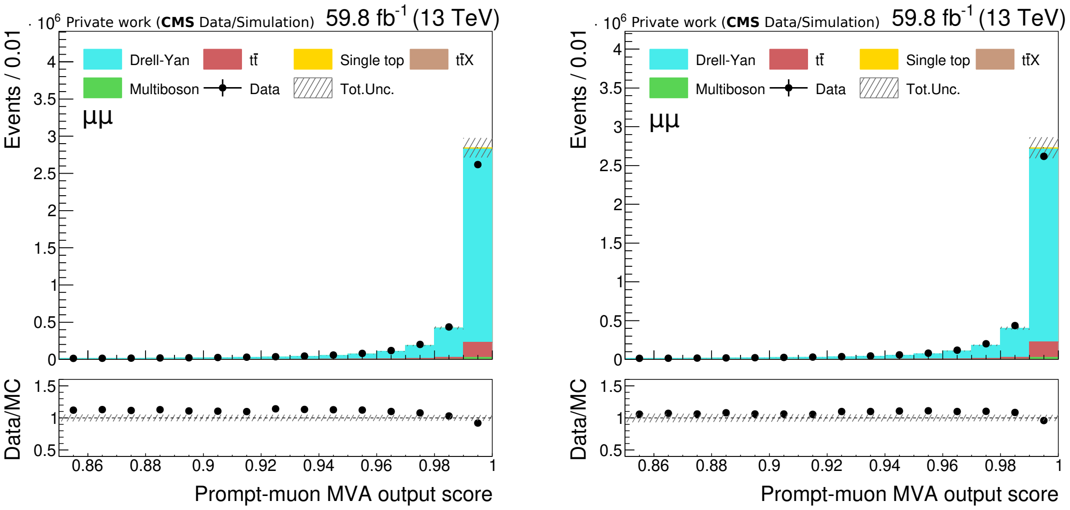


Figure 2.18: Output score distribution of the prompt-muon MVA before (left) and after (right) the application of the continuous SFs. The interpolant is constructed as a function of the discriminator output, the p_{T} and η of the muons.

Another commonly used MVA alogorithm is the TOP LeptonMVA [76], developed by the CMS group of Ghent University in the context of top multilepton analyses. This algorithm aims to select prompt leptons with slightly looser requirements compared to the prompt-muon MVA. It has been trained with variables fulfilling a minimal object selection that achieves a very high lepton efficiency. In the training, alongside the lepton isolation, the p_{T}, $|\eta|$, and variables related to the jet closest to the lepton (e.g. the number of charged particles in the jet or its DeepJet score) are employed. The simulated samples of tZq, $\mathrm{t\bar{t}Z}$ and $\mathrm{t\bar{t}W}$ samples are used for prompt leptons, while nonprompt leptons are taken from $\mathrm{t\bar{t}}$ simulation. The TOP LeptonMVA discriminator is used for the event selection of the main results presented in this thesis, which will be discussed from the next chapter. This choice is based on the samples employed for the training, which for the most part coincide with the events targeted in this work.

CHAPTER 3

Event selection and reconstruction

In this and the following chapters, the main result of this thesis is presented: the production of top quark pairs ($\mathrm{t\bar{t}Z}$), and single top quarks (tWZ and tZq) in association with a Z boson are measured simultaneously for the first time, obtaining both the data points and the full set of correlations. The analysis is performed with data collected with the CMS detector during Run-2, corresponding to an integrated luminosity of 138 fb^{-1}. The data were produced in proton-proton collisions at the LHC, at a center-of-mass energy of 13 TeV. This chapter focuses on the signal event selection and reconstruction and it is structured as follows: first, an overview of the simulated samples employed for the cross section measurement and their features is given. The object and event selection criteria are discussed in Sections 3.2 and 3.3, respectively. Finally, Section 3.4 is dedicated to the nonprompt background estimation.

3.1 Simulation and datasets

The full dataset collected by the CMS experiment during Run-2 is analyzed. The signal processes ($\mathrm{t\bar{t}Z}$, tWZ and tZq) are simulated at NLO in perturbative QCD with MadGraph5_aMC@NLO v2.6.5 [56] interfaced with PYTHIA v8.2 [57] for the parton shower modeling. The top quark is simulated as resonant particle, setting its mass to 172.5 GeV. The measured cross sections reported in this analysis are defined for the invariant mass range within 70 and 110 GeV for lepton pairs with opposite charge and same flavour. For the tZq sample, the four-flavour scheme (4FS) described in Section 1.8 is used: the b quark in the initial state is required to come from gluon splitting at the matrix element level (Figure 3.1). This leads to a better modeling of the final-state particle kinematics. The generation of the $\mathrm{t\bar{t}Z}$ and tWZ samples is performed in the five-flavour scheme (5FS), where the b quark comes directly from the PDF. As explained in Section 1.8.1, this is preferable in order to remove the overlap between the two processes [37], which is handled with the MadSTR plugin [77], MadGraph5_aMC@NLO v2.6.5, and pythia v8.2. As explained in Section 1.8, the tWZ samples are produced with two different diagram removal (DR) methods: DR1, where the resonant terms in the amplitude are discarded, and DR2, where their interference with the non-resonant part in the squared amplitude is kept. The DR1 scheme is taken as nominal, while the DR2 one is used to assign a shape uncertainty to the signal modelling. The diagram subtraction (DS) scheme is not used, as it is found to lie in between DR1 and DR2 [78]. An overlap between tZq and tWZ also arises when considering NLO electroweak corrections, and for s-channel tZq production at NLO QCD [39]. Both these scenarios are not included in the simulations and their effect is considered to be negligible at the current level of precision of the experimentally measured cross sections. All the signal samples are simulated with the NNPDF v3.1 [24] PDF set with NNLO precision.

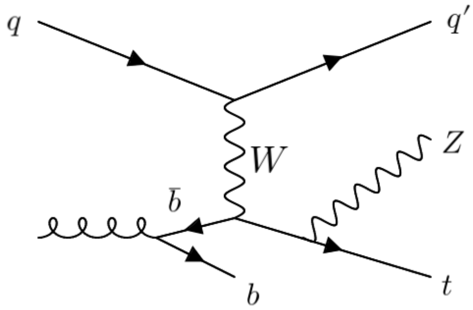


Figure 3.1: LO Feynman diagram of tZq production in the 4FS. The b quark in the initial state comes from gluon splitting rather than from proton PDFs.

Most of the background processes are generated at NLO in perturbative QCD and the the samples generated at leading order are used only for some minor contributions, such as the associated production of top quark pairs with a photon ($\mathrm{t\bar{t}\gamma}$) or with two electroweak bosons ($\mathrm{t\bar{t}XX}$). A summary of the simulated samples and their features can be found in Table 3.1. In this study, the nonprompt contribution mainly arises from processes like $\mathrm{t\bar{t}}$, Drell-Yan (DY), single top production, W+jets or even hadronic decays from $\mathrm{t\bar{t}Z}$ itself and $\mathrm{t\bar{t}W}$. Since this type of events are challenging to model, the nonprompt contribution is estimated with a data-driven procedure explained in Section 3.4.

Table 3.1: Details of MC samples used for this measurement. In the process names, X includes both the W and Z bosons.

Process	Event generator	Perturbative QCD order	Details
Signal samples			
tZq ($\mathrm{Z} \to \ell^+\ell^-$)	MadGraph5_aMC@NLO	NLO	4FS
$\mathrm{t\bar{t}Z}$ ($\mathrm{Z} \to \ell^+\ell^-$)	MadGraph5_aMC@NLO	NLO	5FS
tWZ ($\mathrm{Z} \to \ell^+\ell^-$)	MadGraph5_aMC@NLO	NLO	5FS, DR1 and DR2
Background samples			
WZ, $\mathrm{t\bar{t}W}$, Xγ, multiboson	MadGraph5_aMC@NLO	NLO	5FS
$\mathrm{t\bar{t}\gamma}$, $\mathrm{t\bar{t}XX}$	MadGraph	LO	5FS
ZZ, WW, $\mathrm{t\bar{t}H}$	powheg [79]	NLO	5FS

Several triggers are combined to maximize the selection efficiency in both data and simulations. The HLT algorithms require at least one, two or three leptons with a transverse momentum threshold between 8 and 50 GeV. This choice allows reaching a selection efficiency close to 100% in both data and simulation.

3.2 Object selection

After the trigger selection, additional requirements are imposed on the signature of the reconstructed particle candidates in order to reduce the background contribution.

3.2.1 Leptons

The measurement presented in this thesis is performed using events with three or four leptons (muons or electrons) in the final state. This also includes a small contribution from events involving the leptonic tau decays. From now on, the term lepton will be used exclusively to denote electrons and muons, unless otherwise specified.

The lepton identification method is based on the TOP LeptonMVA algorithm [76] described in Section 2.4.4. First, basic quality cuts are applied to the lepton candidates. For electrons, the track is required to have a hit in each layer of the pixel detector, except for at most one layer. This allows to remove contamination from electron candidates coming from photon conversion. In the case of muons, additional cuts are applied on the quality of the muon track fit. Furthermore, in order to reduce the background contribution from electrons and muons coming from charm and bottom decays, as well as from pions and kaons, isolation cuts are applied to the leptons. Electrons and muons are required to have a mini-isolation (see Section 2.4.4) smaller than 40% of the lepton p_{T}. This requirement also allows to reduce the contribution from pile-up. Furthermore, the impact parameter of the track with respect to the leading primary vertex is required to be smaller than 0.05 cm (0.1 cm) in the transverse (longitudinal) plane.

Leptons satisfying the above selection criteria are further categorized into *tight* and *fakeable* leptons, and the details are given in Table 3.2. The distinction between tight and fakeable is useful to estimate the nonprompt lepton contribution from data, as it will be explained in detail in Section 3.4. Tight leptons are always required to have a classifier output score larger than 0.64. In the case of fakeable leptons, if the score is smaller than 0.64, the relative isolation of the jet matched to the lepton at generator level is required to be smaller than 0.30 or 0.35, depending on the year. If there is no matched jet, e.g. in data or in the case of leptons arising from hadron decays, this variable is set equal to the lepton isolation. The jet that is closest to the lepton within $\Delta R < 0.4$ is required to fail the medium working point of the DeepJet discriminant. Additionally, fakeable and tight electrons have to satisfy a set of conditions on the width of the electron cluster in η-direction ($\sigma_{i\eta i\eta}$), the ratio of energy in the HCAL to the energy in the ECAL associated to electron (H/E), and the difference between the reciprocal of the electron cluster energy and the reciprocal of its track momentum (1/E − 1/p). This emulates the identification criteria applied at trigger level. Fakeable and tight muons are required to pass the medium working point of the cut-based ID described in Section 2.4.4.

The sources of the nonprompt background tend to be different in DY and $\mathrm{t\bar{t}}$ events: for example, heavy flavour hadrons are expected to be more common in processes involving the production of top quarks due to the presence of b quarks in the final states. For this reason, the requirements that allow to distinguish between fakeable and tight leptons have been optimized to make the nonprompt estimation valid for all the different event topologies; the optimization procedure was done in the context of the first tWZ analysis in CMS [3].

3.2.2 Jets

Jets are built from PF candidates and clustered using the anti-k_T algorithm with a distance parameter $R = 0.4$, as already discussed in Section 2.4.2. Selection cuts are then applied on the transverse momentum and pseudorapidity: $p_\mathrm{T} > 25\,\mathrm{GeV}$ and $|\eta| < 5.0$. Loose quality requirements are applied to reject spurious jets from detector noise while keeping almost 99% of the jets that arise from the main collision. In order to prevent double counting of the same object as both a lepton and a jet, a separation of $\Delta R > 0.4$ from the nearest selected lepton is also required for each jet. In order to identify and reject pile-up jets, an MVA discriminant known as pile-up jet ID [81] has been developed combining information on the jet shape and the particle multiplicity. The medium working point of the algorithm is used, which allows to exclude about 40% of the pile-up jets while keeping a 90% efficiency for jets coming from the hard interaction. Jet energy corrections (JECs) are applied to simulated jets in order to remove residual energy contributions from pile-up events and to correct the simulated jet response. These corrections are applied by scaling the jet four-momentum. Since the jet energy resolution (JER) in data is generally worse than in the simulation, a smearing procedure is applied to the jets in the MC in order to describe the data such that both resolutions match [82].

To identify jets originating from b quarks, the medium working point of the DeepJet algorithm described in Section 2.4.2 is used, corresponding to an 85% efficiency and a misidentification rate of 1% for jets originating from the hadronization of u, d, or s quarks and gluons, and 15% for jets from c quarks. This criterion is applied

Table 3.2: Selection criteria used to define fakeable and tight leptons. When data-taking era is not specified, the same cuts are applied to all data-taking periods. Two eras are considered in 2016, referred to as 2016 pre-VFP and 2016 post-VFP: in the former, the silicon strip tracker had fewer hits due to saturation effects in the readout electronics under high-luminosity conditions. This effect was mitigated in the post-VFP era by changing the pre-amplifier bias voltage (VFP) [80].

		Electrons		Muons	
	Data-taking era	Fakeable	Tight	Fakeable	Tight
$\sigma_{i\eta i\eta}$		$< 0.011/0.030$[1]	$< 0.011/0.030$[1]	–	–
$\frac{H}{E}$		< 0.10	< 0.10	–	–
$\frac{1}{E} - \frac{1}{p}$		> -0.04	> -0.04	–	–
Muon cut-based ID		–	–	Medium	Medium
Electron MVA ID		Loose[†]	–	–	–
DeepJet$^{closest\ jet}$	2016 pre-VFP	< 0.15[†]	–	< 0.60[†]	–
	2016 post-VFP	< 0.08[†]	–	< 0.60[†]	–
	2017	< 0.08[†]	–	< 0.60[†]	–
	2018	< 0.06[†]	–	< 0.50[†]	–
Jet relative isolation	2016 pre-VFP	< 0.30[†]	–	–	–
	2016 post-VFP	< 0.35[†]	–	–	–
	2017	< 0.30[†]	–	–	–
	2018	< 0.35[†]	–	–	–
TOP LeptonMVA		–	> 0.64	–	> 0.64

[1] Barrel/endcap

[†] Cuts only applied if TOP LeptonMVA score is < 0.64.

to jets reconstructed within the acceptance of the CMS pixel detector, i.e. within $|\eta| < 2.4$ for data taken in 2016 and $|\eta| < 2.5$ for 2017 and 2018; from now on, these will be referred to as *central* jets. The difference of the simulation in the tagging efficiency is corrected with a per-jet data-to-simulation SF evaluated as a function of the p_{T}, η and flavour of the jet. These SFs are centrally provided within the CMS Collaboration [67] and were extracted from simulated muon-enriched multijet events: due to the muon requirement, this topology is enriched in jets containing heavy-flavour hadrons and can be used to measure the SFs with good precision. The MC b-tagging efficiencies ϵ are derived as a function of p_{T} and η within the analysis phase space, as they strongly depend on the event kinematics. They are defined as the fraction of jets with a certain flavour at generator-level (flav) passing the considered working point:

$$\epsilon(p_{\mathrm{T}}, \eta, \mathrm{flav}) = \frac{\mathrm{N}_{\mathrm{b-tagged\,jets}}(p_{\mathrm{T}}, \eta, \mathrm{flav})}{\mathrm{N}_{\mathrm{total\,jets}}(p_{\mathrm{T}}, \eta, \mathrm{flav})}\,. \tag{3.1}$$

The aforementioned SFs are used to define the probability of a given configuration of jets as:

$$P(\mathrm{MC}) = \prod_{\mathrm{i}}^{\mathrm{tagged\,jets}} \epsilon_{\mathrm{i}}(p_{\mathrm{T}}, \eta, \mathrm{flav}) \prod_{\mathrm{j}}^{\mathrm{not\,tagged\,jets}} (1 - \epsilon_{\mathrm{j}}(p_{\mathrm{T}}, \eta, \mathrm{flav})) \tag{3.2}$$

for simulated events and

$$P(\mathrm{data}) = \prod_{\mathrm{i}}^{\mathrm{tagged\,jets}} \mathrm{SF}_{\mathrm{i}} \cdot \epsilon_{\mathrm{i}}(p_{\mathrm{T}}, \eta, \mathrm{flav}) \prod_{\mathrm{j}}^{\mathrm{not\,tagged\,jets}} (1 - \mathrm{SF}_{\mathrm{j}} \cdot \epsilon_{\mathrm{j}}(p_{\mathrm{T}}, \eta, \mathrm{flav})) \tag{3.3}$$

for data. Each event is then corrected with the factor $P(\mathrm{data})/P(\mathrm{MC})$.

3.3 Signal definition

This measurement makes use of final states with three leptons for all the signal processes, as it is the channel with the highest sensitivity. Figure 3.2 shows some Feynman diagrams with the targeted final states for the three signal processes. Two of the leptons are expected to arise from the decay of the Z boson, while the third originates from the decay of a top quark via a W boson. In the tWZ process, the top quark can also decay hadronically and in this case, the third lepton is generated from the W boson of the associated production. When considering the final state with three leptons, two jets are expected in the tZq process at LO: one b jet coming from the top quark and the so-called *recoil jet*, which is usually emitted in the forward regions of the detector and with a high p_{T}, as it recoils against the whole system. Three and four jets are expected in the final states of tWZ and $\mathrm{t\bar{t}Z}$, respectively.

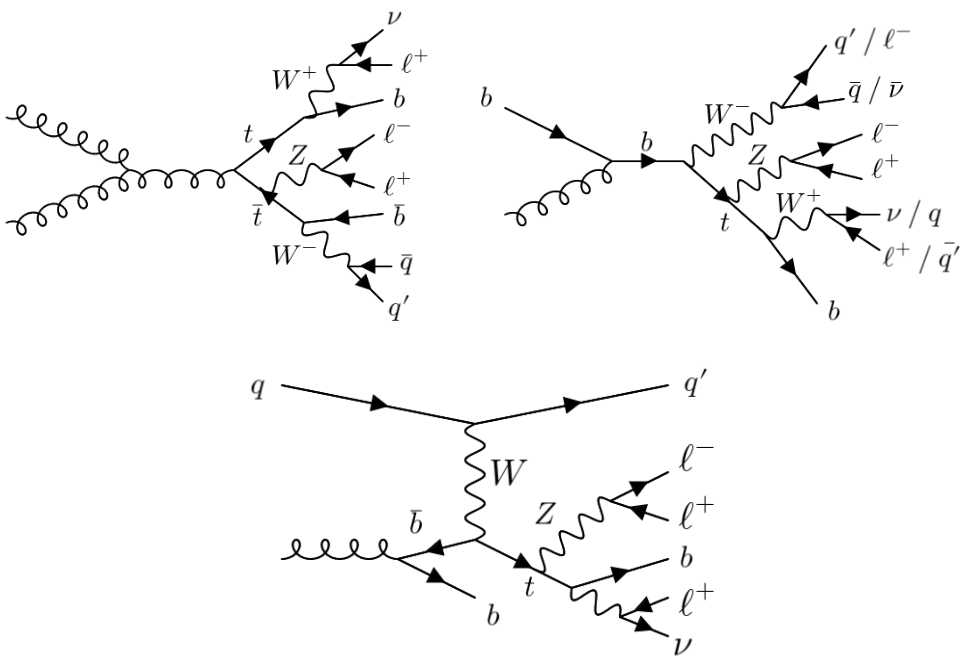


Figure 3.2: Feynman diagrams of the three lepton final states for $\mathrm{t\bar{t}Z}$ (top left), tWZ (top right) and tZq (bottom).

Based on the object identification requirements, a signal region (SR) is defined with the following criteria:

- exactly three leptons, with:
    - $p_{\mathrm{T}} > 25$, 15 and 10 GeV;

  - two leptons with same flavour and opposite charge, and an invariant mass compatible with that of the Z boson within the range $\mid m_{\ell^+\ell^-} - m_{\mathrm{Z}} \mid < 20\,\mathrm{GeV}$;

- at least two jets, with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 5$;
- at least one b jet (included in the previously selected jets).

The lepton with the highest p_{T} will be referred to as *leading* lepton, while the terms *sub-leading* and *trailing* will be used for the leptons with second highest and lowest p_{T}, respectively.

The expected yields from the different processes are reported in Table 3.3 for the full Run-2 data set and some distributions comparing the data with the expectation for signal and backgrounds in the SR can be found in Figures 3.3 and 3.4: it can be seen that a small excess is observed in the region above 60 GeV for the leading lepton p_{T}, and in the central region for the jet pseudorapidity. However, in general, a good data-to-MC agreement is reached in all the distributions, which cover different aspects of the event topology. The main backgrounds arise from the nonprompt contribution and the WZ process. Even if the expected yield of tWZ events is significantly smaller compared to that of $\mathrm{t\bar{t}Z}$, its presence has a substantial impact on the measurement due to the overlap between the two processes. In Figure 3.5 the distribution of the invariant mass of the leptons originating from the decay of the Z boson is shown, split according to the flavours of the leptons in the final states. This check allows to assess the stability of the selection criteria, as a good agreement between data and simulation is observed across all the lepton channels.

Table 3.3: Expected yields in the SR from signal and background processes for the full Run-2 data set. In the process names, V includes both the W and Z boson.

Process	Expected yields
data	3473
tZq	368
$\mathrm{t\bar{t}Z}$	960

Table continues on the next page

Process	Expected yields
tWZ	128
Nonprompt	707
WZ	516
$t\bar{t}X$	155
Multiboson	126
Vγ	290
Signal (total)	1456
Background (total)	1794

3.3.1 Neutrino reconstruction

Exactly one neutrino is expected in the final state, for all the signal processes. The neutrino reconstruction is performed by first identifying the lepton that comes from the W decay, i.e. the only lepton that is not produced in the decay of the Z boson. The identification is performed by looking at the invariant mass of the leptons in the final state, as well as their flavour and charge: the pair of lepton with opposite electric charge and same flavour that minimizes $\mid m_{\ell^+\ell^-} - m_\mathrm{Z} \mid$ is selected as Z boson candidate, while the third lepton is identified as the one produced in the W boson decay and will be denoted here as ℓ_W. The transverse W boson mass m_T^W can be reconstructed by associating the two measured components of the missing energy $\mathrm{E}_\mathrm{T}^\mathrm{miss}$ and ϕ^miss to those of the neutrino ν and using then the transverse components of ℓ_W as

$$m_\mathrm{T}^\mathrm{W} = \sqrt{2 \cdot p_{\mathrm{T},\,\nu} \cdot p_{\mathrm{T},\,\ell_\mathrm{W}}(1 - \cos\Delta\phi(\ell_\mathrm{W},\,\nu))} \tag{3.4}$$

where $\Delta\phi(\ell_\mathrm{W},\,\nu)$ is the difference in azimuthal angle between ℓ_W. The component $p_{z,\,\nu}$ can be determined from the equations of the energy-momentum conservation at the W decay vertex after setting the invariant mass of the $\nu - \ell_\mathrm{W}$ system to the W boson mass [9]. If the masses of ν and ℓ_W are neglected, this gives

$$p_{z,\,\nu}^{\pm} = \frac{\Lambda \cdot p_{z,\,\ell_\mathrm{W}}}{p_{\mathrm{T},\,\ell_\mathrm{W}}} \pm \sqrt{\frac{\Lambda^2 \cdot p_{z,\,\ell_\mathrm{W}}^2}{p_{\mathrm{T},\,\ell_\mathrm{W}}^4} - \frac{E_{\ell_\mathrm{W}}^2 \cdot p_{\mathrm{T},\,\nu}^2 - \Lambda^2}{p_{\mathrm{T},\,\ell_\mathrm{W}}^2}} \tag{3.5}$$

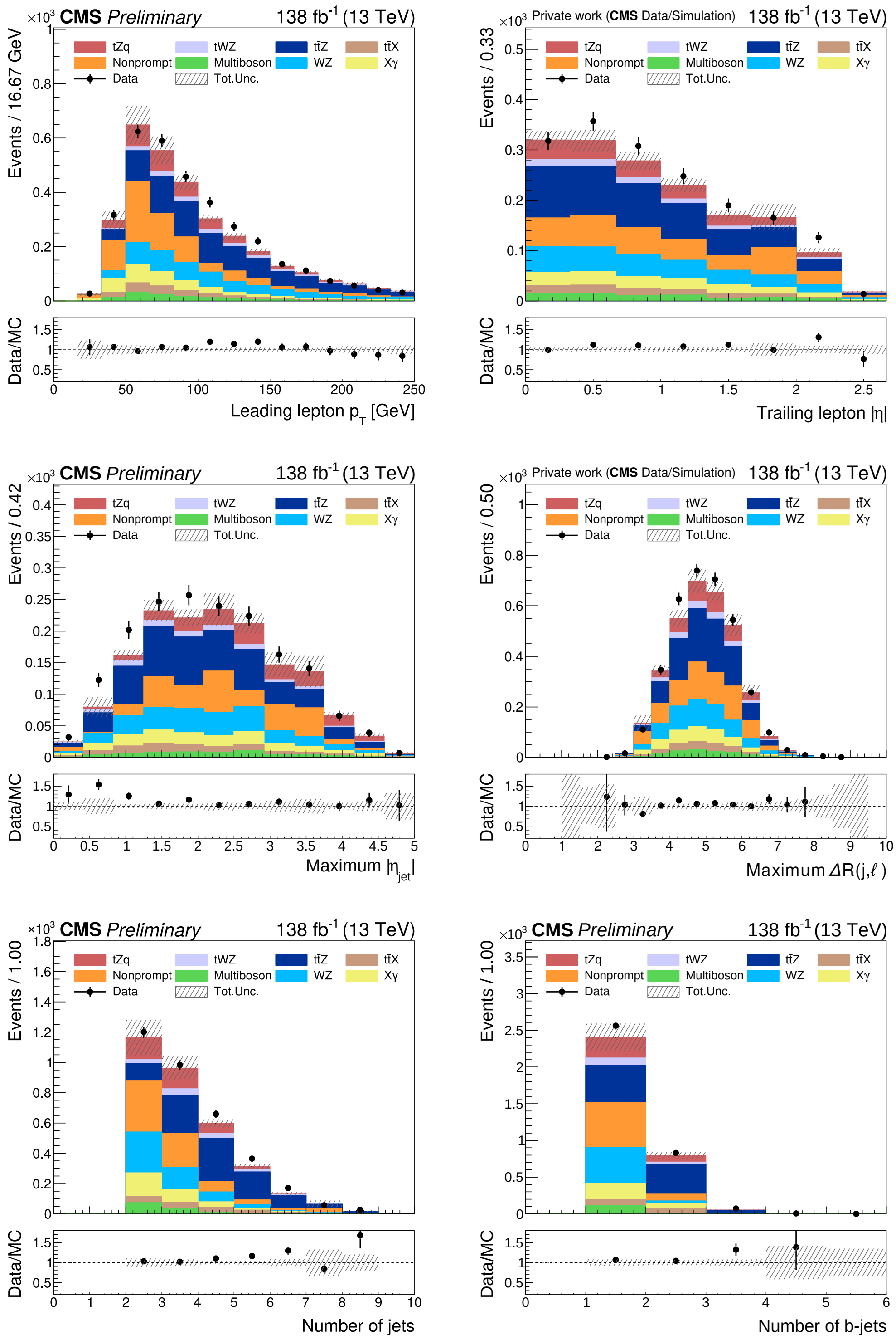


Figure 3.3: Distributions after final event selection for: the p_{T} of the leading lepton (upper left), the η of the trailing lepton (upper right), the η of the jet with maximal η (middle left), the maximum ΔR between a jet and a lepton (middle right), the number of jets (lower left), and the number of b jets (lower right).

