Probing the Standard Model of particle physics with $t\bar{t}H$, WWZ and WZZ multilepton final states

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Andrea Sciandra
aus
Udine (Italien)

Bonn, 22.11.2018

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn http://nbn-resolving.de/urn:nbn:de:hbz:5n-56399 elektronisch publiziert.

1. Gutachter:	Priv.-Doz. Dr. Markus Cristinziani
2. Gutachter:	Prof. Dr. Klaus Desch
Tag der Promotion:	27.05.2019
Erscheinungsjahr:	2020

Contents

CHAPTER 1

Introduction

The Standard Model of particle physics describes the fundamental constituents of matter and their interactions (except gravity) at an impressively high level of accuracy. This theory successfully predicts the existence of both, the electroweak gauge and the Higgs bosons. The electroweak gauge bosons are expected to be involved in triple and quartic self-interactions, while the existence of the Higgs boson allows bosons and fermions to acquire mass in the electroweak gauge theory.

Despite the fact that huge efforts in several experiments had been made, the Higgs boson remained undiscovered for several decades. In July 2012, the ATLAS and CMS experiments at the Large Hadron Collider (LHC), announced the discovery of a particle compatible with the Standard Model Higgs boson. This discovery opened the way for dedicated measurements of its properties: mass, width, spin-parity, decay branching fractions, production cross sections and couplings with other Standard Model particles. The largest Higgs-boson coupling is the top-quark Yukawa coupling. The predictions for this quantity can be directly probed in the associated production of the Higgs boson with a top-quark pair ($t\bar{t}H$).

The measurement of the production of massive triboson processes, i.e. three massive vector bosons, is sensitive to triple and quartic gauge couplings, as well as to Higgs-mediated processes. Therefore, it constitutes a natural portal connecting the pure electroweak theory to the Higgs-boson sector. Massive triboson production is currently among the processes which are the least precisely measured.

Many open problems, as the observation of neutrino flavour oscillations, the matter-antimatter asymmetry in the universe and the existence of dark matter, are not addressed by the Standard Model. Therefore, the quest for inconsistencies of this theory needs to be carried out in all of its sectors. Any deviations from the Standard Model predictions would constitute evidence for the existence of new physics. We know that some kind of "new physics", beyond the Standard Model predictions, should appear at the Planck scale, where the gravity interactions become important. The top-quark Yukawa coupling plays a crucial role within this context, as it dominates radiative corrections to the Higgs self-coupling, determining the structure of the Higgs-boson effective potential. On the other hand, in absence of obvious findings, like sharp peaks, deviations from the Standard Model predictions can be interpreted in terms of an effective theory, which extends the theory by means of higher-dimension operators. This approach begs for an effort in probing all explored and unexplored Standard Model processes at the highest possible level of accuracy.

An overview of the Standard Model theoretical framework is provided in Chapter 2. Particular focus is put on the important role played by the dynamics of gauge theories and the top-quark Yukawa coupling. Chapter 3 discusses the ATLAS experiment and phenomenology of proton–proton interactions at the LHC. Basic concepts of multivariate analysis (MVA) are provided in Chapter 4, as well as a discussion

of the development of a new MVA-based algorithm devoted to the identification of jets originating from the hadronisation of b-quarks.

The work presented in this dissertation aims at probing the Standard Model predictions in multilepton final states, exploring different sectors and, thus, processes. Chapters 5 and 6 discuss searches in multilepton final states for $t\bar{t}H$ and massive triboson production, respectively. These searches are carried out with events collected by the ATLAS detector at the LHC. Chapter 5 is focused on the latest $t\bar{t}H$ four-lepton (4ℓ) analysis, including 2017 data. This analysis is very similar to the $t\bar{t}H \rightarrow 4\ell$ analysis I developed with the 2015 and 2016 data, integral part of the ATLAS combination which provided observation of $t\bar{t}H$ production. In both, the $t\bar{t}H$ and triboson analyses, multivariate techniques are employed in order to enhance the sensitivity to signal processes, with respect to single kinematic discriminants. For the same purpose, the definition of multiple signal and control regions is exploited. The profile likelihood approach, used to extract and interpret information, is also described and the final results are discussed.

CHAPTER 2

The Standard Model of particle physics

This chapter is devoted to the description of the Standard Model of particle physics: a quantum gauge field theory which describes all fundamental constituents of the matter and their interactions (except gravity). The works presented in this dissertation represent tests of this theory, comparing its predictions to experimental evidence in different sectors: the top-quark Yukawa (Chapter 5) and the gauge (Chapter 6) couplings.

A brief introduction to the Lagrangian formalism, the gauge principle and the gauge theory of electroweak interactions is reported in Sections 2.1 and 2.2, with emphasis on triple and quartic gauge couplings. An outline of the electroweak symmetry breaking is provided in Section 2.3. After a non-extensive overview of Higgs-boson physics at the LHC, presenting the main production mechanisms and decay channels (Section 2.4), the crucial sensitivity of the top-quark Yukawa coupling to new physics is discussed in Section 2.5. Sections 2.6 and 2.7 show how to probe beyond Standard Model physics through anomalous gauge couplings and the state of the art in the production of triboson systems at the LHC.

Throughout this chapter, natural units are used, where the speed of light in vacuum c and the reduced Planck constant $\hbar$ are set to $c = \hbar = 1$.

2.1 Foundations of the Standard Model theoretical framework

The current understanding of elementary particles and their fundamental interactions is condensed in the Standard Model of particle physics, elaborated and developed in the 1960s and 1970s [1–7]. The matter is composed of three families of leptons and three of quarks, which interact through the exchange of force-carrying particles (mediators), referred to as "gauge bosons". Both, leptons and quarks are fermions (half-integer spin particles), thus obeying to Fermi-Dirac statistics; the force mediators, instead, are integer-spin particles and, thus, obey to Bose-Einstein statistics. For each of the fermions[1], there exists a corresponding anti-fermion with the same mass but opposite electric charge. Each lepton family is composed of a charged lepton and a neutrino, while there are six "flavours" of quarks, with fractional electric charge and forming three generations. In total the Standard Model has 24 fermion fields: 18 are quarks, 6 flavours ("down", "up", "strange", "charm", "bottom" and "top") times 3 colours[2] ("red", "green" and "blue"), and 6 are leptons, 3 charged (electrons, muons and tauons) and the corresponding

[1] Except, possibly, the neutrinos.
[2] The strong-interaction charge, see Section 2.1.3.

Table 2.1: Fundamental properties of the six leptons and six quark flavours, forming three generations of leptons and quarks, respectively; the electric charge Q, in units of the electron charge e, and the mass (or mass limit), in GeV, are reported. The uncertainty on the mass of charged leptons is omitted, as it is several orders of magnitude smaller than the precision adopted in the table [8].

Generation	lepton/quark	Q/e	mass [GeV]
First	electron (e)	-1	0.511×10^{-3}
	e neutrino (ν_e)	0	$< 2 \times 10^{-9}$
Second	muon (μ)	-1	0.106
	μ neutrino (ν_μ)	0	$< 0.19 \times 10^{-3}$
Third	tauon (τ)	-1	1.777
	τ neutrino (ν_τ)	0	$< 18.2 \times 10^{-3}$
First	up (u)	$+\frac{2}{3}$	$2.2^{+0.6}_{-0.4} \times 10^{-3}$
	down (d)	$-\frac{1}{3}$	$4.7^{+0.5}_{-0.4} \times 10^{-3}$
Second	charm (c)	$+\frac{2}{3}$	1.27 ± 0.03
	strange (s)	$-\frac{1}{3}$	$96^{+8}_{-4} \times 10^{-3}$
Third	top (t)	$+\frac{2}{3}$	173.2 ± 0.9
	bottom (b)	$-\frac{1}{3}$	$4.18^{+0.04}_{-0.03}$

neutrinos. Table 2.1 shows the three lepton and quark generations, along with their charge and mass (or mass limit).

The gauge bosons are the mediators of all fundamental interactions known in nature: the electromagnetic, the weak, the strong and the gravitational. The electromagnetic force is carried by a spin-1 massless vector boson, the photon (γ), and acts between electrically-charged particles. The weak interactions is approximately 10^3 times weaker than the electromagnetic force and is ruled by three gauge vector bosons, the $W^\pm$ and Z bosons, which are massive with spin 1; the weak force is the interaction responsible for phenomena like the absorption and emission of neutrinos, and nuclear β-decays. The strong interaction, responsible of holding together nuclei, is roughly 100 times stronger than the electromagnetic force. Its gauge bosons acting between quarks are eight massless, spin-1 particles called gluons (g). Finally, the gravitational interaction appears between all types of massive particles and is by far the weakest (the gravitational coupling constant is about 10^{42} times smaller than the electromagnetic one); therefore, its

Table 2.2: Fundamental properties of the six gauge bosons mediating the four fundamental forces [8].

Boson	Q/e	mass [GeV]	spin	interaction
photon (γ)	0	$< 10^{-27}$	1	electromagnetic
W boson (W)	± 1	80.385 ± 0.015	1	weak
Z boson (Z)	0	91.1876 ± 0.0023	1	weak
gluon (g)	0	$\lesssim 10^{-3}$	1	strong
graviton (G)	0	$< 6 \times 10^{-41}$	2	gravitational

effect is negligible on a microscopic scale. The graviton (G) is postulated to be the hypothetical spin-2 massless gauge boson carrying the gravitational force. Table 2.2 shows all of the gauge bosons, along with their charge, mass, spin and the respective interaction.

An extensive effort has been spent in the last decades in order to unify the four fundamental interactions, by expressing them as different manifestations of a single fundamental interaction field. This unification is partially achieved by the Standard Model. The Standard Model of particle physics is a renormalisable quantum field theory describing the electromagnetic, weak and strong interactions and defined by the local $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge symmetry. $\mathrm{SU}(3)_C$ represents the non-abelian gauge group, with 8 massless gauge bosons (gluons), which are generators of the gauge group and hold quarks together mediating the strong force; the letter "C" in $\mathrm{SU}(3)_C$ stands for colour. $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ indicates the electroweak symmetry group, which unifies electromagnetic and weak interactions in the so-called "electroweak theory". The "L" in $\mathrm{SU}(2)_L$ stands for "left", as it only involves left-handed fermion fields. On the other hand, the $\mathrm{U}(1)_Y$ symmetry group involves both right- and left-handed fermion fields. The $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ components of the electroweak theory are generated by the weak isospin (T) and the weak hypercharge (Y), respectively. These two quantum numbers are related to the electric charge (Q) through the analogous of the Gell-Mann-Nishijima formula[3]: $Q = T_3 + \frac{Y}{2}$, where T_3 is the third component of the weak isospin.

As it is well known, symmetries have always played a fundamental role in the development of physics. Noether's theorem [9] implies that, if an action is invariant under some group of transformations (symmetry), there exist one or several conserved quantities, called "constants of motion", which are associated to these transformations. This theorem establishes that symmetries directly imply conservation laws; as a matter of fact, the invariance under local gauge transformations of a group implies the dynamics itself, as evident in quantum electrodynamics: the existence and some of the properties of the gauge field (the photon) naturally follow from the gauge invariance under U(1). Local symmetry transformations govern all quantum field theories, quantum electrodynamics, quantum chromodynamics and the Standard Model of electroweak interactions. These theories describe physical phenomena extremely well within the experimentally accessible accuracy.

[3] The Gell-Mann-Nishijima formula relates the baryon number B, the strangeness S and the isospin I_3 of quarks and hadrons to their electric charge: $Q = I_3 + \frac{1}{2}(B + S)$.

2.1.1 Lagrangian formalism and quantum field theory

Quantum Field Theory (QFT) relies on the Lagrangian formalism, inherited from classical mechanics (where it is exploited to describe systems with a finite number of degrees of freedom), and extends quantum mechanics from single localised particles to fields. As opposed to quantum mechanics, where the state of a system is described by a wave function ψ, in quantum field theory particles are defined as excitations of a local field $\phi(x)$. A classical field is a continuous function defined at every point in space-time (for instance, the amplitude of a mechanical wave or an electromagnetic wave are fields).

The Lagrangian density $\mathcal{L}$, in classical mechanics, describes the interactions and the properties of the field $\phi(x)$:

$$\mathcal{L}(x) = \mathcal{L}\left(\phi, \partial_\mu \phi\right), \tag{2.1}$$

as a function of the field itself and its space-time derivatives $\partial_\mu \phi$. The action S is then written as an integral in time and space dimensions of the Lagrangian density:

$$S = \int \mathrm{d}^4 x \; \mathcal{L}\left(\phi, \partial_\mu \phi\right). \tag{2.2}$$

According to the principle of the least action, the system evolves along a path in space-time resulting in a stationary action:

$$\delta S = \delta \left(\int \mathrm{d}^4 x \; \mathcal{L}\left(\phi, \partial_\mu \phi\right) \right) = 0. \tag{2.3}$$

This is equivalent to:

$$\delta S = \int \mathrm{d}^4 x \left[\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi\right)} \right] \delta\phi = 0, \tag{2.4}$$

as, for vanishing field variations at the boundaries, the remaining term of δS is zero; given that the functional form of the field variation $\delta\phi$ is arbitrary, Eq. 2.3 implies the Euler-Lagrange differential equations:

$$\frac{\partial \mathcal{L}}{\partial \phi} = \partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi\right)}. \tag{2.5}$$

All continuous transformations not affecting the equations of motion, defined in Eq. 2.5, constitute the symmetry group of the system and define a *gauge symmetry*. For an infinitesimal symmetry transformation of the field $\phi(x) \to \phi(x) + \delta\phi(x)$ the derivative, consequently, varies as $\partial_\mu \phi(x) \to \partial_\mu \phi(x) + \partial_\mu \delta\phi(x)$. Since under a symmetry transformation the equations of motion remain unchanged, the Lagrangian density can only change up to a total derivative $\partial_\mu X^\mu$. The change in the Lagrangian density is:

$$\delta \mathcal{L} = \partial_\mu X^\mu = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi\right)} \delta\phi \right), \tag{2.6}$$

therefore the current $J^\mu = \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi\right)} \delta\phi - X^\mu$ is conserved, i.e. $\partial_\mu J^\mu = 0$. This proves Noether's theorem [9] for Lagrangian systems which exhibit symmetries. This can be easily generalised to multiple fields $\phi^{(i)}(x)$ and infinitesimal symmetry transformations $\phi^{(i)}(x) \to \phi^{(i)}(x) + \delta\phi^{(i)}(x)$, where the currents

$$J^\mu = \sum_i \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi^{(i)}\right)} \delta\phi^{(i)} - X^\mu \tag{2.7}$$

are conserved ($\partial_\mu J^\mu = 0$). The conserved currents J^μ are as many as the number of independent generators for the transformations $\delta\phi^{(i)}$. To each conserved current, there is a corresponding conserved charge $Q = \int \mathrm{d}^3\vec{x}\, J^0$, i.e. a physical quantity which maintains the same value at all times $\left(\frac{\mathrm{d}Q}{\mathrm{d}t} = 0\right)$, where J^0 is the time component of J^μ.

2.1.2 Quantum electrodynamics

One of the major successes of QFT is the precise description of the interaction between matter and electromagnetic radiation, interaction whose rules are depicted by the theory of Quantum Electrodynamics (QED). This theory relies on very simple principles: *symmetry* and *locality*.

The Lagrangian density of a free electron field $\psi(x)$ of mass m can be written as:

$$\mathcal{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi, \tag{2.8}$$

where $\bar{\psi} = \psi^\dagger\gamma^0$, $\partial\!\!\!/ = \gamma^\mu\partial_\mu$, γ^μ are the 4×4 Dirac matrices satisfying the anti-commutation relation $\{\gamma^\mu, \gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$ and $g^{\mu\nu}$ is the metric tensor. In fact, the Euler-Lagrange equation (Eq. 2.5) for this Lagrangian represents the Dirac equation $(i\partial\!\!\!/ - m)\psi = 0$, the relativistic wave equation describing all spin-$\frac{1}{2}$ massive particles.

The Lagrangian in Eq. 2.8 is invariant under a U(1) global gauge transformation

$$\psi(x) \to \psi'(x) = U\psi(x) = \exp(i\alpha)\psi, \tag{2.9}$$

where α is "global", i.e. it is chosen to be the same at every point in space-time $\left(\frac{\partial\alpha}{\partial x} = 0\right)$. On the other hand, as soon as a phase dependence on the space-time point is considered, the Lagrangian in Eq. 2.8 is not invariant, since the transformation in Eq. 2.9 with $\alpha = \alpha(x)$ implies:

$$\mathcal{L} \to \mathcal{L}' = \mathcal{L} + \bar{\psi}\gamma_\mu\psi\,(\partial^\mu\alpha)\,. \tag{2.10}$$

This is a consequence of the fact that the space-time derivative does not transform simply under the local U(1) transformation; in order to ensure a *local* gauge invariance, it is necessary to define the *gauge field* A_μ such that it transforms as follows:

$$A_\mu \to A'_\mu = A_\mu + \frac{1}{e}\partial_\mu\alpha \tag{2.11}$$

and the *covariant derivative* through the minimal coupling e:

$$D_\mu \equiv \partial_\mu - ieA_\mu, \tag{2.12}$$

where the $-ie$ factor is conventional. Replacing in Eq. 2.8 ∂_μ with the covariant derivative, we obtain a new Lagrangian which is symmetric locally:

$$\mathcal{L} = \bar{\psi}(iD\!\!\!/ - m)\psi. \tag{2.13}$$

The coupling between a fermion field ψ (e.g. the electron) and the gauge field A_μ (photon) arises naturally when the invariance under local gauge transformations of the kinetic-energy term in the free-fermion Lagrangian is required. Since the electromagnetic strength tensor $F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu$ is invariant under the gauge transformation in Eq. 2.11, so is the Lagrangian of the free gauge field $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$. Having invariant terms for a Lagrangian with a fermion field ψ and the gauge field A_μ, the classical QED

Lagrangian reads:

$$\mathcal{L} = \bar{\psi}(i\not{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.14}$$

It is important to remark that a hypothetical mass term for the gauge field, $-\frac{1}{2}A_\mu A^\mu$, is not invariant under the transformation reported in Eq. 2.11, therefore it cannot be introduced, except at the expense of the local gauge invariance.

2.1.3 Quantum chromodynamics

Quantum Chromodynamics (QCD) is a non-abelian theory which describes the behaviour of quarks and mediators of the strong interaction, the gluons. In 1932 Heisenberg [10] suggested that, under nuclear interactions, protons and neutrons can be regarded as degenerate states, since their mass is similar and the electromagnetic interaction is negligible[4]. This means that any arbitrary combination of their wave functions would be equivalent,

$$\psi \equiv \begin{pmatrix}\psi_p \\ \psi_n\end{pmatrix} \rightarrow \psi' = U\psi, \tag{2.15}$$

where U is a unitary transformation. In addition, if $\det|U| = 1$, U represents the Lie group of SU(2):

$$U = \exp\left(-i\frac{\tau^a}{2}\alpha^a\right) \simeq 1 - i\frac{\tau^a}{2}\alpha^a, \tag{2.16}$$

where τ^a, $a = 1, 2, 3$ are the Pauli matrices. Following Yang and Mill's argument, "*once one chooses what to call a proton, what a neutron, at one space-time point, one is then not free to make any choices at other space-time points*" [11], the freedom of choosing what is proton and what is neutron should be preserved independently of where or when we are. This *local gauge isotopic invariance* can be implemented by requiring that the gauge parameters depend on the space-time points, i.e. $\alpha^a \rightarrow \alpha^a(x)$.

Utiyama generalised this idea [12] for any non-abelian group G with generators t_a satisfying the Lie algebra,

$$[t_a, t_b] = if_{abc}t_c, \tag{2.17}$$

where f_{abc} are called "structure constants" of the group. Like in the QED case, the QCD Lagrangian is required to fulfill gauge invariance, and, in particular, to be invariant under the local SU(3) colour group of transformations

$$q(x) \rightarrow q'(x) = Uq(x) = \exp\left(-iT^a\alpha^a(x)\right), \tag{2.18}$$

where $q(x)$ represents the quark fields and T^a is a convenient representation of the generators t^a. The latter are 8, as the generators of SU(N) are $N^2 - 1$.[5] A convenient representation for the generators t^a is provided by $T^a = \frac{\lambda^a}{2}$, where the λ^a are the eight Gell-Mann matrices, which are hermitian and traceless.

One gauge field per generator is introduced and the covariant derivative is defined by $D_\mu \equiv \partial_\mu - igT^aA^a_\mu$; since the covariant derivative transforms exactly like the quark fields (Eq. 2.18, i.e. $D_\mu q \rightarrow U(D_\mu q)$), in order to ensure the invariance under the local non-abelian gauge transformation for terms containing the fields and their derivatives, it is necessary for the gauge fields to transform as follows:

$$T^aA^a_\mu \rightarrow U\left(T^aA^a_\mu + \frac{i}{g}\partial_\mu\right)U^{-1}. \tag{2.19}$$

[4] The strong interaction constant is roughly 100 more intense than the electromagnetic one.

[5] An arbitrary $N \times N$ complex matrix has $2N^2$ real elements. For a unitary matrix $U^\dagger = U^{-1}$, therefore only N^2 elements are independent. The specialty condition $\det|U| = 1$ adds one more constraint, leaving $N^2 - 1$ independent elements.

The strength tensor can be generalised for a non-abelian Lie group as:

$$F^a_{\mu\nu} \equiv \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f^{abc} A^b_\mu A^c_\nu. \tag{2.20}$$

It is the third non-abelian term in Eq. 2.20, which distinguishes QCD from QED, giving rise to triple and quartic gluon self-interactions and ultimately to the property of asymptotic freedom. The QCD field strength tensor is not gauge invariant, because of the gluon self-interaction: the carriers of the colour force are themselves coloured, unlike the electrically-neutral photon. On the other hand, like in QED, a mass term for the gauge bosons, such as $m^2 A^a_\mu A^\mu_a$, is not gauge invariant. The kinetic term for the gauge bosons, $F^a_{\mu\nu} F^{\mu\nu}_a$, is gauge invariant, therefore the QCD Lagrangian reads:

$$\mathcal{L} = \bar{q}(i\not{D} - m)q - \frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a. \tag{2.21}$$

It is important to note that three- and four-gluon vertices are predicted and arising from the gauge bosons kinetic term; in fact, the field strength is of the form $F_{\mu\nu} \propto (\partial_\mu A_\nu - \partial_\nu A_\mu) + g A_\mu A_\nu$, therefore:

$$F_{\mu\nu} F^{\mu\nu} \propto (\partial_\mu A_\nu - \partial_\nu A_\mu)^2 + g(\partial_\mu A_\nu - \partial_\nu A_\mu) A^\mu A^\nu + g^2 A_\mu A_\nu A^\mu A^\nu, \tag{2.22}$$

where the first term is the gluon propagator and the second and third terms represent the triple and quartic gluon self-couplings, respectively.

2.2 Gauge theory of electroweak interactions

There are general principles to build a gauge theory:

- a gauge group G, corresponding to N_G generators, is chosen;
- N_G vector fields, the gauge bosons, are added to the theory in a specific representation of the chosen gauge group G;
- the fundamental representation for the matter fields, the elementary particles, is chosen;
- scalar fields are added to the theory, in order to "give" mass to some of the gauge bosons (see Section 2.3);
- the most general renormalisable Lagrangian, invariant under G and coupling all the fields and the covariant derivative, is defined;
- the scalar fields are shifted to allow the minimum of the potential to be at zero;
- dedicated QFT techniques are exploited to verify the renormalisability and to make predictions.

After developing a gauge theory, following the aforementioned steps, its predictions need to be probed comparing them with the experimental evidence.

The Glashow-Weinberg-Salam [1–3] model is known, at the present time, as the *Standard Model of Electroweak Interactions*, reflecting its great success. Before the full development of this theory, there were several attempts to build a gauge theory of the electroweak interaction, describing electromagnetic and weak interactions as two manifestations of the same (electroweak) force. In 1957, Schwinger [13] suggested a model based on the O(3) group with a triplet gauge field $\left(V^+, V^-, V^0\right)$, where V^0 was identified with the photon and the charged gauge bosons were associated to weak bosons. The model was

proposed before the universal $V-A$ structure of the weak currents had been established [14–16]. In 1958 Bludman [17] attempted, for the first time, to incorporate the $V-A$ structure in a gauge theory of the weak interactions by means of a symmetry under the SU(2) weak isospin group. Also in this case three gauge bosons were required, but the neutral gauge boson was associated with a new massive vector boson, responsible for weak interactions without the exchange of charge (neutral currents). This process was observed experimentally fifteen years later, with the Gargamelle bubble chamber at the CERN neutrino experiment [18].

Glashow [1] noticed that, in order to accomodate both the electromagnetic and the weak interactions, it is necessary to go beyond the SU(2) isospin structure. He suggested the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ gauge group, where the U(1) was associated to the leptonic hypercharge (Y). The theory requires four gauge bosons: a triplet (W^1, W^2, W^3), associated to the generators of SU(2), and a neutral gauge field (B) related to U(1). The charged weak bosons appear as a linear combination of W^1 and W^2, while the photon and the neutral weak boson Z^0 are both given by a mixture of W^3 and B. In 1964, Salam and Ward [19] proposed a similar model. As already discussed in Sections 2.1.2 and 2.1.3, introducing mass terms for the gauge bosons "by hand" would explicitly break the gauge invariance of the theory. Weinberg [3] and Salam [2] independently proposed to exploit the Higgs mechanism (see Section 2.3) to give mass to the weak bosons and, at the same time, to preserve the gauge invariance, allowing the theory to be renormalisable, a result shown later by 't Hooft [20].

The electroweak (EW) theory describes how electromagnetic processes and the weak charged current are invariant under the weak hypercharge U(1) and the weak isospin SU(2) transformations. As already discussed for QED and QCD, also this theory relies on the gauge invariance principle. Fermion fields are grouped into left-handed and right-handed:

$$\psi_L = \frac{1}{2}(1-\gamma_5)\,\psi, \qquad \psi_R = \frac{1}{2}(1+\gamma_5)\,\psi, \tag{2.23}$$

where $\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ and $\frac{1}{2}(1 \mp \gamma_5)$ are the helicity projectors. Left-handed fields ψ_L are doublets with isospin $I = \frac{1}{2}$, whereas right-handed fields ψ_R are electroweak singlets with isospin $I = 0$. In terms of particles this corresponds to:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$$

$$u_R, d_R, e_R \quad c_R, s_R, \mu_R \quad t_R, b_R, \tau_R.$$

It has to be remarked that fermion mass terms mix right- and left-handed fermion components:

$$\bar{\psi}\psi = \bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R. \tag{2.24}$$

On the other hand, the electromagnetic vector current does not mix those components, i.e.

$$\bar{\psi}\gamma^\mu\psi = \bar{\psi}_R\gamma^\mu\psi_R + \bar{\psi}_L\gamma^\mu\psi_L, \tag{2.25}$$

and the $V-A$ weak current can be expressed in terms of the helicity states, i.e.

$$\bar{\psi}_L\gamma^\mu\psi_L = \frac{1}{2}\bar{\psi}\gamma^\mu(1-\gamma_5)\,\psi, \tag{2.26}$$

showing that only left-handed fermions play a role in weak interactions.

The $U(1)_Y$ group transformations change left-handed doublets and right-handed singlets by a phase factor $e^{i\alpha^a(x)\frac{Y}{2}}$. In addition, the left-handed doublets also transform under SU(2) transformations, as

$$\psi_L \rightarrow e^{i\beta^a(x)\frac{\tau^a}{2}}\psi_L, \tag{2.27}$$

where $a = 1, 2, 3$ and $\frac{\tau^a}{2}$ represents the SU(2) generators. The chosen candidate for the gauge group is $SU(2)_L \otimes U(1)_Y$, therefore, in order for the gauge invariance to apply, one gauge field per generator is introduced: three gauge fields, W^i_μ, associated to $SU(2)_L$ and one gauge field, B_μ, corresponding to $U(1)_Y$.

Defining the strength tensors as

$$\begin{aligned} W^i_{\mu\nu} &\equiv \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk}W^j_\mu W^k_\nu \\ B_{\mu\nu} &\equiv \partial_\mu B_\nu - \partial_\nu B_\mu, \end{aligned} \tag{2.28}$$

the free Lagrangian for the gauge fields can be written in a compact form:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}W^i_{\mu\nu}W^{i\,\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \tag{2.29}$$

Indicating with g and g' the coupling constant associated to $SU(2)_L$ and $U(1)_Y$ respectively, the covariant derivatives $\partial_\mu + i\frac{g}{2}\tau^a W^a_\mu + i\frac{g'}{2}YB_\mu$ and $\partial_\mu + i\frac{g'}{2}YB_\mu$ are introduced to respect the gauge invariance principle, and describe the fermion-gauge interaction:

$$\mathcal{L}_{\text{int}} = -\bar{\psi}_L\gamma^\mu\left(\frac{g}{2}\tau^a W^a_\mu + \frac{g'}{2}YB_\mu\right)\psi_L - \bar{\psi}_R\gamma^\mu\frac{g'}{2}YB_\mu\psi_R. \tag{2.30}$$

The first term in Eq. 2.30 involves W^1_μ and W^2_μ and, for $a = 1, 2$ can be written as:

$$-\frac{g}{2}\bar{\psi}_L\gamma^\mu\begin{pmatrix} 0 & W^1_\mu - iW^2_\mu \\ W^1_\mu + iW^2_\mu & 0 \end{pmatrix}\psi_L, \tag{2.31}$$

which suggests the definition of the charged physical gauge bosons as $W^\pm_\mu \equiv \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right)$. In order to obtain the right combination of fields that couples to the electromagnetic current, a rotation should be applied to the neutral fields, i.e.

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}, \tag{2.32}$$

where θ_W is the Weinberg angle and the following relations with the SU(2) and U(1) coupling constants hold:

$$\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} \qquad \cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}. \tag{2.33}$$

From the remaining neutral part of Eq. 2.30 we get the electromagnetic charge, i.e. $e = g\sin\theta_W = g'\cos\theta_W$, and weak interactions through neutral currents, a very successful prediction of the Standard Model.

The EW theory, outlined up to this point, is characterised by the presence of massless fermions and four massless gauge fields, W^i_μ, B_μ or equivalently $W^\pm_\mu$, Z_μ and A_μ.

2.2.1 Triple and quartic gauge couplings

The free Lagrangian for the gauge boson fields entails triple and quartic couplings among the vector gauge bosons. These self-interactions are a direct result of the non-abelian structure of the EW theory. The interacting term of the gauge field kinematical component of the EW Lagrangian density, see Eq. 2.29, can be written as:

$$\mathcal{L}_{\text{gauge int}} = -\frac{g}{2}\epsilon^{ijk}\left(\partial_\mu W_\nu^i - \partial_\nu W_\mu^i\right) W^{j\mu} W^{k\nu} - \frac{g^2}{4}\epsilon^{ijk}\epsilon^{ilm} W_\mu^j W_\nu^k W^{l\mu} W^{m\nu}. \tag{2.34}$$

The first and second term in Eq. 2.34 correspond to triple (TGC) and quartic (QGC) gauge couplings, respectively. The strength of these interactions is given by the same $SU(2)_L$ coupling g which appears in the fermion-gauge interaction Lagrangian (see Eq. 2.30).

The cubic and quartic self-interactions among the EW gauge bosons from Eq. 2.34 can be expanded [21] in terms of the physical gauge bosons $W_\mu^\pm$, Z_μ and A_μ for both TGC:

$$\begin{aligned}\mathcal{L}_{\text{TGC}} = &+ ie\cot\theta_W\left[\left(\partial_\mu W_\nu^- - \partial_\nu W_\mu^-\right) W^{+\mu} Z^\nu - \left(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+\right) W^{-\mu} Z^\nu\right] \\ &+ ie\cot\theta_W W_\mu^- W_\nu^+ \left(\partial^\mu Z^\nu - \partial^\nu Z^\mu\right) \\ &+ ie\left[\left(\partial_\mu W_\nu^- - \partial_\nu W_\mu^-\right) W^{+\mu} A^\nu - \left(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+\right) W^{-\mu} A^\nu + W_\mu^- W_\nu^+ \left(\partial^\mu A^\nu - \partial^\nu A^\mu\right)\right]\end{aligned} \tag{2.35}$$

and QGC:

$$\begin{aligned}\mathcal{L}_{\text{QGC}} = &- \frac{e^2}{2\sin^2\theta_W}\left[\left(W_\mu^+ W^{-\mu}\right)^2 - W_\mu^+ W^{+\mu} W_\nu^- W^{-\nu}\right] \\ &- e^2\cot^2\theta_W\left[W_\mu^+ W^{-\mu} Z_\nu Z^\nu - W_\mu^+ Z^\mu W_\nu^- Z^\nu\right] \\ &- e^2\cot^2\theta_W\left[2W_\mu^+ W^{-\mu} Z_\nu A^\nu - W_\mu^+ Z^\mu W_\nu^- A^\nu - W_\mu^+ A^\mu W_\nu^- Z^\nu\right] \\ &- e^2\left[W_\mu^+ W^{-\mu} A_\nu A^\nu - W_\mu^+ A^\mu W_\nu^- A^\nu\right].\end{aligned} \tag{2.36}$$

The $SU(2)_L$ algebra does not allow for any fully neutral vertex, i.e. a gauge coupling vertex involving only photons and Z bosons, as shown by Equations 2.35 and 2.36. The presence of Levi-Civita symbols (ϵ^{ijk}) in Equation 2.34, also arising from the non-abelian structure of $SU(2)_L$, forbids the tree-level ZZV coupling, where $V = Z, \gamma$; in general, at least a pair of charged W bosons is always present in both, TGC and QGC. The Feynman diagrams representing the predicted self-couplings of the EW gauge bosons are shown in Figure 2.1.

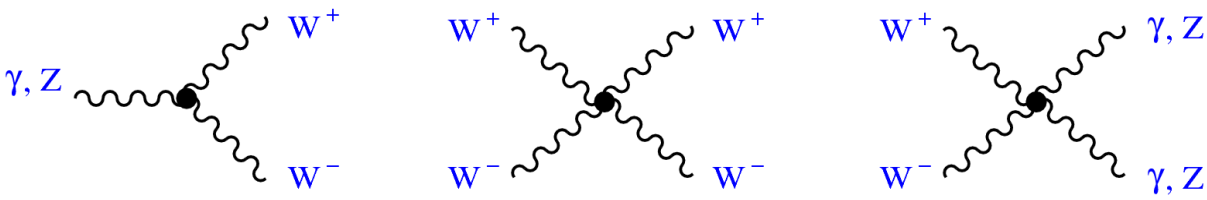


Figure 2.1: Feynman diagrams for the self-couplings of the electroweak gauge bosons [21].

2.3 Higgs mechanism: a spontaneous symmetry breaking

The EW theory, an elegant description of the weak and electromagnetic interactions by means of four massless gauge bosons, has some serious shortcomings. As already emphasised in the previous sections, adding mass terms for fermions, or gauge bosons, to the Lagrangian would break the $SU(2)_L \otimes U(1)_Y$ gauge invariance; this clearly contradicts experimental observation. Furthermore, within the pure EW, predictions for many SM scattering processes, e.g. WW scattering, violate unitarity at high energy (E), as their cross section diverges with E^2. This dependence on the energy clearly makes the theory non-renormalisable.