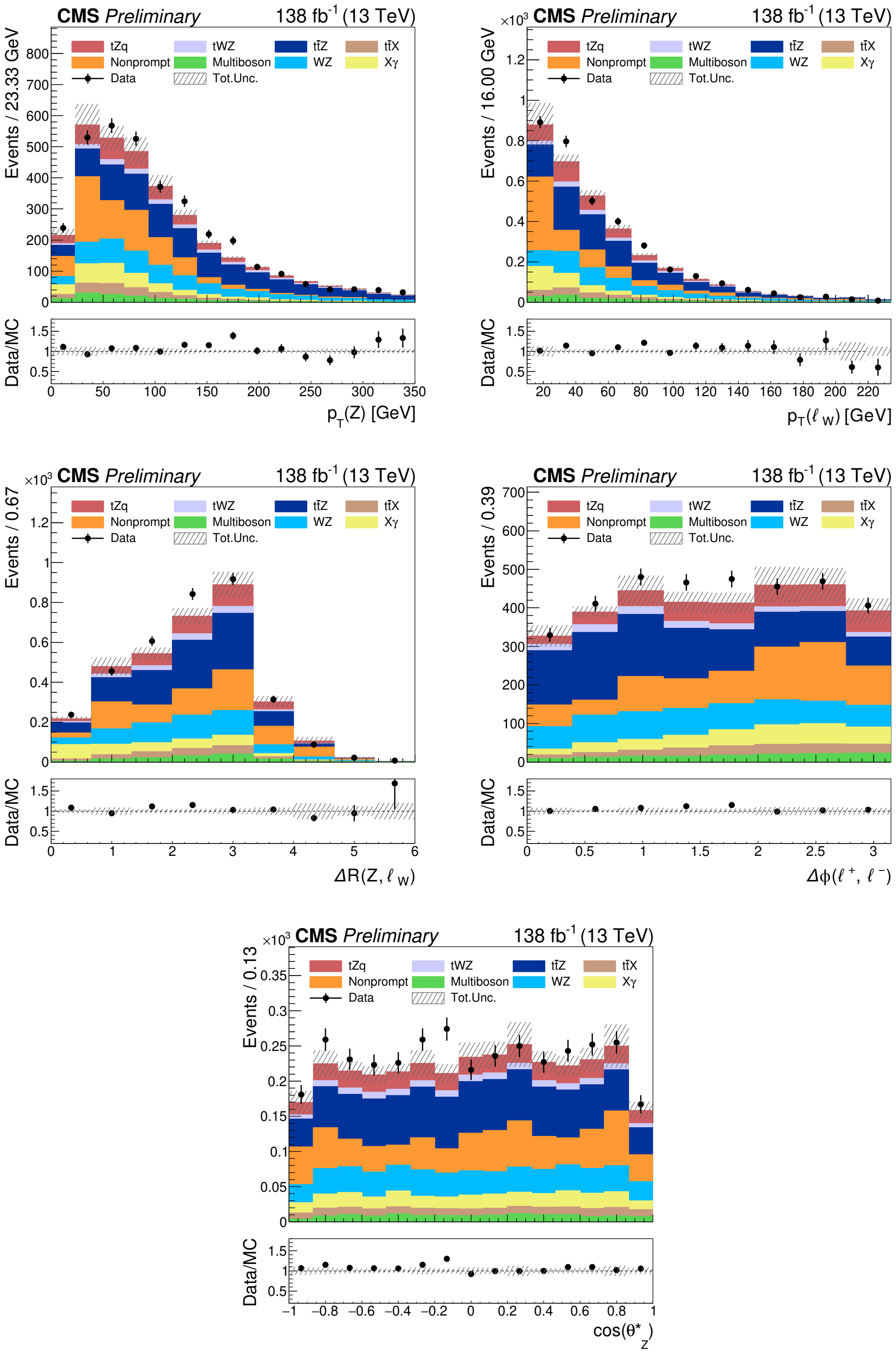


Figure 3.4: Distributions after final event selection for: the p_{T} of the Z boson (upper left), the the p_{T} of the lepton arising from the W boson ℓ_{W} (upper right), $\Delta R\,(\mathrm{Z}, \ell_{\mathrm{W}})$ (middle left), $\Delta\phi(\ell^{+}, \ell^{-})$ (middle right), and the cosine of the polar angle between the Z boson and its negatively charged decay lepton, boosted into the Z boson rest frame (bottom).

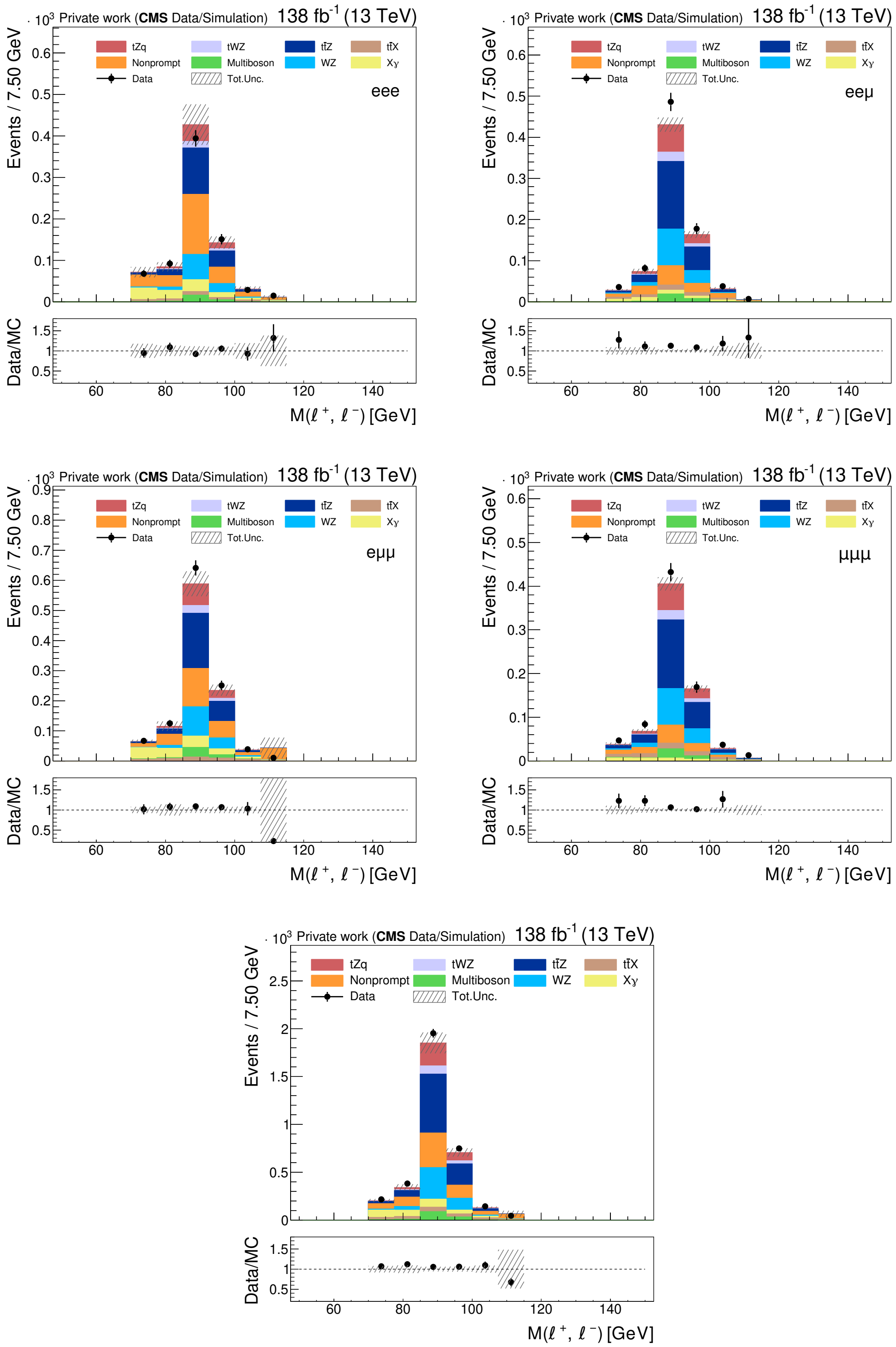


Figure 3.5: Distributions after final event selection for the invariant mass of the two leptons arising from the decay of the Z boson in the different lepton channels: *eee* (top left), *ee*μ (top right), *e*$\mu\mu$ (middle left), $\mu\mu\mu$ (middle right) and all the channels combined (bottom).

where E_{ℓ_W} is the total energy of ℓ_W and

$$\Lambda = \frac{m_\mathrm{W}^2}{2} + p_{\mathrm{T},\ell_\mathrm{W}} \cdot p_{\mathrm{T},\nu} \cdot \cos\Delta\phi(\ell_\mathrm{W},\, \nu)\,. \tag{3.6}$$

When m_T^W is larger than the W boson mass m_W, the term under the square root in Equation 3.5 becomes negative and no real solution exists. In this case, it can be assumed that the measurement of $\mathrm{E}_\mathrm{T}^\mathrm{miss}$ is incorrect due to resolution effects and m_T^W is set to m_W, such that the square root becomes zero and only one real solution exists. If two real solutions for Equation 3.5 exist, then the choice is made based on the top quark reconstruction, which is the next topic of this section. This procedure was developed in previous top quark analyses [83, 84].

3.3.2 Top quark reconstruction

Events coming from the tZq process only have one leptonic top quark, while the top quarks from $\mathrm{t\bar{t}Z}$ and tWZ can decay either hadronically or leptonically. The top quark reconstruction is based on a χ^2 measure to assess the likelihood of correctly assigning the jets to the top quark candidates. Three different cases are considered, depending on the jet mulitplicity of the event:

- two jets and one b jet: in this case only the leptonic top quark is reconstructed and there is no ambiguity.

- Three jets, at least one of which is b-tagged: this time either the hadronic or the leptonic top quark can be reconstructed, but not both simultaneously. Two different χ^2 are evaluated in this case, one for the leptonically decaying top quark

$$\chi^2 = \left(\frac{|m_\mathrm{t} - m_{\ell\nu\mathrm{b}}|}{\sigma_\mathrm{t,lep}}\right)^2 \tag{3.7}$$

  and one for the hadronic top quark

$$\chi^2 = \left(\frac{|m_\mathrm{t} - m_\mathrm{jjb}|}{\sigma_\mathrm{t,had}}\right)^2 , \tag{3.8}$$

  where m_t is the mass of the top quark [9] and $m_{\ell\nu\mathrm{b}}$ (m_jjb) is the invariant mass of the ℓ-ν-b (j-j-b) system; $\sigma_{\mathrm{t,lep}}$ ($\sigma_{\mathrm{t,had}}$) is the resolution of the leptonic (hadronic) top quark. This value was determined from a fit to the distribution of the top quark invariant mass evaluated using the generated particles from MC samples. One global value, 23.08 (31.37), is used in the full data set for the leptonic (hadronic) top quark hypothesis. The top quark candidate with the best χ^2 is reconstructed.

- Four or more jets: in this case, both the leptonic and the hadronic top quarks can be reconstructed simultaneously. Therefore, the χ^2 is given by

  $$\chi^2 = \left(\frac{|m_{\mathrm{t}} - m_{\ell\nu\mathrm{b}}|}{\sigma_{\mathrm{t,lep}}}\right)^2 + \left(\frac{|m_{\mathrm{t}} - m_{\mathrm{jjb}}|}{\sigma_{\mathrm{t,had}}}\right)^2 \tag{3.9}$$

  and in case of ambiguity, the combination of jets that gives the best m_{t} value is kept.

In the cases where one of the two top quarks cannot be reconstructed, either because it was not produced in the event or because there are not enough jets, an unphysical value of -1 is given to the p_{T} and mass of the unreconstructed top quark, while its pseudorapidity is set to zero.

The information of the top quark kinematics as obtained from the reconstruction described above is used to train and evaluate a multiclass neural network which is used to separate $\mathrm{t\bar{t}Z}$ +tWZ, tZq and backgrounds, as described in Section 4. The top quark reconstruction efficiencies are reported in Table 3.4. The values were obtained by taking into account the angular distance ΔR between the generated and the reconstructed top quarks in the simulation of the signal processes: a top quark is considered to be correctly reconstructed if the ΔR between the two particles is smaller than 0.6.

Table 3.4: Top quark reconstruction efficiencies for the signal processes. The top quark is considered correctly reconstructed if the ΔR between the generated and the reconstructed particle is smaller than 0.6. The numbers also include acceptance effects, i.e. only the top quarks reconstructed within the analysis phase space are considered.

Top decay	$\mathrm{t\bar{t}Z}$	tWZ	tZq
hadronic	32%	45%	-
leptonic	44%	48%	45%

3.4 Nonprompt background estimation

The nonprompt lepton contribution in the SR mainly arises from $\mathrm{t\bar{t}}$ and DY events, and it is estimated with a data-driven method known as the *fake factor* (FF) method. First, a region enriched with nonprompt leptons is created by using the following selection criteria on simulated QCD multijet samples:

- exactly one fakeable lepton (as defined in Section 3.2.1),
- at least one jet, with $\Delta R(\ell, \mathrm{j}) > 0.7$;

this is referred to as the measurement region (MR) and it is used in order to evaluate the per-lepton FF f_i, defined as the ratio between the number of leptons passing the tight selection and that passing the fakeable selection:

$$f_i = \frac{N_{\text{tight}}}{N_{\text{tight}} + N_{\text{fakeable}}} . \tag{3.10}$$

The per-lepton FFs are parametrized as a function of the lepton cone-p_{T} and η; the former is used instead of the lepton p_{T} in order to reduce the dependence on the object that fakes the lepton, as this usually originates from a parton. It is defined as

$$\text{cone}-p_{\mathrm{T}} = \begin{cases} p_{\mathrm{T}}^{\ell} & \text{if lepton is tight} \\ x \cdot p_{\mathrm{T}}^{\text{Closest jet}} & \text{otherwise, if jet within } \Delta R < 0.4 \\ x \cdot p_{\mathrm{T}}^{\text{lepton}}(1 + I_{\text{PF}}) & \text{otherwise, if no jet within } \Delta R < 0.4 \end{cases} \tag{3.11}$$

where $p_{\mathrm{T}}^{\text{Closest jet}}$ is the p_{T} of the closest jet within $\Delta R < 0.4$ to the lepton and I_{PF} is the PF isolation with $\Delta R = 0.4$. The value of x is set to 70%, and it was chosen to ensure the continuity of the lepton cone$-p_{\mathrm{T}}$ as a function of the TOP LEPTON-MVA output. The per-lepton FFs employed for this thesis have been evaluated in the context of the first tWZ measurement, which led to the evidence of the process [3].

Subsequently, a second region is defined, called application region (AR). This is done with the same selection criteria as is the SR, but requiring fakeable leptons instead of tight leptons. The used weight in the AR is then

$$(-1)^{n-1} \prod_{i=1}^{3} \frac{f_i}{1 - f_i} \tag{3.12}$$

where n is the number of fakeable leptons that fail the tight selection and the term $(-1)^{n-1}$ is used to avoid double counting. Finally, the nonprompt lepton contribution in the SR is estimated by taking data in the AR and subtracting events with only prompt leptons.

In order to check whether the nonprompt lepton contribution is estimated correctly, the region outside the Z boson peak, i.e. in the range $\mid m_{\ell^+\ell^-} - m_{\mathrm{Z}} \mid > 20$ GeV, is considered. This region generally shows a good data-MC agreement and some distributions are shown in Figures 3.6 and 3.7. It can also be concluded that the nonprompt background estimation is affected by some statistical fluctuations, which come from the requirements applied to fakeable leptons. However, this implies the availability of more statistics for the tight leptons, which translates into a higher precision on the cross section measurements.

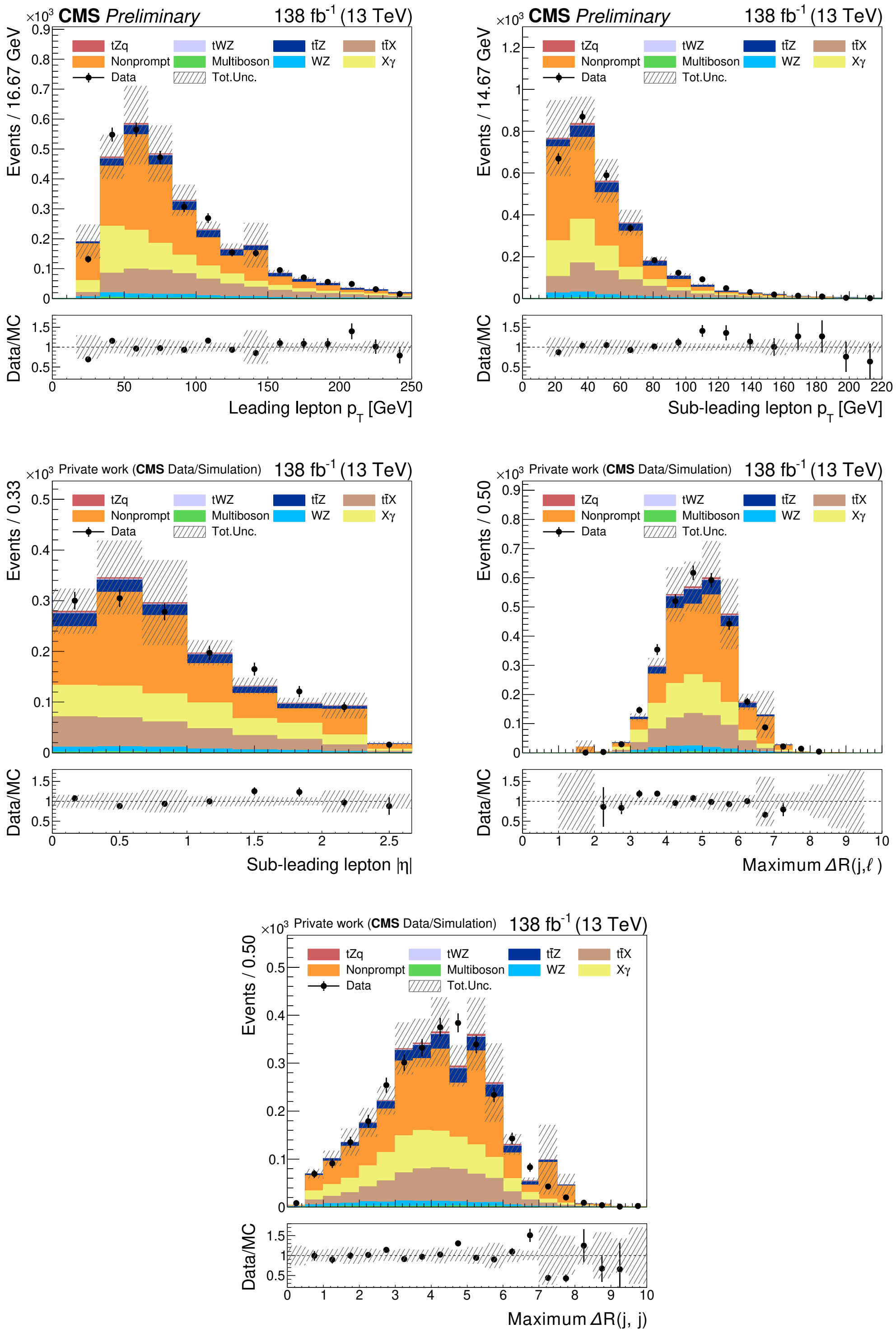


Figure 3.6: Distributions outside the Z peak for: the p_{T} of the leading lepton (top left), the p_{T} and η of the sub-leading lepton (top right and middle left, respectively), the maximum ΔR between a jet and a lepton (middle right) and between any jet pair (bottom).

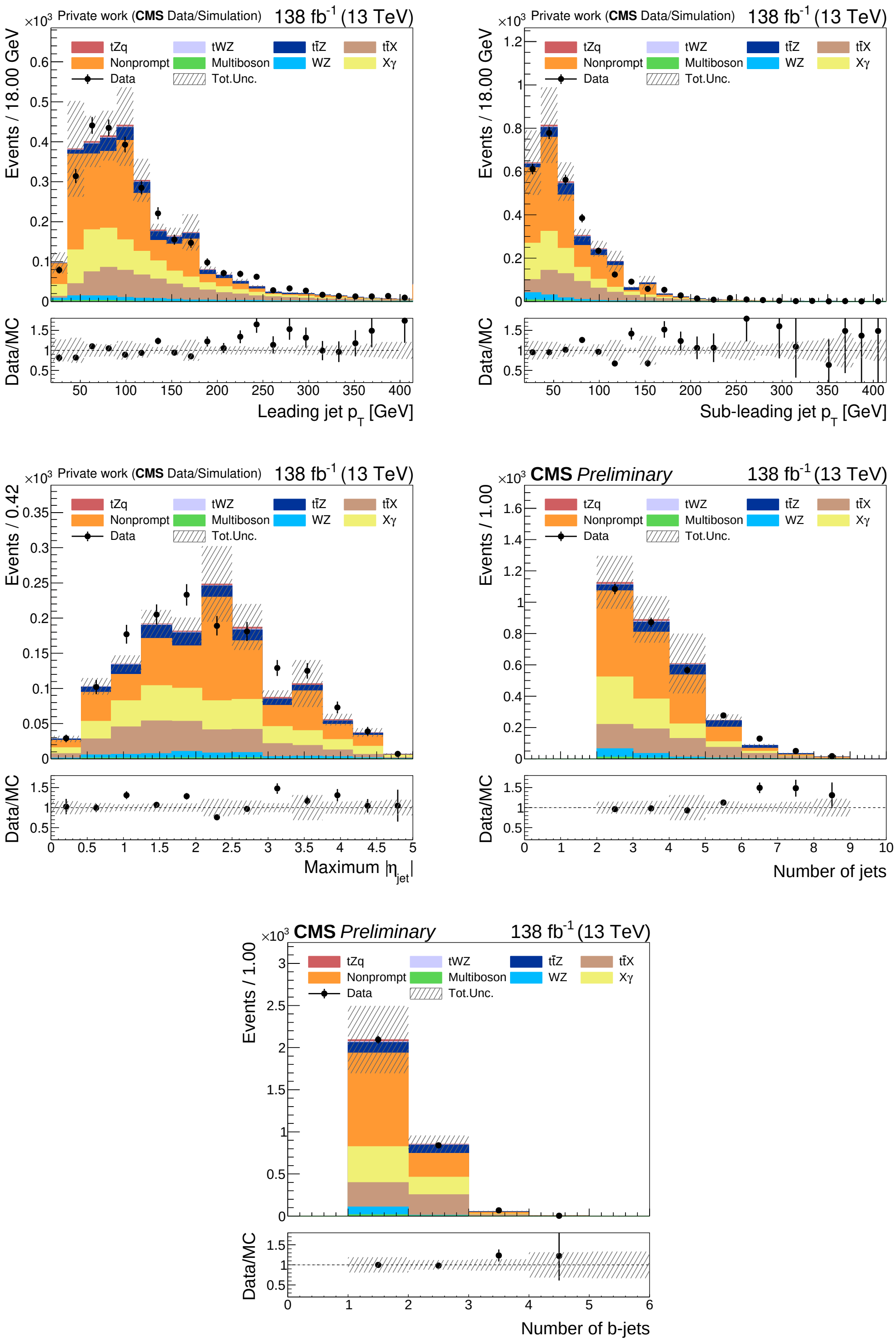


Figure 3.7: Distributions outside the Z peak for: the p_T of the leading and subleading jets (top), the η of the jet with maximal η (middle left), the number of jets (middle right), and the number of b jets (bottom).

CHAPTER 4

Classification of signals and background

This chapter focuses on the methods employed to optimize the discrimination among backgrounds and signals in the analysis. As it is hardly possible to isolate the different processes with a cut-based approach, a multiclass neural network has been trained with the objective to achieve optimal separation between the two signal categories and the background. The model has been designed with KERAS [85] interfaced with TENSORFLOW V2 [86].

4.1 Multiclass neural networks

Multiclass neural networks are mathematical models that consist of interconnected nodes, organized into layers. The initial data are received in the *input* layer of the network and processed through the intermediate (*hidden*) layers using weighted connections and mathematical operations; the final predictions from the different classes are produced and stored in the *output* layer of the network. A multiclass classifier with multiple hidden layers is usually referred to as deep neural network (DNN); such a structure is represented in Figure 4.1. In each of the hidden layers, the input values x_i from N neurons in the preceding layer are multiplied by a set of weights w_i and a bias factor is added. A so-called *activation function*, f, is then

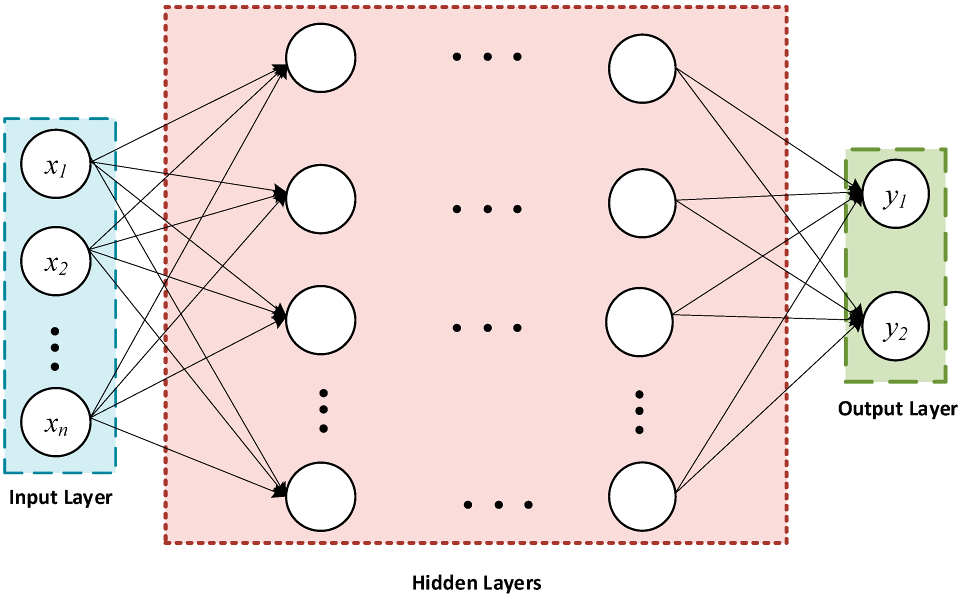


Figure 4.1: Structure of a multiclass neural network. The input layer takes the initial data which are then processed in the hidden layers. The final scores for the different classes are stored in the nodes contained in the output layer. Taken from Ref. [87].

applied before propagating the values to the next layer, and the output value is then

$$y = f\left(\sum_{i}^{N} x_i w_i + b\right) . \tag{4.1}$$

The role of the activation function will be clarified later in this chapter. During the training process, the weights of the network are determined. This is done by providing labeled data to the model which are used to adjust the weights and make the final predictions as accurate as possible. For the algorithm employed in this thesis, the learning process is performed with the Gradient Descent (GD) rule [88], which aims to find a good local minimum of the so-called *loss function*, F: the strategy is based on the observation that if F is differentiable in a neighborhood of a point u, then it decreases fastest when moving from u in the direction of $-\nabla F(u)$:

$$u_{n+1} = u_n - \lambda \nabla F(u_n) \tag{4.2}$$

where λ is the learning rate, a parameter used to control the size of each step towards the minimization of F. The search of the minimum is accelerated by adding a fraction ξ of u_n to the value u_{n+1}:

$$u_{n+1} = u_n + \xi \Delta u_n - \lambda \nabla F(u_n)\,. \tag{4.3}$$

Additionally, the value of the learning rate is updated during the training procedure depending on the mean and the variance of the gradients; in this way, the minimum is identified with higher precision. The optimizer employed in this measurement is a method called Adaptive Moment Estimation ADAM [89] and it includes all the features described so far.

If the gradient becomes really small, the weight might be prevented from changing its value and in the worst case, this can also interrupt the training process. This is known as the *vanishing gradient* problem [90] and it can be prevented by applying an activation function to each of the layers, when computing the output of the node. Two different functions are applied:

- the Scaled Exponential Linear Unit (SELU) function

  $$s(z) = \begin{cases} C \cdot z, \ \text{if } z \geq 0 \\ C \cdot \alpha(e^z - 1), \ \text{if } z < 0 \end{cases} \tag{4.4}$$

  where $C \simeq 1.05$ and $\alpha \simeq 1.67$ [91]. This function introduces self-normalizing properties in the network and generally improves the stability of the model.

- The softmax function $\sigma(z)$, which normalizes the outputs of the network such that their total sum is equal to unity, giving the probability of the input value being in a specific class j; it is defined via the relation:

  $$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{4.5}$$

  where the index *k=1, .., K* refers to the output classes. Since the purpose of this function is to give the probability of a given class to occur, it is applied only on the output layer, while SELU is used in the hidden layers.

The loss function is the multiclass cross-entropy loss, defined as

$$F(\mathbf{q}_i, \mathbf{p}_i) = -\sum_{k=1}^{K} q_{ik} \cdot \log(p_{ik}) \tag{4.6}$$

where $\mathbf{q}_i$ and $\mathbf{p}_i$ are two vectors $(q_{i1}, .., q_{iK})$ and $(p_{i1}, .., p_{iK})$, respectively. In the former, the components are equal to one if the i-th event belongs to the class k in the training data set, and zero otherwise. The latter contains the probabilities of the i-th event being in the class k, evaluated with the softmax function.

The L2 Regularization [92] is applied in each layer in order to avoid overtraining, i.e. when the model starts learning noise or random fluctuations from the input data rather than generalizing the underlying pattern; with this regularization, the complexity of the model is reduced by penalizing high-weight parameters. The parameters used to regulate the learning process and define the architecture of the model, usually referred to as *hyperparameters*, are not learned from the input data but their values have to be set before the training. The hyperparameter optimization (or tuning) is defined as the problem of choosing a set of optimal hyperparameters for the algorithm and it will be discussed in the next Section 4.2.

4.2 Input variables and hyperparameters

The list of the observables used as input data for the training of the DNN is reported in Table 4.1. In addition to kinematic properties of the physics objects, variables coming from the top quark reconstruction are also employed for the training. Some of the corresponding distributions with both simulations and data have been already shown in Figure 3.3, and the remaining ones can be found in Figures 4.2 to 4.5. By looking at the distributions coming from the top quark reconstruction, it can be seen that a good data-to-MC agreement is reached in the cases where the top quark decays leptonically, while a slight excess is observed for the hadronically decaying top. Nevertheless, the overall description of the observables remains good and suitable for the training process.

Table 4.1: List of input variables for the DNN.

Variable
Physics object properties
η of the leading lepton
η of the sub-leading lepton
p_{T} and η of the leading jet
p_{T} and η of the sub-leading jet
Missing transverse energy (MET)
Maximum jet η
Maximum invariant mass of any jet pair
Scalar sum of hadronic p_{T}
Maximum ΔR between leptons
Maximum ΔR between any jet-lepton pair
Maximum ΔR between jets
Charge of the particles in the final state
Transverse mass of the W boson
Number of jets
Number of b jets
Top quark reconstruction
Mass, p_{T} and η of the reconstructed leptonic top quark
Mass, p_{T} and η of the reconstructed hadronic top quark
χ^2 distribution from the top reconstruction

The output layer consists of three output nodes: two for the signals, tZq and the sum of $\mathrm{t\bar{t}Z}$ and tWZ, and one for the backgrounds. One single training is done for the three years of data taking. The simulation samples are divided into two complementary parts, based on even and odd event numbers. In each of the two, 80% of the events are used for training, the remaining 20% are used for testing. The other sample is used to evaluate the model in order to build the output score distributions that are used in the final stage of the analysis; this procedure is commonly used to further decrease the possibility of overtraining. The *dropout*, i.e. the procedure of randomly setting to zero a certain percentage of neurons in each training iteration,

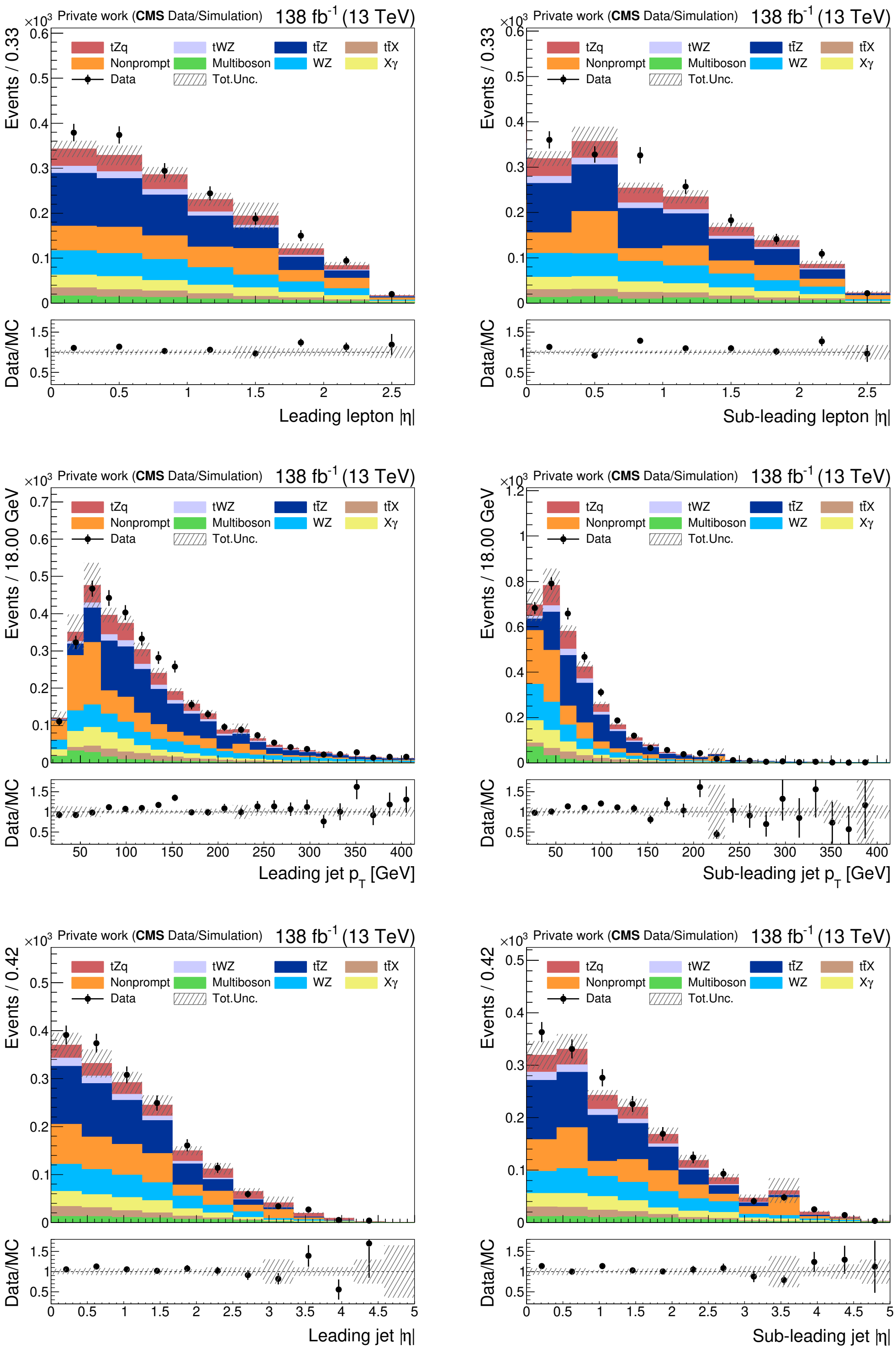


Figure 4.2: Event distributions for the η of the leading and sub-leading leptons (top left and top right, respectively), and the p_{T} (middle row) and η (bottom row) of the leading (left) and sub-leading (right) jets.

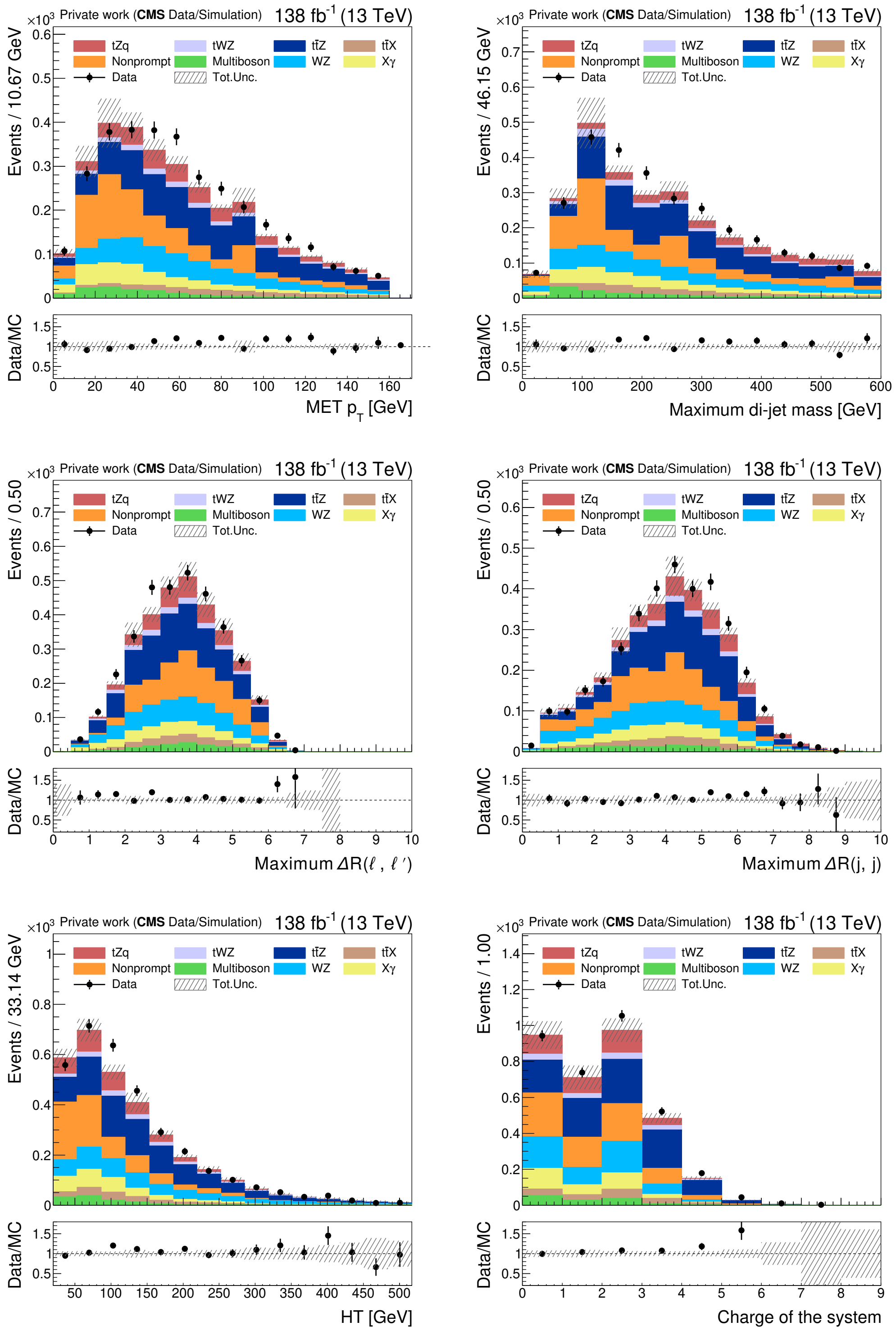


Figure 4.3: Event distributions for the MET (top left), the maximum invariant mass of any jet pair (top right), the maximum ΔR between any lepton (middle left) and jet (middle right) pair, the scalar sum of the p_{T} of all the jets, HT (bottom left) and the sum of all the charge of all the particles in the final state (bottom right).

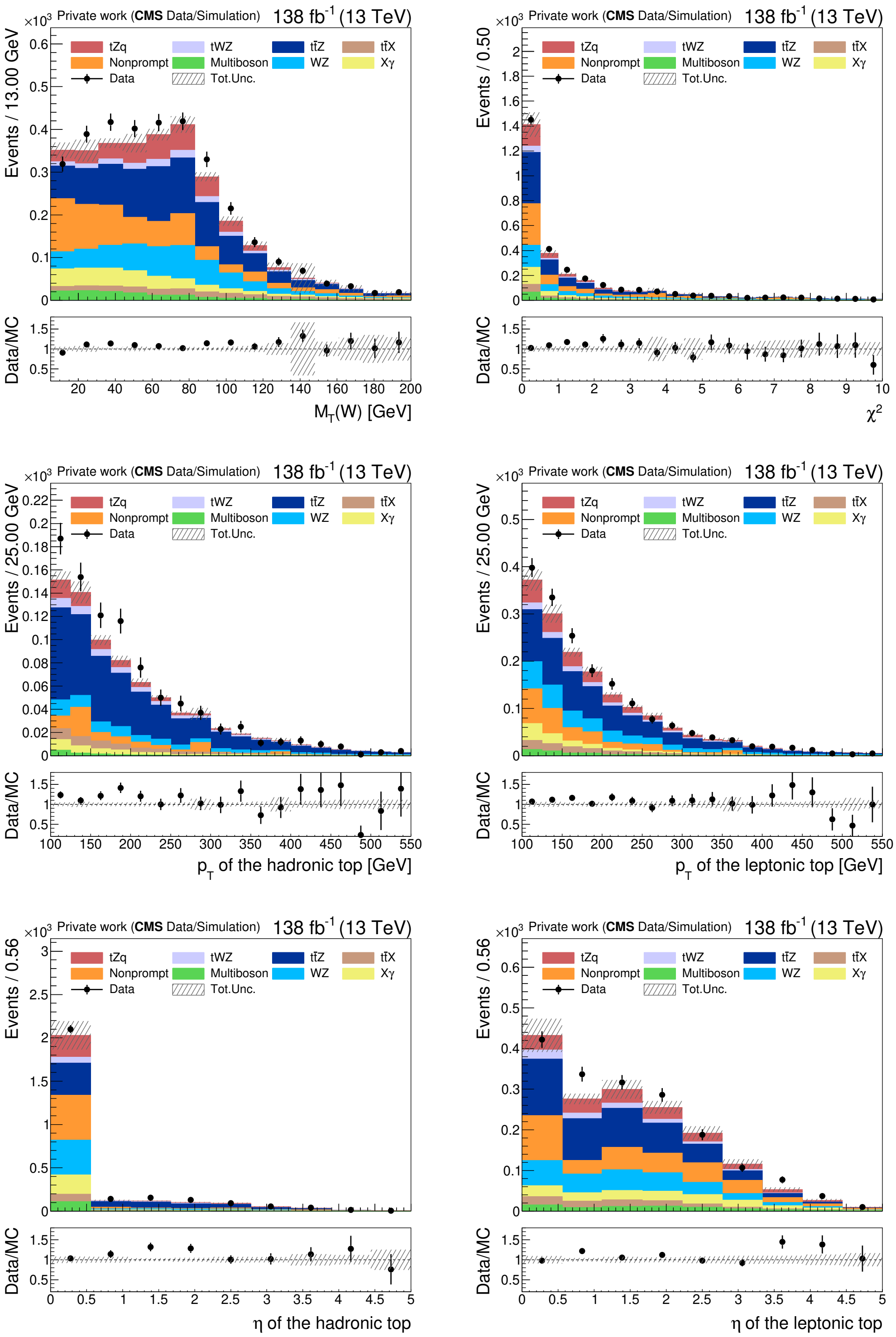


Figure 4.4: Event distributions for the transverse mass of the W boson (top left), the χ^2 distribution from the top quark reconstruction (top right), and the p_{T} (middle row) and η (bottom row) of the leptonically (left) and hadronically (right) decaying top quark.

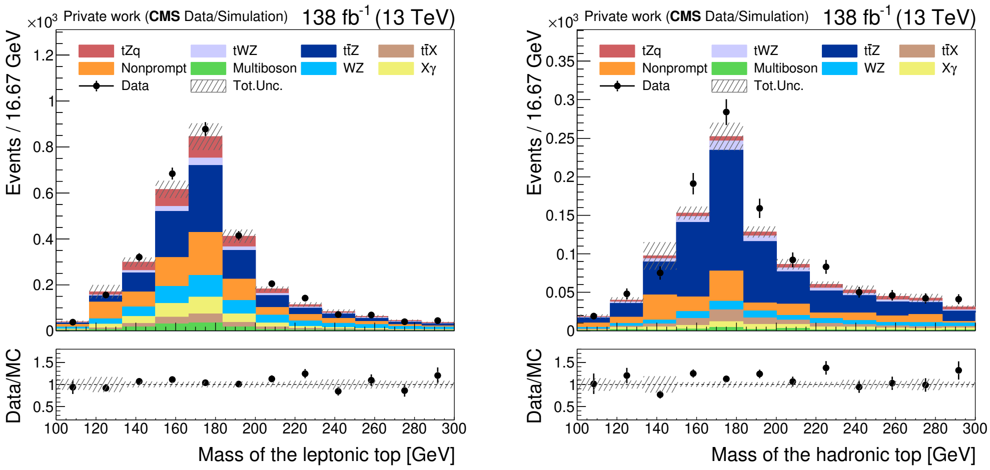


Figure 4.5: Event distributions for the mass of the leptonically (left) and hadronically (right) decaying top quark.

is also applied for the same purpose. The total number of simulated events used to train the model is reported in Table 4.2; the count for background events is lower in comparison to the signals, resulting in class imbalance.