The solution to these inconsistencies is provided by the so-called "*Higgs mechanism*", a theory predicting the existence of an homonymous field: the Higgs field. The introduction of such scalar field with a specific potential allows to achieve a spontaneous symmetry breaking of the local gauge invariant theory, i.e. keep the full Lagrangian invariant under $SU(2)_L \otimes U(1)_Y$, while making the *vacuum* not invariant under this symmetry. In order to allow the Lagrangian to retain all its symmetries, only $SU(2)_L \otimes U(1)_Y$ multiplets can be added to the theory. An additional isospin doublet of complex scalar fields with weak isospin $\frac{1}{2}$ is introduced,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.37}$$

with a corresponding Lagrangian:

$$\mathcal{L}_{\text{scalar}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi), \tag{2.38}$$

where D_μ is the covariant derivative associated to $SU(2)_L \otimes U(1)_Y$ and

$$V(\phi) = \mu^2 \left(\phi^\dagger \phi\right) + \lambda \left(\phi^\dagger \phi\right)^2 \tag{2.39}$$

is the quartic potential associated to the new scalar field. The parameter λ of the potential is assumed to be positive. The case $\mu^2 > 0$ corresponds to a single vacuum at 0 and the exact symmetry of the Lagrangian is preserved in the vacuum. When $\mu^2 < 0$ there is not a single vacuum located at 0, as shown in Figure 2.2, but a continuum of distinct vacua located at $\langle |\phi|^2 \rangle = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}$. Without loss of generality the vacuum expectation value[6] of the Higgs field can be chosen to be $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$:

$$\langle \phi \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{2.40}$$

where $v = \sqrt{-\frac{\mu^2}{\lambda}}$. Any choice of the vacuum breaks the exact $SU(2)_L \otimes U(1)_Y$ symmetry; on the other hand the exact electromagnetic symmetry needs to be preserved in order to maintain the electric charged conserved, i.e. the transformation sub-group $U(1)_{\text{EM}}$ should remain as a symmetry of the vacuum:

$$e^{i\alpha Q} \langle \phi \rangle_0 \simeq (1 + i\alpha Q) \langle \phi \rangle_0 = \langle \phi \rangle_0. \tag{2.41}$$

Therefore the charge operator Q annihilates the vacuum, or, equivalently, the hypercharge of the Higgs doublet has to be $Y = 1$ (from the analogous of the Gell-Mann-Nishijima formula, $Y = 2(Q - I_3)$).

The gauge bosons corresponding to the remaining sub-group SU(2) and its broken generators should

[6] The vacuum expectation value is defined as the absolute value of the field at the minimum of the potential.

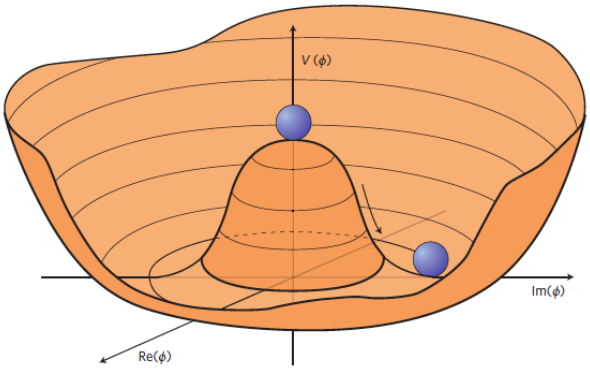


Figure 2.2: Illustration of the shape of the Higgs potential $V(\phi) = \mu^2\left(\phi^\dagger\phi\right) + \lambda\left(\phi^\dagger\phi\right)^2$ for $\mu^2 < 0$ [22].

acquire mass. The Higgs field can be expanded around the minimum:

$$\phi(x) = e^{\frac{i\xi_a(x)\tau^a}{2v}}\begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}, \tag{2.42}$$

where $H(x)$ is the physical scalar Higgs field and $\xi_a(x)$ $(a = 1, 2, 3)$ are new real fields. The invariance of the Lagrangian under SU(2) transformations allows to apply a SU(2) local gauge transformation (with $\alpha_i = -\frac{\xi}{v}$, called the *unitary gauge*) and the field becomes:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.43}$$

The scalar Lagrangian from Eq. 2.38 can be expanded and expressed in terms of the physical gauge fields:

$$\mathcal{L}_{\text{scalar}} = \frac{1}{2}\partial_\mu H\partial^\mu H + \frac{g^2}{4}(v+H)^2\left(W^+_\mu W^{-\mu} + \frac{1}{2\cos^2\theta_W}Z_\mu Z^\mu\right) - \mu^2\frac{(v+H)^2}{2} - \lambda\frac{(v+H)^4}{4}. \tag{2.44}$$

The quadratic terms in the vector fields allow to infer the W- and Z-boson masses, i.e. $M_W = \frac{gv}{2}$ and $M_Z = \frac{gv}{2\cos\theta_W} = \frac{M_W}{\cos\theta_W}$. Since the vacuum is neutral and, therefore, $\mathrm{U(1)_{EM}}$ remains as an exact symmetry of the theory, no quadratic term in A_μ appears and the photon remains massless. The second term in Eq. 2.44 also entails four vertex factors describing the interaction between the Higgs boson and gauge bosons: HWW, HZZ, $HHWW$ and $HHZZ$.

It has to be remarked that, out of the four degrees of freedom introduced by adding the scalar doublet in Eq. 2.37, one scalar boson, the Higgs field H, is remnant of the spontaneous symmetry breaking; the additional three degrees of freedom have been absorbed in order to grant the $W^\pm$ and Z bosons the longitudinal polarisation. The SM predicts a mass term and self-interactions for the Higgs field. In fact from the last two terms in Eq. 2.44:

$$-\frac{1}{2}(-2\mu^2)H^2 + \frac{1}{4}\mu^2v^2\left(\frac{4}{v^3}H^3 + \frac{1}{v^4}H^4 - 1\right), \tag{2.45}$$

where the Higgs-boson mass reads $M_H = \sqrt{-2\mu^2} = \sqrt{2\lambda v^2}$. In spite of predicting the existence of the

Higgs boson, the SM does not predict its mass, as μ^2 is a priori unknown and, even though v is known[7], λ is a free parameter of the theory.

2.3.1 Yukawa couplings and fermion mass

The theory, up to this point, is characterised by massless fermions. In order to give mass to charged leptons in a gauge invariant way, new terms involving the Yukawa coupling (y_ℓ) of the leptons with the Higgs field need to be added to the Lagrangian:

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = -y_\ell \left(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L\right) = -\frac{y_\ell}{\sqrt{2}} \left(v\bar{\ell}\ell + H\bar{\ell}\ell\right), \tag{2.46}$$

where the Higgs-field expansion from Eq. 2.43 is exploited. The introduction of such terms allows to "generate" a mass term for the charged lepton of mass $M_\ell = \frac{y_\ell v}{\sqrt{2}}$ and describes the Higgs-lepton coupling with strength $\frac{y_\ell}{\sqrt{2}} = \frac{M_\ell}{v}$. It has to be remarked that the mass of charged leptons is not predicted since y_ℓ are free parameters of the theory.

The Yukawa term in Eq. 2.46 only introduces mass terms for "down-type" fermions, i.e. only for one of the isospin doublet components. For "up-type quarks" a term involving the charge conjugate of the Higgs doublet, i.e.

$$\tilde{\phi}^c = -i\tau_2 \phi^* = -\frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix}, \tag{2.47}$$

is introduced:

$$\mathcal{L}^{q}_{\text{Yukawa}} = -y_d \bar{Q}_L \phi d_R - y_u \bar{Q}_L \tilde{\phi}^c u_R + \text{h.c.}, \tag{2.48}$$

where $Q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L$. Furthermore, in the most general realisation, the Yukawa couplings $y_{u,d}$ are matrices and this introduces mixing between different flavours; the weak eigenstates (q') are linear superposition of the mass eigenstates (q), i.e.

$$\begin{pmatrix} u' \\ c' \\ t' \end{pmatrix}_{L,R} = U_{L,R} \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R} \quad , \quad \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{L,R} = D_{L,R} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R} , \tag{2.49}$$

where $U_{L,R}$ and $D_{L,R}$ are unitary matrices diagonalising the quark mass matrices $M_{u,d} = \frac{v}{\sqrt{2}} y_{u,d}$ arising from the expansion of Eq. 2.48:

$$U_R^{-1} M_u U_L = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \quad , \quad D_R^{-1} M_d D_L = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix}. \tag{2.50}$$

The mismatch between weak and mass eigenstates leads to transitions between quark generations through flavour changing interactions. The flavour changing interactions are proportional to the Cabibbo-

[7] The Higgs vacuum expectation value can be obtained by taking into account the low-energy phenomenology relation (i.e. $\frac{g^2}{8M_W^2} = \frac{G_F}{\sqrt{2}}$, with G_F indicating the Fermi coupling constant): $v = \left(\sqrt{2}G_F\right)^{-\frac{1}{2}} \simeq 246\,\text{GeV}$.

Kobayashi-Maskawa (CKM) matrix elements. The CKM matrix can be written as:

$$V_{\text{CKM}} \equiv U_L^{\dagger} D_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{2.51}$$

where diagonal elements dominate the flavour changing interactions. There is no mixing in the neutral sector (FCNC, "Flavour Changing Neutral Currents"). The definition of the CKM matrix, up to a non-eliminable phase, leads to CP violation in weak interactions. Analogously, a mixing matrix, the so-called Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix, can be introduced in the neutrino sector.

The introduction of interaction terms between the Higgs and the fermion fields gives mass to the fermions, $M_f = \frac{v}{\sqrt{2}} y_f$, and the Higgs-fermion coupling y_f is proportional to the fermions mass. For this reason fermion couplings to the Higgs boson are very different from each other (from $M_\nu \lesssim 1$ eV up to $M_t \simeq 174$ GeV).

2.4 The Standard Model Higgs boson

The fundamental role played by the Higgs mechanism in the Standard Model theoretical framework is clear from Section 2.3. Despite the fact that huge efforts in several experiments (at LEP [23], Tevatron [24] and LHC [25]) have been made in order to search for the Higgs boson, it remained undiscovered for several decades. The discovery of a particle compatible with the Standard Model Higgs boson in July 2012 by the ATLAS and CMS experiments at the LHC[8] was an important milestone in the history of physics.

A non-extensive review of the different production and decay modes of the Higgs boson in proton–proton (noted pp in what follows) collisions is presented in this section.

2.4.1 Higgs-boson production and decay

Gluon fusion

The gluon fusion, whose leading order diagram is shown in Figure 2.3, is the leading contribution to SM Higgs-boson production at the LHC due to the large contribution from gluon interactions in pp collisions. The top and bottom quarks are the main contributors to the quark loop, while contributions from other fermions are negligible for current studies.

Vector-boson fusion

The leading order diagram for vector-boson fusion is shown in Figure 2.3. Two quarks radiate a vector boson V ($W^{\pm}$ or Z) each and their fusion leads to the production of a Higgs boson. The production of Higgs bosons at the LHC via this mechanism is the second most copious. The presence of diagrams with a vertex connecting the bosons to the Higgs boson without being in a loop is referred to as *direct coupling*. The direct coupling of the Higgs boson to the vector bosons in this production mode allows a direct measurement of the coupling of the Higgs boson to vector bosons.

[8] See Section 3.1 for further details on the LHC.

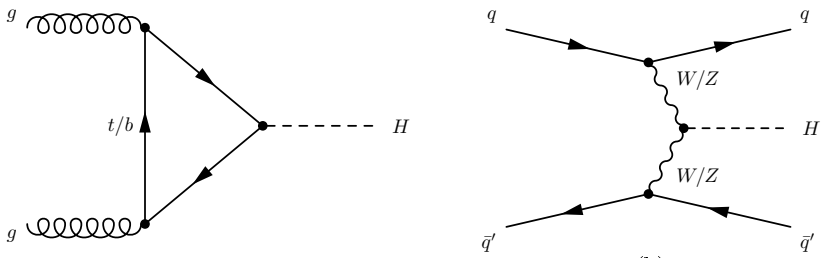


Figure 2.3: Leading order diagrams for the gluon fusion (left) and vector-boson fusion (right) initiated production of the SM Higgs boson.

Higgs-strahlung

The Higgs-strahlung, or associated production of the Higgs boson with vector bosons, leading order Feynman diagrams for qq- and gg-initiated processes are shown in Figure 2.4. These production modes are privileged processes to study the $H \to b\bar{b}$ decay mode, as they benefit from the leptonic decay of the additional vector bosons to reduce the contamination from QCD multijet backgrounds.

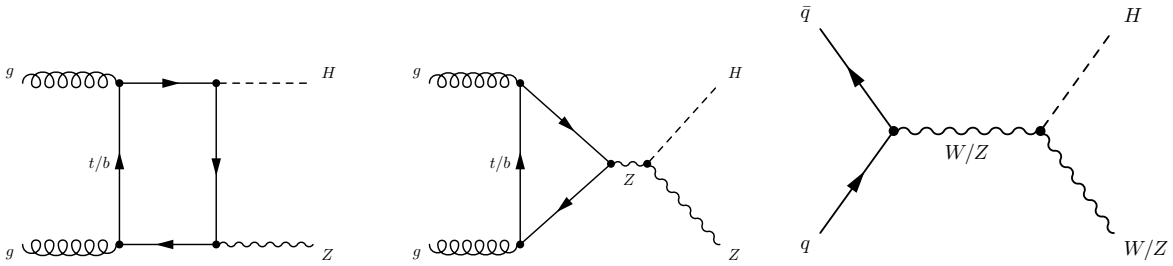


Figure 2.4: Leading order diagrams for the production of a Higgs boson in association with a vector boson.

Associated production of the Higgs boson with top quarks

Figure 2.5 shows a set of Feynman diagrams for the production of a Higgs boson in association with top quarks. These diagrams involve the direct coupling of the Higgs boson to the top quark; therefore these are privileged production modes for the study of the Yukawa coupling of the Higgs boson to the top quark y_t, i.e. the largest Yukawa coupling, as discussed in Section 2.3.1. In particular, the $t\bar{t}H$ production (upper diagram) is the preferred channel for the direct measurement of y_t as it has a higher cross section than the tH processes (bottom diagrams). However, the production of tH, where a Higgs boson can be radiated either from the top quark or from the exchanged W boson, provides a unique opportunity to study the relative sign of the top-quark Yukawa coupling. The measurement of the latter is discussed in Section 2.4.2.

Figure 2.6 summarises the main production cross sections as a function of the centre-of-mass energy ($\sqrt{s}$). The production in association with a top-quark pair is one of the rarest Higgs-boson production modes, but its cross section increases by four times from 8 to 14 TeV, more than the larger production modes.

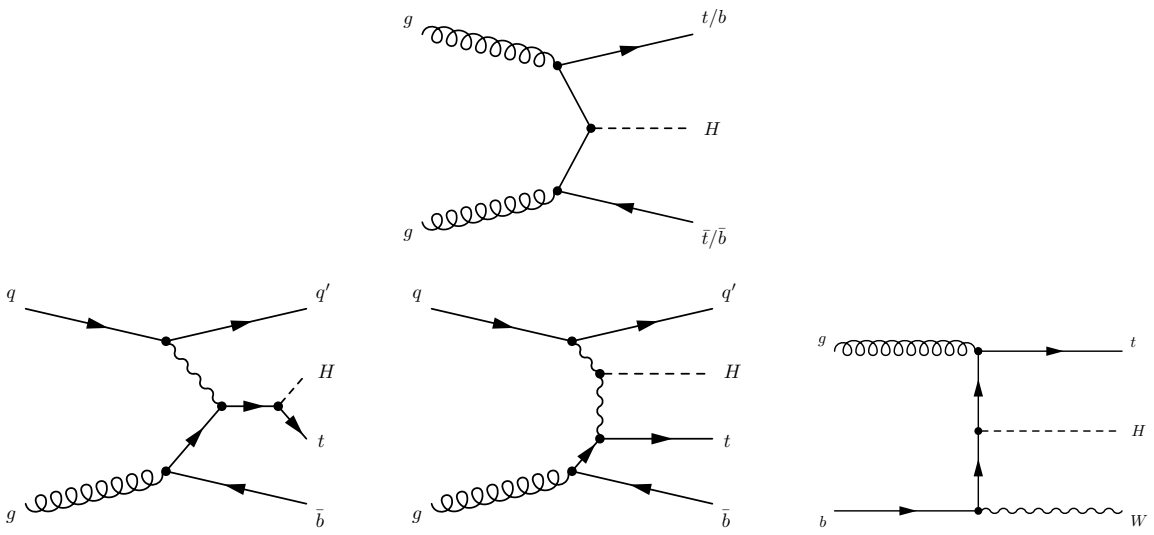


Figure 2.5: Example of leading order diagrams for the production of a Higgs boson in association with top quarks.

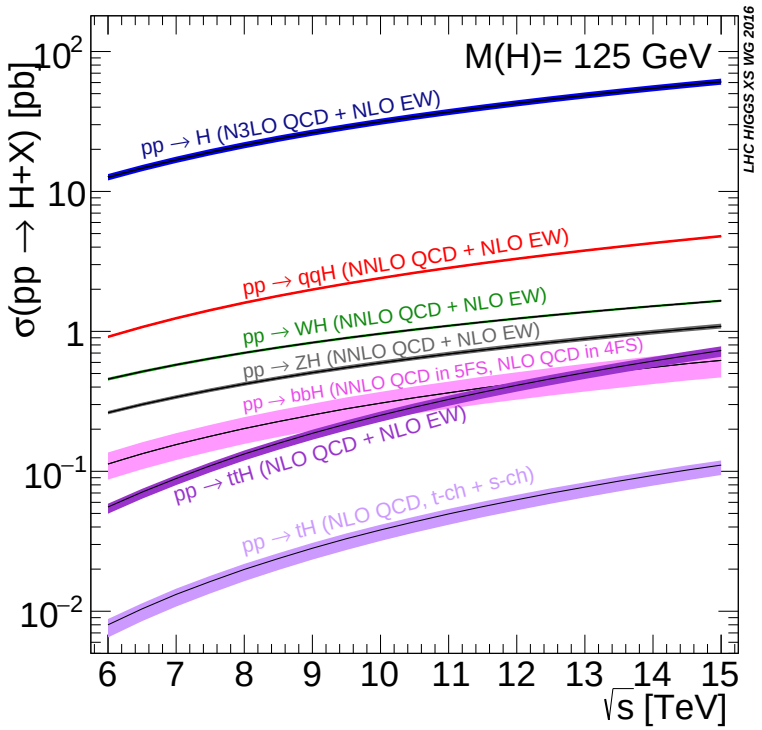


Figure 2.6: The SM Higgs-boson production cross sections as a function of the LHC centre-of-mass energy [26].

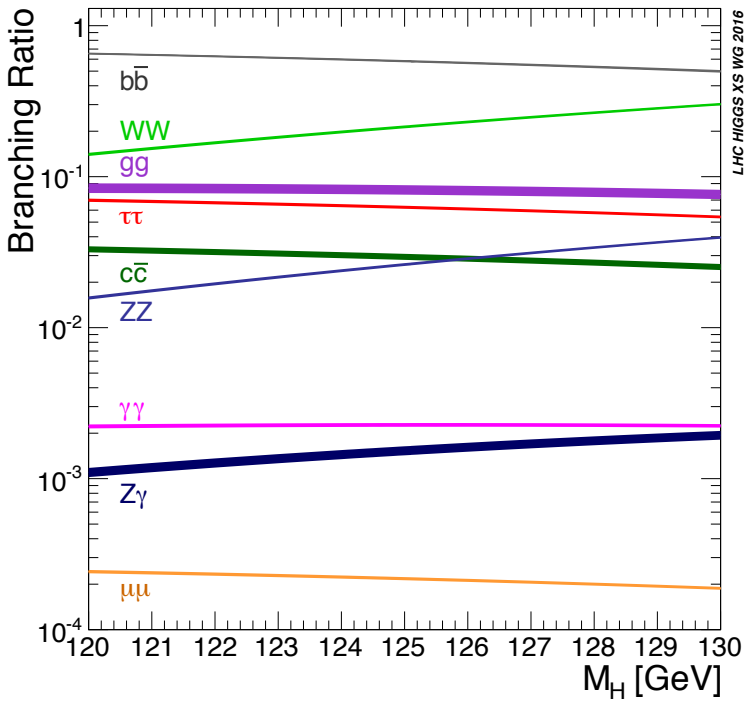


Figure 2.7: The SM Higgs-boson branching ratios and their uncertainties for the mass range around 125 GeV [26].

Higgs-boson decay

The branching ratios of Higgs-boson decay modes are shown as a function of the Higgs-boson mass in Figure 2.7. At the measured mass of $M_H = 125.09 \pm 0.24$ GeV [27] the Higgs boson mainly decays to a $b\bar{b}$ pair ($\sim 58\%$). Due to the large QCD multijet background, the pure gluon fusion decay mode is difficult to detect, but other associated production modes can be exploited.

$H \to WW^*$ is the second dominant decay mode. Because of the high missing energy involved (high p_{T} neutrinos), the only mass that can be reconstructed is the transverse mass of the system. The $H \to ZZ^*/\gamma^* \to 4\ell$, with $\ell = e, \mu$, decay has a very clean signature and is known as the "golden channel"; it allows to fully reconstruct the Higgs-boson mass.

Massless bosons do not couple to the Higgs boson directly but through charged and/or coloured massive particles via loops (see Section 2.4.2). In spite of its low branching ratio, the $H \to \gamma\gamma$ plays a very important role in Higgs-boson searches, since it forms a very narrow invariant mass peak, due to the two high energetic photons. In a hadron collider is very hard to distinguish the $H \to gg$ decay from the huge QCD multijet background. The $H \to Z\gamma$ decay is difficult to measure as the final state can be easily misinterpreted as a Z + jets event, background with a very large cross section.

2.4.2 Direct measurement of the top-quark Yukawa coupling

The top-quark Yukawa coupling can be accessed *directly* and *indirectly*; these two approaches are complementary, as they allow to disentangle hypothetical effects due to physics beyond the SM.

Indirect constraints on y_t are achievable [28] through gluon fusion and Higgs-boson decay to a pair of photons (see Section 2.4.1). One can resolve the loops, shown in Figure 2.8, but an important assumption needs to be made: only SM contributions are entering such production and decay mechanisms; the

presence of new heavy particles carrying electric or colour charge, or both, would significantly change the structure of such loop corrections.

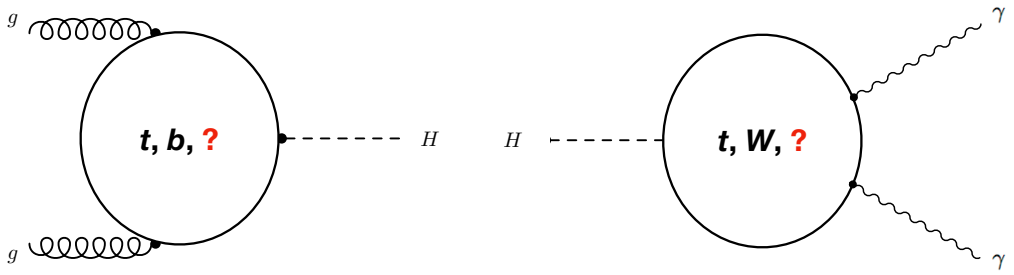


Figure 2.8: Feynman diagrams representing the effective gluon fusion vertex ggH (left) and the effective photon vertex $H\gamma\gamma$ (right).

The production of a Higgs boson in association with a pair of top quarks (the so-called $t\bar{t}H$ channel) is currently the most sensitive direct way to access the top-quark Yukawa coupling. In fact, the production cross section of $t\bar{t}H$ is proportional to $|y_t|^2$ at tree level (see Figure 2.5).

The ATLAS and CMS collaborations have searched for the production of $t\bar{t}H$ in pp collisions at the LHC exploiting data collected at a centre-of-mass energy of 7 and 8 TeV [29–33], where most of the sensitivity was provided by the $H \to b\bar{b}$, WW^*, ZZ^* and $\tau\tau$ decays. The $t\bar{t}H$ signal strength μ, defined as the ratio of the observed to the expected number of $t\bar{t}H$ events, obtained by combining ATLAS and CMS measurements at 7 and 8 TeV yielded $\mu_{t\bar{t}H} = \sigma/\sigma_{\text{SM}} = 2.3^{+0.7}_{-0.6}$ [28]. The "excess" in this result is driven by the multi-lepton final states.

The latest ATLAS and CMS measurements, at a centre-of-mass energy of 13 TeV, allowed, first, to reach the evidence for [34, 35] and, more recently, the sought-after observation of $t\bar{t}H$ production [36, 37]. The latest combination of 7, 8 and 13 TeV data, performed by the ATLAS collaboration, led to a measured $t\bar{t}H$ signal strength of $\mu_{t\bar{t}H} = 1.32^{+0.28}_{-0.26}$ [36], where the excess is still significantly driven by the multi-lepton channels (the latest multi-lepton combination at 13 TeV yields $\mu_{t\bar{t}H} = 1.56^{+0.42}_{-0.40}$). Interestingly, the latest measurement carried out by the CMS collaboration also leads to a $t\bar{t}H$ signal strength above one, corresponding to $\mu_{t\bar{t}H} = 1.26^{+0.31}_{-0.26}$ [37] (where the best-fit value from the combination of leptonic $H \to WW^*$ final states gives $\mu_{t\bar{t}H} = 1.97^{+0.71}_{-0.64}$).

I contributed to the ATLAS search for $t\bar{t}H$ production in the non-resonant four-lepton final state at 13 TeV. The latest analysis, performed with a total integrated luminosity of 80 fb^{-1}, is described in Chapter 5.

2.5 Electroweak vacuum stability and top-quark Yukawa coupling

As soon as radiative corrections are included, all SM couplings and fields undergo an evolution, called renormalisation group (RG) evolution, up to an energy scale Λ, where the theory may become theoretically inconsistent or contradict some observations. The most interesting parameter[9] turns out to be the Higgs-boson self-coupling constant, $\lambda = \frac{M_H^2}{2v^2}$, introduced in Section 2.3. The RG evolution of the Higgs-boson self-coupling constant shows a dependence on the logarithm of the squared energy scale (μ^2) [38]. Since the Higgs-boson coupling is proportional to the mass of the interacting particle, only loops involving

[9] The only other problematic parameter is the U(1) hypercharge which develops a Landau pole, but only at an energy scale significantly exceeding the Planck mass.

massive vector bosons and the top quark are relevant in λ RG evolution; therefore the latter at one loop reads [39]:

$$16\pi^2 \frac{\mathrm{d}\lambda}{\mathrm{d}\ln\mu} = 24\lambda^2 - 9\lambda\left(g^2 + \frac{1}{3}g'^2\right) + \frac{9}{8}g^4 + \frac{3}{8}g'^4 + \frac{3}{4}g^2g'^2 + 12\lambda y_t^2 - 6y_t^4. \tag{2.52}$$

The running of λ shows an interplay between positive contributions from the gauge bosons ($\propto g^4$) and the negative contribution from the top quark ($\propto y_t^4$). Driven by the top-quark loop contributions, the Higgs-boson quartic coupling λ tends to become negative at larger scales μ, making the vacuum unstable and new minima of the potential appear, because of the RG-improved *effective potential* $V(\phi) = \mu^2\phi^2 + \lambda(\mu = \phi)\phi^4$ (where the self-coupling dependence on the energy scale ϕ is intentionally emphasised). Figure 2.9 shows the main top-quark one-loop contribution to the Higgs-field self-coupling and, consequently, entering the effective potential; this contribution is of the form:

$$\delta V_t(\phi) \propto -y_t^4\phi^4 \ln\left(\frac{\phi^2}{\mu^2}\right), \tag{2.53}$$

where the minus sign comes from the fact that the top quark is a fermion.

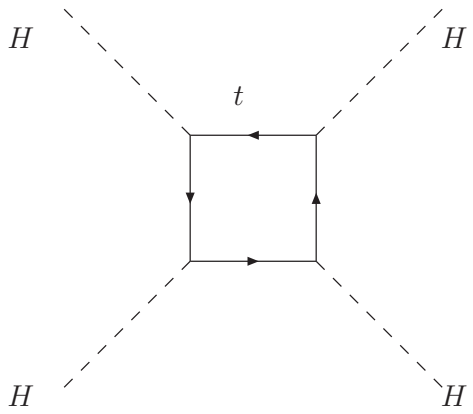


Figure 2.9: Feynman diagram representing the top-quark one-loop contribution, reported in Eq. 2.53, to the Higgs-field effective potential.

It has been demonstrated in Ref. [40] that for $\phi \sim \Lambda$, with $\Lambda \sim 10^{19}$ GeV being the Planck scale, a good approximation for the effective potential is

$$V_{\text{eff}}(\phi) \simeq \lambda(\Lambda)\,\phi^4 + O\left(\lambda^2, g(\Lambda), g'(\Lambda)\right), \tag{2.54}$$

which means that the stability of the SM vacuum is approximately equivalent to the question whether λ stays positive up to the scale Λ [41, 42]. Before the discovery of the Higgs boson it was customary to show instability constraints as a function of M_H, with other parameters fixed by experimental measurements. For a large enough Higgs-boson mass ($M_H \gtrsim 175$ GeV [43]) the Landau pole in λ would occur at energies smaller than the Planck scale. For a small Higgs mass ($M_H \lesssim 113$ GeV [43]) λ would become negative at a relatively low energy scale, as the top-quark loops give an essential contribution to the Higgs effective potential, before eventually reaching a Landau pole.[10] The latter case is characterised by an unstable

[10] The reason for the fact that λ becomes positive again and diverges is the evolution of the gauge couplings and y_t. At large

vacuum, with a life-time smaller than the age of the Universe [44–46]. In the interesting region, around 125 GeV, λ is very close to zero at the Planck scale. Figure 2.10 shows the predicted RG evolution of λ, y_t and the SM gauge couplings.

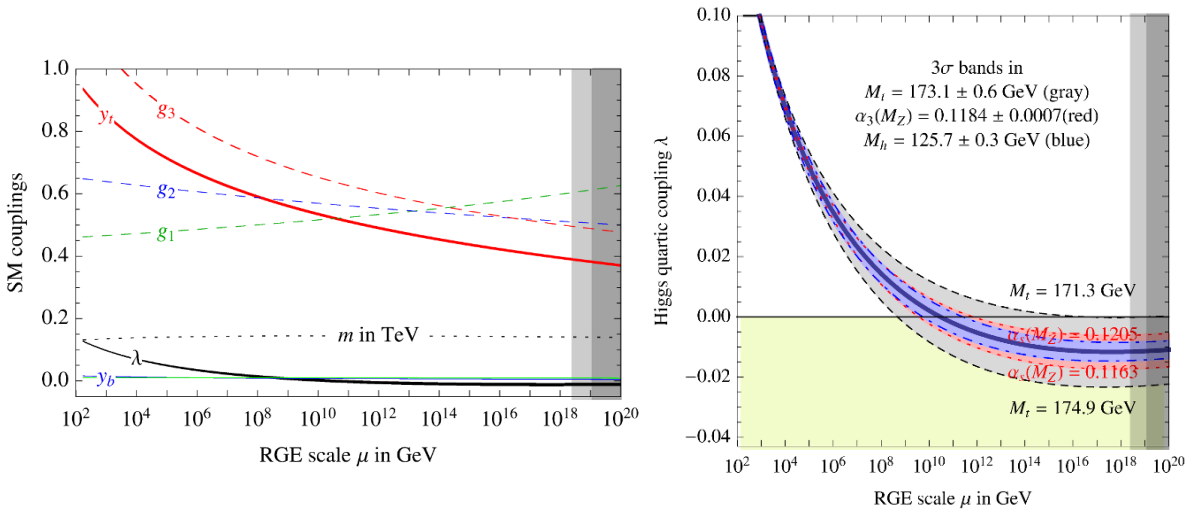


Figure 2.10: Left: SM RG evolution of the gauge couplings ($g_1 = 5/3g'$, $g_2 = g$, $g_3 = \alpha_s$), of the top- and bottom-quark Yukawa couplings (y_t, y_b), and of the Higgs-boson self-coupling (λ). All couplings are defined in the $\overline{\text{MS}}$ ("modified minimal subtraction") scheme. Three-loop RG equations are included and the thickness indicates the $\pm 1\sigma$ uncertainties in M_t, M_H and α_s [47]. Right: RG evolution of λ varying M_t, M_H and α_s by $\pm 3\sigma$ [47].

The Higgs-boson mass is well within the $113 < M_H < 175$ GeV interval,[11] which means that the life-time of the SM vacuum is expected to exceed that of the Universe by several orders of magnitude, see e.g. Ref. [48]. Nevertheless we can assume that there is no new physics up to the Planck scale and verify whether the model runs into any contradiction.

As already discussed (see Eq. 2.52), the contribution of the top quark to the effective potential is very important, as it has the largest Yukawa coupling to the Higgs boson. All relevant parameters of the SM can be fixed to their experimental values except the top-quark Yukawa coupling, which is at the present moment the most uncertain one for the problem under consideration. The RG evolution of the Higgs-boson self-coupling as a function of the energy scale μ is shown in Figure 2.11 for different values of the top-quark Yukawa coupling, together with the behaviour of the effective potential close to the so-called "critical" value y_t^{crit}. This value is defined as the value of the top-quark Yukawa coupling at which our electroweak vacuum is degenerate with a new one at a certain energy scale Λ.

Four different scenarios are possible [39]:

- if $y_t < y_t^{\text{crit}} - \epsilon$, the Higgs potential is monotonic and the EW vacuum is unique;
- if $y_t^{\text{crit}} - \epsilon < y_t < y_t^{\text{crit}}$, the EW vacuum is a global minimum, i.e. deeper than the additional one;
- if $y_t^{\text{crit}} < y_t < y_t^{\text{crit}} + \eta$, the EW vacuum is metastable: a tunnelling through the potential barrier has a life-time larger than the age of the Universe;

scales (above 10^{16} GeV) the electroweak couplings, especially the U(1) coupling g', start to be the dominant contributions in Eq. 2.52. In contrast y_t, which is responsible for the decrease of λ at lower scales, becomes small. This is also the reason why the effective potential goes up again at large field strength ϕ beyond the extra minimum.

[11] See Section 2.4.1 for further details.

- if $y_t > y_t^{\text{crit}} + \eta$, corresponding to roughly $M_t \gtrsim 178\,\text{GeV}$, the life-time of the EW vacuum is smaller than the age of the Universe.

The ϵ and η parameters, strongly dependent on the accuracy of radiative corrections to the Higgs potential being included, are $\epsilon = 1.2 \times 10^{-6}$ and $\eta = 0.04$ in the current most accurate determination [39]. The top-quark critical value is found to be [39]:

$$y_t^{\text{crit}} = 0.9244 + 0.0012 \times \left(\frac{M_H/\text{GeV} - 125.7}{0.4} + \frac{\alpha_s(M_Z) - 0.01184}{0.0007} \right), \tag{2.55}$$

where $\alpha_s(M_Z)$ is the strong coupling constant at the Z-boson mass.