Table 4.2: *Total number of events used for the training of the DNN.*

Process	Events
tZq	918221
$t\bar{t}$Z +tWZ	3349811
Backgrounds	358028

The relative weights of the samples are optimized in the process of hyperparameter tuning in order to obtain an optimal discrimination between the two signal categories together with a good separation from the backgrounds. The hyperparameters are optimized by randomly sampling their values within a sensible range and in the end, the model that gives the best discrimination between the signals is selected. The final choices for the hyperparameters are listed in Table 4.3.

The distributions for the three output nodes are shown in Figures 4.6 and 4.7, together with results of the Kolmogorov-Smirnov (KS) test between histograms obtained from training and testing. These two figures are also showing the loss function

Table 4.3: List of the hyperparameters considered for the tuning, together with their optimized values and the range considered for the tuning.

Hyperprameter	Value	Range of tuning
tZq class weight	200	[5, 500]
$t\bar{t}Z$ + tWZ class weight	109	[5, 500]
Backgrounds class weight	213	[5, 500]
Number of hidden layers	2	[2, 4]
Number of neurons (first layer)	41	[20, 150]
Number of neurons (second layer)	129	[20, 150]
Batch size[1]	2100	[50, 5000]
Dropout	0.3	[0.05, 0.7]
Learning rate	$1 \cdot 10^{-4}$	$[1 \cdot 10^{-2}, 1 \cdot 10^{-7}]$

[1] Number of training examples used in each iteration.

and the ROC curve of the models, which are referring to two different trainings performed on events with even (Figure 4.6) and odd (Figure 4.7) indices. Both trainings show a good performance without signs of overtraining. Even though the hyperparameters are the same, some small differences among the two cases can be noticed; these can be attributed to the random initialization of the weights at the beginning of each training and to the statistical fluctuations affecting the learning process.

The output node distributions with both simulation and data can be found in Figure 4.8. This also shows the case where only events for which the score is the maximum among all the nodes are included in each distribution: this is done in order to make these three categories orthogonal, which will later be needed to correctly perform the measurement of the signal cross sections. Overall, the distributions are showing a good agreement between data and predictions; a small excess is found in the bins enriched with $t\bar{t}Z$, and therefore it can be expected that the measured cross section will be somewhat higher than the prediction, as already observed in previous experimental results [1, 2].

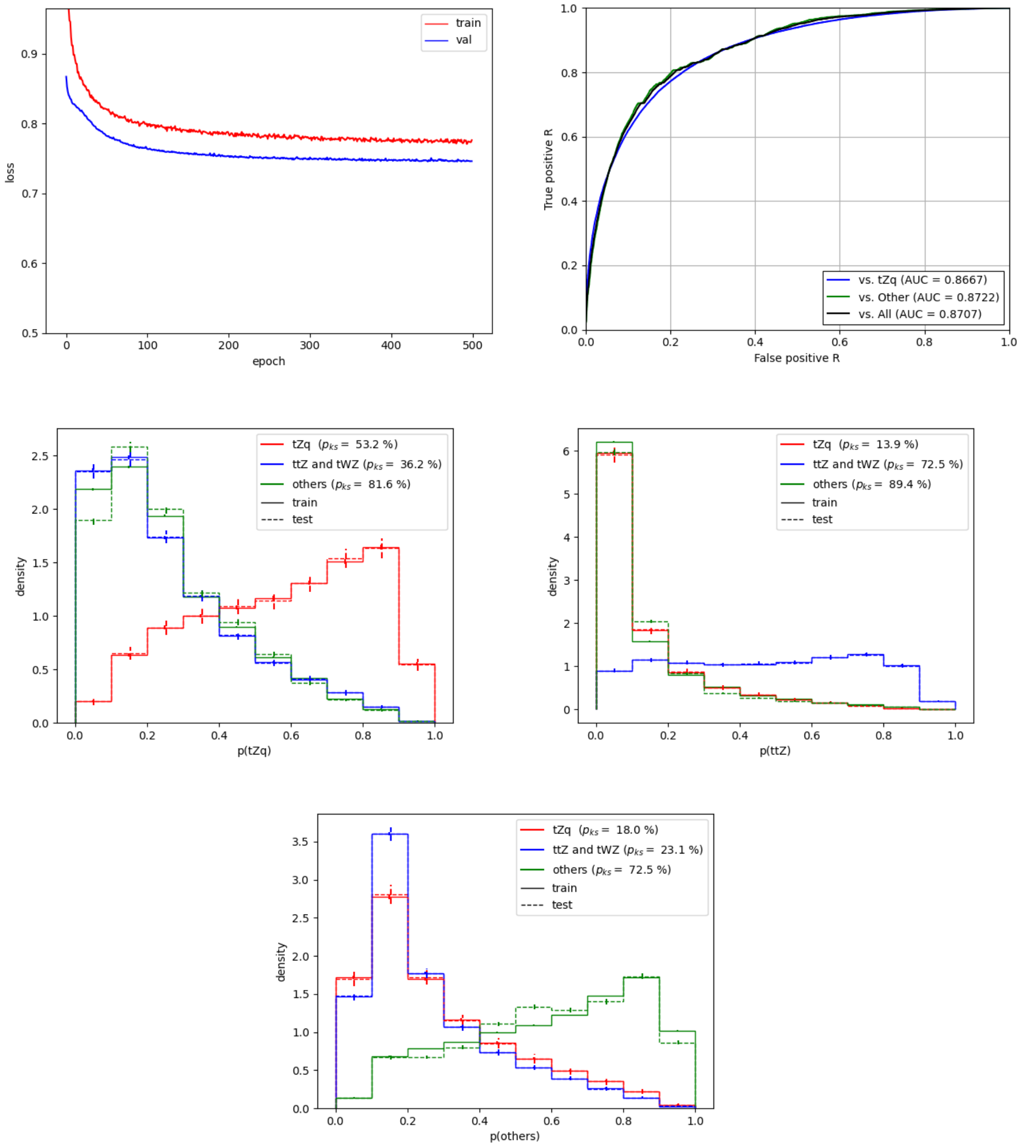


Figure 4.6: Loss function (upper left) and ROC curves (upper right) from the training performed on events with even indexes and distributions of the tZq (lower left), $t\bar{t}Z$ +tWZ (lower center) and background (lower right) output nodes for this half of the training, together with the KS test between histograms obtained from training and testing. The ROC curves show the discrimination of $t\bar{t}Z$ +tWZ against tZq (blue), the backgrounds (green) and all the nodes combined (black).

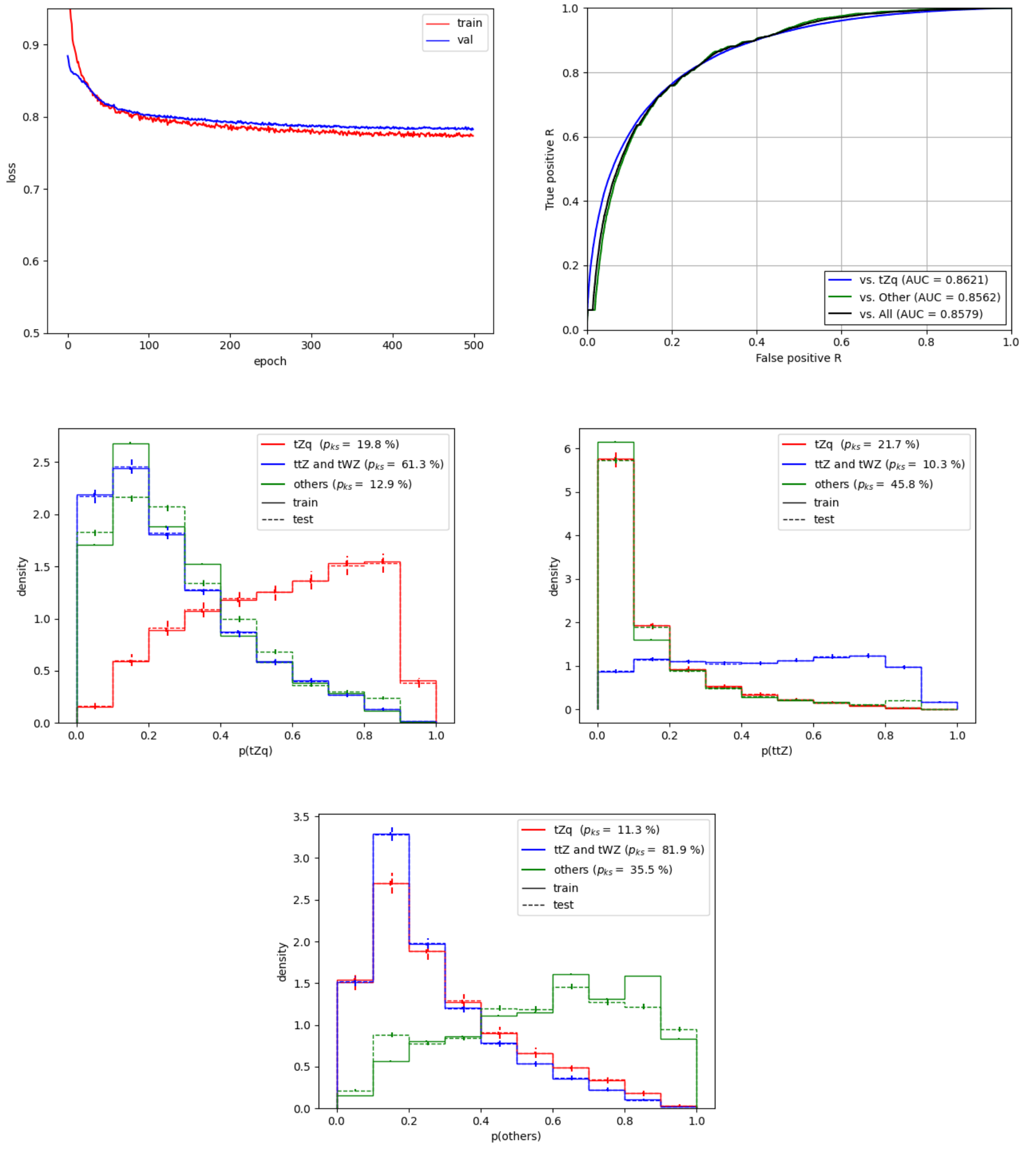


Figure 4.7: Loss function (upper left) and ROC curves (upper right) from the training performed on events with odd indexes and distributions of the tZq (lower left), $t\bar{t}Z$ +tWZ (lower center) and background (lower right) output nodes for this half of the training, together with the KS test between histograms obtained from training and testing. The ROC curves show the discrimination of $t\bar{t}Z$ +tWZ against tZq (blue), the backgrounds (green) and all the nodes combined (black).

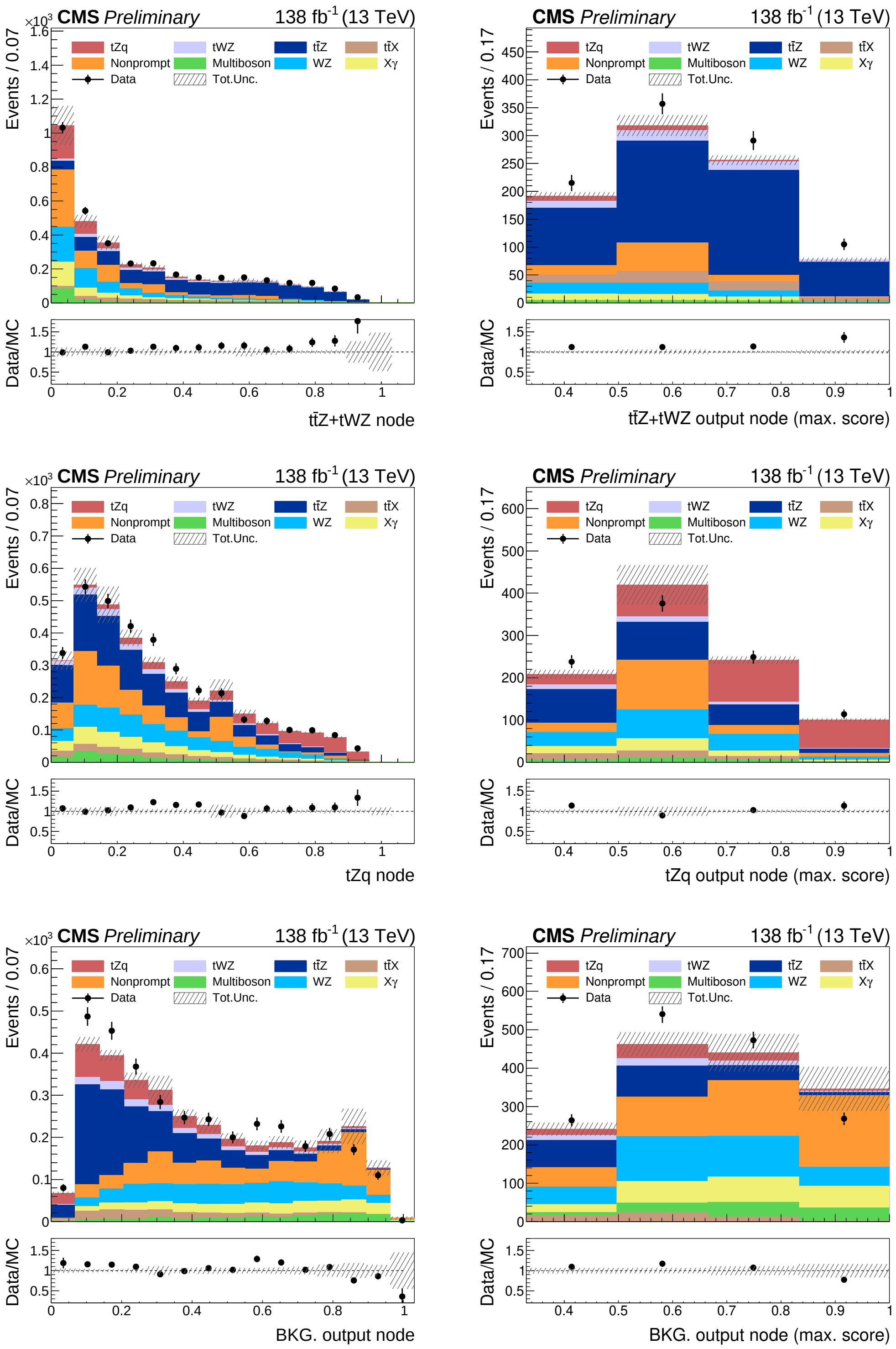


Figure 4.8: Distributions of output values in the three DNN output nodes for the combination of $t\bar{t}Z$ and tWZ (upper), tZq (middle), and background (lower). In the right column, only events for which the output score is larger than that for the other two categories are included in each distribution.

CHAPTER 5

Signal extraction

The main results of this work are presented in this chapter. The signal cross sections are measured both inclusively and differentially in observables that are particularly sensitive to the modeling of the processes and to new physics scenarios. As mentioned in Section 1.8, previous measurements performed by the CMS and ATLAS Collaborations extracted the tZq, $\mathrm{t\bar{t}Z}$ and tWZ cross sections separately, using SM predictions to estimate the contributions from these channels when they are considered backgrounds. Here the processes are simultaneously measured as signals for the first time, improving the sensitivity to possible deviations from the SM and providing a consistent treatment of the uncertainties and the correlations between the signal processes, which is useful as input to global fits.

The signal cross sections are extracted simultaneously with a binned maximum likelihood fit, taking both statistical and systematic uncertainties into account. In the profile of the likelihood ratio, a scan of the signal strength r is done in the numerator and the optimal nuisance parameters $\hat{\theta}_r$ are determined for each of the values, while in the denominator the global minimum for r and $\hat{\theta}$ is used:

$$-2\ln\frac{\mathcal{L}(\mathrm{data}|r,\hat{\theta}_r)}{\mathcal{L}(\mathrm{data}|r=\hat{r},\hat{\theta})}\,. \tag{5.1}$$

The fit is performed simultaneously on the three output nodes of the DNN, for all the years. In order to have orthogonal categories, each event is assigned to the category for which its output node is maximal.

5.1 Systematics

Since the performance in data is only known to a certain degree, and simulations are corrected to match the data, systematic uncertainties in the applied calibrations also need to be considered. Systematic variations are described for every source of systematic uncertainty. A distinction is made between normalization and shape uncertainties. The former expresses possible systematic deviations in normalization, while shape uncertainties are more complex and taken into account by vertical interpolation of the histograms: for each shape uncertainty and for each category affected by it, two additional input histograms are provided. These are obtained by shifting the relevant parameter of uncertainty up and down by one standard deviation. In a process called *template morphing*, the three histograms, i.e. the nominal and the two variations are interpolated. The nuisance parameter then quantifies the deviation from the nominal distribution that provides an optimal fit result.

Experimental uncertainties

The experimental uncertainties can affect both the event selection efficiency and the shape of the distributions. They are applied to all the processes considered for this measurement, with the same treatment between signals and backgrounds.

- **b-tagging**: b-tag SFs are evaluated for jets depending on their flavour and one common multiplicative factor is then applied to the nominal prediction. Separate uncertainties for heavy and light flavour jets are centrally computed within the CMS Collaboration, both sets divided into a part that is treated as uncorrelated across the years and another one treated as correlated.

- **JECs**: as mentioned in Section 3.2.2, JECs are applied to the jets in the MC simulation. The corresponding uncertainties are treated as separate nuisance parameters in the fit. For each of them, the four-momenta of the jets are varied within the uncertainties and the events are then re-analyzed. The b-tagging efficiency SF and the missing transverse energy (MET) are recomputed as well.

- **JER**: uncertainties related to the smearing procedure applied to the JER in the simulated data (see Section 3.2.2) are centrally computed within the CMS experiment and these nuisances are treated as fully uncorrelated across the years.

- **MET**: the p_{T} and the ϕ of the MET vector are recomputed for each of the different JER and JEC sources. Jets are considered only if their electromagnetic energy fraction is less than 0.9; furthermore, muons are removed from the jets. An uncertainty on the unclustered missing energy is considered and treated as uncorrelated between the years.

- **Lepton ID and isolation**: uncertainties from the evaluation of lepton ID and isolation SFs are divided into statistical and systematic sources and they are treated separately for electrons and muons. A single nuisance is considered for both ID and isolation components and the systematic (statistical) uncertainties are treated as correlated (uncorrelated) across the years.

- **Nonprompt background**: three different sources are included in the uncertainty estimate for the nonprompt background. First, the statistical uncertainties coming from the estimation of the per-lepton FFs (σ_{FR}) are propagated into the final templates of the SR; subsequently, the statistical uncertainty from the AR (σ_{AR}) is taken into account by considering the statistical error of the bin before the application of the fake rates; finally, an additional flat bin-by-bin uncertainty of 30% is applied to cover for any residual mismodeling of the nonprompt background. The three contributions are summed in quadrature and set as bin-by-bin uncertainty in the final templates used in the fit.

- **Normalization of backgrounds**: the background processes with similar event topologies are grouped together, following the same distinction as in Table 3.3. For the $\mathrm{t\bar{t}X}$ background, a normalization uncertainty of 11% is assigned; this value comes from the uncertainties on the theory predictions of the processes involved. An uncertainty of 15% is applied on the normalization of the other backgrounds with the exception of $\mathrm{V}\gamma$, for which the value is increased to 40% due to modeling issues. A different procedure is applied to estimate the nonprompt uncertainties, as will be discussed later in this section.

- **Pile-up**: before applying the event selection, event weights are corrected to account for the differences in the distribution of pile-up between data and MC simulation. To estimate the corresponding uncertainty, the pp inelastic cross section is varied by ±4.6%. This uncertainty is considered as fully correlated across the years.

- **L1 prefire**: the term *prefiring* refers to the case in which a trigger decision is assigned to a bunch crossing preceding the one in which the collision actually took place. It depends on the adjustment of the detector signals in time, and the stability of the adjustment over time. It usually occurs at the permille level, reaching the percent level in periods of bad timing adjustment. This leads to a loss of events, i.e. a trigger inefficiency, which was observed during the data taking in 2016 and 2017. A similar issue was found in the muon system, where the assignment of the muon candidates was sometimes wrong due to the limited time resolution of the muon detectors. This effect was most pronounced in 2016, but is non-zero for both 2017 and 2018. To take these effects into account in simulated events, SFs are applied as a function of p_{T} and η of the object candidates. The corresponding uncertainties are obtained by varying these corrections by ±20% and they are considered as correlated between the years.

- **Trigger SFs**: since the trigger efficiency is very close to unity, no trigger SFs are applied. A flat 2% uncertainty is applied in order to cover residual differences; this method was also used in previous CMS analyses using similar trigger paths and criteria for the event selection [1–3].

- **Luminosity**: a normalization uncertainty is assigned to the integrated luminosity as detailed in Table 5.1.

- **HEM1516 issue**: in late 2018, the power supply of two HCAL modules experienced an outage in the middle of the data taking, impacting the jet energy measurement in this region. In order to take this effect into account, the energy is scaled down by 20% for jets with $-1.57 < \phi < -0.87$ and $-2.5 < \eta < -1.3$, by 35% for jets with $-1.57 < \phi < -0.87$ and $-3.0 < \eta < -2.5$. The difference from the nominal calibration is then treated as a systematic uncertainty, which is also propagated to the MET.

Table 5.1: Normalization uncertainties assigned to the integrated luminosity in the different years.

	2016	2017	2018
Uncorrelated 2016	1.0	-	-
Uncorrelated 2017	-	2.0	-
Uncorrelated 2018	-	-	1.5
Correlated	0.6	0.9	2.0
Correlated 2017-2018	-	0.6	0.2

Most of the JEC templates from minor backgrounds (e.g. multiboson processes) suffer from large fluctuations due to the low statistics; for this reason, a smoothing procedure is applied to these variations before including them in the fit. The smoothing procedure is performed in three steps: first, the ratio between the variation and the nominal value is evaluated for each bin; a new smoothed histogram is then created by averaging the bin contents of the ratio with its two neighboring bins; finally, the final template is created by multiplying again each content of the smoothed histogram by the nominal value in that bin. A comparison between some of the shape variations before and after the smoothing is shown in Figure 5.1 for the multiboson background. The plots refer to the multiboson background, one of the processes with lower statistics in the analysis. Statistical fluctuations are removed with the smoothing procedure and the overall shape is preserved, as it can be concluded from the KS test results. Some bins have a negligible statistical uncertainty: this is artificially set to be very small when creating the templates in cases where the original value allows for negative bin contents, as this might create technical problems in the fit procedure.

Theory uncertainties

The uncertainties related to the theory modeling of the processes are listed below. These are normalized to the nominal cross section of the corresponding process

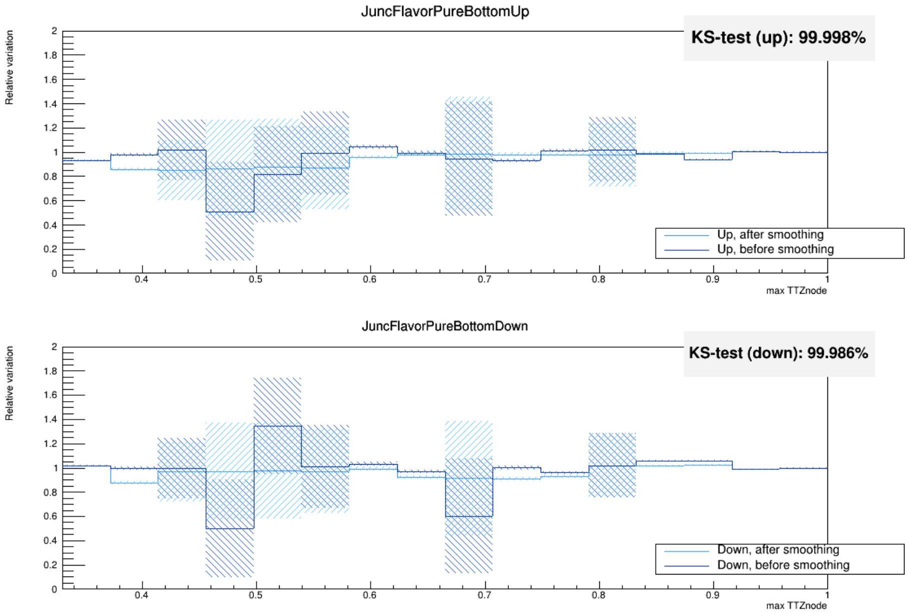


Figure 5.1: Comparison between the histogram templates for the up and down variations (upper and lower row, respectively) of one of the JEC sources before and after the smoothing.

before the selection is applied. This is done to take into account possible acceptance effects introduced by the theory variations during the event selection, without changing the overall cross section.