To summarise, if the measurement of the top-quark Yukawa coupling will result in $y_t < y_t^{\text{crit}} + \eta$, the embedding of the SM without any kind of new physics does not lead to any inconsistency and, thus, no information about the energy scale of new physics can be inferred. In the other case, $y_t > y_t^{\text{crit}} + \eta$, the Higgs-boson self-coupling becomes negative and the Higgs effective potential crosses zero at a certain value of the Higgs field. The presence and intervention of new physics, e.g. new scalars, fermions or thresholds [49–51], would allow to make both the scalar self-coupling and the Higgs potential positive at all energies. Higgs inflation [52], based on the observation that the Higgs field non-minimally coupled to gravity can give rise to inflation, introduces counter-terms, that, in certain conditions, restore the monotonic behaviour of the Higgs effective potential at large Higgs-field values [53].

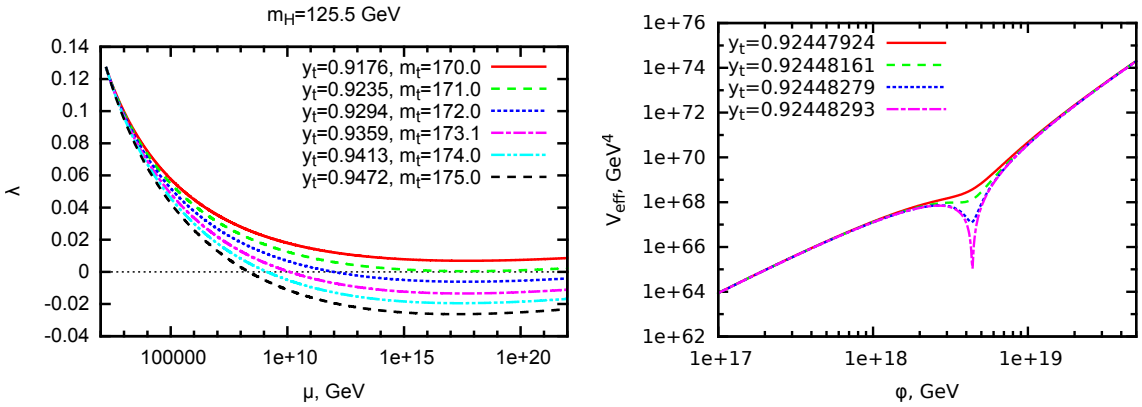


Figure 2.11: Left: RG running of the Higgs-boson self-coupling for several values of the top-quark Yukawa coupling $y_t\,(\mu = 173.2\,\text{GeV})$ and fixed 125.5 GeV Higgs-boson mass [39]. Right: a very small change in the top-quark Yukawa coupling y_t converts the monotonic behaviour of the effective potential for the Higgs field to that with an extra minimum at large values of the Higgs field [53].

As demonstrated in this section, the top-quark Yukawa coupling is a key parameter, determining the structure of the SM at high energy scale.

2.6 Effective Field Theory and anomalous gauge couplings

In absence of obvious findings, like new sharp resonances or strong deviations from the SM prediction, effects due to new physics can be interpreted in terms of Effective Field Theory (EFT) [54] operators; higher-dimension operators and the corresponding *anomalous* gauge couplings can be formulated in a model-independent and systematic way. If the new physics associated with these operators occurs at a

high energy scale Λ, one is motivated to use the formulation of EFT to organise the operators in order of increasing dimensionality. These theories add new operators to the SM Lagrangian with the coupling strengths being free parameters. By constraining the free parameters, it is possible to set limits on the effects of new physics. This more general approach will not point to a concrete and specific model, but may uncover effects providing inputs to new theories explaining the observed deviations from the SM.

If baryon and lepton numbers are conserved, only operators with even dimension can appear in the EFT. Consequently, not to introduce additional fundamental physics,[12] the largest new physics contribution is expected from dimension-six operators. In the following an EFT which includes dimension-six and dimension-eight operators that modify the interactions among electroweak gauge bosons, i.e. [55]

$$\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_{i=WWW,W,B} \frac{c_i}{\Lambda^2} O_i + \sum_{j=0,1} \frac{f_{S,j}}{\Lambda^4} O_{S,j} + \sum_{j=0,\ldots,9} \frac{f_{T,j}}{\Lambda^4} O_{T,j} + \sum_{j=0,\ldots,7} \frac{f_{M,j}}{\Lambda^4} O_{M,j}, \tag{2.56}$$

is considered. Eq. 2.56 shows that in the limit $\Lambda \to \infty$ the EFT Lagrangian reduces to the SM one.

Three different CP conserving,

$$\begin{aligned} O_{WWW} &= \text{Tr}\left[W_{\mu\nu} W^{\nu\rho} W^{\mu}_{\rho}\right] \\ O_W &= (D_\mu \phi)^\dagger W^{\mu\nu} (D_\nu \phi) \\ O_B &= (D_\mu \phi)^\dagger B^{\mu\nu} (D_\nu \phi), \end{aligned} \tag{2.57}$$

and two CP violating dimension-six operators,

$$\begin{aligned} O_{\widetilde{W}WW} &= \text{Tr}\left[\widetilde{W}_{\mu\nu} W^{\nu\rho} W^{\mu}_{\rho}\right] \\ O_{\widetilde{W}} &= (D_\mu \phi)^\dagger \widetilde{W}^{\mu\nu} (D_\nu \phi), \end{aligned} \tag{2.58}$$

affect the TGC and QGC [55] discussed in Section 2.2.1. In Eqs. 2.57 and 2.58 $W_{\mu\nu}$ and $B_{\mu\nu}$ indicate the $\text{SU}(2)_L$ and $\text{U}(1)_Y$ field strength tensors, $\widetilde{W}_{\mu\nu} \equiv \frac{\tau^a}{2} W_a^{\mu\nu}$, D_μ corresponds to the $\text{SU}(2)_L \otimes \text{U}(1)_Y$ covariant derivative (see Section 2.2), and ϕ is the Higgs-doublet field (see Section 2.3). As already seen for the SM, TGC and QGC induced by dimension-six operators are completely related by requiring the gauge invariance to apply.

Table 2.3 shows the vertices induced by each of these dimension-six operators which are entering the production of the $W^\pm W^\mp Z$ and $W^\pm ZZ$ processes, discussed in Chapter 6. This approach is different from the parametrisation exploited at LEP for TGC, which involved five parameters defined such that they equal zero in the SM [56].

Table 2.3: Vertices induced by each dimension-six EFT operator are marked with × in the corresponding column. A indicates the photon. The vertices which are not relevant to $W^\pm W^\mp Z$ and $W^\pm ZZ$ production have been omitted [55].

	WWZ	WWA	$WWZZ$	$WWZA$
O_{WWW}	×	×	×	×
O_W	×	×	×	×
O_B	×	×		
$O_{\widetilde{W}WW}$	×	×	×	×
$O_{\widetilde{W}}$	×	×		

Dimension-six operators giving rise to QGC also generate TGC; in order to disentangle effects of the

[12] The introduction of dimension-four operators would lead to the introduction of new fundamental physics.

QGC we shall consider effective operators exhibiting QGC without a TGC associated to them. Since the dimension of gauge fields is two, the only way to generate genuine quartic vertices is to include dimension-eight (or higher) operators. There are three different families of dimension-eight operators:

- *scalar* (S), operators that are only containing $D_\mu\phi$;
- *mixed* (M), operators involving $D_\mu\phi$ and two field strength tensors $W_{\mu\nu}$;
- *tensor* (T), operators that are only containing field strength tensors $W_{\mu\nu}$.

The scalar operators,

$$\begin{aligned} O_{S,0} &= \left[(D_\mu\phi)^\dagger D_\nu\phi\right] \times \left[(D^\mu\phi)^\dagger D^\nu\phi\right], \\ O_{S,1} &= \left[(D_\mu\phi)^\dagger D^\mu\phi\right] \times \left[(D_\nu\phi)^\dagger D^\nu\phi\right], \end{aligned} \tag{2.59}$$

contain quartic $WWWW$, $WWZZ$ and $ZZZZ$ interactions; the interactions induced by scalar operators are independent of the gauge boson momenta, as no field strength tensor is involved. On the other hand, the mixed operators, combining two EW field strength tensors and two covariant derivatives of the Higgs doublet, depend upon the momenta of the vector bosons due to the presence of the field strength tensor [55]:

$$\begin{aligned} O_{M,0} &= \mathrm{Tr}\left[W_{\mu\nu}W^{\mu\nu}\right] \times \left[(D_\beta\phi)^\dagger D^\beta\phi\right], \\ O_{M,1} &= \mathrm{Tr}\left[W_{\mu\nu}W^{\nu\beta}\right] \times \left[(D_\beta\phi)^\dagger D^\mu\phi\right], \\ O_{M,2} &= \left[B_{\mu\nu}B^{\mu\nu}\right] \times \left[(D_\beta\phi)^\dagger D^\beta\phi\right], \\ O_{M,3} &= \left[B_{\mu\nu}B^{\nu\beta}\right] \times \left[(D_\beta\phi)^\dagger D^\mu\phi\right], \\ O_{M,4} &= \left[(D_\mu\phi)^\dagger W_{\beta\nu}D^\mu\phi\right] \times B^{\beta\nu}, \\ O_{M,5} &= \left[(D_\mu\phi)^\dagger W_{\beta\nu}D^\nu\phi\right] \times B^{\beta\mu}, \\ O_{M,6} &= \left[(D_\mu\phi)^\dagger W_{\beta\nu}W^{\beta\nu}D^\mu\phi\right], \\ O_{M,7} &= \left[(D_\mu\phi)^\dagger W_{\beta\nu}W^{\beta\mu}D^\nu\phi\right]. \end{aligned} \tag{2.60}$$

A consequence is that their Lorentz structure can not be simply reduced to the SM one, as for the scalar operators. Finally, the tensor operators combine four field strength tensors [55]:

$$\begin{aligned} O_{T,0} &= \mathrm{Tr}\left[W_{\mu\nu}W^{\mu\nu}\right] \times \mathrm{Tr}\left[W_{\alpha\beta}W^{\alpha\beta}\right], \\ O_{T,1} &= \mathrm{Tr}\left[W_{\alpha\nu}W^{\mu\beta}\right] \times \mathrm{Tr}\left[W_{\mu\beta}W^{\alpha\nu}\right], \\ O_{T,2} &= \mathrm{Tr}\left[W_{\alpha\mu}W^{\mu\beta}\right] \times \mathrm{Tr}\left[W_{\beta\nu}W^{\nu\alpha}\right], \\ O_{T,5} &= \mathrm{Tr}\left[W_{\mu\nu}W^{\mu\nu}\right] \times B_{\alpha\beta}B^{\alpha\beta}, \\ O_{T,6} &= \mathrm{Tr}\left[W_{\alpha\nu}W^{\mu\beta}\right] \times B_{\mu\beta}B^{\alpha\nu}, \\ O_{T,7} &= \mathrm{Tr}\left[W_{\alpha\mu}W^{\mu\beta}\right] \times B_{\beta\nu}B^{\nu\alpha}, \\ O_{T,8} &= B_{\mu\nu}B^{\mu\nu}B_{\alpha\beta}B^{\alpha\beta}, \\ O_{T,9} &= B_{\alpha\mu}B^{\mu\beta}B_{\beta\nu}B^{\nu\alpha}. \end{aligned} \tag{2.61}$$

Tensor operators give rise to QGC containing only neutral EW gauge bosons, vertices not provided by SM gauge self-interactions (see Section 2.2.1 for further details). Stringent limits on anomalous TGC

have been set at LEP and the LHC. The LHC sensitivity to anomalous QGC has significantly improved in the last few years [57].

Deviations from the SM prediction in the triboson processes might point to effects due to new physics, which can be probed by assuming the previously discussed EFT approach.

2.7 Triboson production at the LHC

The production of three bosons is particularly sensitive to beyond SM physics via anomalous gauge couplings, as discussed in Section 2.6, and narrow resonances. Furthermore, some of the triboson systems are directly connected to Higgs production via the Higgs-mediated $VH(\rightarrow VV)$; therefore, these processes constitute a preferential "portal" connecting the pure EW theory to the Higgs sector. Triboson production is currently among the processes which are the least precisely measured,[13] with several channels still unexplored. This section is devoted to a brief review of the triboson-production searches and measurements carried out at the LHC.

$\gamma\gamma\gamma$

The measurement of the production cross section of three photons provides a test of perturbative QCD in processes with photons in the final state. The measured cross section [58] yields 1.6 times the SM prediction, showing a significant discrepancy larger than 2σ. The predictions underestimate the measurement of the inclusive fiducial cross section and the differential measurements at low photon transverse momenta and invariant masses. On the other hand, an adequate description is observed at high photon energies and invariant mass of the triphoton system.

$V\gamma\gamma$

The first observation of triboson production has been achieved in the search for $Z\gamma\gamma$ [59, 60]. The cross section is measured with leptonic (e^+e^-, $\mu^+\mu^-$, $\tau^+\tau^-$) decays of the Z boson. Events characterised by energetic photons are exploited to look for anomalous TGC, especially fully neutral $ZZ\gamma$ and $Z\gamma\gamma$. No deviations from SM predictions are observed. The $W\gamma\gamma$ search is also carried out in the leptonic ($e\nu_e$, $\mu\nu_\mu$, $\tau\nu_\tau$) final states and provided evidence for its production [60, 61]. The measured $W\gamma\gamma$ cross section is in agreement with the theoretical predictions.

$WV\gamma$

The first attempt to search for the production of triboson at the LHC has been made by the CMS collaboration [62] in the $W(\ell\nu)V(jj)\gamma$ channels, where $V = W, Z$ and $V(jj)$ indicates the hadronic decay of the V boson; the semileptonic final states allow to enlarge the signal events acceptance. The limit obtained on the $WV\gamma$ production cross section is approximately a factor of 3.4 larger than the SM prediction; no evidence for anomalous $WW\gamma\gamma$ and $WWZ\gamma$ couplings is found and the photon spectrum is exploited to set the first limits on the dimension-eight EFT tensor operator T_0 (see Section 2.6).

$W^\pm W^\pm W^\mp$

The search for the production of three W bosons has been the first attempt to probe the production of a three-massive-vector-boson system. The ATLAS result at 8 TeV [63] did not allow to reach the evidence

[13] Apart from very few high purity neutral triboson processes such as $Z\gamma\gamma$.

for its production, as the search in the $3\ell 3\nu$ and $2\ell 2\nu 2j$ final states only yielded a 0.96σ significance for the rejection of the background-only hypothesis.[14] First limits on the EFT scalar operator S_0 and S_1 have been obtained.

The search for the production of $W^{\pm}W^{\mp}Z$ and $W^{\pm}ZZ$ has never been attempted before. Chapter 6 is devoted to a detailed description of the first analysis aiming at probing the SM prediction for the production of $W^{\pm}W^{\mp}Z$ and $W^{\pm}ZZ$ triboson states, as well as reaching the first evidence for the production of three massive vector bosons.

[14] See Section 5.1.1 for details on the definition of the significance.

CHAPTER 3

The ATLAS experiment at the Large Hadron Collider

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [25] at CERN[1] is a superconducting accelerator and collider that was built in the 26.7 kilometers long tunnel, where LEP[2] was operated until its decommissioning in 2000. The tunnel is located between 45 m and 170 m below surface, between the Jura mountains and the Geneva airport. It is currently the highest energy particle collider ever built, whose main goals are to probe the Standard Model of particle physics, the discovery of new particles and new physics beyond the Standard Model. For this purpose, several detectors are located in the accelerator ring. The four largest experiments at the LHC are ALICE, ATLAS, CMS and LHCb. Data recorded by the ATLAS detector [64–66] are exploited in the works presented in this thesis. The layout of the LHC is the same as the one of LEP, with eight straight sections. The LHC is capable of accelerating and colliding hadrons, namely protons and heavy ions. This advanced collider is designed to accelerate protons up to an energy of 7 TeV, starting from an initial energy of 450 GeV.

The rate of produced events (R_{event}), i.e. the number of events produced per second, is described as:

$$R_{\text{event}} = \mathcal{L} \cdot \sigma_{\text{event}}, \tag{3.1}$$

where $\mathcal{L}$ is the instantaneous luminosity of the accelerator and σ_{event} is the cross section of the corresponding physics process; thus, it is important to achieve a high luminosity at the LHC in order to produce a significant amount of interesting physics events which are created very rarely. In the case of two Gaussian beams colliding head-on, the luminosity can be expressed as a function of the beam parameters [67]:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi\sigma_x\sigma_y}, \tag{3.2}$$

where N_1 and N_2 indicate the number of particles per bunch for the two beams, f is the revolution frequency and $\sigma_i,\ i = x, y$, represents the bunch dimension in the plane transverse to the beams (z-axis). The beam in the LHC is segmented into several packages of protons. Each one of them is called "bunch" and N_b represents the number of bunches per beam in Eq. 3.2.

[1] The acronym derives historically from "Conseil Européen pour la Recherche Nucléaire", i.e. the European Organisation for Nuclear Research. The laboratory is known today as CERN.
[2] Large Electron-Positron collider.

At the LHC the transverse beam sizes are 16.7 µm and the number of particles per bunch is $1.15 \cdot 10^{11}$. The bunches are arranged in "trains" of 72 bunches, with 25 ns spacing within the train, and 12 empty bunches between two trains. Given the revolution frequency of 11 kHz and 2808 bunches, a head-on luminosity of $\mathcal{L}_{\text{LHC}} \approx 1.0 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ is obtained. Considering the expected inelastic cross section of $\sigma_{\text{inel}} \approx 60$ mb [68] for proton–proton collisions at 14 TeV, the estimated head-on luminosity and Eq. 3.1, the expected rate of events at the LHC is $R_{\text{event}}^{\text{LHC}} = \mathcal{L}_{\text{LHC}} \cdot \sigma_{\text{inel}} \approx 6 \cdot 10^{8}$ events/s.

LHC is designed to make bunches of protons collide at a centre of mass energy of 14 TeV at a peak instantaneous luminosity of $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ (as mentioned above) in two high-luminosity insertions that are located at opposite sides of the LHC ring, where the two large general-purpose experiments (designed to investigate a wide variety of physical phenomena), ATLAS and CMS, are placed. Concerning the other two large experiments, ALICE primarily focuses on investigating the physical properties of the Quark-Gluon Plasma (QGP), while LHCb is optimised for dedicated measurements of hadrons containing a *b*-quark.

LHC magnets are made with niobium-titaniun (NbTi) cables and are cooled below 2 K with superfluid helium, in order to reach the superconductivity regime (9.2 K), despite the high currents (11080 A) and large magnetic fields (8.33 T). The large magnetic fields bend the 7 TeV proton beams around the LHC ring: in addition dipole, quadrupole and multipole magnets are used respectively to bend, correct and shrink the beam into the small area where collisions take place.

In order to have two counter-circulating proton beams along the same circumference, oppositely oriented magnetic fields are needed. Due to the limited space, only a single cryogenic structure fits in the tunnel. The issue is solved by employing a complex twin-bore design, having both proton rings in the same cryostat. Figure 3.1 shows the cross section of an LHC twin-bore dipole magnet.

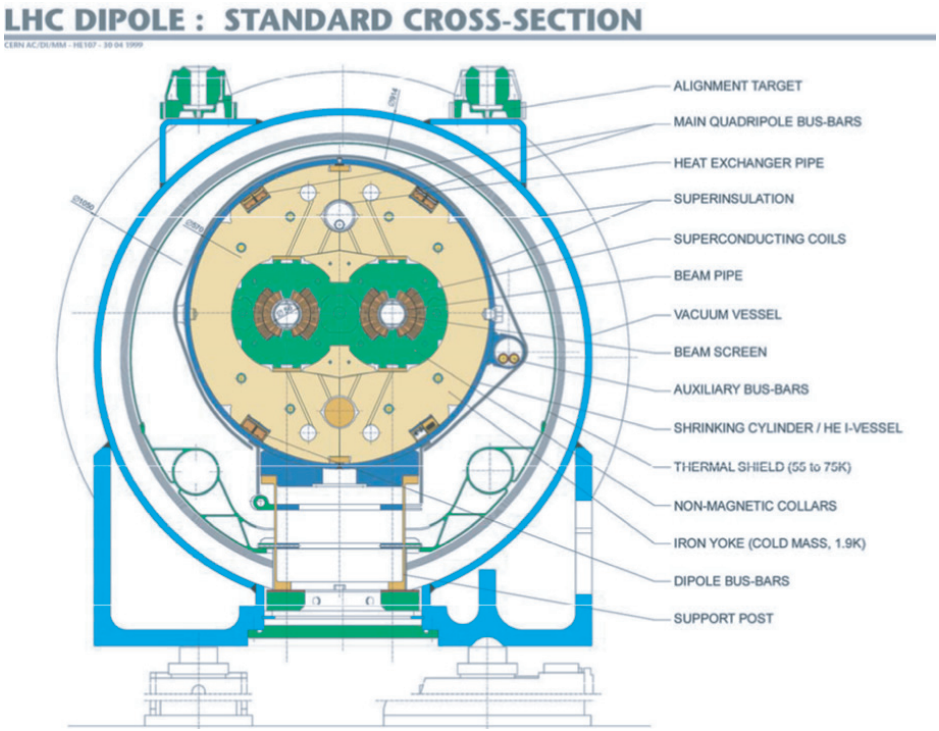


Figure 3.1: Example of an LHC dipole magnet with the twin-bore design [69].

As shown in Figure 3.2, protons start being accelerated in LINAC 2, where their energy is increased up to 50 MeV. The beam is then injected into the Proton Synchrotron (PS) booster and then to PS, where the protons reach an energy of 25 GeV. These protons are then injected into the Super Proton Synchrotron (SPS) that accelerates them to 450 GeV and creates bunches for the LHC injection. Two injection lines exist between the SPS and the LHC, one for each beam pipe of the LHC. Protons are injected in the two counter-rotating lines and they can be accelerated in the LHC up to the final beam energy.

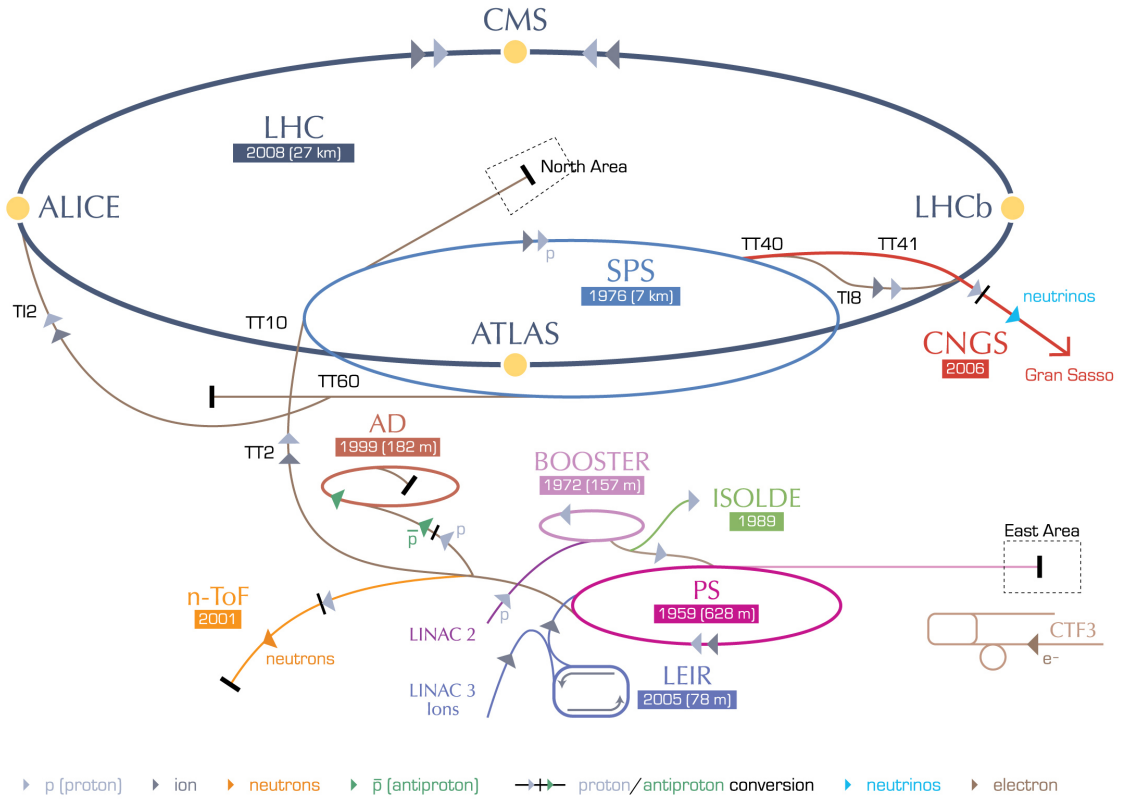


Figure 3.2: Overview of the accelerator complex at CERN [25].

In the LHC, high-luminosity collider par excellence, there is a non-negligible probability for one single bunch crossing to produce several separate inelastic interactions, the so-called “pileup” events; even collisions from preceding and subsequent bunch crossings can contribute to pileup in sub-detectors whose readout time is longer than 25 ns. The density of pileup events has a significant impact on the reconstruction efficiency of the primary vertices at the interaction point, as discussed in Section 3.4.

3.1.1 LHC operation

The LHC began to operate in November 2009 with collisions at a centre-of-mass energy $\left(\sqrt{s}\right)$ of 900 GeV, with the centre-of-mass energy rising to 2.36 TeV by the end of that year. In 2010 the centre-of-mass energy was successfully increased to 7 TeV. During the years 2010 and 2011 the LHC continued to run at $\sqrt{s} = 7$ TeV, with the instantaneous luminosity steadily increasing. In 2010 and 2011 the LHC delivered 48.1 pb^{-1} and 5.46 fb^{-1} of integrated luminosity to ATLAS. In 2012 the centre-of-mass energy

was increased to 8 TeV, and the instantaneous luminosity further increased, leading to a total integrated luminosity of 22.8 fb^{-1} delivered to ATLAS in 2012.

The second phase of LHC, called Run2, started in 2015, after a long shutdown, with collisions at a centre-of-mass energy of 13 TeV. The peak instantaneous luminosity achieved is $2.1 \cdot 10^{34}\,cm^{-2}s^{-1}$, at a bunch crossing of 25 ns. The total integrated luminosity delivered to ATLAS is 158 fb^{-1}, corresponding to $\sim$ 140 fb^{-1} of data good for physics analyses. Figure 3.3 shows the integrated luminosity delivered to ATLAS as a function of time for the 2011-2018 period. The 2015–2017 period of data taking, corresponding to 93 fb^{-1} of integrated luminosity delivered to ATLAS, has been exploited for the searches outlined in Chapters 5 and 6. This corresponds to a dataset of 80 fb^{-1} good for physics analyses.

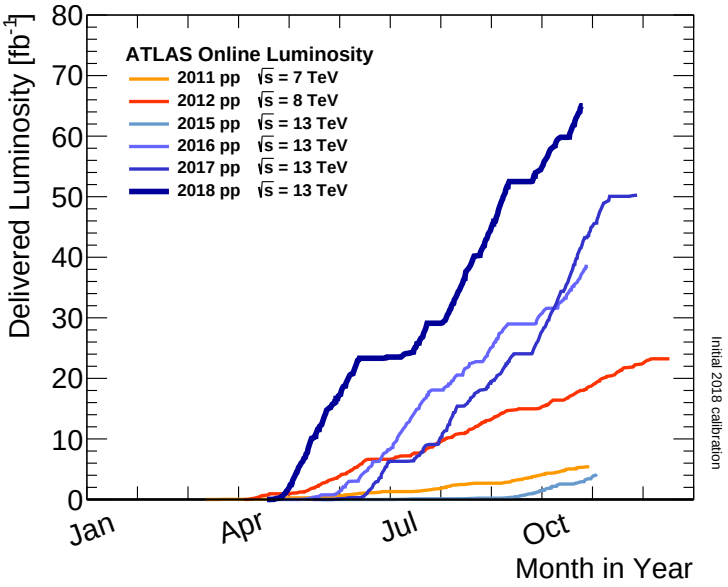


Figure 3.3: Cumulative luminosity versus day delivered to ATLAS during stable beams and for high energy proton–proton collisions [70].

3.2 The ATLAS detector

ATLAS ("**A** **T**oroidal **L**HC **A**pparatu**S**") [64–66] is a multi-purpose detector built around one of the LHC high luminosity interaction points; it is 44 m long, 25 m high, weights over 7000 tons and covers almost the entire 4π solid angle with a forward-backward symmetry. It is designed to do robust pattern recognition, both primary and secondary vertices measurements for charged particles with transverse momentum as low as 100 MeV and to have excellent momentum resolution and energy resolution for neutral particles. Figure 3.4 shows a schematic representation of the ATLAS detector.

The high luminosity and the high centre-of-mass energy of the LHC proton–proton collisions allow to explore physics at the TeV scale. The ATLAS detector has been designed to allow for several types of research in particle physics:

- the Higgs-boson search and the measurement of its fundamental properties;
- the measurement of the fundamental properties of the top quark;

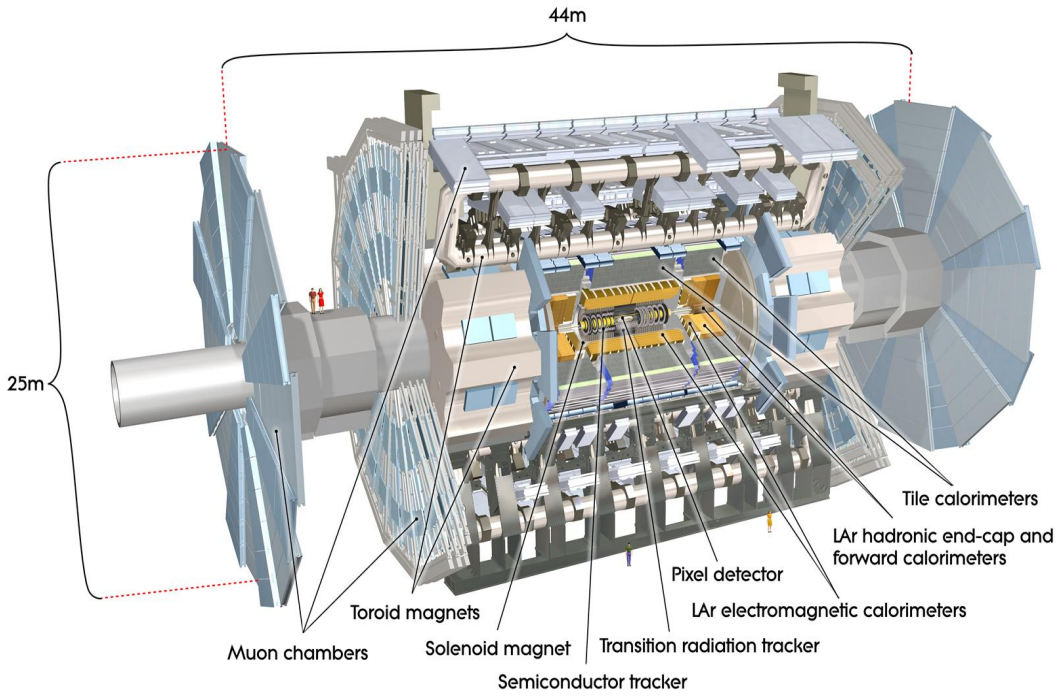


Figure 3.4: Cutaway view of the whole ATLAS detector and its sub-systems [71].

- high precision tests of QCD, flavour physics and electroweak interactions;
- the search for Supersymmetry-like extensions of the SM;
- exotic searches, e.g. searches for new vector bosons and extra-dimensions.

The rate of inelastic proton–proton interactions, as discussed in Section 3.1, is $\sim 6 \cdot 10^8$ events/s. This allows to study rare processes, but on the other hand requires a very fast response of the detector: interesting events, with relatively small cross sections, need to be distinguished from background events, with significantly higher rates. The mean number of interactions per bunch crossing during the 2017 ATLAS data taking period was 38. The ATLAS detector has been designed in order to cope with this high rate of events, as well as with damages caused by the resulting radiation. These challenges have been tackled by ensuring:

- large acceptance in pseudorapidity and full azimuthal coverage, which allows for a representative measurement of missing transverse momentum;
- efficient and fast triggering on low transverse-momentum objects;
- accurate tracking, which allows for high charged particle momentum resolution and reconstruction efficiency, as well as a precise reconstruction of secondary vertices to identify τ leptons and b-hadrons;
- accurate electromagnetic calorimetry, to identify electrons and photons;

- large-coverage hadronic calorimetry, for precise jet measurements;
- accurate and efficient identification of muons.

The following sections are devoted to a more detailed description of the ATLAS sub-systems. A full description can be found in Ref. [66].

3.2.1 Coordinate system

The coordinate system of the ATLAS detector is a right-handed Cartesian system, oriented such that the z-axis is in the beam direction, the x-axis points to the centre of the LHC ring and the y-axis points vertically upwards, as shown in a schematic representation provided in Figure 3.5.

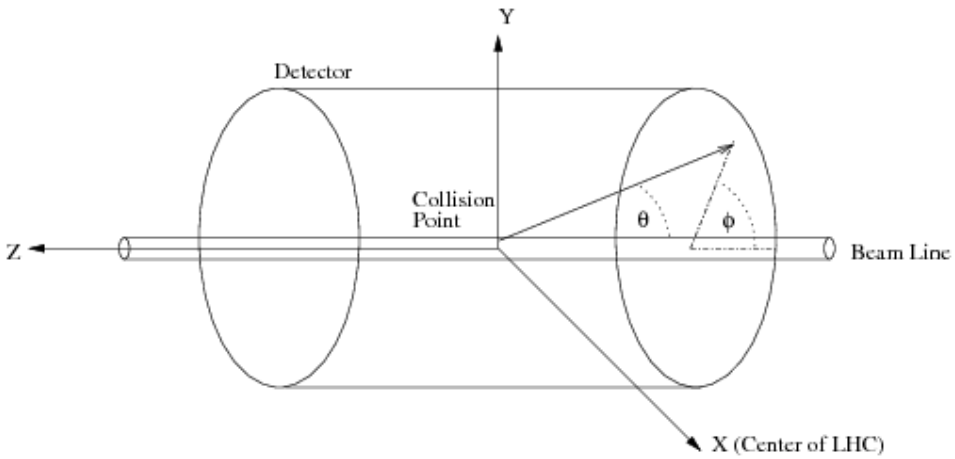


Figure 3.5: Illustration of the ATLAS system of coordinates [72].

Angular distances are often measured in terms of the pseudorapidity, which is defined as:

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

where θ is the polar angle, i.e. the angle measured from the beam axis (z-axis). The two angles, θ and ϕ (the azimuthal angle), are measured from the positive z-axis and from the positive x-axis respectively. The pseudorapidity tends, for relativistic particles with $E >> m$, to the rapidity $\left(y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}\right)$, whose differences are Lorentz invariant under a boost along the longitudinal direction. A commonly used quantity is the angular distance between objects in the pseudorapidity-azimuthal plane, defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3.4}$$

The energy and momentum of the outgoing particles are usually projected onto the transverse $(x-y)$ plane, and are, thus, perpendicular to p_z. The transverse momentum conservation allows to impose a condition on the sum of the transverse momenta of all physical objects, since its initial total component is known to be zero, whereas the initial component along the z axis is unknown. Transverse momentum and transverse energy are defined as $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$ and $E_{\mathrm{T}} = E\sin\theta$, respectively.

3.2.2 Detector sub-systems

The ATLAS detector consists of several sub-detectors which are arranged in an onion-like layered structure to provide an angular uniform coverage. Going from the interaction point to the outside, the sub-systems are:

- the *Inner Detector* (ID), surrounded by a magnet that produces an axial magnetic field of 2 T, dedicated to the reconstruction of the tracks of charged particles and the measurement of their momentum;
- the *Electromagnetic and Hadronic Calorimeters* (ECAL and HCAL), absorbing electrons, photons and hadrons in order to measure their energy;
- the *Muon Spectrometer* (MS), represents the outermost layer, as muons escape the two calorimeters, and provides additional measurements that can be used to independently reconstruct muons.

Figure 3.6 shows each of the ATLAS layers with some sample particle species and their representative signatures in each of the sub-detectors. The dashed lines for a given particle type mean that the particle is not detected in that part of the detector. The trajectories of charged particles such as protons and electrons are bent in the ID, whereas neutral particles such as neutrons are invisible to it. Photons, electrons, protons and neutrons create showers in the calorimeters, while muons reach the muon chambers, and neutrinos are not detected, as they interact very rarely with matter.

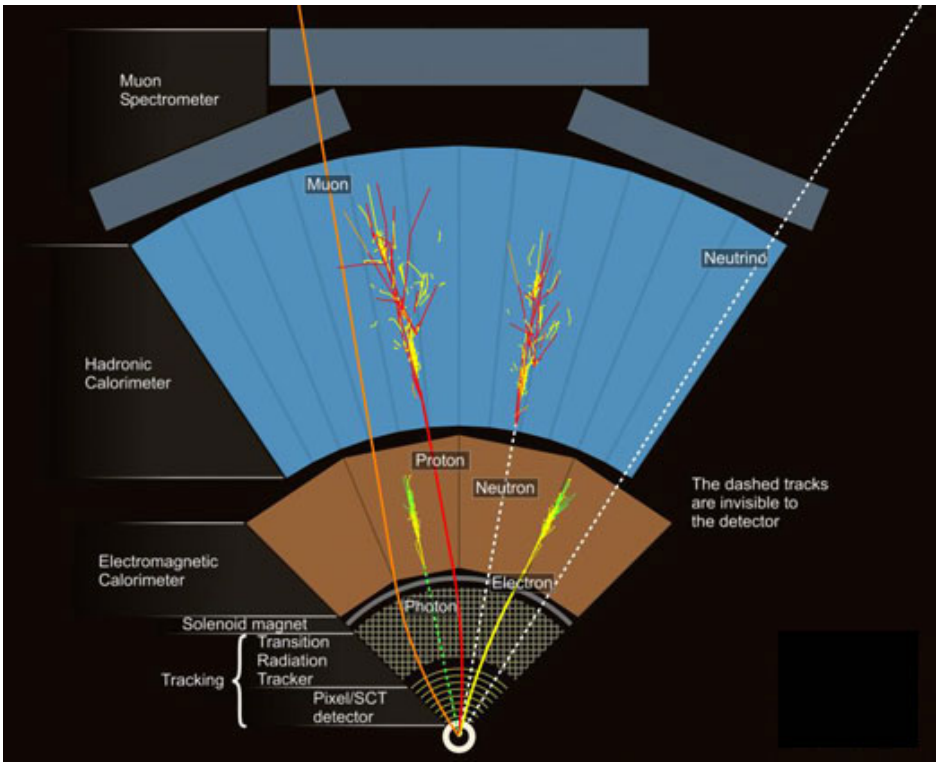


Figure 3.6: Cross-sectional view of the ATLAS detector; simulated particles and their interactions within the detector are overlaid.

3.2.3 Inner Detector

The ATLAS Inner Detector is designed to provide high-precision measurements of the position and momentum of charged particles produced during the collisions. It is contained in a cylindrical enclosure of length 7 m and radius 1.15 m and is surrounded by the barrel solenoid magnet, which produces a magnetic field of 2 T. The ID reconstructs charged tracks, whose p_{T} is greater than 0.5 GeV, and within the pseudorapidity range $|\eta| < 2.5$.

Four sub-systems form the Inner Detector: the Insertable B-Layer (IBL), a Pixel detector and a Semi-Conductor Tracker (SCT), all implemented using silicon sensors, and a Transition Radiation Tracker (TRT), exploiting ionisation caused by the transition of charged particles in straw tubes filled with gas.

Figure 3.7 shows a sketch of a segment of the ATLAS ID barrel modules, while Table 3.1 gives an overview of the fundamental properties of all ID components.

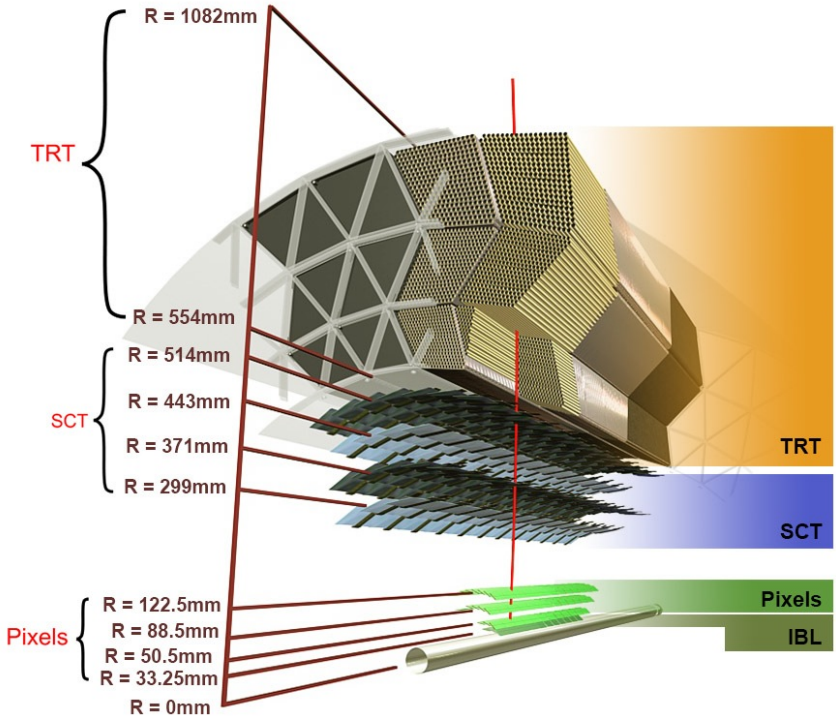


Figure 3.7: A sketch of a segment of the ATLAS ID barrel modules, showing the radial layout of the detection sub-systems. The central grey cylinder is the LHC beam pipe. Visible is the IBL pixel layer, which has been added for Run 2. Not shown are the pixel, SCT, and TRT end-cap modules placed at each end of the barrel, aligned perpendicularly to the beam pipe [73].

Silicon pixel tracker and Insertable B-Layer (IBL)

The silicon pixel tracker and the Insertable B-Layer (IBL) represent the innermost part of the ID. The silicon pixel tracker plays an important role in accurately identifying the multiple collision vertices produced within the proton–proton interaction region, as well as secondary vertices from particles containing a b-quark. It is arranged in ten layers: four cylindrical barrel layers concentrically surrounding the beam pipe and six disk layers (called “endcaps”), three at the end of each of the two sides of the

Table 3.1: ATLAS Inner Detector sub-systems and some of their fundamental properties. R_{min} and R_{max} define detectors extensions in the radial direction $\left(r = \sqrt{x^2 + y^2}\right)$.

Detector	$\lvert\eta\rvert$	R_{min} [mm]	R_{max} [mm]	$r - \phi$ [μm]	z [μm]	Material
IBL	< 2.7	33.2	33.3	8	40	Silicon
Pixel	< 2.5	50.5	122.5	10	115	Silicon
SCT	< 2.5	299	514	17	–	Silicon
TRT	< 2.5	554	1082	130	–	$Xe/CO_2/O_2$ + $Ar/CO_2/O_2$

barrel. Each layer contains pixel modules equipped with a sensor area and various readout electronics. Excluding the innermost barrel layer, there are 1744 pixel modules with dimensions 19 mm × 63 mm each. To meet the stringent specifications on resolution, occupancy and radiation-hardness, the sensors are made of oxygenated n-type silicon wafers of thickness 250 μm. Each sensor contains 47232 pixels with a nominal pixel size of 50 μm in the $r - \phi$ plane and 400 μm along the z-axis. This allows to reach a spatial resolution of $(r - \phi) \times z = 10\,\mu\text{m} \times 115\,\mu\text{m}$.

One of the main concerns before Run 2 was that, with the increasing luminosity, significant radiation damage to the ID could occur; this would lead to a loss in tracking efficiency, especially affecting b-tagging. To cope with this issue, an insertable layer (which can be replaced when damaged), the IBL, has been built; in order to integrate it, the beam pipe outer diameter has been shrunk from 29 to 23.5 mm and the new detector has been inserted into the gap between the Pixel detector and the pipe (see Figure 3.8). The IBL, being on average 33.2 mm away, is currently the closest ATLAS detector to the beam pipe. It assures a high p_{T} and impact parameter resolution, vertex reconstruction and, therefore, is very effective improving b-tagging performance. It uses fast read-out electronics, two different silicon sensor technologies, reduced pixel sizes of $(r - \phi) \times z = 50\,\mu\text{m} \times 250\,\mu\text{m}$ and new carbon foam structures to support the modules. The reduced pixel size provides a spatial resolution of $(r - \phi) \times z = 8\,\mu\text{m} \times 40\,\mu\text{m}$. Including the IBL, the ID has approximately 88.4 million electronic readout channels.

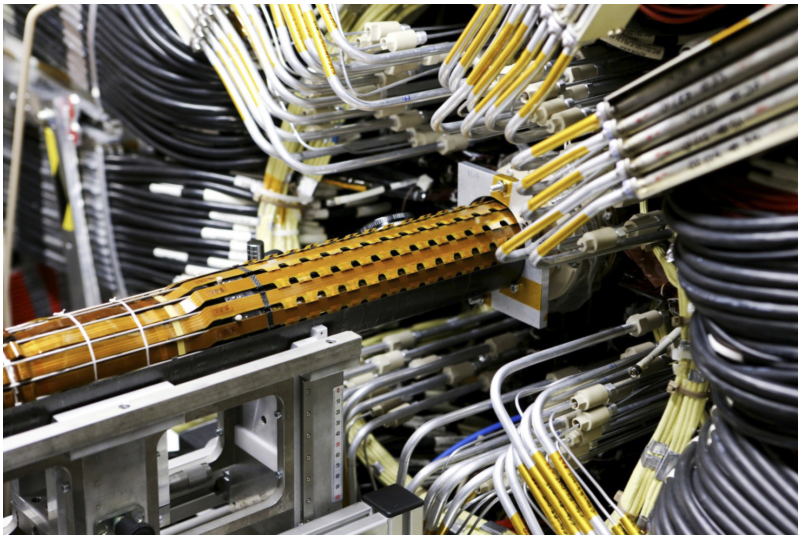


Figure 3.8: Insertion of the IBL into the ATLAS Detector in May 2014 [74].

Semiconductor Tracker (SCT)

The Semiconductor Tracker (SCT) surrounds the pixel detector and is the second layer of the ID. It is arranged in twenty-two layers: four cylindrical barrel layers and eighteen disk layers, nine on each of the endcaps. In the barrel region, the SCT is designed to provide at least four precision space-point measurements in the $(r-\phi)$ and z coordinates, using four pairs of small-angle stereo strips. The stereo silicon modules are created by laying out two individual strips at an angle of 40 mrad. Each layer is made of p-n silicon semiconductor modules of nominal size 6.36 cm × 6.40 cm with 780 readout strips. Each strip is 12 cm long and has a constant pitch of 80 µm. The spatial hit resolution of the strips is $(r-\phi) \times z = 17\,\mu\text{m} \times 580\,\mu\text{m}$. The end-cap modules have a very similar structure, but exploit tapered strips, where one set is aligned radially. The SCT has a total of 6.3 million readout channels.

Transition Radiation Tracker (TRT)

The outermost layer of the ID is the Transition Radiation Tracker (TRT). It is made of 4 mm diameter polyamide tubes filled with a mixture of gases: 70% xenon, 27% CO_2 and 3% O_2. At the centre of each tube, there is a 31 µm diameter tungsten wire plated with 0.5 − 0.7 µm gold held in place with an end-plug. The barrel region has 50 thousand longitudinally-arranged tubes with length 144 cm, and in the end-caps there are 320 thousand radially-arranged tubes with length 32 cm. The dielectric material used to interleave the straw tubes provides transition radiation for traversing relativistic charged particles, that can be used to distinguish electrons from pions based on their energy deposition. The tube wall is kept at a high voltage of −1.5 kV and acts as the cathode, while the wire is kept at ground to act as the anode. As charged particles cross a tube, they ionise the gas, creating electrons that consequently drift to the anode. This drift-time measurement provides a signal proportional to the energy of the particle and, on average, each particle track hits 36 tubes. Each tube provides a spatial hit resolution of 130 µm in a plane perpendicular to the wire. The total number of TRT readout channels is approximately 351 thousand.

3.2.4 Calorimeters

The ATLAS calorimetry system, surrounding the Inner Detector, is dedicated to the measurement of the energy of particles produced during the collisions. It is finely segmented in the η and ϕ directions, and covers both $|\eta| < 4.9$ and the full azimuthal range. It is composed of five sub-systems: the LAr ElectroMagnetic Barrel calorimeter (EMB), the LAr ElectroMagnetic End-Cap calorimeter (EMEC), the Tile barrel hadronic Calorimeter (TileCal), the Hadronic End-Cap calorimeter (HEC) and the LAr Forward Calorimeter (FCal).

The Electromagnetic Calorimeters (ECal) measure the energy of particles that interact electromagnetically producing electromagnetic showers, e.g. electrons and photons (see Figure 3.6). On the other hand, the Hadronic Calorimeters (HCal) measure the energy of particles that interact via the strong force, e.g. pions and kaons. Particles that interact both electromagnetically and strongly, deposit energy in both the ECal and the HCal. Figure 3.9 shows a cutaway view of the calorimeter system of ATLAS. The ATLAS calorimeters are sampling calorimeters, i.e. made of alternating layers of active and passive material. Incoming particles produce a cascade of successively lower-energy particles (forming the so-called "particle shower") by interacting with the dense passive material of the calorimeters. The cascade continues until the entire energy of the incoming particle is exhausted. The active layers collect the energy of particles via ionisation (ECal) or scintillation (HCal) and the passive layers act as pure absorbers.

An overview of the fundamental properties of all ATLAS calorimetry components is provided in Table 3.2.

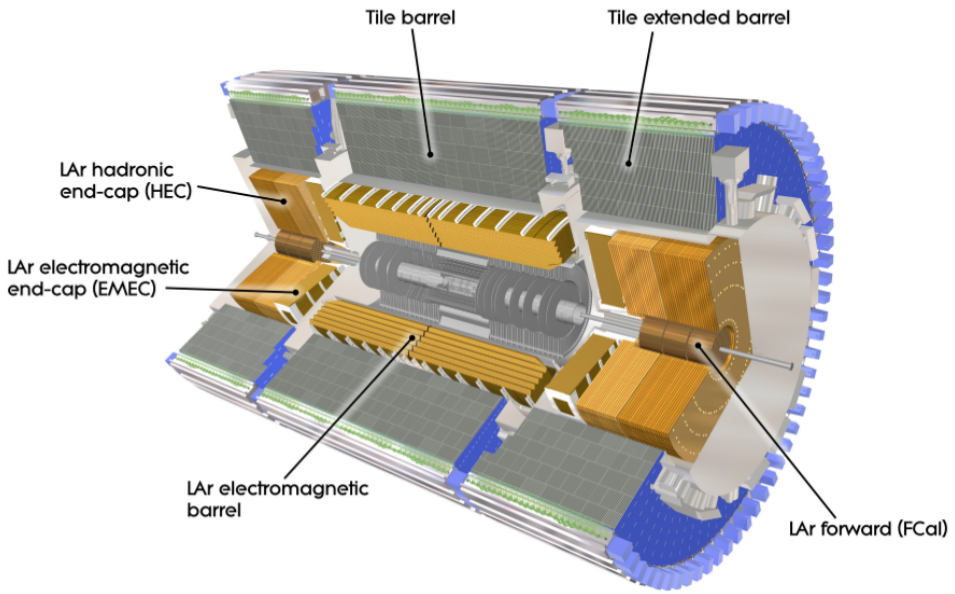


Figure 3.9: Cutaway view of the ATLAS calorimeter system [66].

Table 3.2: ATLAS calorimeters properties; R_{min} and R_{max} define the detectors geometrical extension and E stands for the deposited energy.

Detector	$\lvert\eta\rvert$	R_{min} [m]	R_{max} [m]	Resolution	Material
ECal	< 4.9	1.5	2	$\sigma_E/E = 10\%/\sqrt{E}$	lead (or copper)/liquid argon
HCal	< 4.9	2.25	4.25	$\sigma_E/E = 50\%/\sqrt{E}$	steel/polystyrene scintillators (or liquid argon)

Electromagnetic Calorimeters

The electromagnetic calorimeters directly surround the ID and the barrel solenoid magnet. They are made of lead and Liquid-Argon (LAr) detectors with accordion-shaped Kapton electrodes, where the accordion geometry provides a full, gap-less azimuthal coverage. The liquid argon serves as the active material and was chosen due to its radiation hardness, while the lead absorber plates act as the passive material. The electromagnetic calorimeters are divided into three parts: the barrel (EMB) ($|\eta| < 1.475$), the end-caps (EMEC) ($1.375 < |\eta| < 3.2$) and the first section of the forward calorimeters, known as "FCal1" ($3.1 < |\eta| < 4.9$).[3] The EMB is made of two half-barrels, is 6.4 m long and has an inner and outer diameter of 2.8 m and 4 m, respectively. In total, the EMB is made of 2 048 accordion-shaped absorbers, interleaved with readout electrodes. The electrodes are positioned in the middle of two absorbers (2.1 mm from each absorber) by honeycomb spacers. The EMB is segmented in three layers in depth, as shown in Figure 3.10. The first layer is finely segmented in the η direction. The second layer has square cells of dimension $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ and the third layer has twice the granularity in η. A separate 11 mm deep LAr layer, known as the pre-sampler (PS), is inserted in front of the first layer and provides a coverage of the $|\eta| < 1.475$ region. Including the PS, the EMB has more than 109 thousand readout cells. The PS, three EMB layers and the vast number of cells provide excellent electromagnetic shower

[3] FCal1 has copper absorber plates instead of lead.

sampling. The EMEC consists of two wheels, one on each side of the EMB. It is also segmented into three layers in depth with an additional PS layer covering $1.5 < |\eta| < 1.8$. In total, each end-cap has almost 32 thousand readout channels.

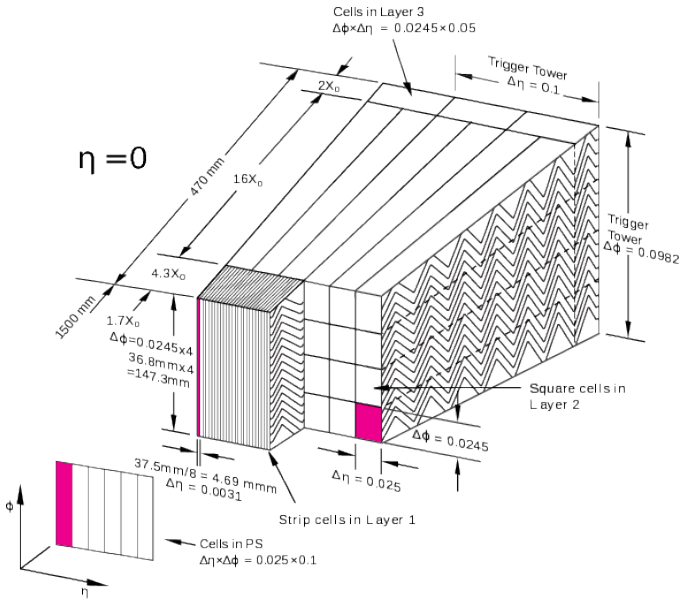


Figure 3.10: Sketch of a barrel module, where the three different layers are shown [66].

Hadronic Calorimeters

The hadronic calorimeters surround the ECal. They consist of the barrel TileCal ($|\eta| < 1.7$), the end-cap HEC ($1.5 < |\eta| < 3.2$) and the two remaining forward calorimeters: FCal2 and FCal3 ($3.1 < |\eta| < 4.9$). The calorimeters use steel as the absorber and polystyrene scintillating tiles as the active medium. The TileCal is divided into two regions: the barrel ($|\eta| < 1.0$) and the extended-barrel ($0.8 < |\eta| < 1.7$) region. Both regions are divided azimuthally into 64 modules that are further split into three layers. The modules extend from an inner radius of 2.28 m to an outer radius of 4.25 m. A single module with alternating steel and scintillating tiles is shown in Figure 3.11. Wavelength-shifting fibers are used to connect the tiles to PhotoMultiplier Tubes (PMT) at the edge of the modules; this allows to match the scintillator wavelength to the PMT sensitivity. The PMTs amplify the scintillator signal produced due to passing particles and convert it to an electrical signal. In total the HCal exploits 9 852 PMTs. The HEC contains two separate wheels per end-cap. They are located directly behind the EMEC. Each wheel is built using 32 wedge-shaped modules that contain copper plates (passive material) interleaved with LAr (active material). The FCal2 and FCal3 detectors use tungsten as the passive material, and LAr as the active material. Each module contains a metal matrix with electrode channels parallel to the beam axis. The HEC and FCals share the cryostat with the EMEC.

3.2.5 Muon spectrometer

The muon spectrometer (MS) is the outermost part of the ATLAS detector. It is a set of detector chambers that are designed to specifically detect and measure the position and momentum of muons passing through

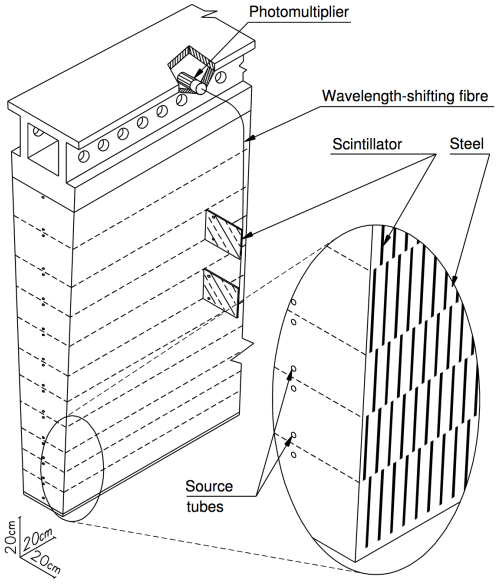


Figure 3.11: Schematic representation, showing how the mechanical assembly and the optical readout of the tile calorimeter are integrated together. The various components of the optical readout, namely the tiles, the fibers and the photomultipliers, are shown [66].

the ID and the calorimeters. The bending power of the toroid magnets, located in the barrel and end-cap sides, allow to bend the muon trajectories over a large distance. In order to measure the curvature of the tracks, a very good hit resolution is needed. In addition to detection, the MS is also designed to trigger on detected particles, as discussed in Section 3.3.

The muon spectrometer contains four different kinds of detectors as shown in Figure 3.12: the Monitored Drift Tube Chambers (MDT), the Resistive Plate Chambers (RPC), the Thin Gap Chambers (TGC) and the Cathode Strip Chambers (CSC). For precision tracking in the barrel region, a combination of MDTs and RPCs is arranged in three concentric cylindrical shells around the beam pipe at radii of approximately 5 m, 7.5 m, and 10 m. The MDT chambers contain three to eight layers of 30 mm diameter, pressurised drift tubes operating with Ar-CO_2 gas (93/7%) at 3 bar. The tube acts as the cathode and contains a 50 μm gold-plated tungsten-rhenium wire which acts as the anode, kept at a potential of 3 kV. Muons ionise the gas mixture in the tubes to create electrons (which are attracted to the wire) and positive ions (which drift towards the cathode). The electrical signal obtained from the wire provides information about the passing muon. Each MDT tube has a space resolution of 80 μm and a time resolution of less than 1 ns. The RPCs consist of parallel electrode-plates made of phenolic-melaminic plastic laminate. Two resistive plates are separated by 2 mm using insulating spacers; the electric field between the plates allows electrical signals, due to electron avalanches produced by ionising muon tracks, to form. The resistive plates are kept at a potential difference of 9.8 kV and the chamber is filled with a gas mixture of $C_2H_2F_4$, C_4H_{10} and SF_6 (94.7, 5 and 0.3% respectively). The RPCs provide good time resolution (less than 2 ns) and are used to trigger on muons.

In the end-cap regions, the muon chambers are arranged in eight large wheels at distances of 7.4 m, 10.8 m, 14 m, and 21.5 m from the interaction point. The chambers used are the MDTs, the TGCs and the CSCs. The MDTs provide precision muon tracking and the TGCs provide fast and efficient tracking information to trigger on (their time resolution is smaller than the bunch spacing, i.e. 25 ns), as well as

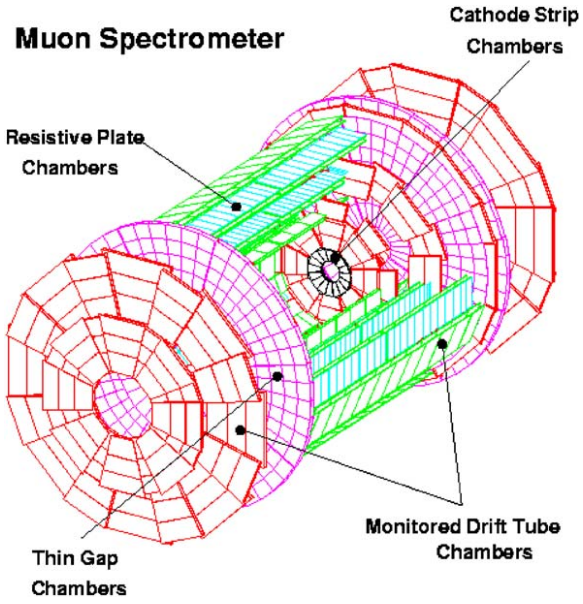


Figure 3.12: Sketch of the ATLAS muon spectrometer [75].

an azimuthal coverage which is complementary to the MDT. TGCs are multi-wire proportional chambers filled with a highly quenching gas mixture of CO_2 and n $-$ C_5H_{12}. The wire-to-wire distance is 1.8 mm and the wire-to-cathode distance is 1.4 mm. The high voltage at which the wires are kept (2.9 kV) and the small distances between the wires and the cathode strips allow to achieve a time resolution of 4 ns.

In the $|\eta| < 2$ region, the innermost wheels contain the CSC chambers. Like the TGCs, the CSC chambers are multi-wire proportional chambers with wires running in the radial direction. The wires are kept at a voltage of 1.9 kV and are filled with a gas mixture of argon-CO_2 (80/20%). These chambers provide good tracking (60 µm) and good timing resolution (below 40 ns). More details on the ATLAS muon spectrometer can be found in Ref. [76].

3.3 Trigger and data acquisition

The total rate of inelastic collisions at the LHC, as ascertained in Section 3.1, is $R_{\text{event}}^{\text{LHC}} \approx 600$ MHz; while the total inelastic cross section is equal to $\sigma_{\text{inel}} \approx 60$ mb [68], the cross section of interesting events is several orders of magnitude smaller, as shown in Figure 3.13. Offline computing power and storage capacity have been improved after the Run 1 data-taking campaign, and are now compatible with an acquisition rate of 1.5 kHz[4] (it was 600 Hz in Run 1); in order to reduce the rate up to the latter value, selections at the different trigger levels must provide sufficient rejection of non-interesting inelastic collisions. Therefore it is really important to have a very efficient and fast trigger selection of the events we are interested in. In the 2015–2017 period, LHC has operated with a bunch crossing every 25 ns, where each bunch crossing lead, on average, to more than 30 inelastic interactions (see Figure 3.14).

Upgrades to the ATLAS trigger and data acquisition (TDAQ) system for Run 2 have led to the development of a two-level trigger for scheme simplification and dynamic sharing of computing resources. Starting from the three levels of Run 1, Level-2 (L2) and Event Filter (EF) triggers have been merged

[4] Corresponding to a $\sim$ 1.5 GB/s data storage bandwidth for a typical event size of 1 MB.

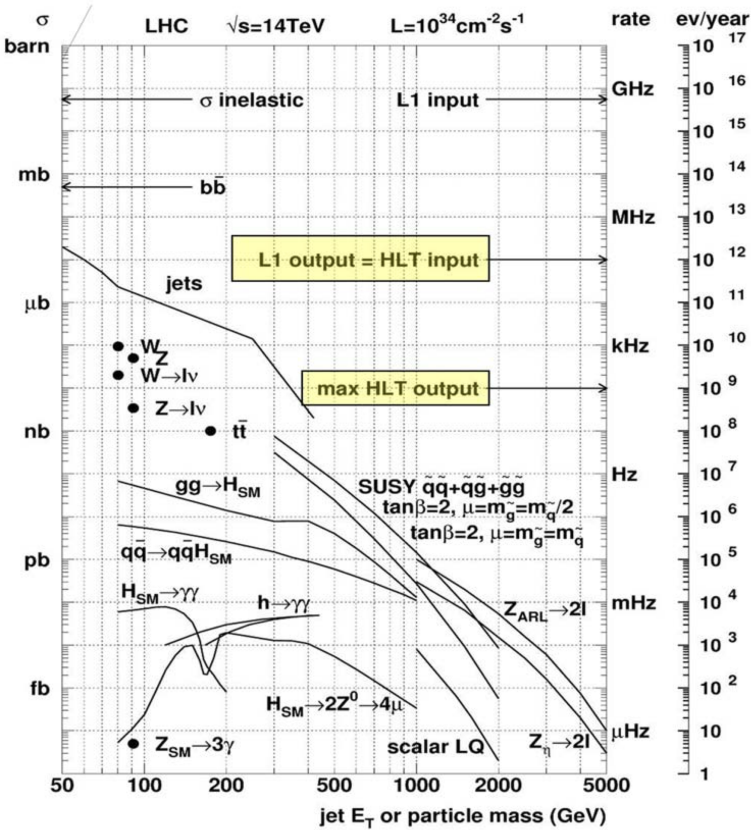


Figure 3.13: pp total inelastic cross section (σ inelastic), cross sections of interesting processes and corresponding production rates at 14 TeV. Input and output rates for L1 ("Level-1") and HLT ("High Level Trigger") triggers are shown [77].

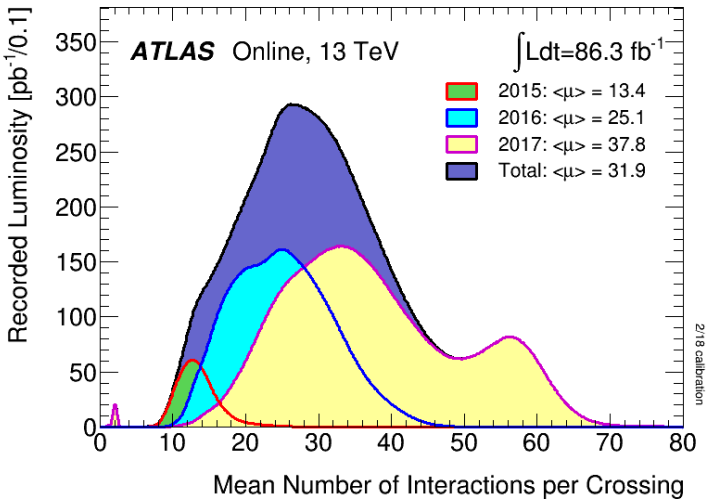


Figure 3.14: Luminosity-weighted distribution of the mean number of interactions per crossing for the 2015–2017 pp collision data; it corresponds to $\mu = \mathcal{L}_{\text{bunch}} \times \sigma_{\text{inelastic}} / f$, where $\mathcal{L}_{\text{bunch}}$ is the per-bunch instantaneous luminosity, $\sigma_{\text{inelastic}}$ is the inelastic cross section and f is the LHC revolution frequency [70].

into a single "High Level Trigger" (HLT) farm.

The first stage, called Level-1 (L1) trigger [78], is a synchronous, pipelined system that operates at the LHC bunch crossing frequency (1/25 ns = 40 MHz). The L1 selection reduces the event rate down to 100 kHz, using information from the calorimeters and from dedicated muon trigger detectors. In particular the L1 calorimeter trigger decision is based on the multiplicities and energy thresholds of the following objects detected by the ATLAS calorimeter sub-system: electromagnetic clusters, taus, jets, missing transverse momentum,[5] scalar sum of transverse momentum ($\sum E_{\mathrm{T}}$) in the calorimeters, and total transverse momentum of reconstructed L1 jets ($\sum E_{\mathrm{T}}$(jets)). These observables are computed by the L1 algorithms, using the measured E_{T} values in trigger towers of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ granularity. As anticipated in Section 3.2.5, L1 also exploits muon information; in fact, the L1 muon triggers exploit the measurement of trajectories in the RPC and in the TGC, which are compared to pre-established templates, to assert the fulfillment of certain p_{T} thresholds. The information coming from the ID tracking system is not used at this stage, since the time needed to reconstruct tracks and vertices exceeds the latency of the L1 system.

The second stage of the trigger system, the HLT, reduces the event rate further up to $\sim$ 1 kHz. It is software-based and uses offline reconstruction algorithms, exploiting information supplied by all sub-detectors in the spatial regions of interests (typically regions identified by the L1 trigger); the size of such regions depends on the type of object being triggered: for instance, a smaller region of interest is used for electron triggers, as compared to jet triggers. With respect to the Run 1 trigger setup, more dedicated (or multi-object) triggers are available in order to cover all interesting event topologies; a large set (called "menu") of trigger selections is implemented, corresponding to $\sim$ 300 L1 and $\sim$ 1000 HLT selections.

Another milestone of the ATLAS operation is the processing of raw data, which consists of two main steps: decoding and reconstruction. The data decoding is the process of transforming raw data from readout electronics, which are written into "bytestream" files, to inputs for the reconstruction process, which produces Event Summary Data (ESD) and Analysis Object Data (AOD) in POOL/ROOT files [79]. It's necessary to have smaller and more customised data files, called derived AOD (xAOD), to be used as input to particular performance studies and physics analyses; in order to produce this reduced data format, event data are filtered in four main steps:

- *skimming*, the selection through event attributes, chosen to support an efficient identification and selection of events of interest to a given analysis;
- *trimming*, the selection of top-level data objects and containers (when fewer top-level data objects are written, fewer proxies are created reading the file);
- *thinning*, the selection of particular data objects in a container, using configurable algorithms that retrieve the original container and copy the selected objects into a new container, which then is written to the output file;
- *slimming*, the selection of interesting information related to objects, done in several ways.