- **Matrix element**: uncertainties on the matrix element are included for the signals, as well as the WZ process, which is the main background. The renormalization and factorization scale are varied independently, i.e. two separate nuisances are included in the fit, and the corresponding uncertainties are treated as uncorrelated between the different processes.
- **PDF and α_S**: the variation of the PDF set are included for the tZq, $t\bar{t}Z$ and tWZ signal samples, as well as for the main background WZ. These uncertainties are treated as correlated across the years and between processes.
- **Parton shower**: uncertainties in the parton shower modeling in the MC generators are computed by varying α_S for initial and final state radiation (ISR and FSR) independently. These uncertainties are considered for tZq,

$t\bar{t}Z$ and tWZ, as well as for the main background WZ. The ISR uncertainty is treated as uncorrelated between the processes, while only two nuisances are considered for FSR, one for the signal samples and one for WZ.

- **tWZ modeling**: two different schemes are used in the MC generation of tWZ in order to account for the removal of the resonant diagrams (DR1 and DR2) [37]. The difference in shape caused by these two different approaches is treated as a shape uncertainty in the fit.
- **Underlying event and colour reconnection modeling**: MC samples with alternative models for both the underlying event [93] and colour reconnection [94], i.e. the reconfiguration of colour charges after the parton shower, are used to estimate the corresponding shape uncertainties. Uncertainties corresponding to the underlying event and colour reconnection are treated as uncorrelated among each other, but correlated among years and processes.
- **Limited sample size**: the Barlow-Beeston-light approach [95] is used to take into account the uncertainties arising from the limited size of the MC samples. In this method, an uncertainty parameter is assigned to the expected yield in each bin of the fit templates.

5.2 Simultaneous inclusive measurement

For the inclusive measurement, the fit also includes two additional regions with the same event selection as that of the SR, but with a difference concerning the lepton and b jet requirements:

- the four lepton region contains $t\bar{t}Z$ events at high purity. The main background for this region is the ZZ process, which is reduced when requiring at least one b jet in the event selection. For this region, the b jet multiplicity is included in the fit.
- The zero b jet region is enriched with WZ events and hence helps to constrain this background. For this region, the jet multiplicity is included in the fit.

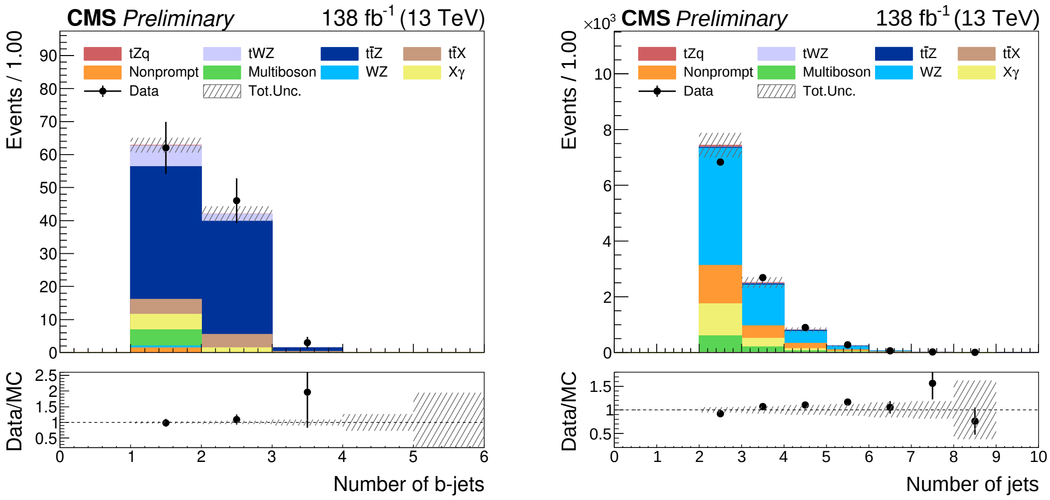


Figure 5.2: The b jet multiplicity in the four-lepton region (left) and jet multiplicity in the region with zero b jets (right). These distributions are added to the fit for the inclusive measurements.

The distributions added for the inclusive fits can be found in Figure 5.2.

The fit yields cross section ratios to the SM of 1.17 ± 0.07 for the combined $t\bar{t}Z$ and tWZ processes and 0.99 ± 0.13 for the tZq process. The inclusive cross sections for top quark production in association with a Z boson are defined in the phase space including resonant and nonresonant production of opposite-sign and same-flavour lepton pairs with an invariant mass $70 < m_{\ell^+\ell^-} < 110$ GeV. The predicted cross sections were evaluated in previous CMS measurements from the signal generator MadGraph5_aMC@NLO. For the $t\bar{t}Z$ and tWZ processes, these are equal to 0.84 ± 0.10 pb and 136^{+9}_{-8} fb, respectively [1, 3]. For the tZq process, the expected cross section was evaluated to be 94.2 ± 3.1 fb [2] in the phase space were the Z boson decays into a pair of leptons. As the calculation also includes nonresonant lepton-pair production with an invariant mass greater than 30 GeV, a transfer factor is evaluated from the simulated samples in order to get the tZq cross section in the correct phase space, and the branching ratio is taken into account as well.

As explained in the beginning of this chapter, the sum of cross sections for $t\bar{t}Z$

and tWZ is considered, since the measurement of this quantity is expected to be more reliable due to the overlap between the two processes. The inclusive cross sections are measured to be

$$\sigma(\mathrm{t\bar{t}Z+tWZ}) = 1.14 \pm 0.05 \text{ (stat) } \pm 0.04 (\text{ syst}) \text{ pb},$$
$$\text{and } \sigma(\mathrm{tZq}) = 0.81 \pm 0.07 \text{ (stat) } \pm 0.06 \text{ (syst) pb}.$$

The post-fit distributions of all the regions included in the fit are shown in Figure 5.3. Both results are found to be in agreement with previous CMS measurements [1–3], as summarized in Table 5.2.

The previous measurement of the tZq cross section [2] was performed on the same dataset as this analysis, but the results were reported in the phase space in which the Z boson decays into two charged leptons (electrons, muons, or tau leptons) with an invariant mass greater than 30 GeV. In this phase space, if $\mathrm{t\bar{t}Z}$ and tWZ are fixed to the SM predictions with normalization uncertainties of 15%, the result of this measurement yields $93.2^{+11.7}_{-10.8}$fb for the tZq process. This is in agreement with the previous result of $87.9^{+10.5}_{-9.5}$ fb [2].

In Ref. [1], the cross section for $\mathrm{t\bar{t}Z}$ was reported for the same phase space considered for this measurement. If only the $\mathrm{t\bar{t}Z}$ cross section is measured, the result strongly depends on the treatment of the tWZ background: if this is fixed to the SM prediction with a normalization uncertainty of 40%, the analysis presented here yields 0.99 ± 0.07 pb for the $\mathrm{t\bar{t}Z}$ cross section, while the value drops to 0.88 ± 0.16 pb if the tWZ process is left freely floating in the fit. In both cases, the tZq process is treated as background with a normalization uncertainty of 15%, and the values are in good agreement with the previous result of 0.95 ± 0.08 pb [1] for the $\mathrm{t\bar{t}Z}$ cross section.

Finally, if the $\mathrm{t\bar{t}Z}$ and tZq processes are treated as backgrounds with normalization uncertainties of 15%, the measurement of the inclusive tWZ cross section yields 0.39 ± 0.16 0b, also in agreement with the recent result of 0.37 ± 0.11 pb [3].

The impacts of the leading systematic uncertainties in the inclusive fits can be found in Figure 5.4. Each systematic uncertainty is included in the fit as a nuisance param-

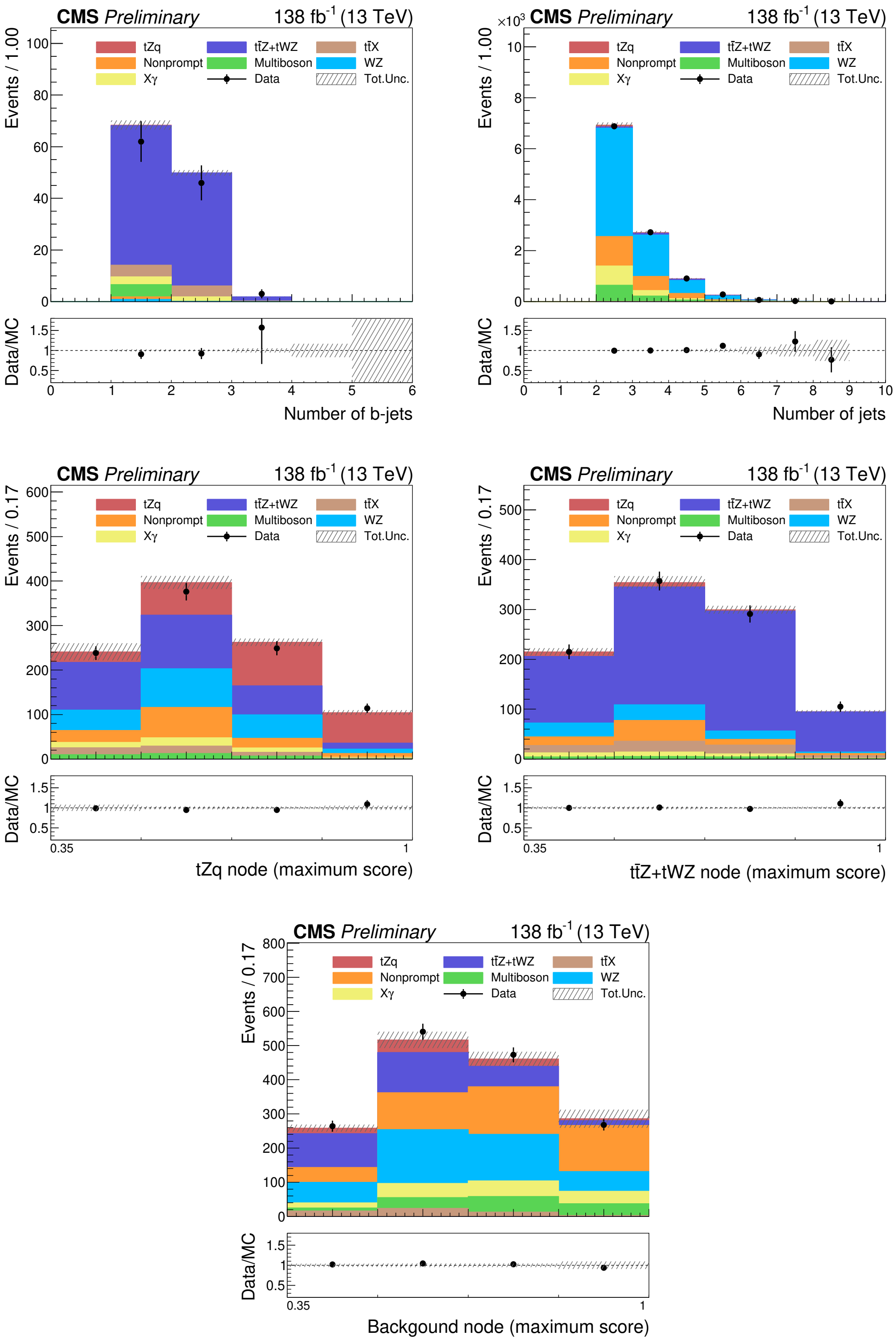


Figure 5.3: Post-fit distributions of the b jet and jet multiplicity in the regions with four leptons and zero b jets (top left and top right, respectively), the tZq (middle left), $t\bar{t}Z$ +tWZ (middle right) and background (bottom) output nodes, where only the events with maximum score are included.

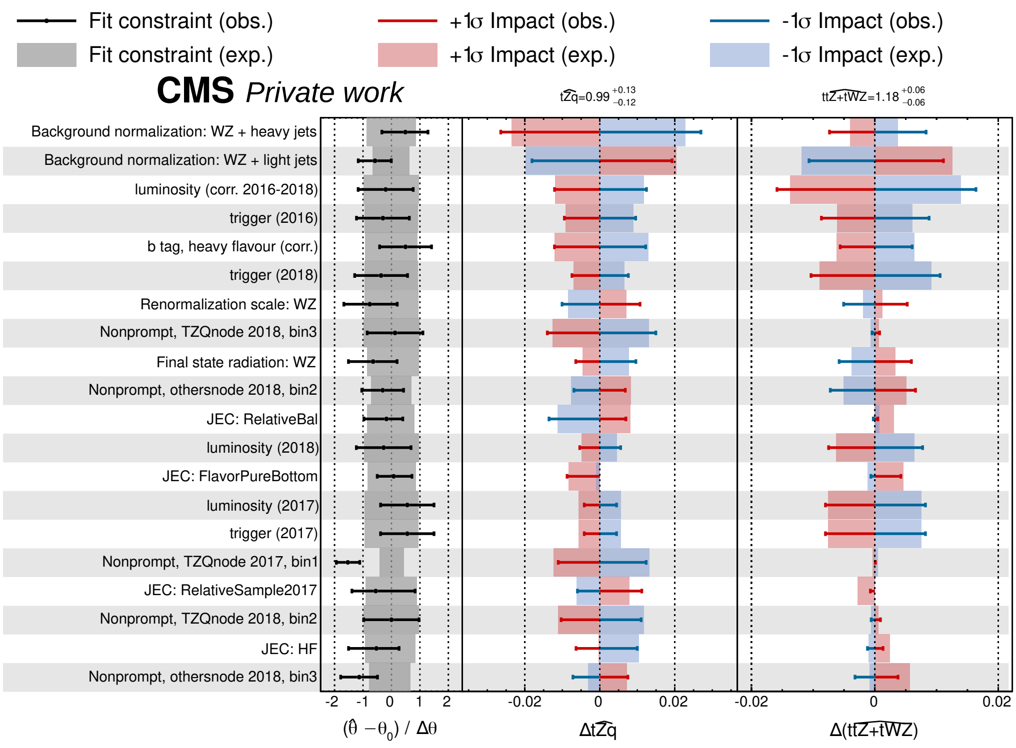


Figure 5.4: Impacts of the leading systematic uncertainties in the inclusive fit of the two signal cross sections. Each uncertainty corresponds to a different nuisance parameter in the fit. The black dots and error bars in the first column show the difference between the post-fit value of the nuisance parameter θ and the corresponding expectation θ_0, divided by the assumed uncertainty $\Delta\theta$. The gray bands are showing the same quantity from a fit performed on Asimov data, i.e. a data set in which the observed quantities are set equal to the expected values. The following columns show the impact of each systematic source on the tZq and $\mathrm{t\bar{t}Z}$ $+\mathrm{tWZ}$ signal strengths for the up (red) and down (blue) variations. As for the first column, the horizontal error bars refer to the case where the fit is performed on real data, while the bands indicate the expected impacts estimated by performing the fit on an Asimov data set.

Table 5.2: Comparison between the results presented in this thesis and the previous CMS measurements.

Signal	Previous CMS result	This result
$\mathrm{t\bar{t}Z}$	0.95 ± 0.08 pb [1]	0.99 ± 0.07 pb (tWZ fixed to SM expectations) 0.88 ± 0.16 pb (tWZ freely floating)
tZq	$87.9^{+10.5}_{-9.5}$ fb [2]	$93.2^{+11.7}_{-10.8}$fb
tWZ	0.39 ± 0.16 pb [3]	0.37 ± 0.11 pb

eter θ; its impact on the signal strength r is defined as the shift Δr that is induced as θ is fixed and brought to its $\pm 1\sigma$ post-fit values, with all other parameters profiled as normal. This is a measure of the correlation between the nuisance parameters and the signal strength, which is useful to determine which systematic uncertainties have the largest effect. The direction of the $+1\sigma$ and -1σ impacts (i.e. when θ is moved to its $+1\sigma$ or -1σ values) on r indicates whether the parameter variation, representing the uncertainty, is correlated or anti-correlated with the measured cross section. The left panel in the impact plot shows the value of $(\theta - \theta_0)/\Delta\theta$, where θ and θ_0 are the post and pre-fit values of the nuisance parameter and $\Delta\theta$ is the pre-fit uncertainty. The error bars show the post-fit uncertainty divided by the pre-fit uncertainty. The leading systematic uncertainties are related to the identification of b jets, matrix element and jet energy scale variation, as they significantly affect the reconstruction of the particles in the final states.

Finally, the impact of the various systematic uncertainties on the measurement of the cross sections is estimated by repeating the fit with the corresponding nuisance fixed to its post-fit value and quantifying how this affects the cross section values; these results are listed in Table 5.3. Both measurements are found to be statistically limited; among the systematic uncertainties, the highest impact on the signal cross sections is coming from the uncertainties related to jets and b-tagging, and from the nonprompt background estimation.

Table 5.3: Systematic uncertainty sources and their relative impact on the inclusive cross section measurements for $\mathrm{t\bar{t}Z}$ +tWZ and tZq. The impacts are estimated by performing the fit with a specific uncertainty source fixed to its post-fit value, and subsequently quantifying the effects on the cross section values.

Source	$\sigma(\mathrm{t\bar{t}Z + tWZ})$ [%]	$\sigma(\mathrm{tZq})$ [%]
Experimental uncertainties		
b-tagging	1%	2%
Jet energy scale	1%	3%
Jet energy resolution	<1%	1%
MET	<1%	3%
Lepton identification efficiencies	1%	2%
Nonprompt background	2%	3%
Other backgrounds	2%	4%
Pileup	<1%	1%
L1 trigger prefiring	<1%	2%
Trigger	2%	2%
Luminosity	2%	2%
HEM1516 issue	<1%	<1%
Theory uncertainties		
Factorization scale	1%	1%
Renormalization scale	1%	2%
PDF and α_S	<1%	<1%
Parton shower	<1%	2%
tWZ modelling	<1%	<1%
Underlying event and colour reconnection	1%	2%
MC statistics	<1%	1%
Statistical	3.7%	9.9%
Total	5.1%	13.1%

The fit can also be performed at fixed signal strengths and, if the procedure is repeated for a sufficient number of points, a two-dimensional likelihood contour are obtained from the fit results in order to understand the correlation among the signal processes, as shown in Figure 5.5. A small positive correlation is observed; this is due to the correlated sources of uncertainty such as JECs, lepton ID, luminosity and trigger.

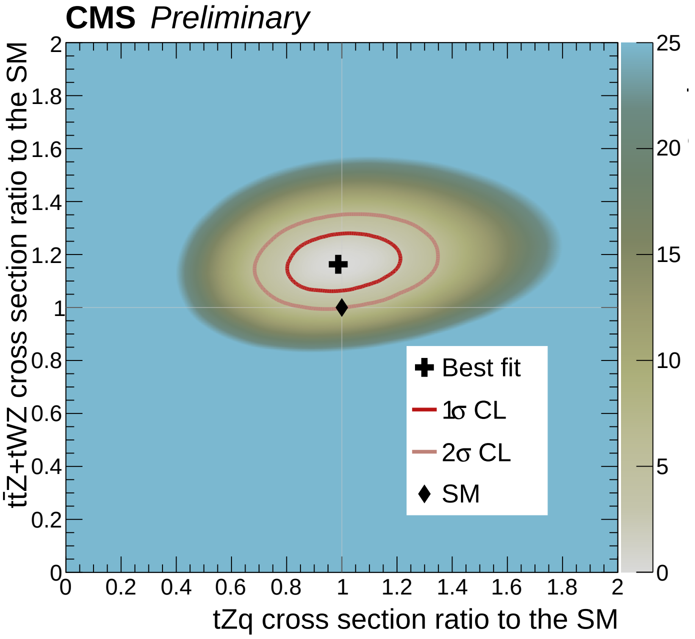


Figure 5.5: Two-dimensional likelihood surface in the simultaneous inclusive fit as a function of the tZq and $t\bar{t}Z$ +tWZ signal strengths. The values are obtained by performing the fit at fixed signal strengths.

5.3 Simultaneous differential measurement

The signal cross sections are measured differentially as functions of the following observables:

- the p_{T} of the Z boson, $p_{\mathrm{T}}(\mathrm{Z})$,
- the p_{T} of the lepton coming from the W boson decay, $p_{\mathrm{T}}(\ell_{\mathrm{W}})$,
- the azimuthal angle between the two leptons coming from the decay of the Z boson, $\Delta\phi(\ell^+,\ell^-)$,

- the ΔR between the Z boson and the lepton coming from the W boson decay, $\Delta R(Z, \ell_\mathrm{W})$ and
- the cosine of the polar angle between the Z boson and the negatively charged lepton coming from its decay, boosted into the Z boson restframe, $\cos(\theta^*_\mathrm{Z})$.

These variables are particularly sensitive to BSM scenarios and can be used for EFT interpretations. For example, $p_\mathrm{T}(\mathrm{Z})$ and $p_\mathrm{T}(\ell_\mathrm{W})$ are sensitive to those EFT operators that describe modified W-t-b couplings [96, 97]. In case of new physics, some discrepancies can be expected in $\Delta\phi(\ell^+, \ell^-)$, as the leptons from the Z boson decay will be more or less collimated depending on the cross section values. The $\Delta R(Z, \ell_\mathrm{W})$ and $\cos(\theta^*_\mathrm{Z})$ variables are sensitive to the modeling of the processes and contain information on the nature of the top quark–Z boson coupling. Furthermore, $\cos(\theta^*_\mathrm{Z})$ can be used to parametrize the polarization of the Z boson.

The measurement is performed at the so-called *parton level*, in order to extract results that can be compared with fixed-order theory calculations: objects are defined based on event generator level particles after initial and final state radiation and before hadronization. All the required information are stored in the simulated samples used for the analysis. The measured cross sections are compared with theory predictions coming from the MC samples. The cross section is defined for an invariant mass of the Z boson (or virtual photon) within 70 and 110 GeV for lepton pairs with opposite charge and same flavour.

The values of the reconstructed observables are affected by the limited acceptance of the detector, as well as hadronization effects, imperfect reconstruction and identification efficiencies. To make comparisons with the theory predictions, these effects are removed with a procedure called *unfolding*: the signal processes are divided into different signal contributions, called generator-level bins

$$\sigma_i = \int_{x_i^{low}}^{x_i^{high}} \frac{\mathrm{d}\sigma(x)}{\mathrm{d}x} \mathrm{d}x, \tag{5.2}$$

where x_i^{low} and x_i^{high} refer to the lower and upper bounds of the i^th generator-level bin. Each bin is then included in the fit as a separate signal template and hence it is scaled by a different signal strength in the fit. The binning of the different observables is chosen by computing a response matrix for both signal samples and maximizing the purity and stability, as well as the number of signal events in each bin. The purity is defined as the fraction of events in a detector-level bin which belong to the corresponding generator-level bin, while the stability is the fraction of events in a generator-level bin that are observed in the corresponding detector-level bin. The response matrices are shown in Figure 5.6 for the case of $p_{\mathrm{T}}(\mathrm{Z})$ and $\cos(\theta^*_{\mathrm{Z}})$, and in Appendix A for all other observables. In each of them, the diagonal elements are found to be large. Alternatively, in order to estimate their accuracy, condition numbers can be considered: these are generally defined as a measure of how sensitive the output value of a function is to changes in the input data. If these numbers are high, typically above 10^4, it can be concluded that the unfolded observables are sensitive to small changes at generator-level and a regularization procedure needs to be applied [98], i.e. a bias term is introduced in the likelihood in order to penalize solutions that are too far from the truth-level distribution. The condition numbers for the different response matrices and signal processes are reported in Table 5.4: as they were all found to be rather close to unity, it can be concluded that no regularization is needed in this analysis.

Table 5.4: Condition numbers of the response matrices.