The primary event processing (called "reconstruction") occurs at the Tier-0 facility hosted by CERN, where one copy of the raw data is archived. Another replica of the raw data and the produced xAOD are distributed to the approximately ten Tier-1 facilities spread around the world. The Tier-1 machines take custodial responsibility for the data, and provide reprocessing capacity and hosting for simulated

[5] The missing transverse momentum is a vector in the transverse plane, resulting from the vectorial sum of all the reconstructed transverse momenta with changed sign. See Section 3.4.5 for further details.

data. Each Tier-1, as illustrated in Figure 3.15, serves several Tier-2 facilities, hosted by universities or laboratories and responsible for providing the simulation capacity for the experiment. Hundreds of smaller institutional resources, at the university or laboratory level, serve as Tier-3 facilities for physics analysis. Event reconstruction is discussed in detail in Section 3.4.

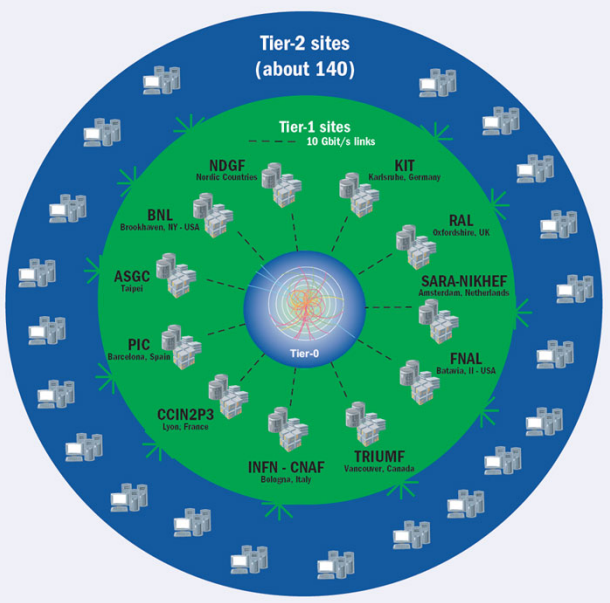


Figure 3.15: Diagram showing the Tier system of Worldwide LHC Computing Grid (WLCG), where the CERN Tier-0 site sends data to the 11 Tier-1 sites and their corresponding Tier-2 sites.

3.4 Reconstruction of physics objects

This section is devoted to a general overview of the reconstruction of physics objects, fundamental ingredient to any physics interpretation, including the works presented in Chapters 4–6; tracks and interaction vertices are reconstructed in the ID and the MS, while clusters of deposited energy are identified in the calorimeter systems. All of this information is combined to reconstruct particles, i.e. muons, electrons, jets, photons and tau leptons, and measure global properties of the event, such as the missing transverse momentum. Tau leptons are not identified as tau reconstructed objects, but their decay products, from leptonic and hadronic decay modes, are reconstructed, identified and treated as electrons, muons, jets and missing transverse momentum.

3.4.1 Track and vertex reconstruction

The tracks of charged particles in the ID are initially reconstructed from hits in the IBL, Pixel detector and SCT, and are used for the object definitions described below. The track reconstruction consists of several steps [80]. All charged particles traversing the ID follow an approximately helical trajectory, because of the presence of the homogeneous magnetic field, and leave hits by interacting with the different components of the ID, as discussed in Section 3.2.3. Pixels and strips in a given sensor, where the deposited energy yields a charge above a threshold, are grouped; these clusters are exploited in order to identify “space-points”, i.e. points in the three-dimensional space where the charged particle traversed

the active material of the ID. A space-point corresponds to a hit in the IBL and Pixel detector, while the SCT space-points correspond to hits on both sides of the module. In dense environments, a single cluster may contain hits from multiple charged particles, called a merged cluster and schematically represented in Figure 3.16.

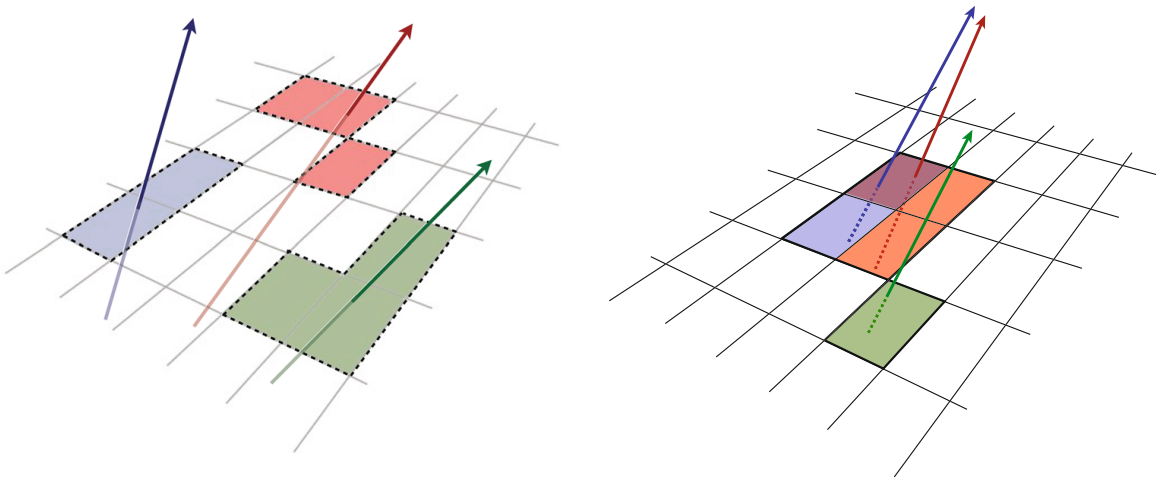


Figure 3.16: Illustration of single-particle pixel clusters on a pixel sensor (left) and a merged pixel cluster (right) due to nearby charged particles. Different colours represent energy deposits from different charged particles traversing the sensors; the particles' trajectories are shown as arrows [80].

The next step is devoted to form "track seeds", by combining sets of three space-points in the four pixel detector layers and the first layer of the SCT. A combinatorial Kalman filter [81] is used to build track candidates from the chosen track seeds, by incorporating additional space-points, which are compatible with the preliminary trajectory and originating from the remaining layers of the Pixel and SCT detectors. The filter creates multiple track candidates per seed if more than one compatible space-point extension is possible on the same ID layer.

As next step, multiple reconstructed track candidates with shared particle clusters are removed according to several track quality requirements, such as: the number of assigned clusters, number of holes (spots without a cluster, where the transit of a charged particle is expected to produce a cluster), the χ^2 of the track fit and the track p_{T}. All of this information is combined to assign a score to each track candidate. When two or more track candidates have a shared cluster, only the candidate having the highest track quality score is kept and other tracks, with a shared cluster and a lower score, are dropped.

The tracks selected through this procedure are then extended into the TRT and re-fitted via a high-resolution fit, taking into account the full information provided from the three sub-detectors [82]. If, according to the fit quality, the extended track is better than the silicon-only track, then the extended one is kept.

The average number of pp interactions per bunch crossing in ATLAS 2015–2017 data events is about 32, as shown in Figure 3.14; thus, each triggered event will be superimposed to several low-p_{T} pp inelastic interactions, commonly labelled as "minimum bias events". Nevertheless it is possible to disentangle effects caused by pile-up collisions from the products of the hard-scattering collision of interest. For this purpose a very accurate knowledge of the position of the hard-scattering vertex, called "primary vertex", is fundamental. This also allows to precisely determine the longitudinal and transverse impact parameters, exploited to select or reject leptons from photon conversion and secondary decays in

jets; the latter case is further discussed in Chapter 4.

In general the reconstruction of vertices can be split into two main stages [83, 84]:

- vertex finding, the association of reconstructed tracks to a given vertex candidate;
- vertex fitting, the reconstruction of the actual vertex position.

The vertex-finding process is run after the reconstruction of ID tracks. These tracks are required to have a transverse momentum of $p_{\mathrm{T}} > 500\,\mathrm{MeV}$, at least nine hits in the IBL, Pixel detector and SCT for $|\eta| < 1.65$ and at least 11 hits for $|\eta| \geq 1.65$. The tracks need to have at least one hit in the first two pixel layers, at most one shared pixel hit or at most two shared SCT hits, exactly zero pixel holes and no more than one SCT hole [85]. Tracks fulfilling these requirements are assigned a seed position for a vertex candidate. An iterative fit procedure between the vertex and the tracks is performed, and tracks are assigned a weight depending on their consistency with the vertex; this process stops when the fit converges. The excluded tracks ("outliers") are used to build a second vertex seed. A fit is performed using the two vertices, and again outlier tracks are used to fit a new vertex. The procedure stops when none of the remaining outliers fits with any vertex gives a χ^2 probability larger than 1%. The primary vertex of each event is chosen, among reconstructed primary vertices, as the vertex with the highest $\sum p_{\mathrm{T}}^2$ of associated tracks.

3.4.2 Jet reconstruction and identification

High p_{T} quarks or gluons, commonly referred to as "partons" are produced in high-energy pp inelastic collisions. Because of the QCD colour confinement (presented in Section 2.1.3) partons cannot propagate as free particles, and spontaneously re-combine, namely "hadronise", with quarks and antiquarks spontaneously created in pairs from the vacuum to form hadrons. The hadronisation process of a single parton gives life to a conglomerate of many particles, condensed in a relatively tight spatial cone, called "jet". The goal of jet reconstruction is the combination of particles produced from the hadronisation in order to form a physics object, whose characteristics allow to infer properties of the initial parton.

The final state measured object is never clearly formed into well separated cone-shaped regions. The jet clustering is affected by several effects, mainly detector resolution, granularity and inefficiencies. It is important to underline that there is neither a single, nor only one correct definition of a reconstructed jet; nevertheless this definition must be consistent with both, theory models and experimental evidence. There are two big families of jet reconstruction algorithms: one based on particle tracking, the other one based on calorimeter energy clusters. The algorithm based on particle tracks is more robust in the presence of high pile-up, even though it does not include information related to neutral particles. The jet reconstruction exploited in the works presented in this thesis is based on the combination of energy deposits (clustering) in the calorimeters; in particular, identified energy deposits are clustered to form a single physics object: a reconstructed jet.

Among all jet definitions based on calorimetric clusters, the two most common ones are the cone jet- and the k_{T} jet-based algorithms.

Anti-k_{T} algorithm

Cone jet algorithms try to find almost circular stable regions of energy in the (η, ϕ) angular plane of calorimeters measurements. This definition leads to an "infrared sensitivity" in perturbation theory [86]. In addition, sometimes nearby clusters of energy that are expected to be part of the reconstructed jet are excluded from the stable jet found by the algorithm (this is on account of the regularity of circular

boundaries). Finally, there is also another issue to tackle: how to share the energy clusters, when two (or more) cones overlap. In the last years this jet definition has been widely replaced by the k_T algorithms, as they allow to cope with all of these problems.

The anti-k_T jet algorithm relies upon the observation that, in a shower, final state particles are predominantly collinear, i.e. have small relative (between their constituent particles) transverse momentum. A k_T algorithm starts by defining an ensemble of distance measures [87]:

$$d_{ij} = \min\left((k_T)_i^{2p}, (k_T)_j^{2p}\right)\frac{\Delta_{ij}^2}{R^2}\,, \tag{3.5}$$

$$d_{i,\text{beam}} = (k_T)_i^{2p}\,, \tag{3.6}$$

where y_i, ϕ_i, and $(k_T)_i$ are respectively the rapidity, azimuth angle and transverse momentum of the particle i and $\Delta_{ij}^2 = \left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2$; R is the resolution parameter and it roughly defines the spatial dimension (ΔR) of the reconstructed jet. The k_T algorithm is iterative: it seeks the minimum among all these distance measures and, if the smallest one is found to be a d_{ij} (the distance between two particles), it sums the four momenta of the two particles, it updates distances and proceeds finding the new smallest one. On the other hand, if the smallest one is a $d_{i,\text{beam}}$ (the distance between particle i and the beam axis) it stops: the jet reconstruction is complete and the final jet object is removed from the event. This procedure continues until all the energy clusters in the event are grouped into jets. Figure 3.17 shows a comparison between typical geometrical structures of jets reconstructed via cone and k_T algorithms.

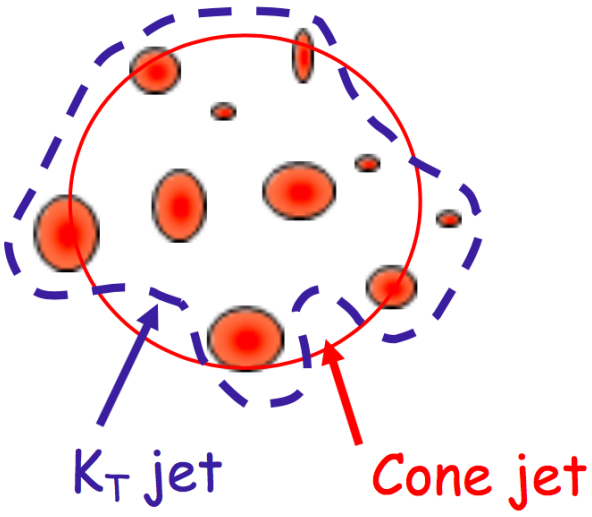


Figure 3.17: Different jet structure for cone and a generic k_T algorithm. The cone algorithm uses a rigid boundary, while the k_T algorithms show a good resilience, allowing to reconstruct a more amorphous jet [86].

The parameter p is introduced to tune the relative influence of energy (k_T) and geometrical (Δ_{ij}) scales; a very interesting case is $p < 0$, which corresponds to the *anti-k_T* algorithm. Anti-k_T algorithms, in particular with $R = 0.4$ and $p = -1$, are very reliable and, thus, currently widely exploited algorithms within the ATLAS Collaboration; this jet definition is assumed in the following chapters. The reason of this success is due to the fact that taking $p < 0$ yields an algorithm that is infrared- and collinear-safe: in the presence of soft radiation the clustering algorithm always reconstructs the same number of jets in the event and the splitting of one parton into two partons does not change the result of the jet clustering.

Also, the shape of anti-k_T jets is approximately conical, and as such calibration is easier. Further details on the anti-k_T jet algorithm can be found in Ref. [87].

Local fluctuations in the pile-up activity may result in spurious jets, originating from a pile-up vertex. In order to reduce this contamination a discriminant, called the jet-vertex-tagger (JVT) [88], is exploited; the JVT combines track-based variables, such as the fraction of the total momentum of tracks in the jet which is associated with the primary vertex, via a k-nearest neighbour (k-NN) algorithm (see Section 4.2.3).

The identification of jets originating from the hadronisation of a b-quark (known as "b-tagging") is described in detail in Section 4.3.

3.4.3 Electron reconstruction and identification

Electrons are produced in many interesting physics processes, but are also subject to a large contamination from hadrons and electrons originating from photon conversion and heavy-flavour decay. Therefore, it is of vital importance to efficiently identify electrons originating in the primary vertex and reject background sources. For this purpose, the reconstruction of electrons is based on a combination of information from the tracking and calorimeter systems. The silicon detectors and the TRT are used to identify the electron track, while the electromagnetic calorimeter system allows to measure the energy deposition of the electron candidate; finally, hadronic calorimeters are used to veto particles characterised by an intense hadronic activity.

The track reconstruction described in Section 3.4.1 is exploited in order to reconstruct the track of the electron candidate, while energy clusters in the electromagnetic calorimeter are split in the $\eta - \phi$ space, namely in cells of size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ (i.e. the granularity of the EM calorimeter middle layer). According to the latter units, a sliding window with a size of 3×5 is used to search for electron cluster "seeds" as longitudinal towers with total cluster transverse energy above 2.5 GeV. Seeds are merged across the longitudinal layers to form clusters, using the *sliding-window algorithm* [89]. The cluster kinematics are reconstructed using an extended window, whose dimensions depend on the cluster position in the calorimeter. The clustering efficiencies are 95% for transverse energy above 7 GeV and 99% for a transverse energy above 15 GeV.

After the reconstruction of EM clusters, the "loose" track reconstruction proceeds from the track seeds produced by the inner detectors (see Section 3.4.1). This reconstruction has two steps. The first one is the pattern recognition using the energy loss information. Two pattern recognitions corresponding to the pion hypothesis and electron hypothesis are considered. If the track seeds have transverse momentum above 1 GeV, the pion pattern recognition algorithm is discarded. In this case, the electron pattern recognition algorithm, which allows large energy loss, is employed. In case the track reconstruction passes the electron hypothesis, the specific track re-fit algorithm proceeds. This algorithm requires the match between tracks having reconstructed EM clusters and more than four silicon detector hits. These tracks are reconstructed by means of a Gaussian Sum Filter (GSF) [90], which takes the non-linear bremsstrahlung into account. The reconstruction efficiency for electrons associated to good quality tracks varies from 97% to 99%.

A non-negligible fraction of reconstructed electron candidates corresponds to photon conversion, non-isolated electrons from in-jet decays and jets mimicking electrons. In order to reduce the contamination from these background sources, a set of discriminating quantities (i.e. the shape of the electromagnetic shower, the quality and length of the inner detector track and the track-calorimeter matching) is used to build a likelihood-based (LH) discriminant. A reference set of three qualities is available: LooseLH, MediumLH and TightLH. Each criterion is determined by using the relation between background rejection and electron identification efficiency. The efficiency for prompt electrons is $\sim$ 80% ($\sim$ 95%) for the TightLH (LooseLH) selection, considering a transverse electron energy of 40 GeV [91].

3.4.4 Muon reconstruction and identification

The reconstruction and identification of muons relies on accurate and independent track reconstruction in the ID and the MS [92]; the final muon track is obtained by combining these track candidates. The muon reconstruction in the ID is the same as the general track reconstruction described in Section 3.4.1. The MS tracks are reconstructed by using the hit pattern from the MDT and trigger chambers through a Hough transform algorithm [93]. This algorithm finds at least two seed-segments in the middle layers of the MDT; the muon tracks are reconstructed by employing a global χ^2 fit, which takes seed-segments and hits as inputs. Muons are categorised according to the available information from the detector sub-systems.

Combined (CB) muons

A combined track is formed with a global re-fit that exploits the hits from both the ID and MS sub-detectors. During the global fit procedure, MS hits may be added to or removed from the track, in order to maximise the fit quality. Most CB muons are reconstructed following an "outside-in" pattern recognition: muons are first reconstructed in the MS and then extrapolated inward and matched to an ID track. A complementary approach, i.e. "inside-out" combined reconstruction is used as alternative reconstruction. The acceptance of CB muons is limited by the ID coverage, namely $|\eta| < 2.5$.

Segment-tagged (ST) muons

A seed track reconstructed in the ID is considered a muon candidate if, once extrapolated to the MS, it is associated with at least one local track segment in the MDT or CSC chambers. The reconstruction of ST muons happens when muon candidates cross only one layer of the MS chambers. This muon category allows to enhance the muon reconstruction efficiency at low p_{T}. The track properties to describe the reconstructed muon are taken from the ID reconstructed track only.

Calorimeter-tagged (CT) muons

The ID track of a muon candidate is matched to an energy deposit in the calorimeter compatible with a minimum-ionising particle. CT muons have the lowest purity, but allow to recover acceptance in the region ($|\eta| < 0.1$) where the MS is only partially instrumented to allow for cabling and services to the ID and calorimeter system. The identification criteria for CT muons in this region ($|\eta| < 0.1$) are optimised for a transverse momentum of $15 < p_{\mathrm{T}} < 100\,\mathrm{GeV}$.

Extrapolated (ME) muons

Extrapolated (ME) muons are reconstructed exploiting only the MS track and requiring this to be compatible with originating from the primary interaction point. The estimated energy loss of the muon in the calorimeters is taken into account to define the parameters of the muon track with respect to the primary vertex. ME muons are required to traverse at least two layers of MS chambers, in order to provide a track measurement, and three layers in the forward region ($2.5 < |\eta| < 2.7$); ME muon reconstruction allows to recover acceptance in the latter region, which is not covered by the ID.

Muon identification

The muon identification is based on quality requirements optimised to suppress background muons, while selecting prompt muons with high efficiency and guaranteeing a precise and robust measurement of their

momentum. Some of the variables used in muon identification are the relative difference between the p_{T} measurements in the ID and MS (see, for instance, the "q over p ratio" defined in Eq. 4.1), the number of hits in the ID and MS and the χ^2 of the combined track fit. Four muon identification selections (Loose, Medium, Tight, and High-p_{T}) are available [92]; corresponding efficiencies are reported in Table 3.3 for prompt muons from W-boson decays and decays of hadrons in flight.

Table 3.3: Efficiency for prompt muons from W-boson decays ($\epsilon_{\mu}^{\mathrm{MC}}$) and decays of hadrons in flight ($\epsilon_{\mathrm{Hadrons}}^{\mathrm{MC}}$), misidentified as prompt muons, computed using a $t\bar{t}$ simulation. The results are shown for the four identification selection criteria separating low ($4 < p_{\mathrm{T}} < 20\,\mathrm{GeV}$) and high ($20 < p_{\mathrm{T}} < 100\,\mathrm{GeV}$) momentum muons for candidates with $|\eta| < 2.5$ [92].

Selection	$4 < p_{\mathrm{T}} < 20\,\mathrm{GeV}$		$20 < p_{\mathrm{T}} < 100\,\mathrm{GeV}$	
	$\epsilon_{\mu}^{\mathrm{MC}}[\%]$	$\epsilon_{\mathrm{Hadrons}}^{\mathrm{MC}}[\%]$	$\epsilon_{\mu}^{\mathrm{MC}}[\%]$	$\epsilon_{\mathrm{Hadrons}}^{\mathrm{MC}}[\%]$
Loose	96.7	0.53	98.1	0.76
Medium	95.5	0.38	96.1	0.17
Tight	89.9	0.19	91.8	0.11
High-p_{T}	78.1	0.26	80.4	0.13

3.4.5 Missing transverse momentum

A precise measurement of the missing transverse momentum is crucial in many physics studies at the LHC, and also a key element in the searches described in Chapters 5 and 6. The missing transverse momentum is defined as the momentum reconstructed in the transverse plane by all detector systems, thus it is explicitly affected by the acceptance, resolution and efficiency of the ATLAS detector components. The missing transverse momentum $\left(\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}\right)$ is a vector in the transverse plane, resulting from the vectorial sum of all the reconstructed transverse momenta with changed sign; as for the works presented in this dissertation the physics objects considered are electrons, muons and jets, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as follows:

$$\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{\mathrm{jets}} \vec{p}_{\mathrm{T}}^{\,\mathrm{jet}} - \sum_{\mathrm{softjets}} \vec{p}_{\mathrm{T}}^{\,\mathrm{softjet}} - \sum_{\mathrm{electrons}} \vec{E}_{\mathrm{T}}^{\,e} - \sum_{\mathrm{muons}} \vec{p}_{\mathrm{T}}^{\,\mu} - \sum_{\mathrm{clusters}} \vec{E}_{\mathrm{T}}^{\,\mathrm{cluster}}, \tag{3.7}$$

where the second term represents the contribution reconstructed from cells in jets with $7 < p_{\mathrm{T}} < 25\,\mathrm{GeV}$. The last terms is calculated from the cells in clusters which are not included in the reconstructed objects; this estimation is improved by adding tracks from low-p_{T} particles which do not reach the calorimeter or do not seed a cluster [94].

3.5 Proton–proton collision phenomenology and simulation at the LHC

The physics interpretation of data events collected at the LHC urges an accurate comparison to theoretical models, in order to test the compatibility between the latter and the experimental observation. This is achieved by generating *Monte Carlo* (MC) simulations, which reproduce the production and decay mechanisms of both signal and background processes in pp collisions. Simulations play a crucial role for most of the physics analyses and provide predictions about the expected event yields for a given physics process, as well as the corresponding decay signatures and kinematic distributions.

Many interesting processes at the LHC involve a large transfer of momentum, leading, for instance, to the production of jets and heavy particles with high transverse momenta. The simulation of sub-processes with large invariant momentum transfer constitutes the core of any simulation of collision events in contemporary experiments. As discussed in Section 2.1.3, QCD quanta are asymptotically free and, therefore, such interactions can be described by perturbation theory. The simulation does not fully reproduce the theoretical predictions of QED and QCD, but rather includes numerical calculations and parametrisations of analytical results.

Figure 3.18 shows the different steps involved in the simulation of pp collisions: the generation of the hard sub-process according to the Parton Distribution Functions, the parton shower, the hadronisation and the decay of the final-state products. In addition the interaction of the final-state particles with the detector is simulated.

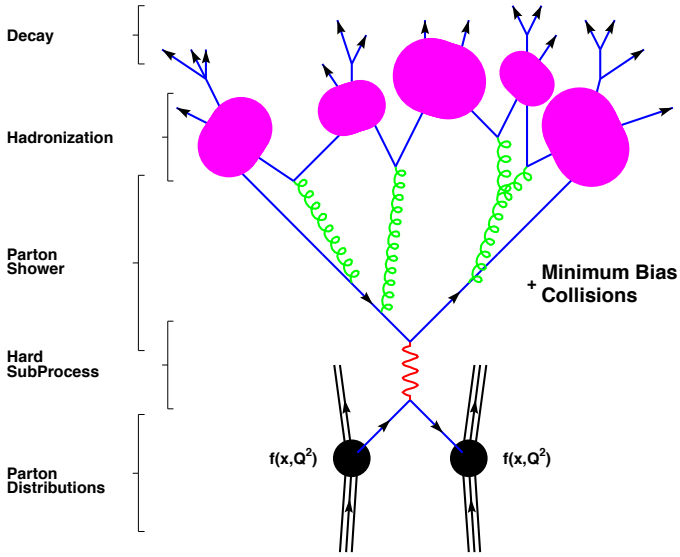


Figure 3.18: Schematic representation of the basic structure of a generated event, including showering and hadronisation. The time evolution of the event goes from bottom to top. Two protons, each indicated by three solid lines to denote their valence quark content, collide and a parton is resolved at energy scale Q and momentum fraction x [95].

3.5.1 Parton Distribution Function (PDF)

The Parton Distribution Functions (PDFs) represent the probability density functions to find a parton carrying a fraction x of the longitudinal proton momentum at a certain momentum scale μ^2. They are extracted by fitting deep inelastic lepton–nucleon scattering events [96] and constrained by exploiting information contained in precision measurements from the LHC Run 1 data [97]. The cross-section calculation for hard processes from pp collisions has to take into account the dominant low-energy QCD effects in the parton structure; for this purpose the QCD *factorisation theorem* (FT) is employed. The FT allows to disentangle the calculable short-distance process and the universal long-distance contribution, which can be derived from a fit to data; the factorisation scale μ_F is used to separate the low-energy and high-energy effects. The evolution of PDFs is described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [98–100]. Global fits combining several QCD measurements, such as deep inelastic scattering and hadron–hadron collision data (HERA, Tevatron), are exploited at the LHC: PDF4LHC [97], NNPDF [101], CT14 [102], MSTW [103].

3.5.2 Proton–proton hard scattering at the LHC

Cross sections for a scattering subprocess $ab \to n$ at hadron colliders can be computed in collinear factorisation through [104]:

$$\sigma = \sum_{a,b} \int_0^1 \int_0^1 \mathrm{d}x_a \, \mathrm{d}x_b \int f_a^{h_1}(x_a, \mu_F) f_b^{h_2}(x_b, \mu_F) \, \mathrm{d}\hat{\sigma}_{ab\to n}(\mu_F, \mu_R), \tag{3.8}$$

where:

- $f_a^h(x, \mu)$ are the PDFs, which depend on the momentum fraction x, with respect to the parent hadron h, and on the factorisation scale[6] μ_F;
- $\hat{\sigma}_{ab\to n}$ indicates the parton-level cross section for the production of the final state n, given the initial-state partons a and b. It depends on the momenta allowed by the final-state phase space Φ_n, on the factorisation scale μ_F and on the renormalisation scale μ_R, reflecting the basic issue that all perturbative calculations involve a truncated expansion [105].

The non-interacting partons are known as the *spectator partons* as they do not participate in the hard scattering interaction (namely they do not enter the parton-level cross section). The parton-level cross section $\hat{\sigma}_{ab\to n}$ can be expressed in terms of the matrix element squared $|\mathcal{M}_{ab\to n}|^2 (\Phi_n; \mu_F, \mu_R)$, the parton flux $\hat{s} = x_a x_b s$ (where s is the hadronic centre-of-mass energy squared) and the final-state phase space Φ_n:

$$\mathrm{d}\hat{\sigma}_{ab\to n}(\mu_F, \mu_R) = \mathrm{d}\Phi_n \frac{1}{2\hat{s}} |\mathcal{M}_{ab\to n}|^2 (\Phi_n; \mu_F, \mu_R), \tag{3.9}$$

where the matrix element squared can be written as a sum over Feynman diagrams and computed in different ways [106]. The differential phase space element is defined in terms of the initial-state momenta:

$$\mathrm{d}\Phi_n = \prod_{i=1}^{n} \frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^{(4)} \left(x_a P + x_b P - \sum_{i=1}^{n} p_i \right), \tag{3.10}$$

[6] In principle two factorisation scales can be introduced, one for each hadron. This becomes relevant for certain processes, such as the Higgs-boson production via vector-boson fusion, where, at leading order, the two hadrons do not interact through the exchange of colour.

where P is the fixed initial-state momentum of the hadron and the four-dimensional delta function ensures the four-momentum conservation between initial and final states.

In the processes considered, it is possible to have additional QCD and/or QED radiation from the initial-state partons or the final-state products. These additional contributions are known as as initial state radiation (ISR) and final state radiation (FSR).

The generators exploited in the works presented in this thesis are: Sherpa [107], MadGraph [108], MadGraph5_aMC@NLO [109] and Powheg-box [110–112]. MadGraph5_aMC@NLO and Powheg-box generators are specialised in the production of the hard process and need to be interfaced to a supervising generator, which takes care of other simulation steps (described in the following subsections) needed for the final state particles to be produced. Sherpa is a general purpose generator, including the implementation of all of the needed simulation steps.

3.5.3 Underlying events

In hadron–hadron collisions, unlike e^+e^- or ep collisions, it is possible to have interactions between spectator partons, an additional activity that is not directly associated with the hard interaction. These are commonly called underlying events and can be classified in two main categories: the beam remnants and the multiple parton soft QCD interaction (pile-up events). The impact of underlying events cannot be ignored for the performance of energy or momentum measurements, and thus several MC models, whose parameters can be tuned by using experimental results, are usually employed in the estimation of such effects.

3.5.4 Parton shower and hadronisation

Partons generated from the hard scattering usually have a sizable energy and emit additional partons via three main mechanisms: $q \to qg$, $g \to q\bar{q}$ and $g \to gg$, where q and g represent a quark and a gluon respectively. These processes are described by the parton shower (PS) algorithms, whose purposes are mainly two:

- providing estimates of higher-order corrections, enhanced by large kinematic logarithms. These occur in the phase-space regions of collinear parton splitting and/or soft gluon emission;
- generating high-multiplicity partonic states which can readily be converted into observed hadrons through a soft-hadronisation mechanism, i.e. involving modest transfers of quantum numbers or momentum between neighbouring regions of the phase space.

The probability distributions of parton branching can be computed exploiting their relation with the DGLAP splitting functions [95]. The PS procedure faces the ambiguous overlap between the components of the event which belong to the "hard process" (calculated using a multi-parton amplitude, as discussed in Section 3.5.2) from those developed during its evolution (described by the PS algorithms). A given $(n+1)$-jet, in fact, can be obtained in two different ways: from the collinear/soft radiation evolution of a $(n+1)$-parton final state, or from an n-parton configuration where hard, large angle emission during its evolution leads to the extra jet. This potential double counting of jet configurations is avoided through a "matching" and "merging" procedure; algorithms widely used for this purpose are the Catani-Krauss-Kühn-Webber (CKKW) [113, 114] and the Michelangelo L. Mangano (MLM) algorithms [115]. A special feature of the Sherpa generator is the implementation of the CKKW algorithm that allows to combine real emissions of the matrix element with the parton shower and the Sudakov form factors [116]. There are two common schemes for subtracting the overlap, from the matrix-element calculation (MadGraph5_aMC@NLO) or from the PS (Powheg-box), while still maintaining the required accuracy in the *matched* simulation.

The parton showering continues until the partons reach the hadronisation energy scale, at approximately 1 GeV. At this energy scale the hadronisation phase starts and partons combine into colourless hadrons. Several phenomenological models are exploited to describe the hadronisation and the following decay of hadrons into the final state particles: non-physical parameters are usually calibrated using experimental data; each set of these calibration parameters is named "tune". The most common model describing hadronisation is the Lund string model [117].

3.5.5 Detector simulation

In order to account for detector acceptance, inefficiencies and resolution, the interaction of the generated stable particles with all ATLAS detector components is simulated. This includes hits and clusters in active parts of the detector, as well as the interaction with passive material, such as cables. The particles' interaction throughout the ATLAS detector is commonly simulated using the GEANT-4 software package [118, 119]. Nevertheless this detector simulation is computationally very expensive, therefore in some cases the calorimeter response is reproduced through a faster parametrisation, which simplifies longitudinal and lateral energy profiles of showers; this is the so-called ATLAS *fast simulation II* (AFII) [120].

CHAPTER 4

A new Soft Muon Tagger for ATLAS Run 2

4.1 The key role of the bottom quark in hadron collider physics

One of the most important topics among the studies of QCD predictions is understanding the production of the b-quark, therefore a concise summary of its characteristics, production and decay mechanism is given in the following.

The relevance of the b-quark relies on the fact that, together with the top quark, it belongs to the heaviest quark family, with a bare mass of approximately 4.2 GeV [8]. Because of the high b-quark mass, hence a large accessible phase space, it is possible to describe b-hadron decays with the "spectator model". According to this model, quarks are treated as non-interacting particles and the lifetime of ground-state hadrons is determined by the weak decay of the b-quark. Under this assumption all b-hadrons would have identical mean lifetimes. The dominant decay mode of a b-quark is $b \to cW^-$, where the virtual W boson materialises either into a pair of leptons, $\ell\bar{\nu}_\ell$ (semileptonic decay), or into a pair of quarks, which then hadronise. The decays in which the spectator quark, $\bar{q}$ in Figure 4.1, combines with one of the quarks from the W boson to form one of the final state hadrons, are suppressed by a factor $\sim 1/9$, because the colours of the two quarks from different sources have to match in order to form a colourless bound state ("colour suppression") [121].

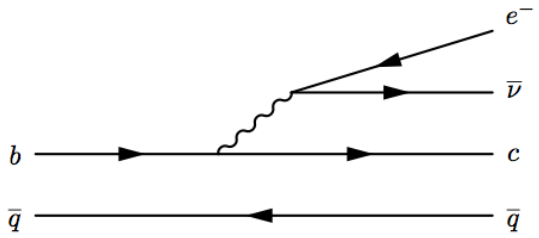


Figure 4.1: Example of a b-hadron decay, where a virtual W boson decays leptonically.

The B^0 meson, a bound state composed of a bottom antiquark and a down quark, has a relatively short mean lifetime of $1.5 \cdot 10^{-12}$ s [8]. Exploiting the spectator model it is possible to explain this; in fact, as it can be inferred from Figure 4.2, the calculations for the Feynman diagram for the B meson width and the muon width are very similar. The main differences in the calculation are the b-quark mass (much larger than the muon one), the different coupling between $\mu - \nu$ and $b - c$, and, above all, a phase space nine

times larger for b-quarks. Thereby the lifetime of the b-hadron is well predicted from the muon lifetime: predictions lead to a range of $1.3 < \tau_B[\text{ps}] < 1.7$ [122].

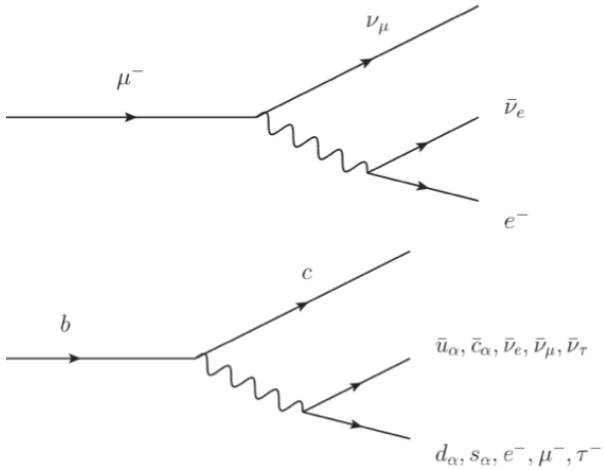


Figure 4.2: Comparison between semileptonic b-quark and μ decays.

Studying heavy flavour production in pp collisions at the Large Hadron Collider (LHC) provides a way to test calculations based on the perturbative factorisation approach of QCD processes at high energy scales. In this scheme, the cross sections are computed as a convolution of the parton distribution functions of the incoming protons, the partonic hard scattering cross sections and the fragmentation functions [123].

Leading-order Feynman diagrams showing the dominant b-quark production modes at hadron colliders are shown in Figure 4.3. In these cases b-quarks are produced as quark-antiquark pairs. b-quarks hadronise into mesons, a pair of a quark and an anti-quark, as well as baryons, composite particles made of three quarks. During the hadronisation the b-hadron keeps in general most of the momentum of the b-quark, unlike other hadrons, composed of lighter quarks only.

Typically b-hadrons produced at the LHC can fly a few centimeters before decaying. Furthermore, as

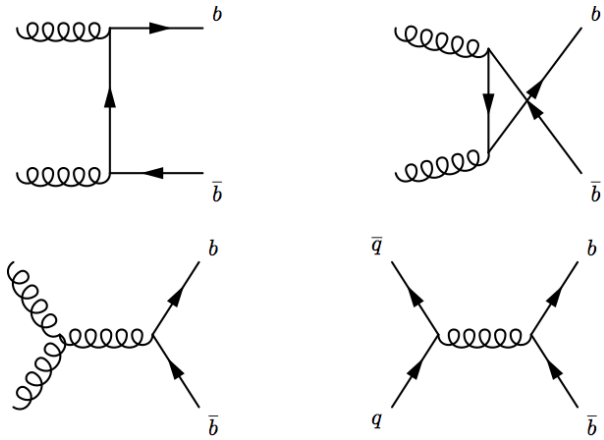


Figure 4.3: Feynman diagrams illustrating the production of b-quarks via gluon-gluon fusion and quark-antiquark annihilation.

shown in Figure 4.4, the production cross section of $b\bar{b}$ bounded quark systems (bottomonium) increases very steeply with the centre-of-mass energy of the colliding pp system. Thus measurements in the new energy domain of the LHC can contribute to a deeper understanding of the physics behind the hadron production processes.

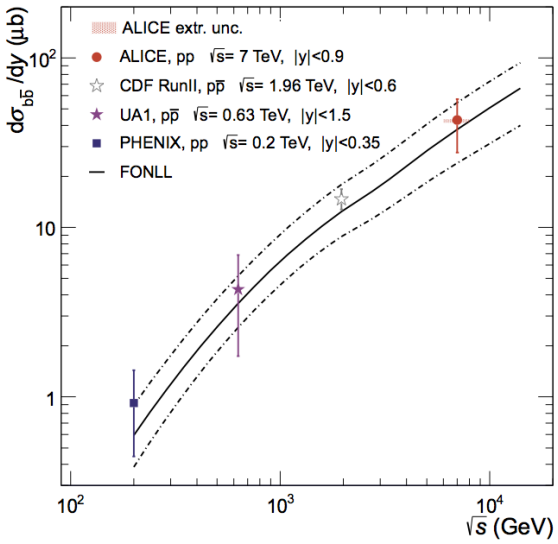


Figure 4.4: Bottomonium ($b\bar{b}$) production cross section as a function of the pp or $p\bar{p}$ centre-of-mass energy [124].

Important signatures of physics processes that are studied at the LHC also often contain a b-quark. For precise studies of these processes it is important to identify the b-quarks that are produced in the decay chain and many studies rely on this identification. For example top-quark and Higgs-boson properties are of great interest and b-quarks are often produced during their decay. Two Feynman diagrams of the most common top-quark and Higgs-boson decay modes are shown in Figure 4.5.

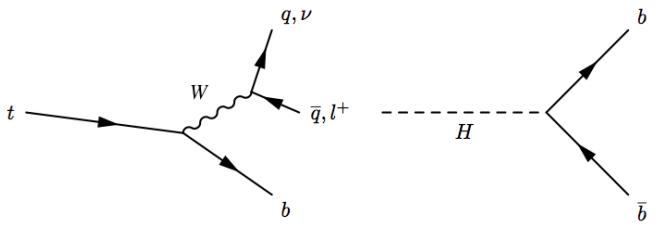


Figure 4.5: Example of physics processes involving b-quarks in the final state. The decay of a top quark is shown on the left, while an example of a Higgs-boson decay is shown on the right.

A fundamental parameter of the Standard Model is the CKM[1] matrix element $|V_{tb}|$. The only known way a top quark can decay is through the weak interaction producing a W boson and a down-type quark (down, strange or bottom) and the branching ratio $\Gamma(W^+b)/\Gamma(W^+q)$, with $q = b, s, d$, is 0.957 ± 0.034 according to the best current determination [8]. Since this ratio is equal to $|V_{tb}|^2$ according to the Standard

[1] Cabibbo-Kobayashi-Maskawa.

Model, this gives a way of determining $|V_{tb}|$.[2] Alternatively in combination with the determination of $|V_{tb}|$ from single top production it provides a way to test for the assumption that the CKM matrix is unitary.[3]

In addition to its relationship with the top quark, the b-quark is a crucial element in the Higgs physics studies; in fact, as anticipated, from Figure 2.7 it is clear that, for $m_H = 125$ GeV, $H \to b\bar{b}$ is the most favoured Higgs-boson decay channel. In particular, as discussed in Section 2.4.1, $H \to b\bar{b}$ is predicted in the SM to have a branching ratio of 58% for $m_H = 125$ GeV [125]; observation of the production of the Higgs boson decaying to a $b\bar{b}$ pair in association with a W or Z boson has been recently achieved [126, 127].

The preferential relationship between the b-quark and both the top quark and the Higgs boson provides a crucial role to the identification of jets from b-hadronisation within the context of searches for associated Higgs production with a top-quark pair; see Chapter 5 for further details.

4.2 Multivariate techniques

This section illustrates the set of algorithms we explored in order to develop a new Soft Muon Tagger (SMT), using complementary information to other existing b-tagging algorithms, and discusses the principles of a Boosted Decision Tree, exploited to separate $t\bar{t}H$ (Chapter 5) and WVZ (Chapter 6) from background processes.

4.2.1 TMVA framework

TMVA[4] is a standalone package providing a ROOT-integrated [79] machine learning environment for the evaluation of multivariate classification techniques [128]. It builds discriminating variables from a weighted combination of several input variables; it is possible to select the algorithm to be used in the analysis (e.g. rectangular cuts, k-NN, Likelihood or BDT). The algorithm uses input variables, weighted in a specific way, to assign a number to the event, which represents the likelihood of that event being a signal type. This number is commonly called the MVA[5] score.

Before the algorithm can be used with the data to separate signal from background, it has to be trained. To this purpose usually a set of simulated events, where it is known whether they belong to signal or background processes, is used to train the algorithm in correctly selecting the events (supervised learning). Along with the training, a test is performed to avoid "overtraining" (further details in Section 4.2.4).

4.2.2 Likelihood method

The individual likelihoods ($\mathcal{L}$) are products of the corresponding probability densities of the discriminating input variables used. In practice, this method uses polynomial splines fitted to histograms, or unbinned Gaussian kernel density estimators, to evaluate the probability density functions (PDF) obtained from the distributions of the input variables.

A likelihood ratio, R, is defined for an event by the ratio of the signal to the signal plus background likelihoods:

$$R = \frac{\mathcal{L}_s}{\mathcal{L}_s + \mathcal{L}_b}.$$

[2] As, a priori, there could be more than three quark generations, it has to be assumed that $\sum_{i=d,s,b} |V_{ti}|^2 = 1$.

[3] $V_{CKM} V_{CKM}^{\dagger} = I$.

[4] Toolkit for Multivariate Data Analysis.

[5] MultiVariate Analysis.

4.2.3 Artificial Neural Network

Artificial Neural Networks (ANN) are a family of statistical learning models inspired by biological neural networks (the central nervous systems of animals, in particular the brain) and are used to estimate or approximate functions, that can depend on a large number of inputs and that are generally unknown.

A statistical model may commonly be called "neural" if it possesses the following characteristics:

- it contains sets of adaptive weights, i.e. numerical parameters that are tuned by a learning algorithm;
- the capability of approximating non-linear functions of their inputs.

The adaptive weights can be thought of as connection strengths between neurons, which are activated during training and prediction.

What has attracted most of the interest in neural networks is the possibility of learning; given a specific task to solve, learning means using a set of observations to solve the task. For a more general treatment of neural networks see Refs. [129] and [130].

The k-Nearest Neighbour method

The TMVA k-Nearest Neighbour method (k-NN) compares an observed (or test) event to reference events from a training dataset and it searches for a fixed number of adjacent events, which then define a volume for the metric used. The k-NN method has best performance when the boundary that separates signal and background events has irregular features that cannot be easily approximated by parametric learning methods.

4.2.4 Boosted Decision Tree

A BDT, or Boosted Decision Tree, is a multivariate binary-tree structured classifier. Several decisions, corresponding to nodes, based on one variable at a time are taken until specific preset conditions, for instance the minimum percentage of training events required in a node or the maximal allowed depth of the tree, are fulfilled. The whole phase space is therefore divided into several subregions (leaves), as schematically shown in Figure 4.6; this learning process goes under the name of "training" of the decision tree. Each of the resulting subregions is then classified as signal-like (S) or background-like (B) according to the class of the majority of events populating it.

Decisions taken during the training of a tree can be based on various separation criteria [128], which are symmetric with respect to the event classes, since a cut mostly selecting background is as helpful as one efficiently selecting signal events. Standard criteria are the statistical significance, $S/\sqrt{S+B}$, and the Gini index, $p \cdot (1-p)$, where p is the fraction of events belonging to the same event class after the decision.

For all separation criteria the separation index reaches its maximum when the samples are fully mixed, i.e. at $p = 0.5$, and goes down to zero when the sample reduces to one event class only. Each node corresponds to a cut on a single variable which maximises the increase in separation index between the parent node and the sum of the indices of the two daughter nodes, weighted by their relative fraction of events. The granularity used to scan over the variable range and optimise the cuts can be set as an option. The training of the decision tree will stop when either the number of events in one leaf node goes below a certain threshold or the maximal depth is reached.

Several trees are sequentially trained with a boosting process between each decision tree training. Boosting consists in adjusting the weights, w_i, of individual events according to whether the previously trained tree classifies them correctly; it helps stabilising the response of the decision trees and significantly

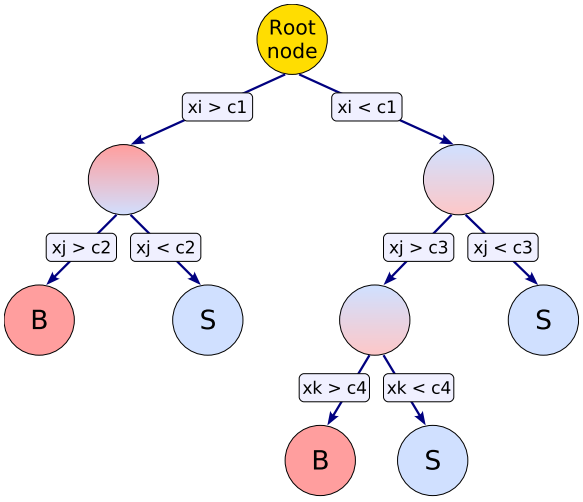


Figure 4.6: Schematic view of a decision tree. Starting from the root node, a sequence of binary decisions based on the discriminating variables (x_i) are taken. At each step a cut on the variable giving the best signal-to-background separation is performed, splitting the sample into two subsamples. The same variable may thus be used at several nodes, while others might not be used at all. Depending on the majority of events that end up in the leaves at the bottom of the tree, leaves are labelled "S" for signal or "B" for background [128].

enhances the performance of the final algorithm. Among many boosting algorithms available in TMVA the most relevant are described in what follows. A specific treatment of boosting can be found in Refs. [131] and [132].

AdaBoost, the Adaptive Boost algorithm; after a tree is trained, it is possible to compare the training sample scores, s_i, with the training sample true identities, y_i. We can define a characteristic function to indicate whether an event is incorrectly classified, $I(s, y) = 0$ if $s = y$, and 1 otherwise. The algorithm calculates the error rate for the tree as follows:

$$e = \frac{\sum_i w_i I(s_i, y_i)}{\sum_i w_i}$$

and the boost factor for the tree is computed as:

$$\alpha = \beta \cdot \ln\left(\frac{1-e}{e}\right),$$

where β is a user-specified boost strength (typically between 0 and 1). Once the boost factor for the tree has been evaluated, event weights are adjusted accordingly:

$$w_i \to w_i \cdot e^{\alpha \cdot I(s_i, y_i)}.$$

At this point the weights are renormalized so that $\sum_i w_i = 1$. The new weights are used to train the next tree. After that, the weights are boosted again. Boosting is cumulative: the weights are never reset to their original values. Finally, given a set of trees with indices n, the boost factor α_n calculated during boosting becomes the weight of that tree during scoring. The final score of an event is a weighted linear

sum of all decision tree scores:

$$s(\mathbf{x}) = \frac{1}{N_{\text{collection}}} \cdot \sum_{n}^{N_{\text{collection}}} \alpha_n \cdot c_n(\mathbf{x}),$$

where the sum is over all classifiers in the collection, $\mathbf{x}$ is the tuple of input variables and $c_n(\mathbf{x})$ the result of an individual classifier, encoded as $c_n(\mathbf{x}) = +1$ (-1) for signal (background) events.

Gradient boost uses the model response $F(\mathbf{x})$, a weighted sum of weak classifiers, similar to $s(\mathbf{x})$. The boosting procedure employs a loss-function $L(F, y) = \ln\left(1 + e^{-2F(\mathbf{x})y}\right)$ to adjust the parameters of the weak classifiers, in order to minimise the deviation obtained from the training sample between the model response $F(\mathbf{x})$ and the true value y. The Gradient boost algorithm slightly differs from the one used in the AdaBoost method and has the advantage of being more robust with respect to statistical fluctuations. The final classification corresponds to the minimal model response $F(\mathbf{x})$.

Randomised trees, where each tree is grown in such a way that at each split only a random subset of all variables is considered. Moreover, each tree in the forest is grown using only a (re-sampled) subset of the original training events.

Bagging denotes a re-sampling technique, where a classifier is repeatedly trained using re-sampled training events, such that the combined classifier represents an average of the individual classifiers. A priori, bagging does not aim at enhancing a weak classifier in the way adaptive and gradient boosting do, and is thus not a boosting algorithm in a strict sense.

In principle, the learning process could continue until each leaf node contains a few events, or even one event only, belonging to one event class. The resulting decision tree may be considered as a very powerful classifier, but it would be harmfully affected by the so-called "overtraining". This occurs when the algorithm becomes too specific to the particular set of events it is trained on and loses its unbiased discriminant power. To avoid overtraining a reasonable tree depth limit has to be set or long decision trees must be pruned after training. As an illustration, Figure 4.7 shows the selection performed by two different algorithms, trained in a two-dimensional phase space, i.e. trained exploiting two input variables; one region is selected by a regularised algorithm (black line) and the other one by an overtrained algorithm (green line). Although the algorithm affected by overtraining is very effective in separating signal (blue dots) and background (red dots) events, its response on a statistically independent set of events is subject to misclassification due to statistical fluctuations in the signal and background datasets and is, thus, prone to have a higher error rate.

4.3 Identification of b-jets in ATLAS

The identification of jets, reconstructed with a certain algorithm (see Section 3.4.2), containing a b-quark or, rather, originating from the hadronisation of a b-quark, is a key ingredient to many fundamental measurements and searches, as discussed in Section 4.1. Several algorithms to identify b-jets have been developed in ATLAS. They range from relatively simple algorithms, based on impact parameters of tracks and secondary vertices to the more refined JetFitter algorithm, which exploits the topology of weak b- and c-hadron decays. A first step, common to some algorithms, consists in attempting to reconstruct V^0 decay vertices[6], in order to reject them since they can mimic tracks from b-hadron decays.

[6] The appearance of the decay of an undetected neutral strange particle, such as K_s^0 or Λ, into two observed charged daughter particles gives rise to the terminology "V^0" to describe the decay topology.

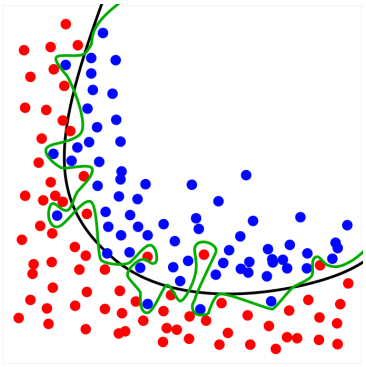


Figure 4.7: The blue and red points are signal and background events, respectively. The green line represents an overfitted model and the black line represents a regularised model. Although the green line best follows the training events, it is too dependent on the specific features of these events and it is likely to have a higher error rate on an independent set of events as compared to the black line.

4.3.1 Impact parameter-based algorithms

Combining the impact parameter significances of all reconstructed tracks (see Figure 4.8) in the jet is the basis of the first method to tag b-jets. A simple tagging algorithm doing this combination, JetProb, has been devised to be used for early data and has been extensively used in 2010 [133].

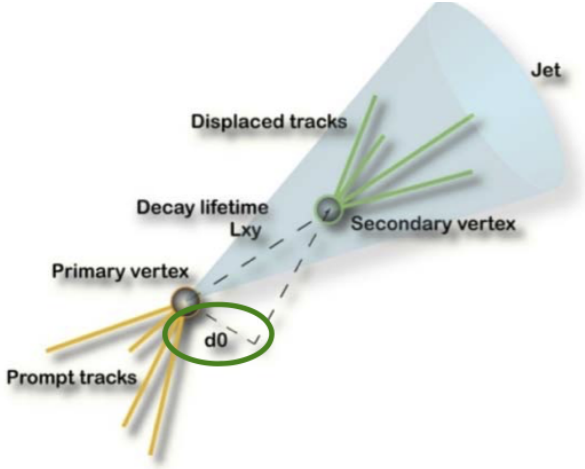


Figure 4.8: Representation of the transverse impact parameter definition (d_0).

The IP2D tagger makes use of the transverse impact parameter significance, d_0/σ_{d_0}, whereas IP3D takes advantage of both the transverse and the longitudinal impact parameter significance, $z_0 \sin\theta/\sigma_{z_0 \sin\theta}$, in a two-dimensional template to account for their correlation [134]. These high-performing impact parameter-based algorithms use a likelihood ratio discriminant (LLR, see Section 4.2.2), whose input variables are compared to pre-defined smoothed and normalised distributions ("templates") for both the b- and light-jet hypotheses, obtained from MC simulation. The LLR discriminant is built as the sum of

per-track contributions:

$$\sum_{i=1}^{N} \log\left(p_b(t_i)/p_u(t_i)\right),$$

where p_b and p_u are the template probability density functions for the b- and light-flavour jet hypotheses, t_i represents the i^{th} track and N is the number of tracks for the considered jet.

The impact parameters of different tracks originating from the same secondary (or even tertiary) vertex are intrinsically correlated with each other. A new b-tagging algorithm, the Impact Parameter-based Recursive Neural Network (RNNIP) [134], has been developed in order to exploit such correlation. Recurrent neural networks [135] have been trained feeding transverse and longitudinal impact parameter significances, the angular distance between the track and the jet axis and the fraction of transverse momentum carried by the single track relative to the jet p_{T}.

4.3.2 Secondary vertex-based algorithms

The discrimination between b- and background jets can be also attained by seeking a single displaced inclusive vertex formed by the decay products of the b-hadron, including the products of a possible subsequent charm-hadron decay (see Figure 4.9). The reconstruction starts by building all two-track pairs that form a good vertex by means of a χ^2 fit, using only tracks associated to the jet and far enough from the primary vertex. Vertices compatible with a V^0 decay or material interaction are rejected. All tracks from the remaining two-track vertices are combined into a single inclusive vertex, using an iterative procedure to remove the worst track until the χ^2 of the vertex fit is good.

The SV1 algorithm [136] takes advantage of eight input variables, including the invariant mass of all tracks associated to the vertex, the number of tracks associated to a secondary vertex, the ratio of the sum of the energies of the tracks in the vertex to the sum of the energies of all tracks in the jet and the three-dimensional decay length significance. These variables, as for IP2D and IP3D, are combined using a likelihood ratio technique. SV1 relies on a 2D-distribution of the two first variables and a 1D-distribution of the number of two-track vertices. Finally, the distance ΔR between the jet axis and the line joining the primary vertex to the secondary one is used as input variable.

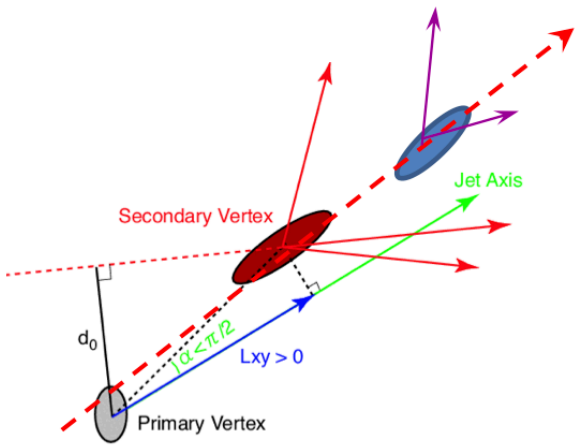


Figure 4.9: Representation of a secondary and a tertiary vertex.

4.3.3 Decay-chain reconstruction with JetFitter

A third algorithm, called JetFitter, exploits the topology of weak b- and c-hadron decays inside the jet. A modified Kalman filter [133] is used to find a common line on which the primary vertex and the b- and c-vertices lie, as well as their position on this line, giving an approximated flight path for the b-hadron. Through the implementation of such approach, the b- and c-hadron vertices are not necessarily merged, even when only a single track is attached to each of them. The discrimination between b-, c- and light-jets is based on a likelihood using similar variables as for the SV1 algorithm, see Section 4.3.2, and additional variables such as the flight length significances of the vertices. The current version of this algorithm is the result of several improvements, which allow to increase the reconstruction efficiency of tertiary vertices and reduce the impact of pile-up tracks [134].

4.3.4 High-level taggers: MV2 and DL1

The most discriminating variables resulting from the previously described algorithms (IP3D, RNNIP, SV and JetFitter) are combined and output weight probability densities evaluated separately for b-, c-, and light-flavour jets. In Run 1 the combination was done through a Neural Network resulting in the MV1 tagging algorithm [133]. For Run 2, the combination is done either using a gradient Boosted Decision Tree (MV2) or training a Deep Neural Network (DL1) through Keras and Theano [137].

The MV2 tagging algorithm is trained on the hybrid samples described in Section 4.5 with a c- and a light-jet fraction of 7% and 93% respectively; this background results in a suitable compromise between c-jet and light-jet rejection. Several variants of the MV2 taggers have been developed, adding new input variables in order to maximise its performance:

- **MV2**: a reference option [138], trained with 24 input variables including the standard impact parameter (IP2D and IP3D) and secondary vertex (SV1 and JetFitter) inputs;
- **MV2Mu**: a tagger including the SMT output (see Section 4.8) on top of the standard inputs;
- **MV2MuRnn**: a tagger including both, the SMT output and RNNIP inputs in addition to the standard input variables.

More details about the development and optimisation that we performed for the MV2Mu tagger are provided in Section 4.10.

The DL1 tagger has a multidimensional output corresponding to the likelihood for a jet of being b-, c- or light-flavoured. It is trained on the same set of input variables used for MV2, with the addition of the JetFitter c-tagging variables [134] and the full set of SMT observables, described in Section 4.7. In contrast to MV2, all flavours are equally treated, therefore the trained algorithm can be used to tag b-, as well as c-jets.

4.4 Principles of Soft Muon Tagging

The Soft Muon Tagger (SMT) is based on the reconstruction of muons coming from semileptonic decays of heavy-flavour hadrons. An earlier version of the SMT was based on a simple requirement on a χ^2 distribution [139]. These muons usually have a sizable transverse momentum (though smaller than the typical p_{T} of leptons from electroweak bosons decays, hence the label "Soft"), as well as a large transverse momentum relative to the jet axis, $p_{\mathrm{T}}^{\mathrm{rel}}$. The presence of a muon is enhanced in b-jets with respect to c- and light-flavour jets due to the significant semileptonic decay branching ratio of b-hadrons,

$BR(b \to \mu\nu X) \approx 11\%$, and c-hadrons produced by the b-hadron decay, sequential semileptonic decay, $BR(b \to c \to \mu\nu X) \approx 10\%$) [8]. The SMT, while intrinsically limited by the semileptonic branching ratio, provides a useful complement to the impact parameter- and vertex-based taggers, described in Sections 4.3.1 and 4.3.2, for jets whose performance is affected by the semileptonic decay itself. In fact, when the latter occurs, the multiplicity, the invariant mass and the energy of charged decay particles inside the reconstructed jet is significantly reduced.

4.5 Simulated samples

The results discussed in the following sections are based on MC simulation samples of $t\bar{t}$ and Z' produced in proton–proton collisions with a centre-of-mass energy of 13 TeV.

The $t\bar{t}$ sample is simulated with Powheg [110] and interfaced with Pythia8 [140] for parton shower and hadronisation. Only $t\bar{t}$ decays with at least one lepton from a subsequent W boson decay are simulated. The CT10 [141, 142] set of PDFs is used. Events with the associated production of a single top quark and a W boson, referred to as Wt events, form a significant background to $t\bar{t}$ production and are also simulated with Powheg interfaced to Pythia6 [143].

In order to optimise the performance at high jet p_{T}, the SMT has been trained using a mixture of the previously described $t\bar{t}$ simulation and a dedicated sample of Z' decaying to jets. Pythia8 is used to generate the latter with the A14 [144] set of tuned parameters for the underlying event and the leading-order NNPDF2.3 [145] parton distribution function. The cross section of the hard-scattering process is modified by applying an event-by-event weighting factor to broaden the natural width of the resonance and widen the transverse momentum distribution of the jets produced in its hadronic decays. The branching fractions of these decays are set to be one-third each for the bb, cc and light-flavour quark pairs, to give a p_{T} spectrum uniformly populated in all jet flavours.

The labelling scheme used to define the flavour of the jet (b-, c- or light-flavour jets) is performed by matching the jets to b- and c-hadrons with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ within a cone of radius $\Delta R = 0.3$ around the jet axis. First, b-hadrons within the cone are searched for; if no b-hadron is found, the same procedure is repeated for c-hadrons and for τ leptons. The remaining jets with no such matching hadrons or τ leptons are assigned to the light-flavour jet category.

4.6 Selection and composition

The new SMT algorithm is based on calorimetric jets and combined (CB) muons, i.e. muon candidates with matching tracks reconstructed in the ATLAS Inner Detector (ID) and Muon Spectrometer (MS) systems, see Section 3.4.4.

Jets are reconstructed from topological clusters in the calorimeters using the anti-k_t algorithm [87] with radius parameter $R = 0.4$. The selection applied in these studies requires jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$. In addition, a requirement on the output of the Jet Vertex Tagger algorithm (JVT) [88] is also applied. As anticipated in Section 3.4.2, this algorithm allows jets originating from pile-up interactions to be suppressed using track-based variables included in a multivariate discriminant. Jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$ are removed if the JVT output is smaller than 0.59. This cut is 92% efficient for jets originating from the hard scatter vertex, with a residual rate for pile-up jets of approximately 2%.

Muons are associated to the closest selected jet by requiring an angular separation between the muon and the jet-axis of $\Delta R < 0.4$. Muons are required to have $p_{\mathrm{T}} > 5\,\mathrm{GeV}$, $d_0 < 4$ mm and $|\eta| < 2.5$. Since minimum ionising particles lose on average $\sim 3\,\mathrm{GeV}$ in the ATLAS calorimeter system, the efficiency for reconstructing muons with p_{T} below this value is small. The fraction of b-jets in the $t\bar{t}$ simulated

sample described in Section 4.5 with a reconstructed muon passing the requirements listed above is ∼12%. There are three main background sources in light jets that give rise to muon candidates passing these requirements and they are evaluated in the $t\bar{t}$ simulated sample:

- prompt muons from the nearby W boson, randomly associated to light jets (∼ 1% contamination);
- energetic hadrons ("punch-through") that travel through the calorimeter system and reach the MS (∼ 1% contamination);
- muons coming from the decay in flight of light hadrons, mostly pions and kaons (∼ 0.1% contamination).

4.7 Input variables

The final set of six SMT input variables is the result of an optimisation, based on the power of physical quantities in separating b-jets from background jets [146]. This set includes both kinematic and track properties, with three variables in each of these two categories.

Three kinematic variables separating muons in light jets from those from b- or c- hadron decays are used:

- $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$: angular distance between the muon and the associated jet;
- $p_{\mathrm{T}}^{\mathrm{rel}}$: orthogonal projection of the muon p_{T} onto the jet axis;
- d_0: muon impact parameter measured with respect to the interaction primary vertex.

The $p_{\mathrm{T}}^{\mathrm{rel}}$ value of decay products keeps memory of the decaying parent particle mass. Therefore, muons from direct b decays tend to be more boosted in a plane transverse to the jet axis, i.e. have a larger $p_{\mathrm{T}}^{\mathrm{rel}}$ than the ones from background sources, especially than muons from light hadron decays. Muons from the $b \to c$ cascade decay are more difficult to identify since they tend to be softer in the $p_{\mathrm{T}}^{\mathrm{rel}}$ spectrum, as compared to muons from direct b decays. Muons from π and K decays in flight have tracks that do not in general extrapolate close to the primary vertex. Nevertheless, if the decay occurs at a small angle, the decaying hadron and the muon can be reconstructed as a single track and the muon candidate can pass the track selection, since the muon candidate would not have a significant impact parameter.

In addition to ΔR, d_0 and $p_{\mathrm{T}}^{\mathrm{rel}}$, the SMT algorithm makes use of three observables defining the quality of the muon track. The scattering neighbour significance ($\mathcal{S}$) is computed by considering pairs of adjacent hits along the track (treating them as neighbouring scattering centres) and evaluating the significance of the angular difference $\Delta\phi$ between the two half tracks ending/starting at each of them; the final discriminant is obtained by summing up these significances along the whole track times the particle charge (q):

$$\mathcal{S} = q \times \sum_i \frac{\Delta\phi^i_{\mathrm{scat}}}{\sigma_{\Delta\phi^i_{\mathrm{scat}}}}.$$

This variable is introduced to quantify the significance of a kink along the track, a change in trajectory expected in the presence of a π or K decaying to a muon. It is a measurement of how many hits in the tracker do not fit well with the reconstructed track, thus a higher absolute value of the significance is more likely to correspond to a decay in flight.

The momentum imbalance significance ($\mathcal{M}$) is defined as follows:

$$\mathcal{M} = \frac{p_{\mathrm{ID}} - p^{\mathrm{extr}}_{\mathrm{MS}}}{\sigma_{E_{\mathrm{loss}}}},$$

where p_{ID} is the muon momentum measured by the Inner Detector, $p_{\mathrm{MS}}^{\mathrm{extr}}$ is the momentum measured by the MS and extrapolated to the vertex and $\sigma_{E_{\mathrm{loss}}}$ is the uncertainty on the energy loss measured by the calorimeters.

Another quantity sensitive to muons originating from these decays through the p_{ID} to p_{MS} comparison is:

$$\mathcal{R} = \frac{(q/p)_{\mathrm{ID}}}{(q/p)_{\mathrm{MS}}}, \tag{4.1}$$

where $(q/p)_{\mathrm{ID}}$ is the charge-to-momentum-ratio, i.e. the track curvature, measured by the Inner Detector and $(q/p)_{\mathrm{MS}}$ is the same measured by the MS.

Figure 4.10 shows the shapes of the SMT discriminating input variables for b-, c- and light-flavour jets in simulated $t\bar{t}$ events.

4.8 Training and performance of the SMT algorithm

The efficiency of muon reconstruction and association to jets is ~65% for b-jets containing a muon, with a light-jet mis-identification probability of ~1.8%, as evaluated in $t\bar{t}$ events. The performance achievable via simple cuts on the discriminating variables is limited, e.g. the optimal cut on the momentum imbalance reduces the muon reconstruction efficiency inside b-jets to ~55%, with a corresponding decrease in mis-identification efficiency to ~0.8%. This mis-identification efficiency can be further reduced retaining a good signal efficiency by adopting a multivariate approach.

The full set of six variables shown in Figure 4.10 is used as input to a dedicated MVA. The performance from different MVA methods, described in Section 4.2, with default settings [128] has been compared, as shown in Figure 4.11; since the gradient-boosted BDT (BDTG) and the adaptive-boosted BDT are the most performant, a BDTG has been chosen according to the robustness arguments discussed in Section 4.2.4. The final configuration of the BDTG, as a result of balance between expected performance and overtraining bias, has the following settings:

- number of trees: 300;
- maximum depth: 6;
- minimum node size: 0.05%;
- number of cuts (granularity): 200.