Variable	tZq	$\mathrm{t\bar{t}Z}$ +tWZ
p_{T} (Z)	1.15	1.25
p_{T} (ℓ_{W})	1.08	1.18
$\Delta\phi(\ell^+, \ell^-)$	1.05	1.11
$\Delta R(Z, \ell_{\mathrm{W}})$	1.13	1.21
$\cos(\theta^*_{\mathrm{Z}})$	1.09	1.1

The correlations between the observables considered for the unfolding and the DNN input variables is studied in order to avoid possible biases in the cross section measurement. Some distributions that were initially used, e.g. the p_{T} of the leptons, were removed from the training as they were found to be highly correlated with the p_{T} of the Z boson and ℓ_{W}. In this way, it can be ensured that biases connected to detector features and MC modeling do not affect the measurement. Furthermore, as the DNN is only trained on SM assumptions, such a bias would lead the DNN to

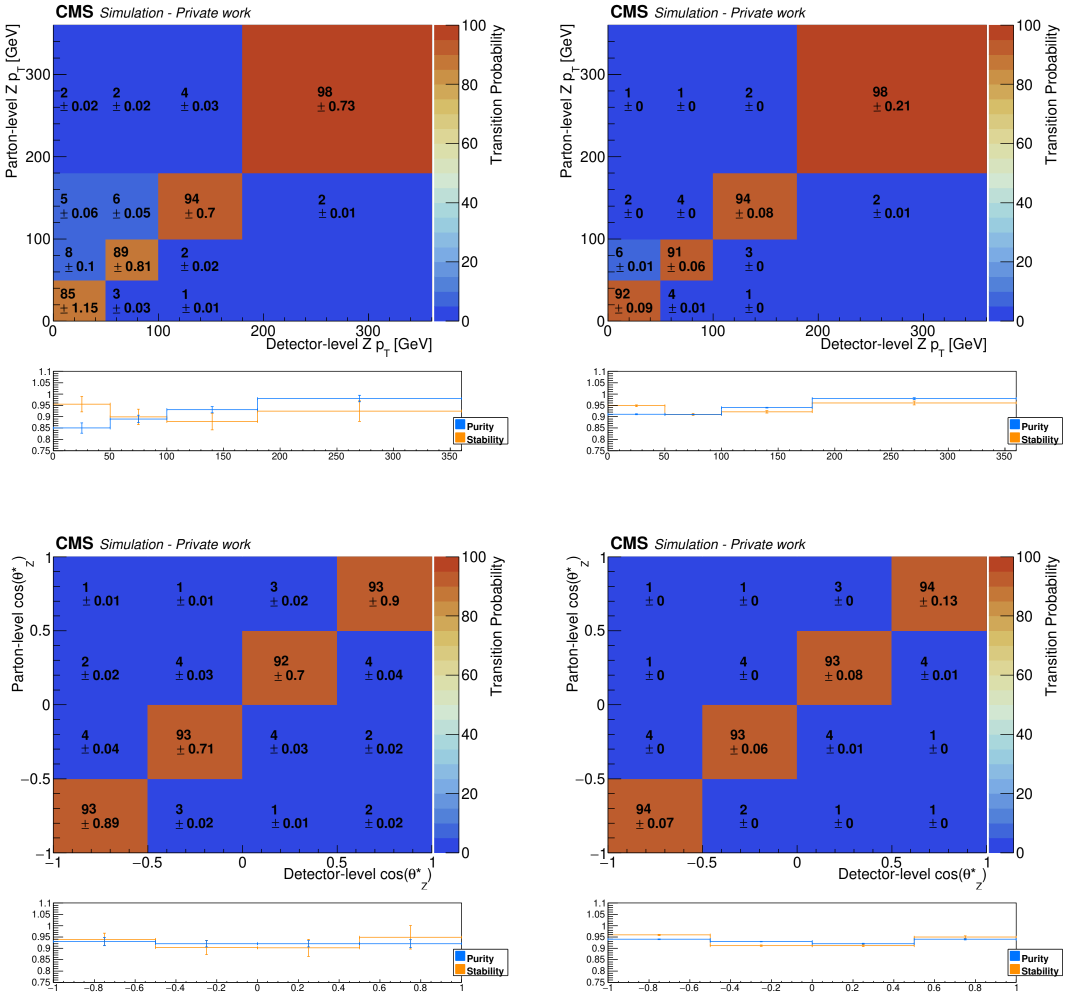


Figure 5.6: Response matrices for the sum of the $t\bar{t}Z$ and tWZ (left) and tZq (right) samples. Plots refer to $p_{T}(Z)$ (top) and $\cos(\theta^{*}_{Z})$ (bottom).

only select SM-like events.

The samples are divided into sub-samples corresponding to the different generator-level bins of the observable to unfold, and each of these is scaled by a different signal strength. In order to perform the simultaneous differential measurement, events in the signal output nodes are further split into different categories corresponding to the detector-level bins, while only one category is used for the background node. As an example, the resulting pre-fit and post-fit plots are shown in Figures 5.7 and 5.8 for the p_{T} of the Z boson, while the corresponding impact plot can be found in Figure 5.9. As for the inclusive measurement, it can be seen that data are found to be somewhat higher than the expectation for the sum of the $\mathrm{t\bar{t}Z}$ and tWZ cross sections, while results are in good agreement with the SM prediction in the case of tZq. Some nuisances, mainly related to the background normalization and the JECs, have a significant impact on the signal strength in the first bin of the differential measurement, and their contribution is gradually reduced in the following bins. This trend arises from the higher background contamination in the first bin: jets and b jets in background processes have higher chances to be misidenitfied, and therefore the yields in the first bin are subject to more changes when varying the SFs. Additionally, energies in the first bin are generally lower, which also causes the uncertainties to increase.

The unfolded distributions are provided in Figures 5.10 and 5.11 for all the observables. These plots also include theory predictions as obtained from the MC generator, together with their uncertainties; these are estimated by summing in quadrature the systematics coming from matrix element scale (renormalization and factorization), PDFs and parton showers. For the matrix element scale, the renormalization and factorization component are considered separately and summed in quadrature. In the sum of $\mathrm{t\bar{t}Z}$ and tWZ, the uncertainties related to the overlap removal are also included. The differential cross sections show in which regions the data have different trends compared to the prediction, which can be useful to identify the source of the discrepancies. This is especially clear in the unfolded distribution of $p_{\mathrm{T}}(\ell_{\mathrm{W}})$: the excess is quite significant for low p_{T} values and it reduces towards the end of the distribution. As most of the time, this variable refers to the lepton

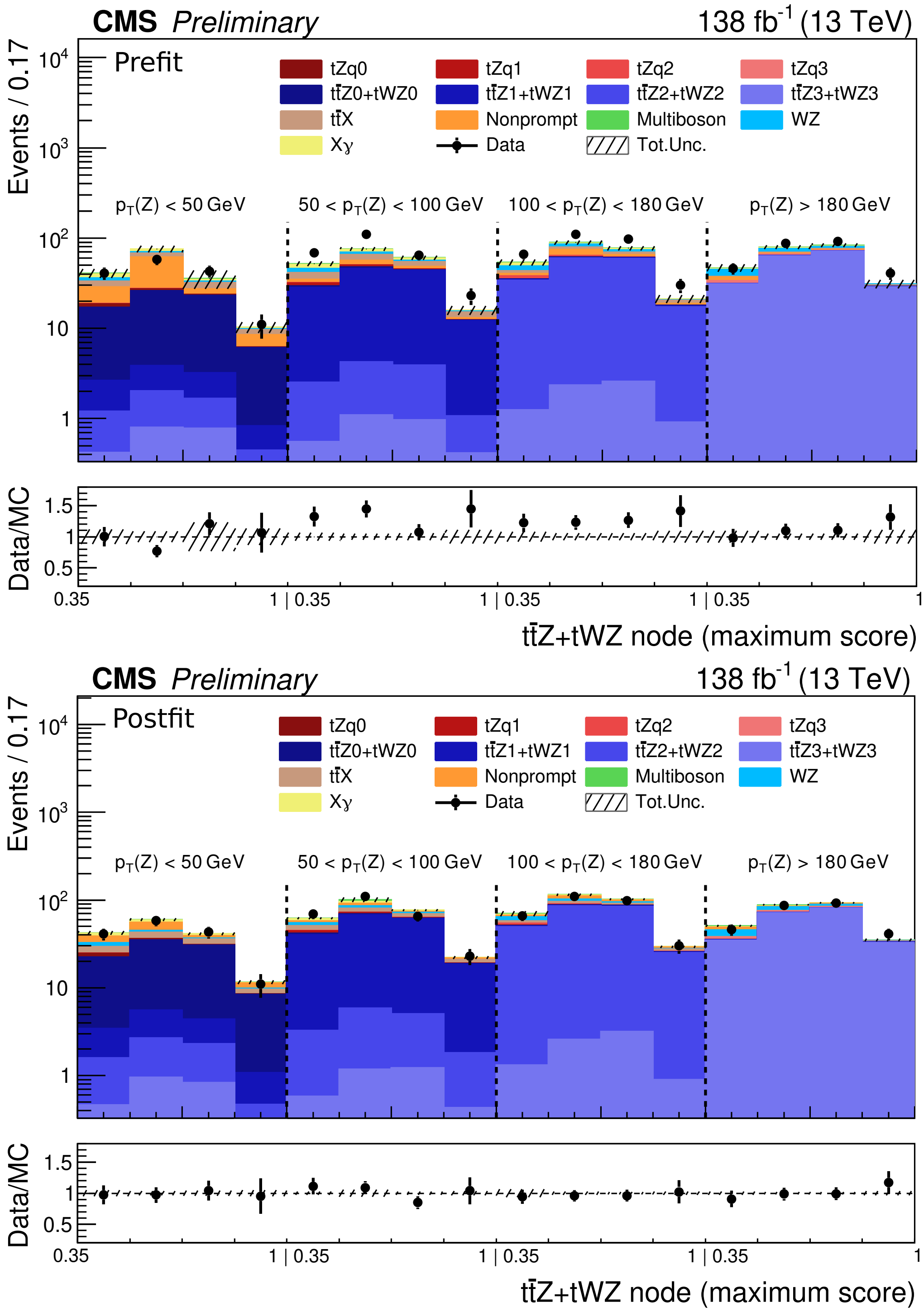


Figure 5.7: Pre-fit (top) and post-fit (bottom) distributions from the $t\bar{t}Z$ +tWZ output node. For each, four different templates are created according to the value of the reconstructed p_T of the Z boson. In the legend, the numerical suffices refer to the different generator-level bins.

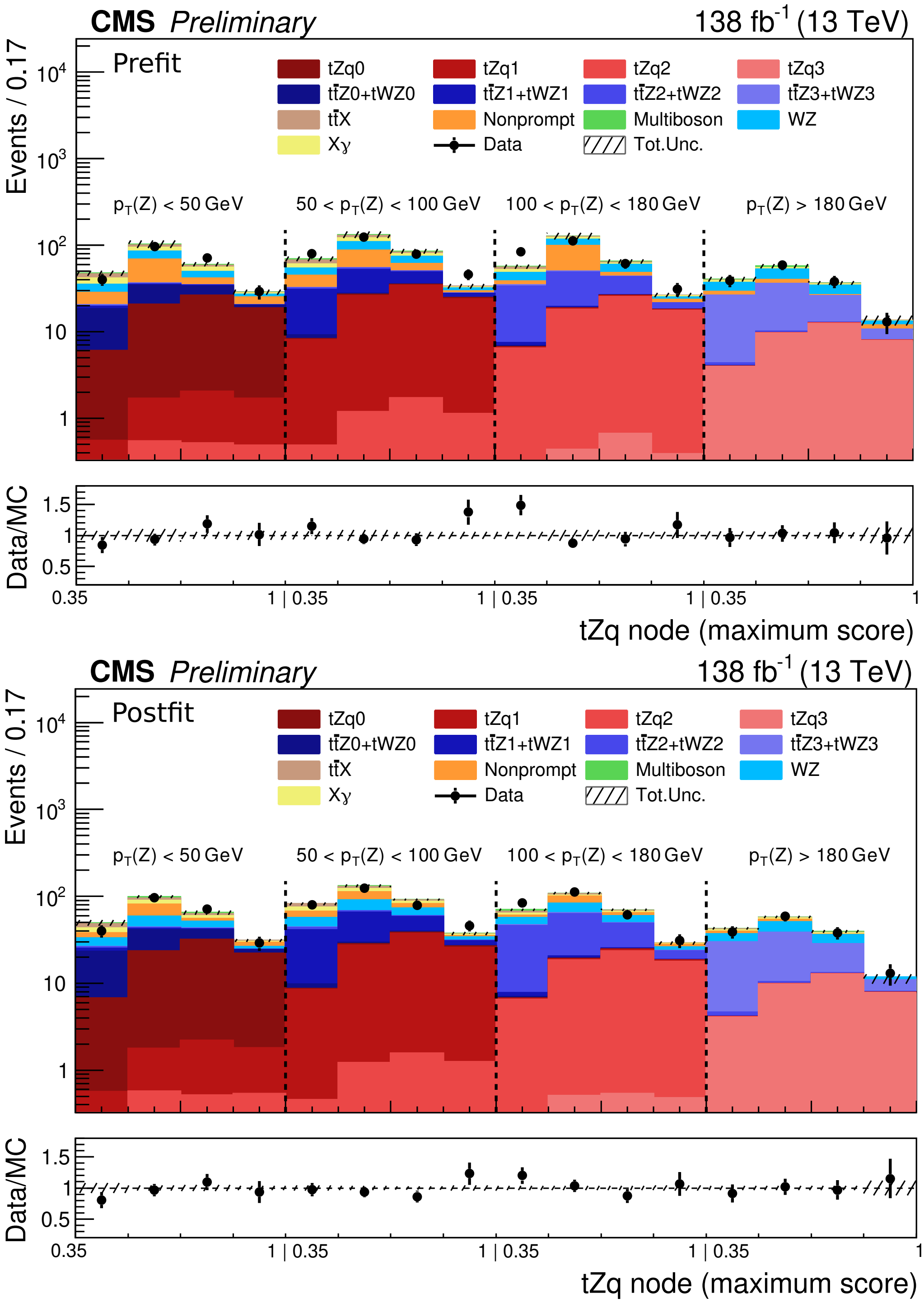


Figure 5.8: Pre-fit (top) and post-fit (bottom) distributions from the tZq output node. For each, four different templates are created according to the value of the reconstructed p_{T} of the Z boson. In the legend, the numerical suffices refer to the different generator-level bins.

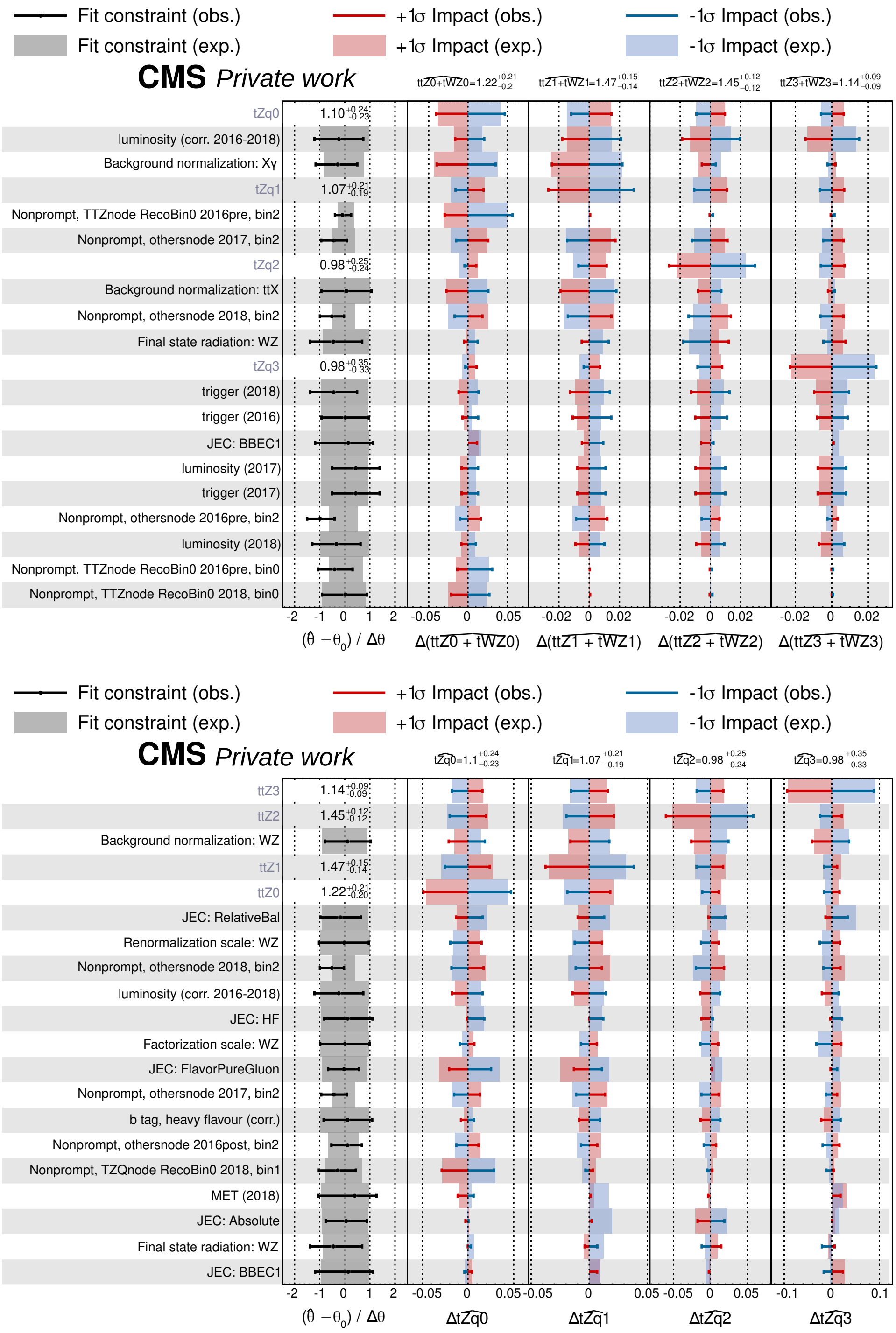


Figure 5.9: Impacts of the leading systematic uncertainties on the signal strengths of $\mathrm{t\bar{t}Z}$ +tWZ (top) and tZq (bottom). These results refer to the case where the two cross sections are measured directly and simultaneously as a function of the p_T of the Z boson. A description of the various objects can be found in Figure 5.4.

arising from the top quark decay, it is interesting to note that the same trend has been observed in differential measurements of the $\mathrm{t\bar{t}}$ cross section [99].

Figures 5.12 and 5.13 show the normalized unfolded distributions, which are constructed in order to make shape effects more visible. Since most normalization uncertainties cancel out, systematic uncertainties are significantly reduced; however, the data are dominated by statistical uncertainties. The correlation matrix is used to evaluate the uncertainties in the normalized differential cross sections via the Gaussian error propagation.

The covariance matrices resulting from the fits are provided in Figure 5.14 for the p_{T} of the Z boson, and in Appendix B for the other observables. The results are still dominated by the statistical uncertainties of the data, and thus largely uncorrelated across the bins. However, understanding the correlations among the signal strengths is essential to correctly make theory interpretations of the results. This is especially true in the case of EFT interpretations, where the sum of the signal cross sections is needed in this case.

5.4 Outlook

Precise measurements are crucial to test the SM predictions and identify possible deviations that might lead to the discovery of new physics. The results presented in this thesis are limited by the statistical uncertainties. Including additional data recorded at higher luminosities can bring substantial improvements to the precision of the extracted cross sections. The LHC Run-3 started in July 2022 bringing proton-proton collisions at a center-of-mass energy of 13.6 TeV and a total luminosity of $250\,\mathrm{fb}^{-1}$ is expected to be recorded. The statistical uncertainty on the inclusive cross sections measured in this thesis is expected to become subdominant already with Run-3 data, as it is expected to be scaled down by the inverse square root of the increased luminosity: a level of about 5% and 2% can then be expected for the statistical uncertainty on the inclusive tZq and $\mathrm{t\bar{t}Z}$ +tWZ cross sections, respectively.

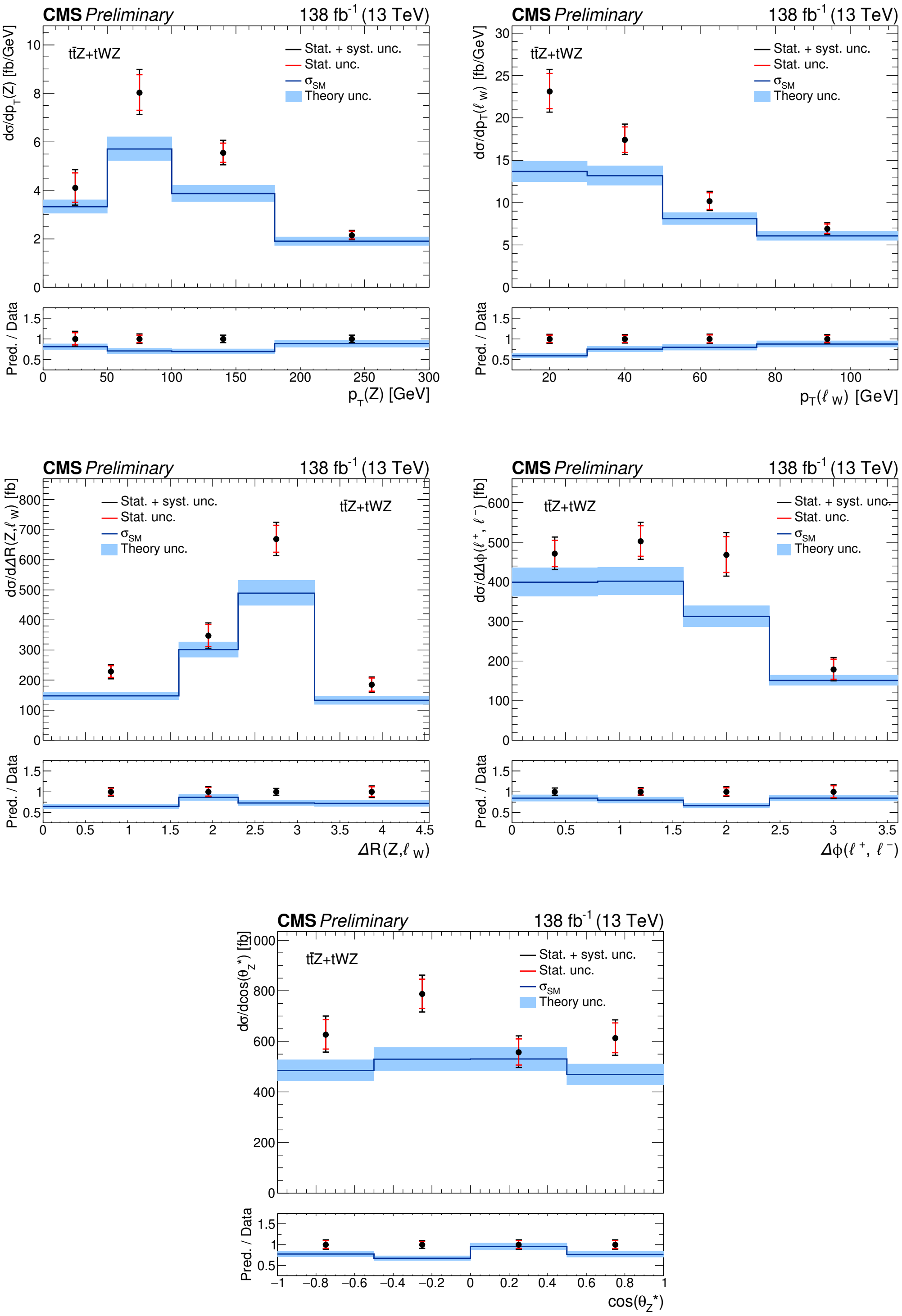


Figure 5.10: Unfolded distributions of the sum of the tWZ and $\mathrm{t\bar{t}Z}$ cross sections in the simultaneous measurement as a function of the Z boson p_T (top left), ℓ_W p_T (top right), $\Delta R(Z, \ell_\mathrm{W})$ (center left), $\Delta\phi(\ell^+, \ell^-)$ (center right) and $\cos(\theta^*_\mathrm{Z})$ (bottom).