The muon transverse impact parameter (d_0) is the most discriminating input, according to the preliminary ranking from TMVA [128], which is independent of the algorithm and relying on the inherent discriminating power of a given variable. Figure 4.12 shows the overtraining check for the final SMT algorithm: the BDTG responses on the training sample and a statistically independent set of events are compatible. Overtraining effects can be spotted by employing the so-called Kolmogorov-Smirnov (K-S) test [147] on the classifier. The Kolmogorov-Smirnov test measures the maximum distance between the cumulative distribution of two samples (here train and test) to assess whether their distributions differ. The null distribution of the K-S statistic is calculated under the null hypothesis that the two samples are drawn from the same distribution. The probability that the maximum K-S distance λ_α is larger than the observed may be calculated using K-S statistics for each parameter α that has the meaning of statistical significance level. When $\lambda < \lambda_\alpha$ then the two distributions are equivalent with a $1-\alpha$ significance level: a typical significance value is 95%, which means $1-\alpha = 0.05$. The TMVA toolkit reports the K-S test converted to significance levels $(1-\alpha)$ for overtraining checks. Values larger than 0.05 imply good

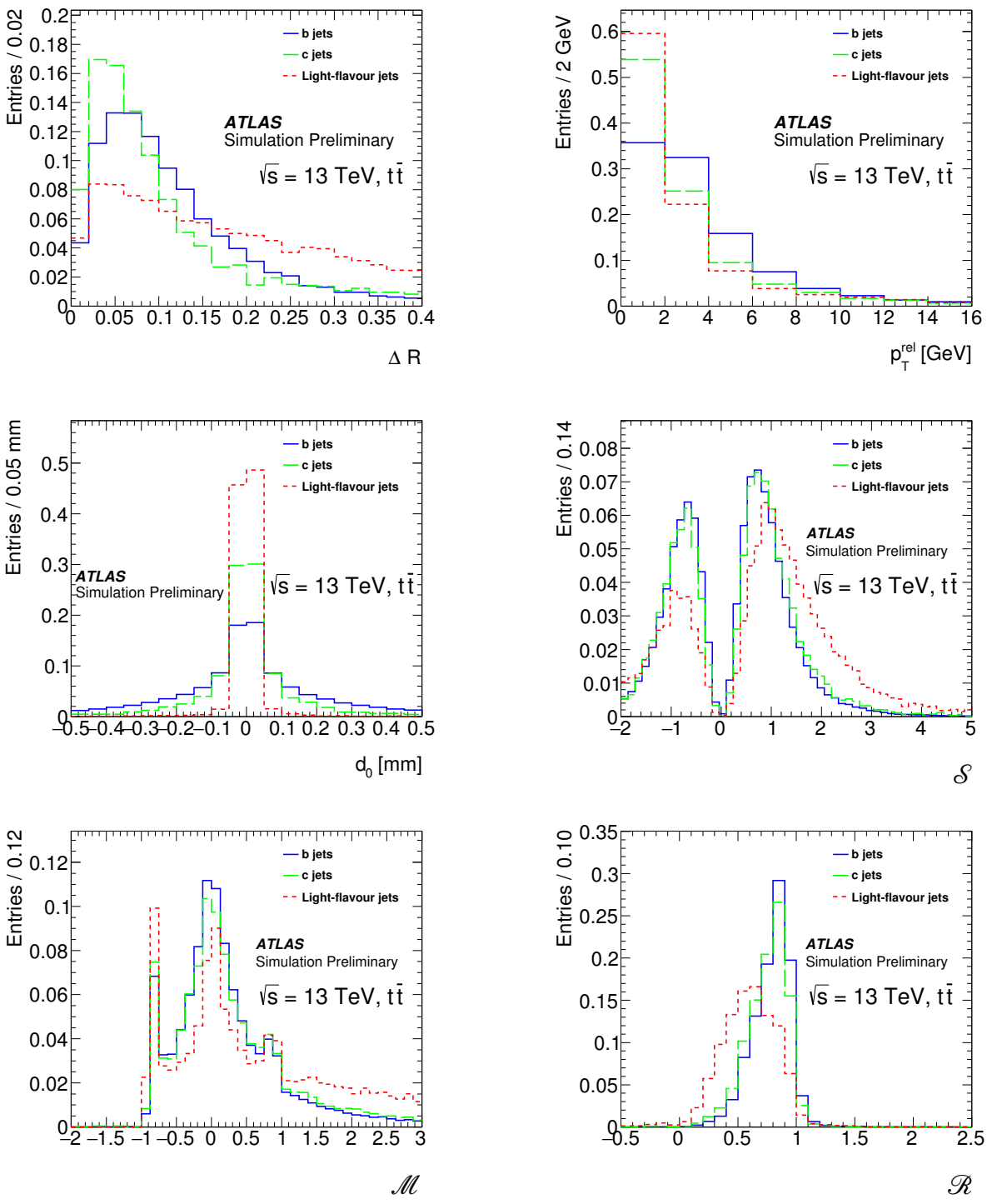


Figure 4.10: Normalised distributions of the six SMT input variables for reconstructed CB muons associated to b-jets (blue), c-jets (green) and light-flavour jets (red) in simulated $t\bar{t}$ events [134].

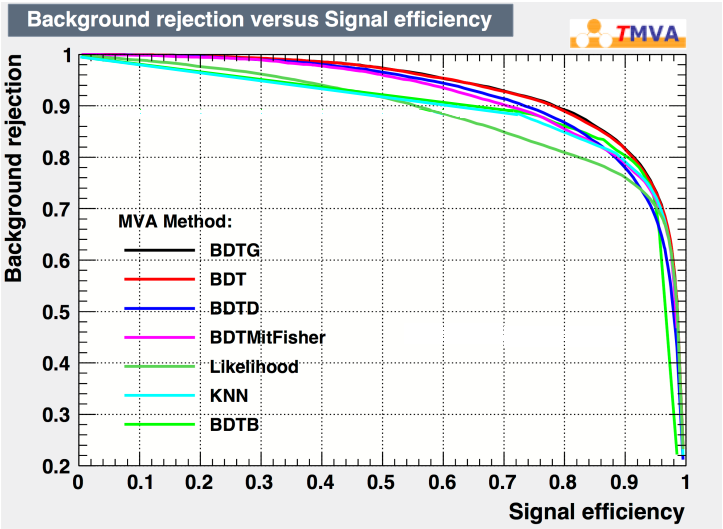


Figure 4.11: Performance of SMT algorithms trained through different MVA methods with default settings. The MVA methods described in Section 4.2 are explored: BDTG (gradient boost), BDT (adaptive boost), BDTD (decorrelation and adaptive boost), BDTMitFisher (Fisher discriminant used for node splitting), Likelihood (Likelihood method), KNN (k-Nearest Neighbour method) and BDTB (bagging re-sampling) [128].

agreement between the train and test samples, whereas smaller values would reveal some overtraining issues.

Figure 4.13 shows the new discriminant output for b-, c- and light-flavour jets containing a muon candidate passing the SMT selection. By cutting on the SMT discriminant (BDT > -0.15), the efficiency for accepting a jet having a candidate SMT muon is 85% for b-jets and 15% for light-flavoured jets in $t\bar{t}$ events. This translates to an overall b-jet tagging efficiency of 10%, and light-jet mistag rate of 0.02%.

4.9 Validation of SMT input variables and MVA output

Given the complexity of the SMT input variables, probing different kinematic and topological properties of muons in jets, it is particularly important to ensure that the simulation correctly models the relevant features. The modelling of variables is explored using an e-μ sample dominated by $t\bar{t}$ events which are a rich source of b-jets.

The data/MC comparisons shown in the following are meant to illustrate the overall agreement and to identify localised effects and possible mis-modellings. Only statistical uncertainties are taken into account, while systematic uncertainties due to the different flavour compositions in data and MC simulation are neglected.

A subset of the data collected by the ATLAS detector from proton–proton collisions in 2016 at a centre-of-mass energy of 13 TeV is used for this study. This subset is chosen to uniformly sample the data collected at different pile-up conditions. Events with two charged leptons are selected online using either a single-electron or single-muon trigger. The thresholds are set to be almost fully efficient for leptons with $p_{\mathrm{T}} > 25$ GeV passing offline selection requirements on the lepton identification and isolation (see Sections 3.4.3 and 3.4.4). The selected $e\mu$ + jets events correspond to a total integrated luminosity of 2.5 fb^{-1}. Simulated events have been reweighted to reproduce the pile-up distributions of the data samples (see Section 3.3).

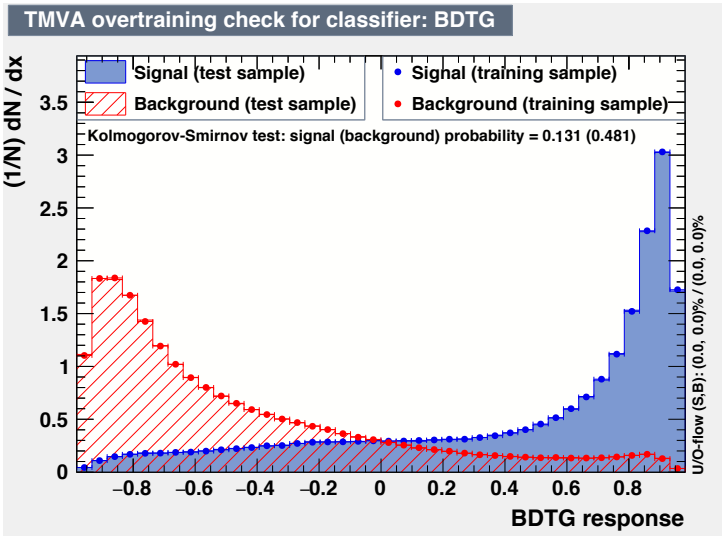


Figure 4.12: BDTG expected response for b- (blue) and background (red) jets in Monte Carlo simulation; the comparison between the BDTG distributions on the training sample (dots) and a statistically-independent one (histograms), for both signal and background jets, shows that no overtraining bias is present.

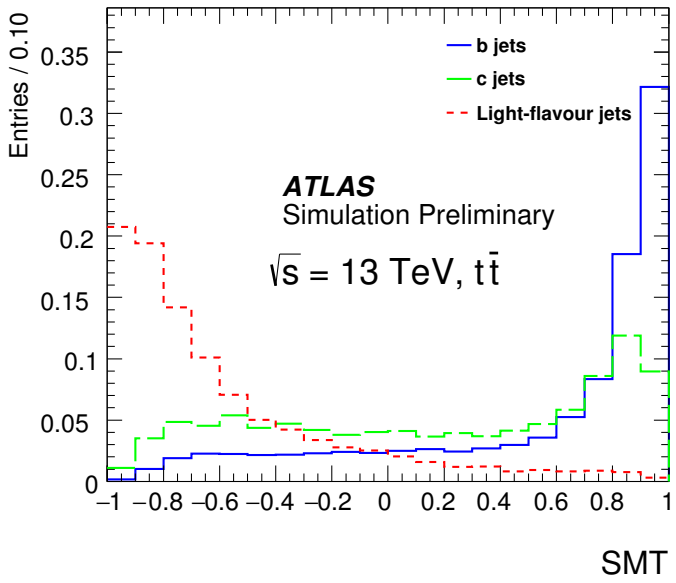


Figure 4.13: Normalised BDT response in simulated $t\bar{t}$ events of the SMT algorithm for reconstructed muons associated to b-jets (blue), c-jets (green) and light-flavour jets (red) [134].

The $t\bar{t}$-dominated e-μ sample is selected by requiring an opposite charge sign e-μ pair, with the leading lepton $p_{\mathrm{T}} > 25\,(20)$ GeV, for electrons (muons) and the p_{T} of the second lepton > 15 GeV. The invariant mass of the dilepton pair is required to be greater than 10 GeV. Events are required to have at least two and at most seven jets passing the selection discussed in Section 4.6. The subleading jet is required to be identified as a b-jet using the MV2 discriminant (see Section 4.3.4) at a working point with an identification efficiency of 77%. The other selected jets are used to study the b-tagging algorithm response. Simulation predicts b-jets to account for ~25 % of the other selected jets. Selected data events are compared with $t\bar{t}$ and Wt simulated samples, properly rescaled to their relative cross sections.

Figure 4.14 shows data/MC comparison for two discriminating SMT input variables (ΔR and $p_{\mathrm{T}}^{\mathrm{rel}}$) and the BDT output using $t\bar{t}$ events; an overall good agreement between data and MC simulation is observed in these SMT input variables, as well as in the final BDT discriminant. Some localised discrepancies between data and simulation are found in the low BDT region (BDT < -0.3), where the observed difference is of the order of 20–25%.

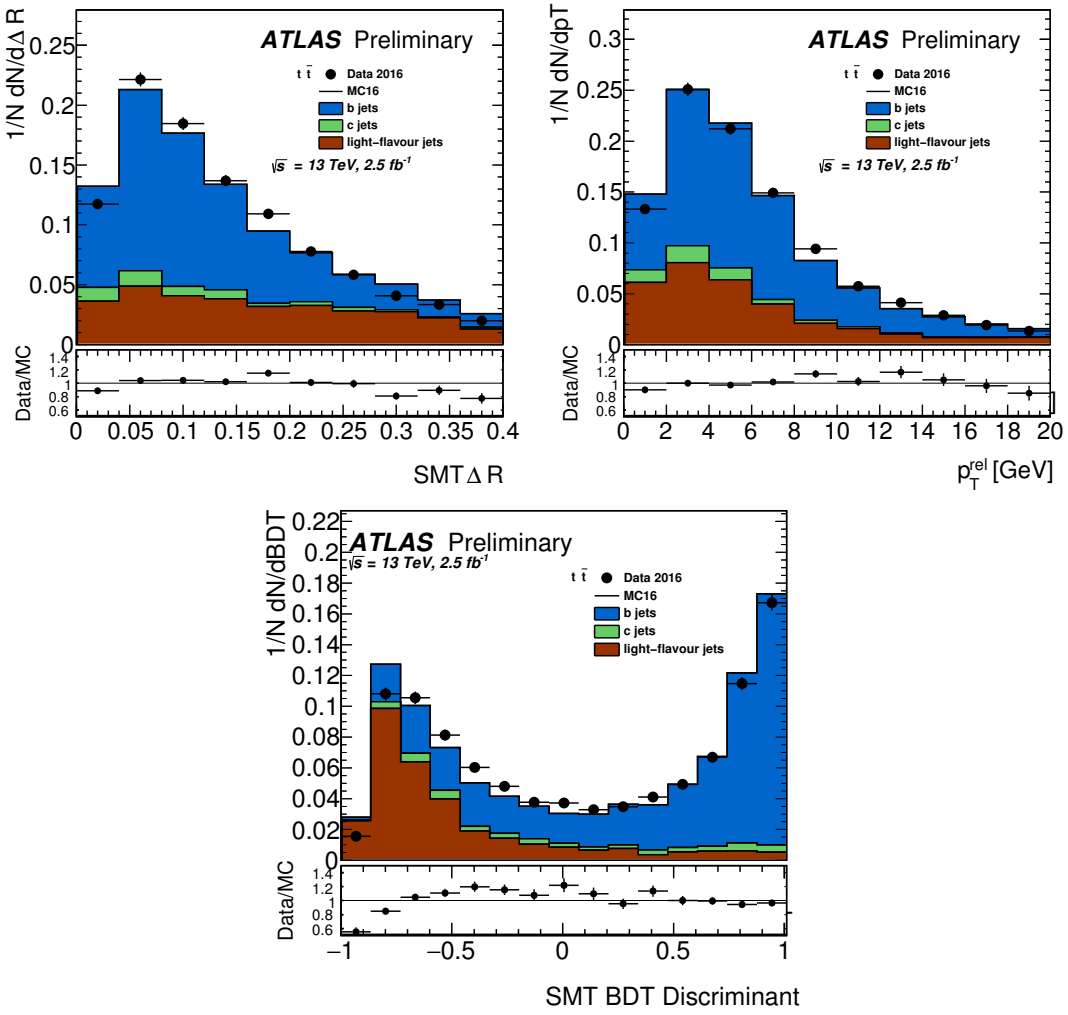


Figure 4.14: Data/MC comparison for the two most discriminating input variables (ΔR and $p_{\mathrm{T}}^{\mathrm{rel}}$) and the SMT final discriminant in a $t\bar{t}$ dilepton-dominated phase space [134].

Figure 4.15 shows the event display of a $t\bar{t}$ dilepton candidate event where a soft muon is associated to a b-jet candidate tagged by the new BDT-based Soft Muon Tagger algorithm.

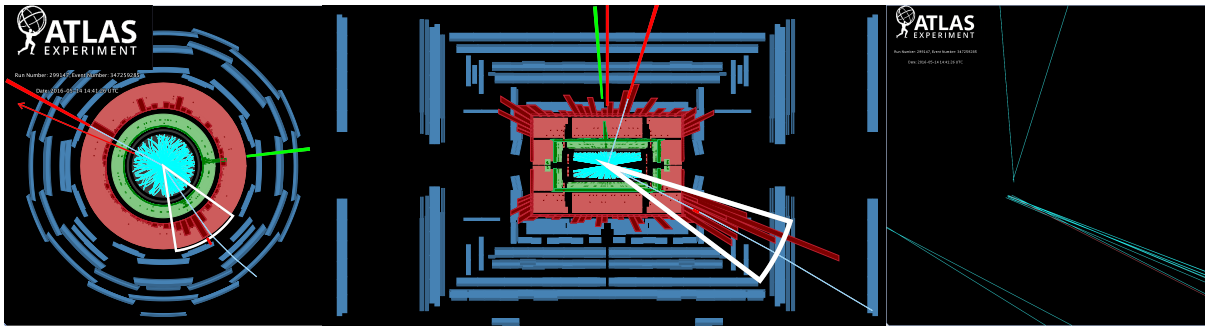


Figure 4.15: Event display of a $t\bar{t}$ dilepton candidate event from proton–proton collisions recorded by ATLAS with LHC stable beams at a collision energy of 13 TeV. In addition to two prompt leptons (an electron of 107 GeV and a muon of 268 GeV) a soft muon (7 GeV) is associated to a b-jet candidate (within $\Delta R < 0.4$) tagged by the new BDT-based Soft Muon Tagger algorithm. The right panel shows a zoom on the reconstructed primary vertex (PV, two tracks) and secondary vertex (SV, six tracks) in the $z - \rho$ plane. The track of the muon within the tagged jet is shown in red. The distance between PV and SV is about 2 mm [148].

4.10 Implementation of SMT in high-level taggers

The MV2 b-tagging algorithm combines 24 input variables based on properties of track impact parameter, secondary vertex and weak decay topology algorithms into a BDT. As anticipated in Section 4.3.4, among the MV2 tagger variants a new option has been developed, including in addition the SMT output ("MV2Mu"), and added for the 2017 data taking and processing campaigns. The usage of both the full set of six SMT input variables (Section 4.7) and the SMT output (Section 4.8) has been tested for the MV2 training. The MV2 BDT configuration is significantly simplified by using a single output variable from SMT instead of six inputs and, furthermore, it is found that this approach leads to a better rejection of light- and c-jets than when using the full set of SMT input variables.

As shown in Figure 4.16, the implementation of the SMT output itself as additional input variable to MV2 leads to a 20 – 25% improvement in light-jet rejection in the 70 – 85% b-jet efficiency range, relevant for most of ATLAS physics analyses. The inclusion of the SMT BDT output improves the light-jet rejection in the low-medium jet p_{T} range (below 250 GeV) where the additional information of the presence of the muon from the semileptonic decay in jets is beneficial to enhance the b-to-light jet separation. Figure 4.17 shows that the b-to-c separation does not profit from the muon-in-jet information at efficiencies larger than $\sim$ 50%, since leptons are produced in b- and c-decays alike.

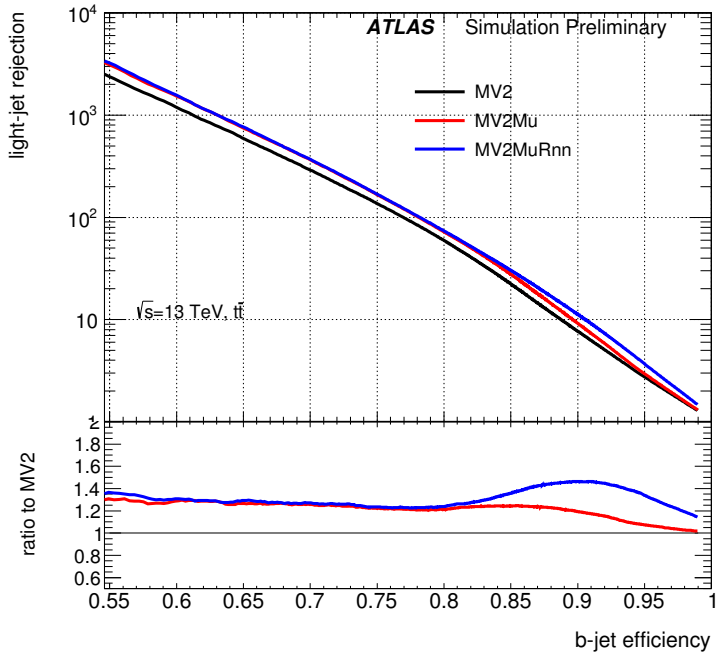


Figure 4.16: Light-jet rejection as a function of b-jet efficiency for MV2 (black line), MV2Mu (red line), MV2MuRnn (blue line). The evaluation of algorithms is performed on $t\bar{t}$ events and the ratio reported on the bottom of the figure is calculated for the two MV2 variants with respect to MV2. The MV2-to-MV2Mu ratio shows that a $20-25\%$ improvement in light-jet rejection in the $70-85\%$ b-jet efficiency range is expected from the addition of SMT to the MV2 input variables [134].

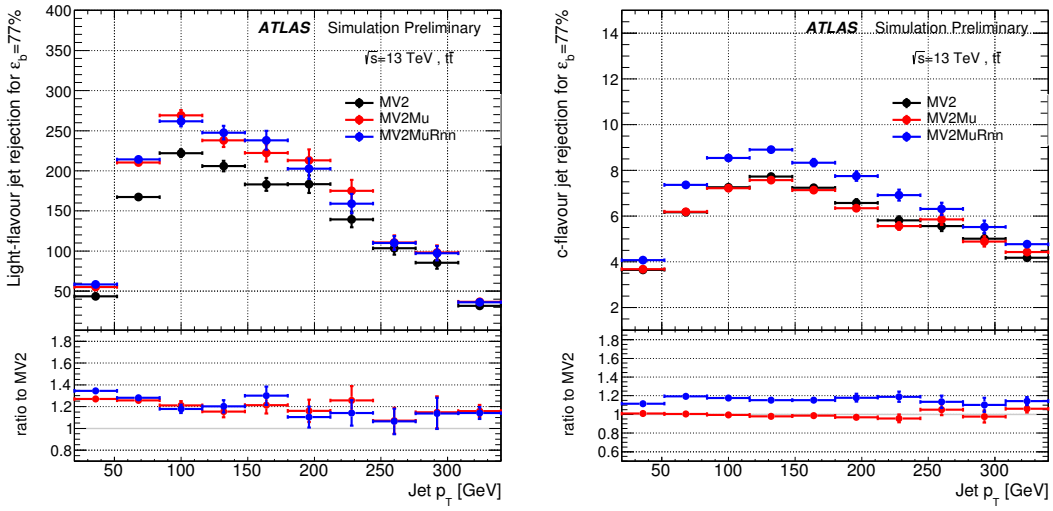


Figure 4.17: Light-flavour and c-jet rejection as a function of the jet transverse momentum for MV2 (black), MV2Mu (red), MV2MuRnn (blue). The algorithm evaluation is performed on $t\bar{t}$ events for a flat b-jet efficiency of 77% for each p_{T} bin. The ratio reported on the bottom of the figure is calculated for each MV2 variant (MV2Mu, MV2MuRnn) with respect to MV2 [134].

CHAPTER 5

Observation of $t\bar{t}H$ production: the four-lepton final state

The top-quark Yukawa coupling, as already discussed in Sections 2.3.1 and 2.5, is a key parameter, determining the structure of the Standard Model at high energy scales; therefore its measurement is among the LHC milestones. This chapter illustrates the search for the Higgs-boson production in association with a top-antitop quark pair, the so-called $t\bar{t}H$ production, in multilepton final states. Particular focus is put on the non-resonant four light lepton final state ($t\bar{t}H \to 4\ell$). This phase space is limited by the small branching ratios involved, but is very pure in signal.

The first section is devoted to a concise description of the profile likelihood fit, the fit model and the construction of hypothesis tests; all of these are fundamental ingredients for both, the search described later on in this chapter and the measurement discussed in Chapter 6. After a brief introduction to the $t\bar{t}H$ multilepton channels, in Section 5.2, a description of the set of data and Monte Carlo simulation samples used is given in Section 5.3. The definition of physics objects and the event selection are provided in Section 5.4, followed by a detailed treatment of the $t\bar{t}H \to 4\ell$ analysis. In Section 5.5 the signal, control and validation regions exploited to carry out the analysis are defined; the background composition of the signal region is described in Section 5.6, while the estimation of the contribution from non-prompt and fake leptons is explained in Section 5.7. Event reconstruction techniques are exploited in order to build discriminating observables (Section 5.8) and combine them with kinematical properties through a multivariate approach, in order to suppress background contamination (Section 5.9); a validation of these observables is performed and discussed in Section 5.10. As shown in Section 5.12, the $t\bar{t}H \to 4\ell$ sensitivity has been optimised taking into account the systematic scheme presented in Section 5.11. Expected and observed fit results for the $t\bar{t}H \to 4\ell$ channel, the latest multilepton and ATLAS overall $t\bar{t}H$ combinations are shown in Section 5.13. The last section is devoted to prospects for the $t\bar{t}H \to 4\ell$ channel sensitivity with the full Run 2 integrated luminosity.

5.1 Statistical analysis: profile likelihood fit and fit model

The profile likelihood technique is widely employed in high energy physics as statistical test for the discovery or the exclusion of a hypothesis; this frequentist approach allows to determine confidence intervals (CL) for a given measurement or test the compatibility of an observation with a certain hypothesis.

Sections 5.1.1-5.1.6 follow the discussion reported in Ref. [149], a necessary overview of the frequentist

approach exploited for the fit results shown in this chapter and Chapter 6.

5.1.1 Experiment sensitivity and p-value

In a search the statistical significance of an observed signal can be quantified by evaluating the compatibility of the observed data with a given hypothesis H; this is achieved by computing a p-value, i.e., the probability, under the assumption of H, of finding observed data in equal or worse disagreement with the predictions of H. Given a test statistic q, the p-value corresponding to the hypothesis H is defined as:

$$p_H = \int_{q_{\text{obs}}}^{\infty} f(q|H)\, dq, \tag{5.1}$$

where q_{obs} is the observed value of the test statistic q in data and $f(q|H)$ represents the probability density function (p.d.f.) of q under the assumption of the H hypothesis.

The p-value is a measurement of the discrepancy between the observed data and a given hypothesis, i.e. the smaller the p-value the less the observed data is compatible with the hypothesis being tested.

Usually two hypotheses are defined to be tested against each other:

- H_0, the null hypothesis, often referred to as "background-only hypothesis" (b), since it constitutes the assumption of having only known and probed physics processes;
- H_1, the alternative hypothesis, often referred to as "signal-plus-background hypothesis" ($s + b$), including both the established physics processes and the sought-after signal process.

The p-value for the signal-plus-background hypothesis is:

$$p_{s+b} = P(q \geq q_{\text{obs}}|s + b) = \int_{q_{\text{obs}}}^{\infty} f(q|s + b)\, dq, \tag{5.2}$$

where $f(q|s + b)$ is the p.d.f. of q under the assumption of the alternative hypothesis. Equivalently for the null hypothesis:

$$p_b = P(q \leq q_{\text{obs}}|b) = \int_{-\infty}^{q_{\text{obs}}} f(q|b)\, dq. \tag{5.3}$$

Figure 5.1 shows the distribution of the test statistic q under the $s + b$ and b hypotheses.

In particle physics the p-value is commonly converted to an equivalent significance, Z, defined as:

$$Z = \Phi^{-1}(1 - p), \tag{5.4}$$

where Φ^{-1} is in the inverse of the cumulative distribution (quantile) of the standard Gaussian. Within the particle physics community a threshold for the p-value of 0.05, corresponding to $Z = 1.64$, is often used to exclude a signal hypothesis, while the *discovery* (often called "observation") corresponds to a significance of at least $Z = 5$, corresponding to a p-value of 2.87×10^{-7}, for the rejection of the background-only hypothesis. The *evidence* for the production of a signal process corresponds to a p-value of 1.349×10^{-3} for the rejection of the background-only hypothesis, which gives a significance of $Z = 3$.

The sensitivity of an experiment is usually quantified by reporting the expected significance obtained under the assumption of different hypotheses. The sensitivity to discover a given signal process s, for instance, is characterised by the expectation value, under the assumption of $s + b$, of the value of Z obtained from a test of the b hypothesis.

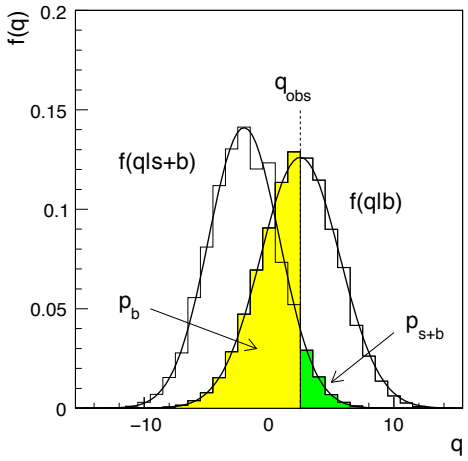


Figure 5.1: The distribution of the test statistic q under the $s + b$ and b hypotheses; the corresponding p-values for the test statistic observed in data (q_{obs}) are also shown [149].

5.1.2 The profile likelihood ratio

The p.d.f $f(x|H)$ refers to the probability density for the observable x for a single event, while $L(\boldsymbol{x}|H)$ describes the probability density for a dataset $\boldsymbol{x}$ with many events.

Given a certain hypothesis H, the likelihood function is obtained by reinterpreting the probability density function $L(\boldsymbol{x}|H)$ as the probability function of H (i.e. parameters of the considered model) given the data; the likelihood function is not a p.d.f., in fact it's not necessarily normalised to unity and it may not be integrable at all. The likelihood function describes the plausibility of a certain hypothesis given specific observed data events. The full structure of $L(\boldsymbol{x}|H)$, with both $\boldsymbol{x}$ and H fixed, is referred to as model.

In the searches described later on in this chapter and in Chapter 6, the alternative hypothesis corresponds to the SM prediction, while the null hypothesis excludes SM $t\bar{t}H$ and triboson processes, respectively. The signal strength μ is defined as

$$\mu = \frac{\sigma_s}{\sigma_s^{\text{SM}}}, \tag{5.5}$$

where σ_s corresponds to the measured cross section for the signal process s (usually in a fiducial phase space defined by a specific selection) and σ_s^{SM} is the corresponding prediction from SM. The $\mu = 1$ case describes the signal-plus-background hypothesis, while $\mu = 0$ represents the background-only hypothesis.

A measurement is usually affected by both, statistical and systematic uncertainties, which are taken into account by introducing a set of *nuisance parameters* $\boldsymbol{\theta} = (\theta_1, \theta_2, ...)$ in the likelihood function, i.e. $L(\boldsymbol{x}|\,\mu, \boldsymbol{\theta})$. When the likelihood function depends on many parameters, the achievable constraints on (or confidence intervals for) μ might be weak; naively μ could be rejected only if it was out of the confidence set for all of the nuisance parameters $\boldsymbol{\theta}$. Usually, and within this context, the main aim is not determining the true values of $\boldsymbol{\theta}$, but rather obtaining tight confidence intervals for μ; this is achieved by considering as test statistic the profile likelihood ratio $\lambda(\mu)$:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}. \tag{5.6}$$

In the numerator, called conditional maximum-likelihood (ML) estimator of $\boldsymbol{\theta}$, $\hat{\hat{\boldsymbol{\theta}}}$ represents the value of $\boldsymbol{\theta}$ maximising L for a specific value of the signal strength μ; the denominator is called maximised unconditional likelihood, corresponding to the global maximum of L in the $(\mu, \boldsymbol{\theta})$ space.

A detailed description of the profile likelihood construction and the fit model is provided in Section 5.1.7.

5.1.3 The test statistics t_μ and $\tilde{t}_\mu$

The definition of $\lambda(\mu)$, given in Eq. 5.6, shows that the profile likelihood is ranging between 0 and 1, where values close to 1 imply good agreement between the data and the hypothesised value of μ. It is convenient to define the test statistic

$$t_\mu = -2 \ln \lambda(\mu), \tag{5.7}$$

where large values of t_μ correspond to an increasing incompatibility between the data and μ.

As illustrated in Section 5.1.1, the level of discrepancy between a hypothesised value of μ and the observed data can be quantified by means of a p-value:

$$p_\mu = P(t_\mu \geq t_{\mu,\text{obs}}|\mu) = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu|\mu)\, dt_\mu, \tag{5.8}$$

where $t_{\mu,\text{obs}}$ is the value of the test statistic t_μ observed in data and $f(t_\mu|\mu)$ represents the p.d.f. of t_μ under the assumption of the signal strength μ.

Minimising the statistic t_μ, or, equivalently, maximising λ_μ, allows to determine the value of the signal strength μ which is the most compatible with the observed data. The uncertainty on μ is retrieved by profiling t_μ as a function of μ: the 68% confidence level error band is defined by the values of μ where t_μ is increased by 0.5 with respect to its minimum.

The sensitivity of an experiment is characterised by the compatibility of a hypothesis μ, when data originate from a different model with $\mu' \neq \mu$. The p-value corresponding to the median t_μ assuming the alternative value μ' (med$[t_\mu|\mu']$) measures the sensitivity of the experiment. Since the relation between t_μ and the p-value is monotonic, it is equal to the median p-value assuming μ'.

An additional constraint on the signal strength can be determined, as a new signal will, usually, increase the event rate beyond the background-only expectation. The alternative test statistic $\tilde{t}_\mu$ is defined in order to take this into account: when data results into a negative $\hat{\mu}$, the best level of agreement between the physical value of μ and the observed data occurs for $\mu = 0$. An alternative λ_μ, $\tilde{\lambda}(\mu)$, can be defined accordingly:

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} & \hat{\mu} \geq 0, \\ \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0, \hat{\boldsymbol{\theta}}(0))} & \hat{\mu} < 0, \end{cases} \tag{5.9}$$

where $\hat{\boldsymbol{\theta}}(0)$ and $\hat{\boldsymbol{\theta}}(\mu)$ are the conditional ML estimators of $\boldsymbol{\theta}$ for a signal strength of 0 and μ, respectively. As for t_μ, a test statistic $\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu)$ can be defined and the level of disagreement between data and the hypothesised value of μ can be quantified by computing the p-value, as reported in Eq. 5.8.

5.1.4 The test statistic q_0 for the discovery of a new process

The test statistic $\tilde{t}_\mu$ with $\mu = 0$ is commonly exploited to test against the background-only hypothesis. From Eq. 5.9 the resulting test statistic is defined as follows:

$$q_0 = \tilde{t}_0 = \begin{cases} -2 \ln \lambda(0) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \tag{5.10}$$

where $\lambda(0)$ is the profile likelihood ratio defined in Eq. 5.6 for $\mu = 0$.

The test statistic t_0, i.e. Eq. 5.7 for $\mu = 0$, may lead to the rejection of the null hypothesis in case of both an upward and a downward fluctuation of the observed data. This is appropriate when the presence of a new phenomenon or physics mechanism, for instance neutrino oscillations, could lead either to an increase or decrease in the mean event rate.

The level of disagreement between the null hypothesis and the observed data can result in the exclusion of the background-only hypothesis, i.e. the discovery of a new positive signal process. The median expected significance of the experiment is computed through the p.d.f. $f(q_0|1)$ of the test statistic q_0, as discussed in general in the previous section; it quantifies how sensitive the analysis is to the signal-plus-background model ($\mu = 1$).