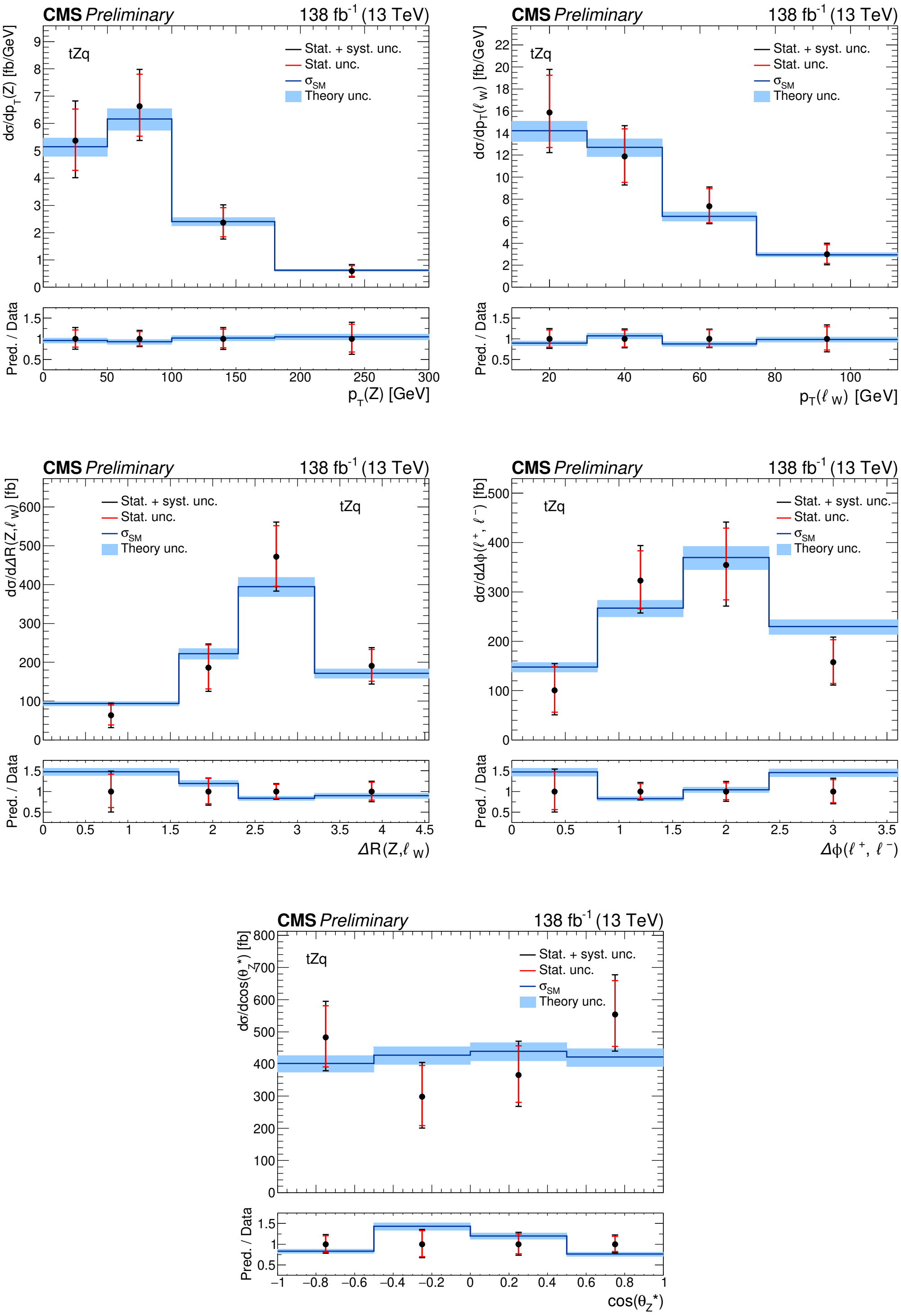


Figure 5.11: Unfolded distributions of the tZq cross section in the simultaneous measurement as a function of the Z boson p_T (top left), ℓ_W p_T (top right), $\Delta R(Z, \ell_W)$ (center left), $\Delta\phi(\ell^+, \ell^-)$ (center right) and $\cos(\theta^*_Z)$ (bottom).

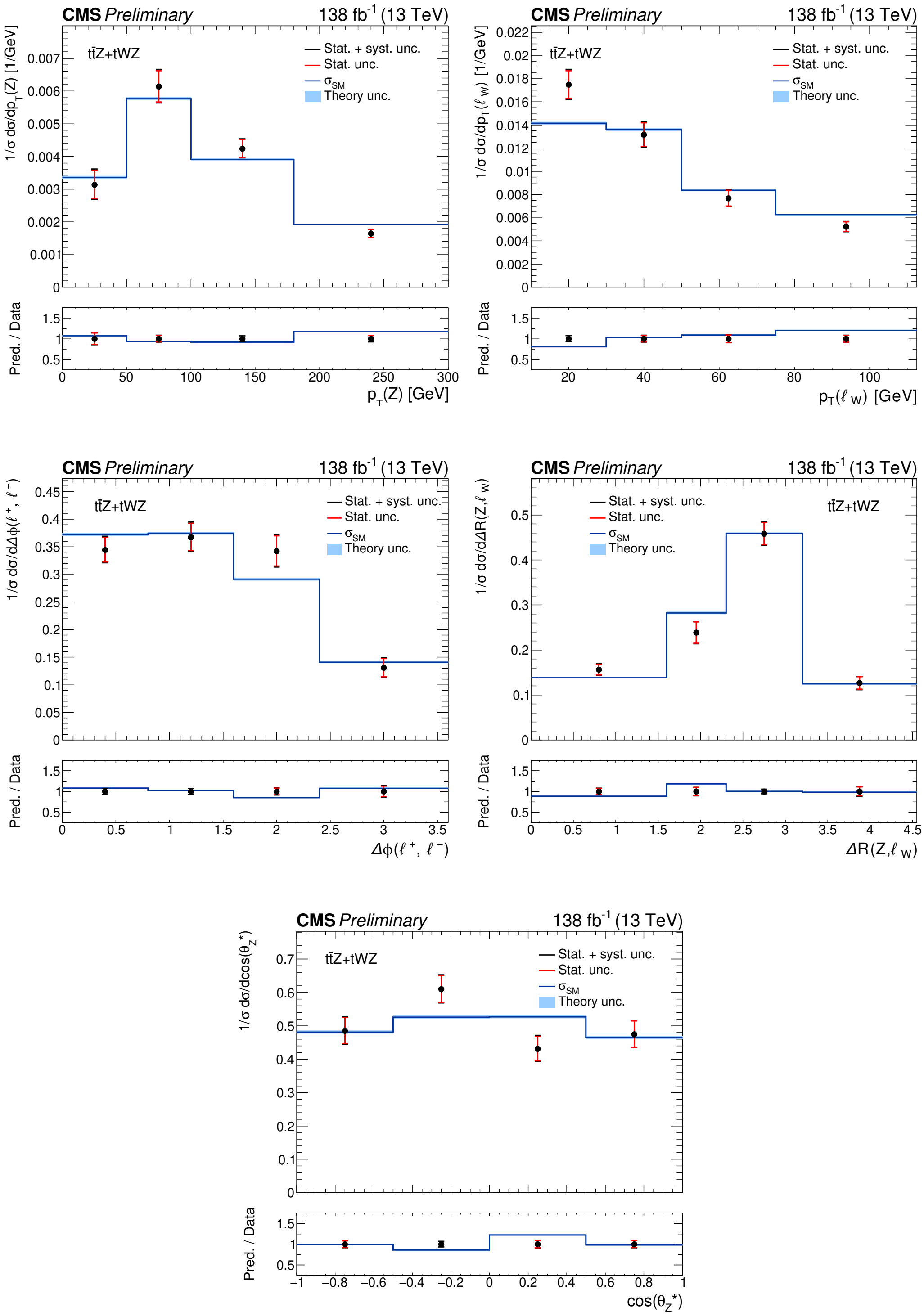


Figure 5.12: Normalized unfolded distributions of the sum of the tWZ and $t\bar{t}$Z cross sections in the simultaneous measurement as a function of the Z boson p_{T} (top left), ℓ_{W} p_{T} (top right), $\Delta R(Z, \ell_{\mathrm{W}})$ (center left), $\Delta\phi(\ell^{+}, \ell^{-})$ (center right) and $\cos(\theta^{*}_{\mathrm{Z}})$ (bottom).

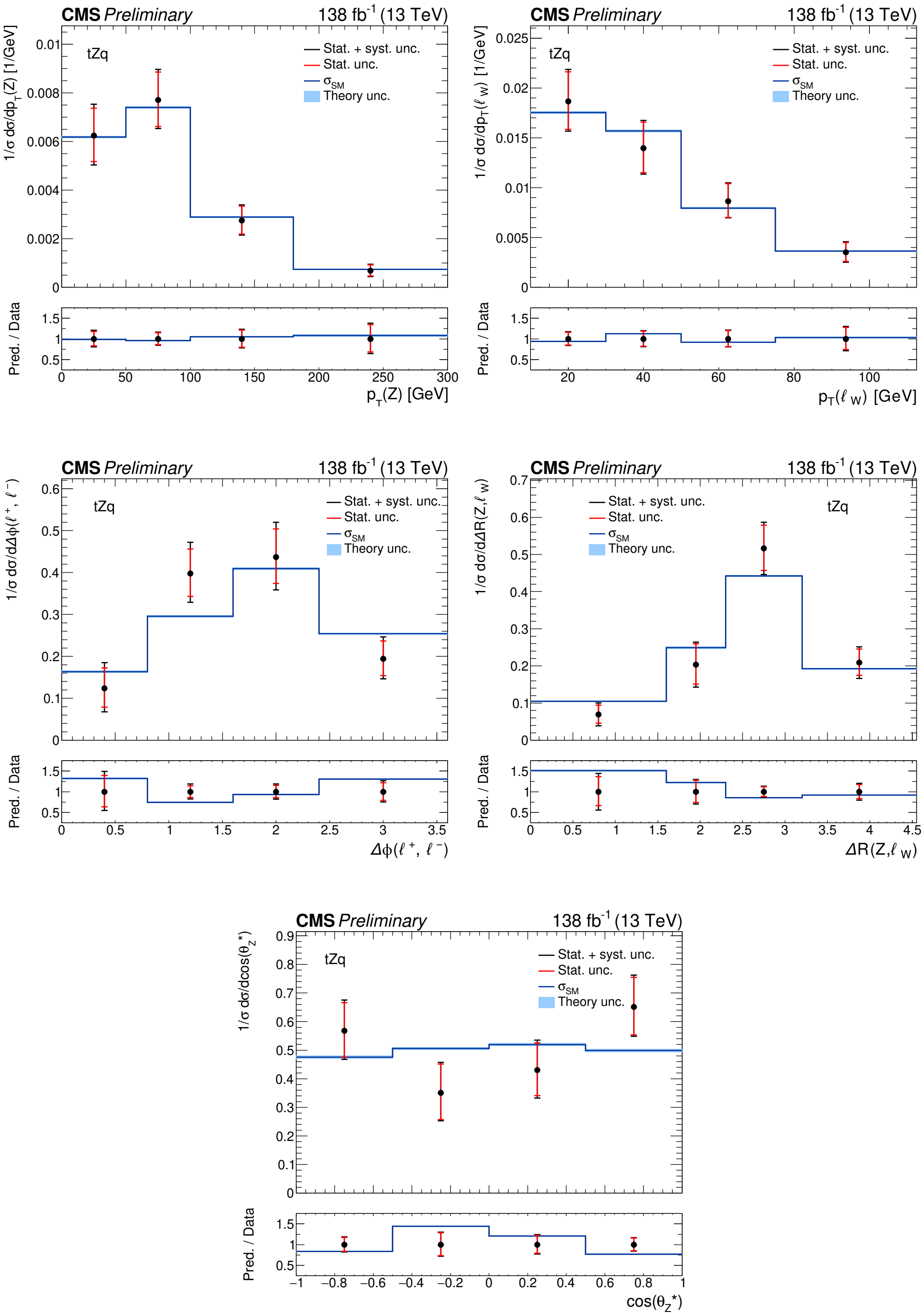
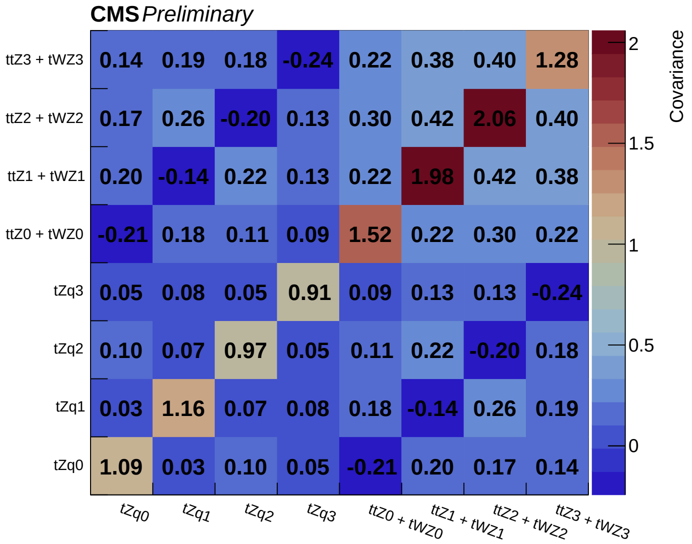


Figure 5.13: Normalized unfolded distributions of the tZq cross section in the simultaneous measurement as a function of the Z boson p_{T} (top left), ℓ_{W} p_{T} (top right), $\Delta R(Z, \ell_{\mathrm{W}})$ (center left), $\Delta\phi(\ell^{+}, \ell^{-})$ (center right) and $\cos(\theta^{*}_{\mathrm{Z}})$ (bottom).



	tZq0	tZq1	tZq2	tZq3	ttZ0 + tWZ0	ttZ1 + tWZ1	ttZ2 + tWZ2	ttZ3 + tWZ3
ttZ3 + tWZ3	0.14	0.19	0.18	-0.24	0.22	0.38	0.40	1.28
ttZ2 + tWZ2	0.17	0.26	-0.20	0.13	0.30	0.42	2.06	0.40
ttZ1 + tWZ1	0.20	-0.14	0.22	0.13	0.22	1.98	0.42	0.38
ttZ0 + tWZ0	-0.21	0.18	0.11	0.09	1.52	0.22	0.30	0.22
tZq3	0.05	0.08	0.05	0.91	0.09	0.13	0.13	-0.24
tZq2	0.10	0.07	0.97	0.05	0.11	0.22	-0.20	0.18
tZq1	0.03	1.16	0.07	0.08	0.18	-0.14	0.26	0.19
tZq0	1.09	0.03	0.10	0.05	-0.21	0.20	0.17	0.14

Figure 5.14: Covariance matrices for the simultaneous differential fit, for the case where the measurement is performed as a function of the Z boson p_{T}.

A long shutdown is planned after Run-3 to prepare both accelerator and experiments for the high luminosity LHC (HL-LHC) [100]. The project is planned to start in 2029 and it aims to achieve a total integrated luminosity of $3000\,\mathrm{fb}^{-1}$ over a period of twelve years. This large amount of events will allow to significantly increase the accuracy of the measurements performed on Run-2 data. A projection for the expected performance of a differential measurement of the Z boson associated top quark production is shown in Figure 5.15, for different hypotheses on the Wilson coefficients introduced in Section 1.8 and the resulting shapes: the distributions are expected to become sensitive even to small deviations from the SM and it will be possible to look for non-standard top quark electroweak couplings with much better precision.

An increase in statistics can also allow for the inclusion of more signal processes in the simultaneous measurement, like the associated production of top quarks with a W boson or a photon. This would generally increase the sensitivity to the top quark electroweak couplings. The addition of new signal processes might also help to highlight the distinctions between single and pair production of top quarks, allowing for

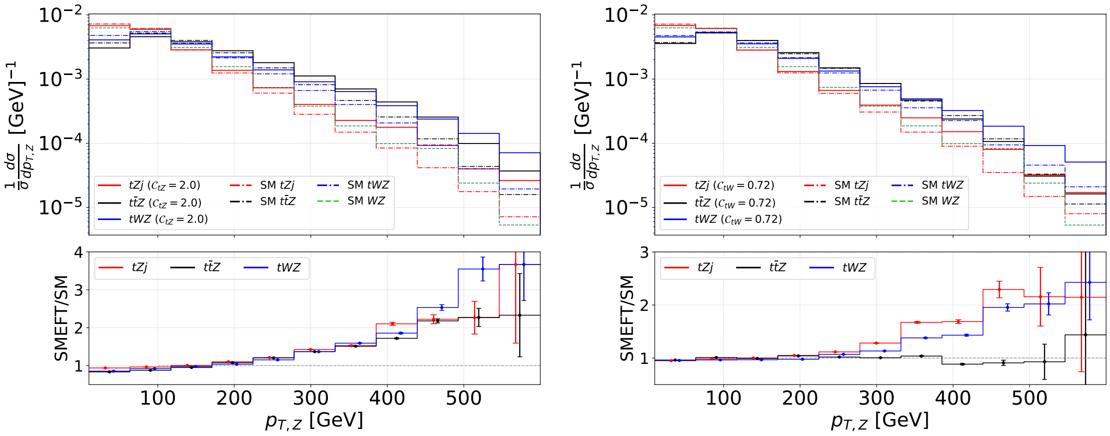


Figure 5.15: Differential t$\bar{t}$Z, tWZ and tZq cross sections as a function of the p_T of the Z boson. The SM expectations (dashed lines) are compared with the shapes obtained from different hypotheses on the Wilson coefficients of the EFT operators that modify the couplings of the top quark with the Z and W boson (C_{tZ} and C_{tW}, respectively). The continuous lines are obtained with C_{tZ}=2 (right) and C_{tW}=0.72. Taken from Ref. [101].

the identification of new physics scenarios that specifically impact one production mode over the other.

Furthermore, the differential measurement of the signal cross sections as a function of new observables, like those involving the top quark kinematics, can bring complementary information to the ones provided in this work. This might require a tighter event selection, as the presence of hadronic variables in the top quark reconstruction is expected to increase the bin-to-bin migrations in the response matrices.

Conclusion

In this thesis, the first simultaneous measurement of single and pair production of top quarks in association with a Z boson (tZq, tWZ and $\mathrm{t\bar{t}Z}$) has been presented. The results of this measurement are currently under review by the CMS Collaboration and have been presented in March 2024 at the conference "Moriond/EW2024: 58th Rencontres de Moriond on Electroweak Interactions & Unified Theories". Additionally, a novel approach to evaluate the muon SFs in MVA-based ID has been described. Results have been obtained analyzing data recorded with the CMS experiment during Run-2 at the LHC.

Physics analyses targeting multilepton final states usually make use of MVA-based IDs for the identification of lepton candidates, which allow to better discriminate the signal against the nonprompt backgrounds compared to the cut-based approaches. A novel technique has been developed to correct the full shape of these discriminators in the simulation and match the one observed in data. The muon SFs are first evaluated with the tag-and-probe method for different ranges of the MVA discriminator and subsequently interpolated as a function of the muon's kinematic properties. The implementation of this method is now centrally available within the CMS Collaboration and can be exploited to use the full discriminator output distribution in physics analyses. For example, its shape can be used as input to train MVA algorithms and improve the discrimination between signals and backgrounds.

The cross sections of single and pair production of top quarks in association with a

Z boson have been measured simultaneously for the first time. Events recorded with the CMS experiment in proton-proton collisions at 13 TeV in the years 2016–2018 have been analyzed targeting final states with three leptons for all the signal processes. Two leptons are expected to originate from the Z boson, while the third one appears from the decay of a W boson. The event selection is performed using a set of optimized criteria to reduce the background contributions and the top quarks in the final states are reconstructed from the objects fulfilling these requirements. While most of the backgrounds are estimated with simulated events, the nonprompt contribution has been determined with a data-driven method. The separation between signals and backgrounds is further optimized using a multiclass neural network and the output node distributions are exploited to build different fit categories for the signal extraction. The signal cross sections are extracted from a maximum likelihood fit taking into account both statistical and systematic uncertainties. The inclusive measurement gives

$$\sigma(\mathrm{t\bar{t}Z+tWZ}) = 1.14 \pm 0.05\ \mathrm{(stat)}\ \pm 0.04\mathrm{(\ syst)\ pb},$$
$$\text{and}\ \ \sigma(\mathrm{tZq}) = 0.81 \pm 0.07\ \mathrm{(stat)}\ \pm 0.06\ \mathrm{(syst)\ pb}.$$

where the $\mathrm{t\bar{t}Z}$ and tWZ cross sections are measured together, as the overlap between the two processes makes their separation challenging on both theoretical and experimental fronts. In the inclusive measurement, events with four leptons in the final state are included in the fit, and the region with zero b jets is used to constrain the WZ background. The tZq cross section is found to be in good agreement with SM predictions, while an excess is measured for the combination of $\mathrm{t\bar{t}Z}$ and tWZ. The results are consistent with previous CMS measurements [1–3] and there is indication that the prediction for the tWZ process is underestimated. Compared to previous measurements, in which the signals have been measured individually, the simultaneous measurement is less dependent on the modeling of each process. Even though the results are still statistically limited, they can be used to constrain the impacts of possible deviations from the SM predictions across the different processes.

The signal cross sections are also measured differentially as a function of five observables describing the kinematic properties of the leptons in the final states: the unfolded distributions of the tZq cross section show a good agreement with the predictions from the simulated samples, while the $\mathrm{t\bar{t}Z}$ and tWZ sum tends to be slightly higher then predicted, as for the inclusive measurement. The trend in the p_{T} of the lepton arising from the W boson decay is reminiscent of that observed in inclusive $\mathrm{t\bar{t}}$ production, as predictions are underestimated at low p_{T} values.

The limitations arising from statistical uncertainties are expected to be significantly reduced already with the full Run-3 data set, and even more during HL-LHC. The increase in statistics will also bring the possibility to extend the simultaneous measurement to more processes, like the associated production of top quarks with a W boson or a photon.

APPENDIX A

Response matrices

The response matrices for the remaining unfolding observables are shown in Figures A.1 to A.3.

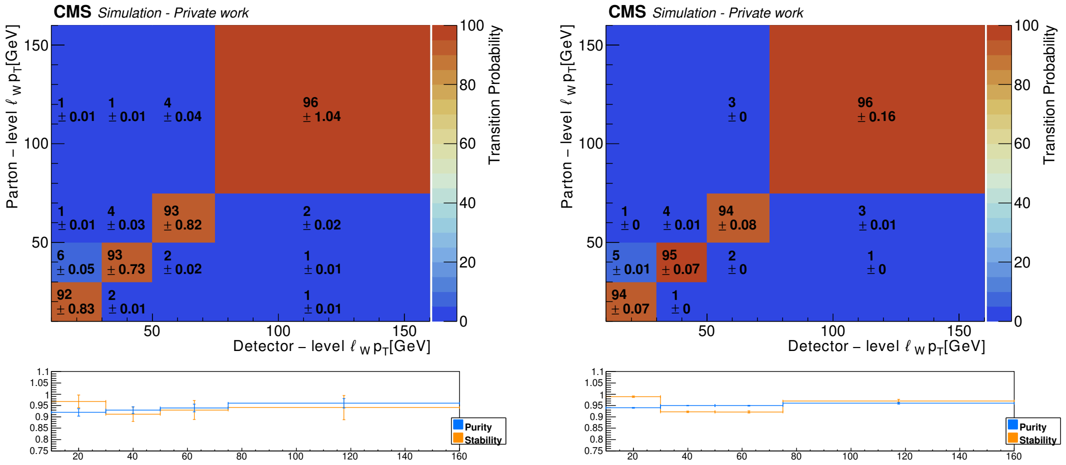


Figure A.1: Response matrices for the sum of the $t\bar{t}Z$ and tWZ (left) and tZq (right) samples. Plots refer to $p_T(\ell_W)$.

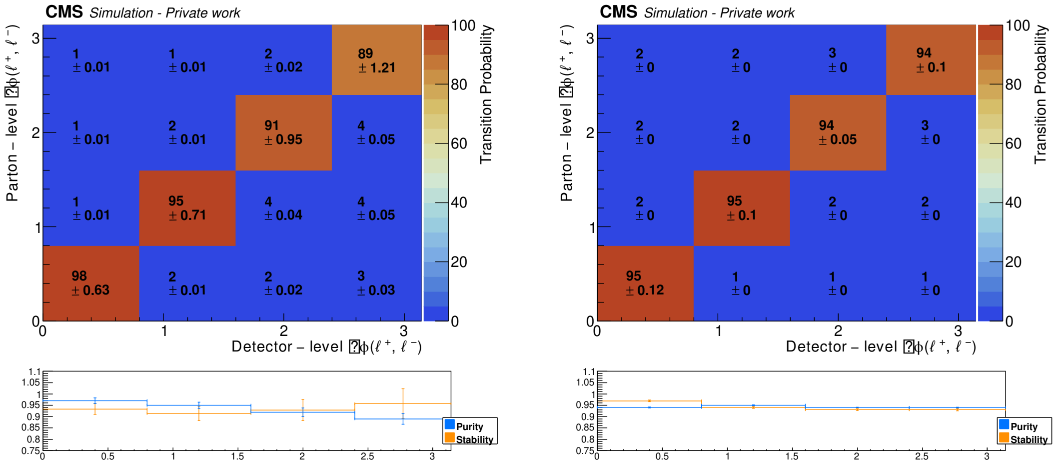


Figure A.2: Response matrices for the sum of the $t\bar{t}Z$ and tWZ (left) and tZq (right) samples. Plots refer to $\Delta\phi(\ell^+, \ell^-)$.

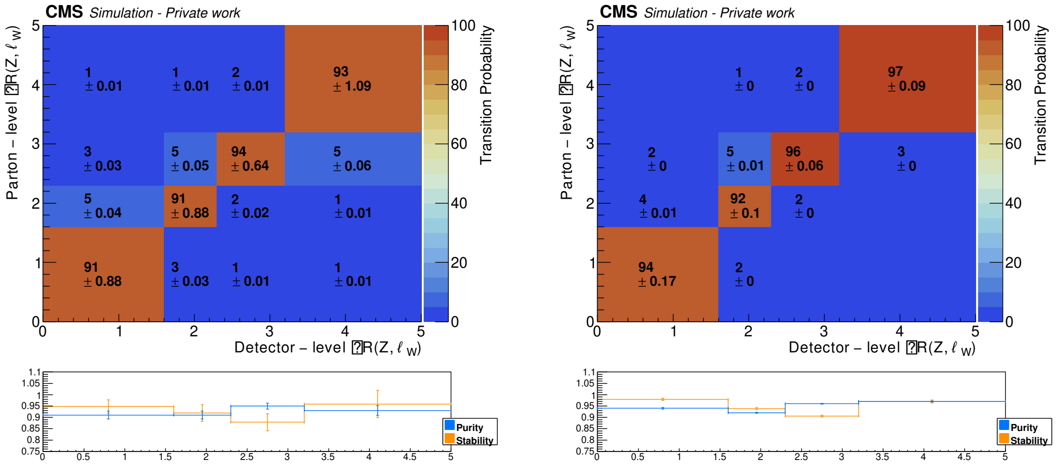
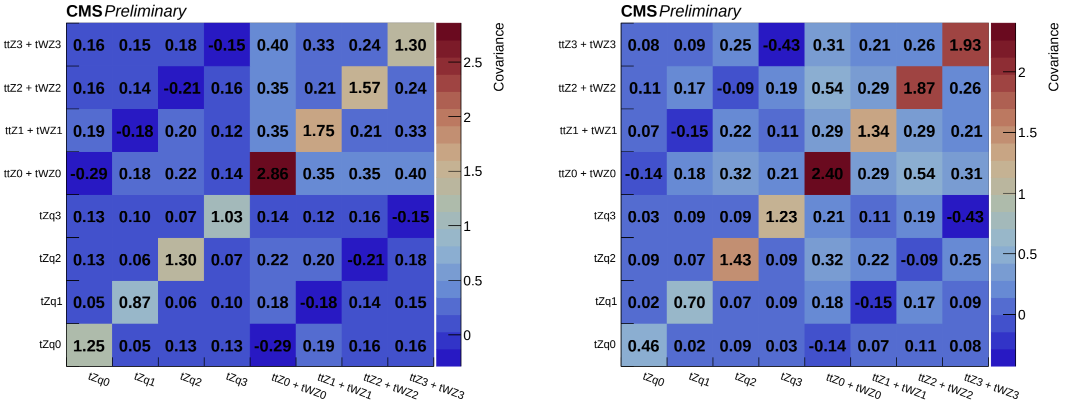


Figure A.3: Response matrices for the sum of the $t\bar{t}Z$ and tWZ (left) and tZq (right) samples. Plots refer to $\Delta R(Z, \ell_W)$.

APPENDIX B

Covariance matrices

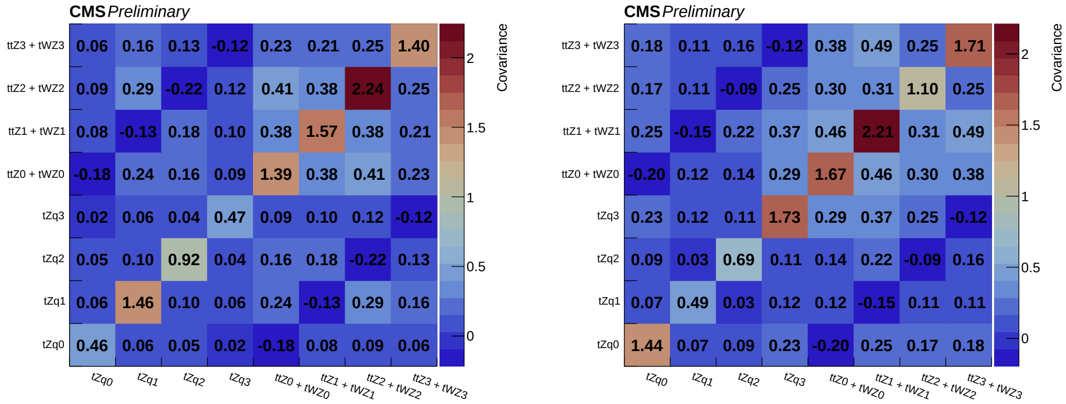
The remaining covariance matrices resulting from the simultaneous differential fits are shown Figures B.1 and B.2.



	tZq0	tZq1	tZq2	tZq3	ttZ0 + tWZ0	ttZ1 + tWZ1	ttZ2 + tWZ2	ttZ3 + tWZ3
ttZ3 + tWZ3	0.16	0.15	0.18	-0.15	0.40	0.33	0.24	1.30
ttZ2 + tWZ2	0.16	0.14	-0.21	0.16	0.35	0.21	1.57	0.24
ttZ1 + tWZ1	0.19	-0.18	0.20	0.12	0.35	1.75	0.21	0.33
ttZ0 + tWZ0	-0.29	0.18	0.22	0.14	2.86	0.35	0.35	0.40
tZq3	0.13	0.10	0.07	1.03	0.14	0.12	0.16	-0.15
tZq2	0.13	0.06	1.30	0.07	0.22	0.20	-0.21	0.18
tZq1	0.05	0.87	0.06	0.10	0.18	-0.18	0.14	0.15
tZq0	1.25	0.05	0.13	0.13	-0.29	0.19	0.16	0.16

	tZq0	tZq1	tZq2	tZq3	ttZ0 + tWZ0	ttZ1 + tWZ1	ttZ2 + tWZ2	ttZ3 + tWZ3
ttZ3 + tWZ3	0.08	0.09	0.25	-0.43	0.31	0.21	0.26	1.93
ttZ2 + tWZ2	0.11	0.17	-0.09	0.19	0.54	0.29	1.87	0.26
ttZ1 + tWZ1	0.07	-0.15	0.22	0.11	0.29	1.34	0.29	0.21
ttZ0 + tWZ0	-0.14	0.18	0.32	0.21	2.40	0.29	0.54	0.31
tZq3	0.03	0.09	0.09	1.23	0.21	0.11	0.19	-0.43
tZq2	0.09	0.07	1.43	0.09	0.32	0.22	-0.09	0.25
tZq1	0.02	0.70	0.07	0.09	0.18	-0.15	0.17	0.09
tZq0	0.46	0.02	0.09	0.03	-0.14	0.07	0.11	0.08

Figure B.1: Covariance matrices for the simultaneous fit. The plot refers to the case where the measurement is performed as a function of ℓ_{W} p_{T} (left) and $\Delta R(Z, \ell_{\mathrm{W}})$ (right).



	tZq0	tZq1	tZq2	tZq3	ttZ0 + tWZ0	ttZ1 + tWZ1	ttZ2 + tWZ2	ttZ3 + tWZ3
ttZ3 + tWZ3	0.06	0.16	0.13	-0.12	0.23	0.21	0.25	1.40
ttZ2 + tWZ2	0.09	0.29	-0.22	0.12	0.41	0.38	2.24	0.25
ttZ1 + tWZ1	0.08	-0.13	0.18	0.10	0.38	1.57	0.38	0.21
ttZ0 + tWZ0	-0.18	0.24	0.16	0.09	1.39	0.38	0.41	0.23
tZq3	0.02	0.06	0.04	0.47	0.09	0.10	0.12	-0.12
tZq2	0.05	0.10	0.92	0.04	0.16	0.18	-0.22	0.13
tZq1	0.06	1.46	0.10	0.06	0.24	-0.13	0.29	0.16
tZq0	0.46	0.06	0.05	0.02	-0.18	0.08	0.09	0.06

	tZq0	tZq1	tZq2	tZq3	ttZ0 + tWZ0	ttZ1 + tWZ1	ttZ2 + tWZ2	ttZ3 + tWZ3
ttZ3 + tWZ3	0.18	0.11	0.16	-0.12	0.38	0.49	0.25	1.71
ttZ2 + tWZ2	0.17	0.11	-0.09	0.25	0.30	0.31	1.10	0.25
ttZ1 + tWZ1	0.25	-0.15	0.22	0.37	0.46	2.21	0.31	0.49
ttZ0 + tWZ0	-0.20	0.12	0.14	0.29	1.67	0.46	0.30	0.38
tZq3	0.23	0.12	0.11	1.73	0.29	0.37	0.25	-0.12
tZq2	0.09	0.03	0.69	0.11	0.14	0.22	-0.09	0.16
tZq1	0.07	0.49	0.03	0.12	0.12	-0.15	0.11	0.11
tZq0	1.44	0.07	0.09	0.23	-0.20	0.25	0.17	0.18

Figure B.2: Covariance matrices for the simultaneous fit. The plot refers to the case where the measurement is performed as a function of $\Delta\phi(\ell^+, \ell^-)$ (left) and $\cos(\theta^*_\mathrm{Z})$ (right).

Bibliography

[1] CMS Collaboration, "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV." *JHEP* **03** (2020) 056, [`arXiv:1907.11270`].

[2] 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." *JHEP* **02** (2022) 107, [`arXiv:2111.02860`].

[3] CMS Collaboration, "Evidence for tWZ production in proton-proton collisions at $\sqrt{s} = 13$ TeV in multilepton final states." `arXiv:2312.11668`.

[4] R. P. Feynman, "Space-time approach to quantum electrodynamics." *Phys. Rev.* **76** (1949) 769–789.

[5] M. Thomson, *Modern Particle Physics.* Cambridge University Press (2013).

[6] A. Bettini, *Introduction to Elementary Particle Physics.* Cambridge University Press (2012).

[7] S. Mitra, "Exponential Resummation of QCD at finite chemical potential.".

[8] H. Burkhardt and J. Steinberger, "Tests of the electroweak theory at the Z resonance." *Annual Review of Nuclear and Particle Science* **41** (1991), no. 1 55–96, [`https://doi.org/10.1146/annurev.ns.41.120191.000415`].

[9] Particle Data Group Collaboration, "Review of particle physics." *Phys. Rev. D* **98** (Aug, 2018) 030001.

[10] R. Gupta, "Introduction to Lattice QCD." (1998).

[11] L. Landau, "On the conservation laws for weak interactions." *Nuclear Physics* **3** (1957), no. 1 127–131.

[12] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory.* West-view Press (1995).

[13] J. Schwichtenberg, *Physics from symmetry.* Springer (2018).

[14] J. Ellis et al., "A Historical Profile of the Higgs Boson." `arXiv:1504.07217`.

[15] S. Weinberg, "A Model of Leptons." *Phys. Rev. Lett.* **19** (1957) 1264–1266.

[16] UA1 Collaboration, "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$=540 GeV." *Physics Letters B* **122** (1983), no. 1 103–116.

[17] UA2 Collaboration, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider." *Physics Letters B* **122** (1983), no. 5 476–485.

[18] UA1 Collaboration, "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider." *Physics Letters B* **126** (1983), no. 5 398–410.

[19] UA2 Collaboration, "Evidence for $Z^0 \to e^+e^-$ at the CERN pp collider." *Physics Letters B* **129** (1983), no. 1 130–140.

[20] CDF Collaboration, "Observation of Top Quark Production in $\bar{p}-p$ collisions." *Physical Review Letters* **74** (apr, 1995) 2626–2631.

[21] D0 Collaboration, "Observation of the top quark." *Physical Review Letters* **74** (apr, 1995) 2632–2637.

[22] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." *Physics Letters B* **716** (Sept., 2012) 1–29.

[23] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.” *Physics Letters B* **716** (Sept., 2012) 30–61.

[24] NNPDF Collaboration, “Parton distributions from high-precision collider data.” *Eur. Phys. J. C* **77** (2017) 663.

[25] Y. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics..” *JETP* **46** (1977).

[26] V. Gribov and L. Lipatov, “Deep inelastic electron scattering in perturbation theory.” *Physics Letters B* **37** (1971), no. 1 78–80.

[27] G. Altarelli and G. Parisi, “Asymptotic freedom in parton language.” *Nuclear Physics B* **126** (1977), no. 2 298–318.

[28] M. Kobayashi and T. Maskawa, “CP-Violation in the Renormalizable Theory of Weak Interaction.” *Progress of Theoretical Physics* **49** (02, 1973) 652–657.

[29] CMS Collaboration, “Measurement of the Top Quark Pair Production Cross Section in Proton-Proton Collisions at $\sqrt{s}$ =13 TeV.” *Physical Review Letters* **116** (Feb., 2016).

[30] ATLAS Collaboration, “Measurement of the $\mathrm{t\bar{t}}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector.” *Physics Letters B* **772** (Sept., 2017) 879.

[31] CMS Collaboration, “Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton–proton collisions at $\sqrt{s}$ =13 TeV.” *The European Physical Journal C* **80** (May, 2020).

[32] ATLAS Collaboration, “Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector.” *Journal of High Energy Physics* **2017** (Apr., 2017).

[33] CMS Collaboration, "CMS Standard Model summary plots ." `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined`. Accessed on January 11th, 2024.

[34] ATLAS Collaboration, "Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a Z boson at $\sqrt{s} = 13$ TeV with the ATLAS detector." *Eur. Phys. J. C* **81** (2021), no. 8 737, [`arXiv:2103.12603`].

[35] ATLAS Collaboration, "Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." *JHEP* **07** (2020) 124, [`arXiv:2002.07546`].

[36] L. M. F. Maltoni and K. Mimasu, "Top-quark electroweak interactions at high energy." *J. High Energ. Phys.* **10** (2019) 004.

[37] H. E. Faham et al., "Single top production in association with a WZ pair at the LHC in the SMEFT." *J. High Energ. Phys.* **01** (2022) 100.

[38] S. Frixione et al., "Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons." (2015).

[39] D. Pagani et al., "NLO QCD+EW predictions for tHj and tZj production at the LHC." (2020).

[40] W. Buchmüller and D. Wyler, "Effective lagrangian analysis of new interactions and flavour conservation." *Nucl. Phys. B* **268** (1986) 621.

[41] C. P. Burgess, "Introduction to effective field theory." *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 329, [`hep-th/0701053`].

[42] J. A. A. Saavedra et al., "Interpreting top-quark LHC measurements in the standard-model effective field theory." (2018).

[43] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report.* CERN Yellow Reports: Monographs. CERN, Geneva (2004).

[44] "Facts and figures about the LHC." `https://home.cern/resources/faqs/facts-and-figures-about-lhc`. Accessed on January 11th, 2024.

[45] ALICE Collaboration, "The ALICE experiment at the CERN LHC." *Journal of Instrumentation* **3** (aug, 2008) S08002.

[46] LHCb Collaboration, "The LHCb Detector at the LHC." *Journal of Instrumentation* **3** (aug, 2008) S08005.

[47] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider." *Journal of Instrumentation* **3** (aug, 2008) S08003.

[48] CMS Collaboration, "The CMS Experiment at the CERN LHC." *Journal of Instrumentation* **3** (aug, 2008) S08004.

[49] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment." *Journal of Physics: Conference Series* **513** (June, 2014) 022032.

[50] CMS Collaboration, "CMS slice image view (transverse/longitudinal/3-D)." `https://cds.cern.ch/record/2204863`. (2016).

[51] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker." *Journal of Instrumentation* **9** (Oct., 2014) P10009–P10009.

[52] CMS Collaboration, "Performance and operation of the CMS electromagnetic calorimeter." *Journal of Instrumentation* **5** (Mar., 2010) T03010–T03010.

[53] CMS Collaboration, "Performance of CMS hadron calorimeter timing and synchronization using test beam, cosmic ray, and LHC beam data." *Journal of Instrumentation* **5** (Mar., 2010) T03013–T03013.

[54] "Muon drift tubes." `https://cms.cern/index.php/detector/detecting-muons/muon-drift-tubes`. Accessed on 11th June, 2024.

[55] N. Metropolis and S. Ulam, "The Monte Carlo Method." *Journal of the American Statistical Association* **44** (1949), no. 247 335–341. PMID: 18139350.

[56] 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." *JHEP* **07** (2014) 079, [`arXiv:1405.0301`].

[57] T. Sjöstrand et al., "An introduction to PYTHIA 8.2." *Comput. Phys. Commun.* **191** (2015) 159, [`arXiv:1410.3012`].

[58] C. Bierlich et al., "A comprehensive guide to the physics and usage of PYTHIA 8.3." (2022).

[59] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO." *Journal of High Energy Physics* **2012** (Dec., 2012).

[60] S. Agostinelli, J. Allison, et al., "Geant4—a simulation toolkit." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3 250–303.

[61] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker." *Journal of Instrumentation* **9** (Oct., 2014) P10009–P10009.

[62] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector." *Journal of Instrumentation* **12** (Oct., 2017) P10003–P10003.

[63] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm." *Journal of High Energy Physics* **2008** (Apr., 2008) 063–063.

[64] G. P. Salam, "Towards jetography." *The European Physical Journal C* **67** (May, 2010) 637–686.

[65] S. Sapeta, "QCD and Jets at Hadron Colliders." *Prog. Part. Nucl. Phys.* **89** (2016) 1–55, [`arXiv:1511.09336`]. 77 pages, 37 figures; v2: text developments in several places, references added, version published in Prog. Part. Nucl. Phys. 2016.

[66] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, "Jet flavour classification using DeepJet." *Journal of Instrumentation* **15** (dec, 2020) P12012.

[67] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV.” *Journal of Instrumentation* **13** (may, 2018) P05011–P05011.

[68] CMS Collaboration, “Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV.” *Journal of High Energy Physics* **2011** (Jan., 2011).

[69] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC.” *Journal of Instrumentation* **16** (May, 2021) P05014.

[70] “Efficiency studies using the tag and probe method.” `https://cms-opendata-workshop.github.io/workshop-lesson-tagandprobe/aio/index.html`. Accessed on 30th June, 2024.

[71] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV.” *Journal of Instrumentation* **13** (June, 2018) P06015–P06015.

[72] CMS Collaboration, “Muon identification using multivariate techniques in the CMS experiment in proton-proton collisions at $\sqrt{s} = 13$ TeV.” (2023).

[73] A. Frankenthal, “Computing tag-and-probe efficiencies with Apache Spark and Apache Parquet.”.

[74] P. Virtanen et al., “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” *Nature Methods* **17** (2020) 261–272.

[75] C. Barber, D. Dobkin, and H. Huhdanpaa, “Qhull: Quickhull algorithm for computing the convex hull.” *Astrophysics Source Code Library* (04, 2013) 04016–.

[76] CMS Collaboration, “Observation of four top quark production in proton-proton collisions at $\sqrt{s}$=13 TeV.” *Physics Letters B* **847** (2023) 138290.

[77] S. Frixione et al., “Automated simulations beyond the standard model: supersymmetry.” *Eur. Phys. J. C* **12** (2019) 008, [`arXiv:1907.04898`].

[78] F. Demartin et al., “tWH associated production at the LHC.” *Eur. Phys. J. C* **77** (2017) 34, [`arXiv:1607.05862`].

[79] S. Frixione et al., “Matching NLO QCD computations with parton shower simulations: the POWHEG method.” *JHEP* **11** (2007) 070, [`arXiv:0709.2092`].

[80] CMS Collaboration, “Simulation of the Silicon Strip Tracker pre-amplifier in early 2016 data.”.

[81] CMS Collaboration, “Performance of the pile up jet identification in CMS for Run 2.”.

[82] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV.” *Journal of Instrumentation* **12** (feb, 2017) P02014–P02014.

[83] CMS Collaboration, “Measurement of the t-channel single top quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV.” *Phys. Rev. Lett.* **107** (2011) 091802, [`arXiv:1106.3052`].

[84] T. Chwalek, *Messung der W-Boson-Helizitätsanteile in Top-Quark-Zerfällen mit dem CDF II Experiment und Studien zu einer frühen Messung des $t\bar{t}$ -Wirkungsquerschnitts mit dem CMS Experiment*, Ph.D. thesis, Karlsruhe U. (2010).

[85] F. Chollet et al., “Keras.” (2015).

[86] M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.” (2016).

[87] Z. Ahmad et al., “Anomaly Detection Using Deep Neural Network for IoT Architecture.” *Applied Sciences* **11** (2021), no. 15.

[88] S. Ruder, “An overview of gradient descent optimization algorithms.” (2017).

[89] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization.” (2017).

[90] S. Basodi, C. Ji, H. Zhang, and Y. Pan, "Gradient amplification: An efficient way to train deep neural networks." (2020).

[91] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks." (2017).

[92] C. Cortes, M. Mohri, and A. Rostamizadeh, "L2 regularization for learning kernels." (2012).

[93] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements." *The European Physical Journal C* **80** (Jan., 2020).

[94] CMS Collaboration, "CMS PYTHIA 8 colour reconnection tunes based on underlying-event data." *Eur. Phys. J. C* **83** (2023), no. 7 587, [`arXiv:2205.02905`].

[95] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples." *Computer Physics Communications* **77** (1993), no. 2 219–228.

[96] C. Degrande et al., "Single-top associated production with a Z or H boson at the LHC: the SMEFT interpretation." *Journal of High Energy Physics* **2018** (Oct., 2018).

[97] Q.-H. Cao et al., "A general analysis of Wtb anomalous couplings." *Chinese Physics C* **41** (June, 2017) 063101.

[98] A. Höcker and V. Kartvelishvili, "SVD approach to data unfolding." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **372** (Apr., 1996) 469–481.

[99] CMS Collaboration, "Differential cross section measurements for the production of top quark pairs and of additional jets using dilepton events from pp collisions at $\sqrt{s} = 13$ TeV." (2024).

[100] O. Aberle et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report.* CERN Yellow Reports: Monographs. CERN, Geneva (2020).

[101] R. K. Barman and A. Ismail, “Constraining the top electroweak sector of the SMEFT through Z associated top pair and single top production at the HL-LHC.” (2022).

Acknowledgements

I begin by expressing my biggest thanks to Andreas Meyer, for giving me the possibility to work at DESY and for his guidance throughout the entire duration of my PhD: this work wouldn't have been possible without his support and patience. I extend my gratitude to Johannes Haller, for his valuable comments and input shared during our meetings, which contributed to the development of my research.

At DESY, the entire CMS Top Group has always provided support and created a friendly atmosphere, which significantly enriched my experience. A special thanks goes to Nadjieh Jafari, for contributing to my research since my master's thesis, and to Maria Aldaya, for her humanity and support throughout various projects. Working in an international environment alongside many brilliant PhD students and postdocs has not only improved the quality of my research, but has also been a source of personal growth for me. Since I joined DESY for my master's thesis, David Walter has been giving me constant support by answering all my endless questions and providing new ideas, and I am sincerely thankful for his help. I would also like to thank Alessia Saggio, who not only provided assistance with my analysis but also constantly showed an admirable strength in addressing challenges and has been an example for me to learn from.

Thank you to Soham Bhattacharya for helping me solve my doubts in physics and for his punctuality in proposing tea breaks during our working days. In addition, I want to thank my office mates Konstantin Sharko and Iakov Andreev for joining me in the effort of keeping our office plants alive, sometimes even succeeding. I also would like to express special thanks to my friends Beatriz Ribeiro Lopes and Ana Ventura Barroso for their support throughout this journey, starting from the challenging times of the pandemic: I was lucky to have you by my side, thank you for the many meaningful moments we shared and for your patience, especially when waiting for me during my occasional delays.

I could not have undertaken this journey without the encouragement of my close friends and family. I am profoundly grateful to all of them, particularly to my par-

ents, Gabriella and Giuseppe, whose unconditional love and understanding gave me the strength I needed to overcome challenges and reach my goals.

And last but not least, I would like to express my warmest thanks to my friend and colleague Nicolas Tonon, who has always given me precious help since the day we met. Thank you for being such a positive person and for helping me integrate into the group, from holding a place for me at the canteen to inviting me to your cocktail parties. I am lucky to have met you and you are truly missed.