5.1.5 The test statistic q_μ: upper limit on a signal hypothesis

In order to estimate an upper limit on the signal strength μ, we may define the test statistic q_μ as

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} \tag{5.11}$$

where $\lambda(\mu)$ is the profile likelihood ratio defined in Eq. 5.6. Setting $q_\mu = 0$ for $\hat{\mu} > \mu$ allows not to regard data compatible with $\hat{\mu} > \mu$ as representing less compatibility with μ than the data obtained; thus this region is not considered as part of the test. It is important to emphasise that q_0 does not correspond to the special case $\mu = 0$ for q_μ, in fact q_μ is zero if the data fluctuate upward, i.e. $\hat{\mu} > \mu$, while q_0 is zero if the data fluctuate downward, $\hat{\mu} < 0$.

Regarding the signal significance, in order to compute the expected and observed upper limit, the probability distribution of the test statistic $f(q_\mu|\mu')$ needs to be evaluated. For example, the 95% confidence level expected upper limit on the hypothesis $\mu = 0$, using a data model $\mu' = 1$, correspond to the value μ'' such that the p-value of $f(q_0|1)$ using the median of $f(q_{\mu''}|1)$ is equal to 5%. The observed upper limit, instead, corresponds to the value μ'' such that the p-value of $f(q_0|1)$ using $q_{\mu'',\text{obs}}$ is equal to 5%.

If a positive signal strength is expected ($\mu \geq 0$), $\tilde{\lambda}(\mu)$ (see Eq. 5.9) can replace $\lambda(\mu)$ and the test statistic $\tilde{q}_\mu$ is obtained:

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu, \end{cases} = \begin{cases} -2 \ln \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0, \hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{\mu} < 0, \\ -2 \ln \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} & 0 \leq \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \tag{5.12}$$

Numerical examples show that the difference between tests based on q_μ and $\tilde{q}_\mu$ is usually negligible [149], but using q_μ leads to important simplifications.

5.1.6 Asymptotic limit: approximate distribution of the profile likelihood ratio and Asimov dataset

The computation of the p-value of a given hypothesis, using Eq. 5.8, passes through the determination of the full test statistic p.d.f. $f(t_\mu|\mu')$. In principle this can be achieved by generating a large number of pseudo-experiments (often called "MC toys") of the hypothesis μ' and computing the resulting test statistic t_μ; this approach is very expensive on the computational side, as, for instance, a refined definition of the tails of the distribution up to $p_0 \sim 10^{-7}$ requires about 10^8 simulated pseudo-experiments.

In the limit of large statistics, the asymptotic limit, based on Wald's approximation [150], facilitates this process, allowing to determine an approximation of the $f(t_\mu|\mu')$ distribution. Wald showed that, for the case of a single parameter of interest, $f(t_\mu|\mu')$ can be approximated by:

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2} + O(1/\sqrt{N}), \tag{5.13}$$

where $\hat{\mu}$ follows a Gaussian distribution, with a mean μ' and a standard deviation σ, and N represents the size of the data sample. In the asymptotic limit the standard deviation σ can be determined from the "Asimov dataset" [149], described below.

The asymptotic limit corresponds to the case where $\hat{\mu}$ is Gaussian-distributed and the $O(1/\sqrt{N})$ term in Eq. 5.13 can be neglected. In this case the test statistic $t_\mu = -2\ln\lambda(\mu)$ follows a non-central chi-square distribution for one degree of freedom,

$$f(t_\mu;\Lambda) = \frac{1}{2\sqrt{t_\mu}}\frac{1}{\sqrt{2\pi}}\left[\exp\left\{\left(-\frac{1}{2}\left(\sqrt{t_\mu}+\sqrt{\Lambda}\right)^2\right)\right\} + \exp\left\{\left(-\frac{1}{2}\left(\sqrt{t_\mu}-\sqrt{\Lambda}\right)^2\right)\right\}\right], \tag{5.14}$$

where the non-centrality parameter Λ is:

$$\Lambda = \frac{(\mu-\mu')}{\sigma^2}. \tag{5.15}$$

As a result, shown by Wilks in Ref. [151], for the special case of $\mu' = \mu$ the non-centrality parameter Λ is zero and the t_μ test statistic approaches a chi-square distribution for one degree of freedom.

As an example, the level of agreement between the q_0 (introduced in Section 5.1.4) distribution approximated assuming the asymptotic limit as valid (black solid curve) and the MC toys method (histograms) is shown in Figure 5.2. It has to be remarked that the prediction for the test statistic from the asymptotic limit distribution is independent of the number of expected background events, a property that does not apply to the evaluation obtained via MC pseudo-experiments.

The Asimov dataset is defined as an artificial dataset such that, when used to evaluate the estimators for all parameters, the true parameter values are obtained. It replaces the collection of MC pseudo-experiments and allows to determine the asymptotic parametrisation of $f(t_\mu;\Lambda)$ and to study the expected constraints on the nuisance parameters included in the fit model.

5.1.7 Construction of profile likelihood ratios, fit models and nuisance parameters

In order to illustrate the usage of the profile likelihood ratio, let's consider an experiment where for each selected event the value of a specific observable is measured; the expectation value of the i^{th} bin of the resulting histogram $\boldsymbol{n} = (n_1, n_2, ..., n_N)$ can be written as

$$E[n_i] = \mu s_i + b_i, \tag{5.16}$$

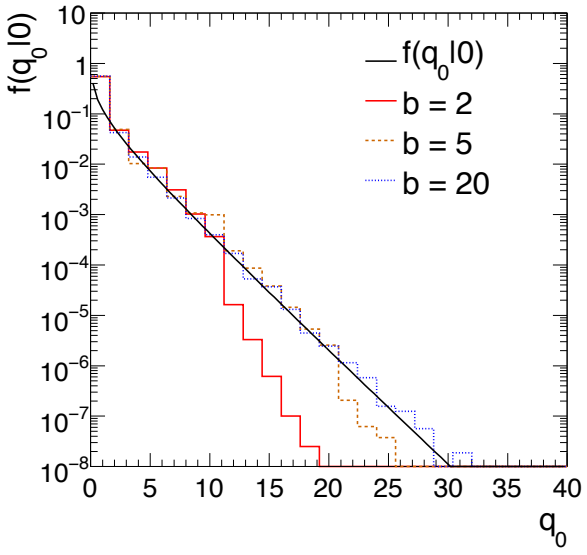


Figure 5.2: The p.d.f. $f(q_0|0)$ for an example counting experiment is shown. The solid curve represents the $f(q_0|0)$ asymptotic approximation, while the histograms are from MC pseudo-experiments, using different numbers of expected events (b) from background processes. As the number of expected background events increases, the asymptotic limit approximation gets more and more reliable, also for small and large values of the test statistic [149].

where s_i and b_i represent the number of expected signal and background events, respectively. These numbers of events are affected by both, statistical and systematic uncertainties, as anticipated in Section 5.1.2. These effects are introduced as nuisance parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_P)$ in the likelihood function.

Usually, nuisance parameters are not known a priori, but they can be rather fitted to the data. Their introduction broadens the profile likelihood as a function of μ, relative to what it would be if their values were assumed to be fixed, and results in a loss of sensitivity. On the other hand, the additional flexibility introduced to parametrise systematic effects will improve the capability of the fit model to describe the observed data.

In the fit model a nuisance parameter is allowed to vary within a given range (commonly determined by auxiliary studies) and granted a p.d.f. $\rho(\theta|\tilde{\theta})$, where $\tilde{\theta}$ is the auxiliary measurement. According to Bayes' theorem, the posterior p.d.f. arises from the measurement $\tilde{\theta}$ as

$$\rho(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta), \tag{5.17}$$

where $\pi_\theta(\theta)$ is the "hyperprior" for those measurements and is often taken as a uniform distribution. As a consequence of Eq. 5.17, if $p(\tilde{\theta}|\theta)$ is a Gaussian or a Poisson distribution, then $\rho(\theta|\tilde{\theta})$ is a normal or a gamma distribution, respectively.

In the following, the functional form for the prior $\rho(\theta|\tilde{\theta})$ is assumed to be one of the three described below:

- Log-normal function, exploited for normalisation systematic uncertainties, where the θ parameter can only assume positive values:

$$\rho(\theta) = (2\pi)^{-1/2}(\ln(\sigma)\theta)^{-1} \exp\left(-\ln^2(\theta/\tilde{\theta})/2\ln^2(\sigma)\right); \tag{5.18}$$

- Gaussian function, assumed for systematic uncertainties which affect the shape of the histograms taken as inputs to the fit model:

$$\rho(\theta) = (2\pi\sigma)^{-1/2} e^{-(\theta-\tilde{\theta})^2/2\sigma^2}; \tag{5.19}$$

- Gamma function, employed to describe the statistical uncertainties due to the finite number of simulated events in a given bin. These nuisance parameters will be referred to as "gamma parameters (γ)" in the following.

The available information from the sample $\boldsymbol{n}$ might not lead to a satisfactory constraint for nuisance parameters having a large impact on the sensitivity. In this case, control samples dominated by background events may be introduced in order to improve the knowledge on the nuisance parameters considered. Calling $\boldsymbol{m} = (m_1, ..., m_M)$ the histogram corresponding to values of an observable in a control sample, it is possible to define the expectation value of m_i as

$$E[m_i] = u_i(\theta), \tag{5.20}$$

where the u_i are computable quantities depending on the nuisance parameters θ, as the number of expected events in a background-dominated region.

The fit model assumes data events to follow a Poisson distribution around the expected number of events. The full likelihood function can be built as the product of Poisson probabilities for all N and M bins:

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k} \prod_{l=1}^{P} \rho(\theta_l), \tag{5.21}$$

where $\rho(\theta_l)$ represents the functional form of the priors for the l^{th} nuisance parameter. Equation 5.21 assumes that all systematic uncertainties are uncorrelated, so that their p.d.f.'s can factorise in the profile likelihood construction.

5.2 $t\bar{t}H$ multilepton decay channels

The $t\bar{t}H$ multilepton decay channels offer a large and heterogeneous set of final states to probe the Standard Model top-quark Yukawa coupling. The combination of ATLAS and CMS results in Run 1 did not give a conclusive answer concerning possible anomalous deviations of this parameter from the SM prediction: the combined fit result of all $t\bar{t}H$ measurements yielded $\mu_{t\bar{t}H} = \sigma_{t\bar{t}H}/\sigma_{t\bar{t}H}^{\text{SM}} = 2.3 \pm 0.7$, where the excess was primarily driven by $t\bar{t}H$ multileptonic final states [28]. Thus, the multilepton measurement in Run 2 could play a crucial role to finally prove or exclude possible anomalies in the top-quark Yukawa coupling.

Figure 5.3 shows the branching ratios of the top-quark and the Higgs-boson decay. The multilepton search effort comprises seven final states, distinguished by the number and flavour of leptons:

- two same-charge light leptons ($\ell = e, \mu$) and no hadronically-decaying τ lepton candidates (2ℓ SS);
- two same-charge light leptons and one hadronically-decaying τ lepton candidate (2ℓ SS+$1\tau_{\text{had}}$);
- three light leptons (3ℓ);
- four light leptons (4ℓ);

- two opposite-charge light leptons and one hadronically-decaying τ lepton candidate (2ℓ OS+$1\tau_{\text{had}}$);
- three light leptons and one hadronically-decaying τ lepton candidate (3ℓ+$1\tau_{\text{had}}$);
- one light lepton and two hadronically-decaying τ lepton candidates (1ℓ+$2\tau_{\text{had}}$).

These signatures are primarily sensitive to $H \to WW^\star$ (with subsequent decay to $\ell\nu\ell\nu$ or $\ell\nu jj$, where j indicates a jet), $H \to ZZ^\star$ and $H \to \tau\tau$ decays.

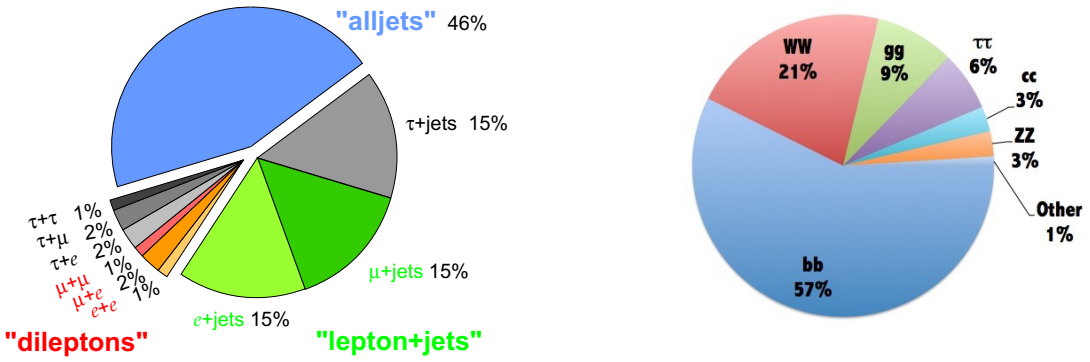


Figure 5.3: The pie chart on the left hand-side represents the branching ratios (BR) of a top-antitop quark pair. The blue slice represents the fully hadronic decay channel, with a BR of 46% (56% when including hadronically-decaying τ); in 2 shades of green is the single lepton (e or μ) + jets BR of 30% (36% when including leptonically decaying τ); the dileptonic BR (excluding τ lepton) is shown in red and orange, summing up to a total of 4% (6.4% when including leptonically decaying τ). The pie chart on the right hand-side shows the branching ratios of the Higgs boson, assuming a mass of 125 GeV.

5.3 Data and Monte Carlo simulation

5.3.1 Dataset

The analysis described in this chapter uses data events collected from proton–proton (pp) collisions recorded by the ATLAS detector at $\sqrt{s} = 13$ TeV. These data events were collected by the ATLAS detector during 2015, 2016 and 2017 with a peak instantaneous luminosity of $\mathcal{L} = 2.1 \times 10^{34}$ cm^{-2}s^{-1} and the Inner B-Layer (IBL) on. The maximum average number of pp interactions per bunch crossing (average pile-up, $< \mu >$) in the dataset is 79.8 and the bunch spacing is 25 ns. After the application of beam and data quality requirements (corresponding to the "Good Run List" [152]), the total integrated luminosity considered corresponds to 80 fb^{-1}.

5.3.2 Signal and background modelling

The $t\bar{t}H \to 4\ell$-like phase space is populated by a non-negligible amount of $t\bar{t}Z$, ZZ and $t\bar{t}$ events. A very large fraction of the overall background contamination is due to these three processes. Example Feynman diagrams representing the production of $t\bar{t}Z$, ZZ and $t\bar{t}$ are shown in Figure 5.4. The four leptons produced in $t\bar{t}Z$ and ZZ events are mostly produced in the hard-scattering primary vertex, whereas in $t\bar{t}$ events at least two of them are produced in secondary vertices (mostly heavy-flavour hadron decays).

The background composition of the phase space sensitive to the non-resonant[1] $t\bar{t}H \to 4\ell$ production is discussed in greater detail in Section 5.6.

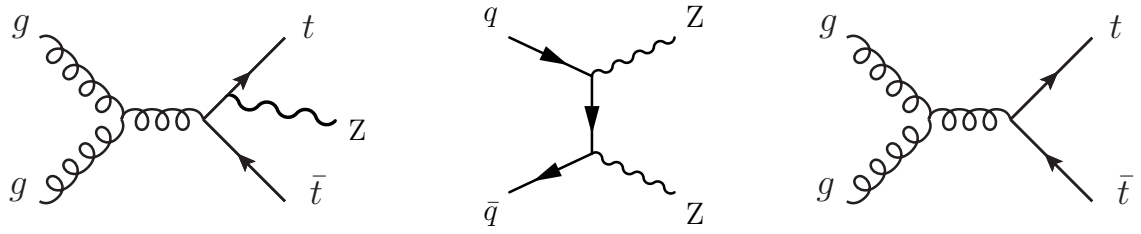


Figure 5.4: Example leading order diagrams for the production of $t\bar{t}Z$, ZZ and $t\bar{t}$.

This section summarises the set of MC simulation samples used in this analysis. Interesting properties of, and differences between, generators used in this analysis are provided in Section 3.5. The simulation of the $t\bar{t}H$ process is performed with the Powheg-BOX [110] generator, while $t\bar{t}W$ and $t\bar{t}Z$ are generated with a next-to-leading order (NLO) QCD matrix element computed by MadGraph5_aMC@NLO [109] and interfaced to Pythia 8 [140] for parton showering and fragmentation into particles. In the case of $t\bar{t}Z$, the inclusive $t\bar{t}\ell^+\ell^-$ matrix element is computed, including off-shell Z and γ^* contributions up to a lepton-pair invariant mass of 5 GeV ($m(\ell^+\ell^-) > 5$ GeV).

The overall $t\bar{t}H$ cross section is 507.1 fb, computed at NLO in both QCD and electroweak couplings [153–159] as reported in Ref. [160, 161]. The relative uncertainties are $^{+5.8\%}_{-9.2\%}$ from QCD renormalisation and factorisation scale choice and ±3.6% from parton distribution function (PDF) uncertainties (uncertainties on α_s are included).

The cross sections for $t\bar{t}W$ and $t\bar{t}Z$ production, including the process $pp \to t\bar{t}\ell^+\ell^-$ over the full Z/γ^* mass spectrum, are computed at NLO in QCD and electroweak couplings using the configuration described in Refs. [109, 159]. The total cross section used for $t\bar{t}\ell^+\ell^-$ (with $m(\ell^+\ell^-) > 5$ GeV) is 123.7 fb and 600.8 fb for $t\bar{t}W$. QCD scale and PDF+α_s uncertainties on these cross sections are summing up to 12% and 3 – 4%, respectively.

Diboson processes are generated with Sherpa [107]. Matrix elements for the full diboson production, where both bosons decay to a pair of charged leptons (ee, $\mu\mu$ or $\tau\tau$), have been generated with up to three additional partons in the final state. Diboson events are generated at NLO accuracy for 0 and 1 additional partons and at LO accuracy for 2 and 3 additional partons. A generator-level cut of 5 GeV on the transverse momentum of the two highest p_{T} leptons is imposed for all samples and any opposite-charge lepton-pair invariant mass is required to be above 4 GeV. The loop-induced gg-initiated diboson production is also simulated with Sherpa with zero or one additional parton at LO. Diboson production through vector boson scattering (VBS) is generated at LO with up to 1 additional parton using the same generator.

$t\bar{t}$ events are generated with Powheg-BOX and interfaced with Pythia 8 for the parton showering and fragmentation. The overall $t\bar{t}$ cross section is 832^{+46}_{-51} pb, calculated at next-to-next-to-leading order (NNLO) in perturbative QCD including resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms [162]. The same generator and showering setup has been used to model other backgrounds involving the production of top quarks, such as single top (s- and t-channel, and Wt). A $t\bar{t} + \gamma$ sample is used in addition to $t\bar{t}$ to improve the modelling of the $t\bar{t}$ background with a hard scattered photon. The sample is generated with MadGraph5_aMC@NLO interfaced to the Pythia 8 parton shower: the matrix

[1] By "non-resonant" here is meant that the four leptons are not coming from a $H \to ZZ^* \to 4\ell$ decay.

element is generated at LO including the decays of the top quarks. The photons can be emitted either in the production or in the decay stage. In order to avoid infrared and collinear singularities, a set of kinematic requirements are applied ($p_{\mathrm{T}}(\gamma) > 10\,\mathrm{GeV}$, $|\eta(\gamma)| < 5$ and ΔR between the photon and any charged particle above 0.2). The cross section is normalised to the NLO reference calculation [163]. The partial overlap between the $t\bar{t}$ and $t\bar{t}+\gamma$ samples is removed by vetoing in the $t\bar{t}$ simulation the component characterised by a hard photon ($p_{\mathrm{T}} > 15\,\mathrm{GeV}$) produced at the matrix-element level. In the following $t\bar{t}+\gamma$ is incorporated in the $t\bar{t}$ background.

The Z + jets background is simulated using the Sherpa generator, with matrix elements calculated for $\ell^+\ell^-$ with 0, 1 or 2 additional partons at NLO and $\ell^+\ell^-$ with 3 or 4 additional partons at LO accuracy. These matrix elements are merged with the Sherpa parton shower using the MEPS@NLO prescription [164].

The Wt channel production of a single top quark together with a Z boson (WtZ) is generated with MadGraph5_aMC@NLO and showered with PYTHIA8. The Z boson is required to decay to two charged leptons. The generation is performed at NLO in QCD. Diagram removal is employed to remove the overlap of WtZ with $t\bar{t}Z$ and with $t\bar{t}$ production followed by a three body top decay ($t \to WZb$) [165]. The procedure also removes the interference between WtZ and these two processes.

Rare backgrounds, such as the associated production of a single top quark and a Higgs boson (i.e. $tHqb$ and WtH), $t\bar{t}W^+W^-$, $t\bar{t}t$ and $t\bar{t}t\bar{t}$, are also simulated and the cross sections are computed using the MadGraph5_aMC@NLO generator at NLO in QCD. The full leptonically decaying on-shell VVV samples are generated with Sherpa, and with 0 additional partons at NLO in QCD and 1 or 2 additional partons at LO in QCD.

Matrix-element and parton-shower generators used for simulating the signal and background processes are summarised in Table 5.1. In all simulations the Higgs-boson (top-quark) mass is assumed to be exactly 125(172.5) GeV. All Monte Carlo samples are processed through a complete simulation of the ATLAS detector response based on Geant 4 [118]. Additional simulated pp collisions have been overlaid to the main event, in order to model the effects from additional ("pile-up") pp collisions in the same and nearby bunch crossings. All simulated events were processed using the same reconstruction algorithms and analysis chain as data events, as described in Section 3.3. Simulated events, as discussed in greater detail in Section 5.11, are corrected so that the object reconstruction and identification efficiencies, energy scales and energy resolutions match those determined in data events. All MC simulation samples described above are generated in two statistically independent sets:

- *mc16a*, approximately reproducing the 2015 and 2016 pile-up spectrum;
- *mc16d*, whose simulated pile-up spectrum is similar to the one in 2017 data.

Reproducing the same pile-up spectrum allows to optimise the effective statistics for the simulated samples and their description of the two data sub-sets. For this purpose, mc16a (mc16d) events are reweighted in order to perfectly match the pile-up profile observed in the 2015+2016 (2017) dataset.

5.4 Object definition and event preselection

This section describes the basic definition of objects, such as leptons and jets, and the event preselection underlying the region definitions provided in Section 5.5.

The primary vertex of each event is chosen, among reconstructed primary vertices, as the vertex with the highest $\sum p_{\mathrm{T}}^2$ of associated tracks. Events with significant noise in the calorimeters or data corruption are removed.

Table 5.1: The table shows the configurations used for the generation of signal and background events. If only one parton distribution function (PDF) is shown, the same one is used for both the matrix-element (ME) and the parton-shower generators; if two are shown, the first is used for the matrix-element calculation and the second for the parton shower. "V" refers to production of an electroweak boson (W or Z boson). "Tune" refers to the underlying-event tune of the parton shower generator. "MG5_aMC" refers to MadGraph5_aMC@NLO 2.2.1 [109]; "Pythia 6" refers to version 6.427 [166]; "Pythia 8" refers to version 8.2 [140]; "Herwig++" refers to version 2.7 [167]. Samples using Pythia 6 or Pythia 8 have heavy-flavour hadron decays modeled by EvtGen 1.2.0 [168]. All samples include leading-logarithm photon emission, either modelled by the parton-shower generator or by PHOTOS [169].

Process	ME generator (alternative)	Parton shower (alternative)	PDF set	Tune
$t\bar{t}H$	Powheg-BOX [110]	Pythia 8 [140]	NNPDF 3.0 NLO [101]/ NNPDF 2.3 LO [170]	A14 [144]
	(-)	(Herwig++ [167])		
	(MG5_aMC [109])	(-)		
$tHqb$	MG5_aMC	Pythia 8	CT10 [141]	A14
WtH	MG5_aMC	Herwig++	CT10 /CTEQ6L1 [172]	UE-EE-5 [171]
$t\bar{t}W$	MG5_aMC	Pythia 8	NNPDF 3.0 NLO /NNPDF 2.3 LO	A14
	(Sherpa [107])	(Sherpa)		
$t\bar{t}(Z/\gamma^*)$	MG5_aMC	Pythia 8	NNPDF 3.0 NLO /NNPDF 2.3 LO	A14
	(Sherpa)	(Sherpa)		
$t(Z/\gamma^*)$	MG5_aMC	Pythia 8	CTEQ6L1	Perugia2012 [173]
$Wt(Z/\gamma^*)$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}t$, $t\bar{t}t\bar{t}$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}W^+W^-$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}$	Powheg-BOX	Pythia 8	CT10/CTEQ6L1	Perugia2012
$t\bar{t}\gamma$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
s-, t-channel, Wt single top	Powheg-BOX	Pythia 6 [166]	CT10 /CTEQ6L1	Perugia2012
VV, $qqVV$, VVV	Sherpa	Sherpa	NNPDF 3.0 NLO	Sherpa default
$Z \to \ell^+\ell^-$	Sherpa	Sherpa	NNPDF 3.0 NLO	Sherpa default

Events are required to have been selected by at least one of the dilepton triggers (ee, $e\mu$, $\mu\mu$). This logical OR combination of dilepton triggers provides a relatively large signal acceptance, since it allows the minimum lepton-p_{T} threshold for the leptons firing the trigger to be lower than what allowed by single-lepton triggers.

Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that are associated with charged particle tracks reconstructed in the inner detector. Only candidates with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ are considered. The electron η reconstructed from the electromagnetic cluster is required to be smaller than 2.47 in absolute value. Candidates in the "crack" region (transition region between different electromagnetic calorimeter components, see Section 3.2.4), namely $1.37 < |\eta| < 1.52$, are rejected. A multivariate likelihood discriminant ("electron LH") combining shower shape and track information is used to distinguish real electrons from hadronic showers mimicking electron calorimetric showers ("fake" electrons). For the object preselection, a loose electron discriminant working point is used, and no isolation is required. To further reduce contributions from electrons originating from secondary vertices ("non-prompt" electrons), the track is required to be consistent with originating from the primary vertex; the transverse impact parameter significance ($|d_0|/\sigma_{d_0}$) and the longitudinal impact parameter ($|z_0 \sin\theta_\ell|$) are required to be smaller than 5 and 0.5 mm, respectively.

Muon candidates are reconstructed by combining inner detector tracks with track segments or full tracks in the muon spectrometer. In the region $|\eta| < 0.1$, where the muon spectrometer coverage is reduced, muon candidates are also reconstructed from inner detector tracks matched to isolated energy deposits in the calorimeters consistent with the energy loss of a minimum-ionising particle. Candidates are required to satisfy $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.5$, to account for the ATLAS detector coverage. The longitudinal impact parameter selection is the same as for electron candidates, while the transverse impact parameter requirement for muon candidates is tighter: $|d_0|/\sigma_{d_0} < 3$. No isolation is required in the preselection of muon candidates.

Jets are reconstructed from calibrated topological clusters built from energy deposits in the calorimeters, using the anti-k_t algorithm with a radius parameter $R = 0.4$ (see Section 3.4.2). Jets with energy contributions likely arising from noise or detector effects are removed, and only jets satisfying $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$ are used in this analysis. For jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$, a jet-track association algorithm is used to reject jets arising from pile-up collisions. The average efficiency of this association is 92% per jet.

Jets containing b-hadrons are identified (b-tagged) via the MV2 algorithm, described in Section 4.3.4. The working point used for this search corresponds to an average efficiency of 70% for b-jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$ in $t\bar{t}$ simulated events. The expected rejection factors against light and c-jets are 313 and 8, respectively.

Hadronically-decaying τ lepton candidates (τ_{had}) are reconstructed from clusters in the calorimeters and associated ID tracks. The candidates are required to have either one or three associated tracks, with a total charge of ± 1. Candidates with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$, excluding the electromagnetic calorimeter transition region, are considered. A BDTG discriminant exploiting calorimeter and tracking-based variables is used to identify τ_{had} candidates and reject generic jet backgrounds.

To avoid double counting of reconstructed objects and to remove leptons likely originating from hadron decays, ambiguities are resolved in the following logical order: any electron candidate within $\Delta R = 0.1$ of another electron candidate with higher p_{T} is removed; any electron candidate within $\Delta R = 0.1$ of a muon candidate is removed; any jet within $\Delta R = 0.3$ of an electron candidate is removed; if a muon candidate and a jet lie within $\Delta R = \min\left(0.4,\ 0.04 + 10/p_{\mathrm{T}}^{\mu}[\mathrm{GeV}]\right)$ of each other, the jet is kept and the muon is removed. Any τ_{had} candidate within $\Delta R = 0.2$ of an electron or a muon candidate is removed and any jet within $\Delta R = 0.3$ of a τ_{had} candidate is considered only as a τ_{had} candidate in events with two light leptons. This algorithm is applied to the preselected objects and all further requirements on

the leptons and jets start with the candidates passing this selection. Table 5.2 gives an overview of the overlap-removal procedure adopted for this analysis.

Events are then distributed among the seven multilepton channels, based on the multiplicity of objects after the preselection. No events are, therefore, shared between different channels.

Table 5.2: Summary of the overlap-removal procedure between electrons, muons, jets and hadronically-decaying tau leptons.

Keep	**Remove**	**Cone size (ΔR)**
electron	electron (lower p_{T})	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	$\min\left(0.4,\ 0.04 + 10/p_{\mathrm{T}}^{\mu}[\mathrm{GeV}]\right)$
electron	tau	0.2
muon	tau	0.2
tau	jet	0.3

5.5 The four-lepton channel: signal and validation regions

This section describes the final selection of reconstructed leptons and the definition of regions exploited for the analysis targeting the 4ℓ final state. This selection is applied on top of the preselection described in Section 5.4, common to all $t\bar{t}H$ multilepton channels.

The 4ℓ channel requires exactly four light leptons (leptonically-decaying tau leptons included) such that $p_{\mathrm{T}} > 10$ GeV, while no requirement on the multiplicity of hadronically-decaying tau candidates is made; any jet also reconstructed as τ_{had} candidate is treated as a jet. In the following, leptons are assumed to be sorted by decreasing transverse momentum.

The sum of the electric charge of leptons reconstructed and selected in each event must be zero, as expected for signal events characterised by four light leptons produced in the primary vertex. A dedicated optimisation of the lepton isolation definition, based on the sensitivity to signal events, has been performed in the $t\bar{t}H$ four-lepton channel; as a result, all four leptons are required to pass an ATLAS standard isolation working point. This selection corresponds to requiring muons (electrons) to fulfil $p_{\mathrm{T}}^{\text{cone 30(20)}}/p_{\mathrm{T}}^{\mu(e)} < 0.15$, where $p_{\mathrm{T}}^{\text{cone 30(20)}}$ is the sum of p_{T} of tracks in the $\Delta R = 0.3(0.2)$ cone around the muon (electron). The isolation requirement has been optimised in order to allow a significant reduction of fake and non-prompt lepton contributions (namely reduced by a factor 10), while retaining 72% of signal events. Applying tighter isolation selections can lead to a very significant (60 – 80%) loss in signal acceptance and, consequently, in expected sensitivity. This is due to two main reasons: $t\bar{t}H \to 4\ell$ events tend to be "busy" environments, given the presence of four light leptons and two b-jets, and to produce soft leptons; in fact, isolation requirements tend to be less efficient at low lepton p_{T}.

A 4ℓ pre-MVA signal region is defined. In order to reject $t\bar{t}Z$ and ZZ background processes, all same-flavour $\ell^+\ell^-$ pairs in the event must satisfy $|m(\ell^+\ell^-) - 91.2\ \mathrm{GeV}| > 10\ \mathrm{GeV}$ (usually called "Z veto"). To remove leptons from leptonic decays of quarkonia resonances, all same-flavour $\ell^+\ell^-$ pairs must satisfy $m(\ell^+\ell^-) > 12$ GeV. To reduce contamination from other Higgs-boson production processes (such as gluon or vector-boson fusion production) and to ensure statistical independence from dedicated resonant $H \to ZZ^* \to 4\ell$ measurements [36], a Higgs-boson veto ($m(4\ell) < 115\ \mathrm{GeV}$ and $m(4\ell) > 130\ \mathrm{GeV}$) is

applied. Since two b-jets are expected to be produced in the $t\bar{t}H \to 4\ell$ final state, events are required to have at least two reconstructed jets, of which at least one must be identified as coming from the hadronisation of a b-quark. The 4ℓ pre-MVA region is split[2] into *Z-enriched* and *Z-depleted* signal regions, according to whether same-flavour $\ell^+\ell^-$ pairs are selected or not, respectively.

Events entering the $t\bar{t}Z \to 4\ell$ validation region are also required to have exactly four light leptons sharing the same object definition used in the pre-MVA signal region; leading and subleading leptons are also required to pass $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $p_{\mathrm{T}} > 15\,\mathrm{GeV}$, respectively, there has to be exactly one same-flavour $\ell^+\ell^-$ lepton pair within the Z-boson mass window $\left(\left|m(\ell^+\ell^-) - 91.2\,\mathrm{GeV}\right| < 10\,\mathrm{GeV}\right)$ and at least two reconstructed jets, of which at at least one b-tagged.

5.6 Backgrounds

The background events entering the $t\bar{t}H \to 4\ell$ pre-MVA signal region can be split into two main categories:

- irreducible background, processes in which all of the reconstructed leptons in the final state are directly coming from the hard-scattering primary vertex (prompt leptons);
- reducible background, namely processes where at least one of the leptons reconstructed in the final state is arising from a different source. Leptons originating from hadron decays (non-prompt leptons), interactions with the detector (fake leptons or charge misreconstruction) and misreconstruction of objects (fake leptons) belong to this category.

Figure 5.5 shows the expected composition of the 4ℓ pre-MVA signal region in simulation; in this region, 25% of the events is expected to be $t\bar{t}H$ events, while 69% and 6% of the events are expected to be due to irreducible and reducible background processes, respectively.

5.6.1 Irreducible backgrounds

The irreducible background events have four prompt leptons reconstructed in the final state, thus with properties very similar to the signal; all of these processes are estimated exploiting the prediction from simulation. The largest contribution is due to the $t\bar{t}Z/\gamma^\star$ production, where two leptons are originating from the leptonic decays of the top-quark pair and the other two from the decay of the Z or $\gamma^\star$ boson. This signature is, therefore, very similar to the signal. Contributions from ZZ (the only diboson process providing four prompt leptons) and rare processes (mainly where a top quark is produced, e.g. $t\bar{t}t\bar{t}$ and $t\bar{t}W^+W^-$) are also contributing to the expected yields in the 4ℓ pre-MVA region. The Higgs-boson production in association with a single top quark is treated as a background process, since its expected contribution to the four lepton, and more in general multilepton, final state is limited by both the small SM cross section and the branching ratios of the decays of interest.

The modelling of the major backgrounds, namely $t\bar{t}Z/\gamma^\star$ and ZZ, is explored in dedicated validation regions, non-overlapping with the 4ℓ pre-MVA signal region and enriched with the processes under study. The $ZZ \to 4\ell$ validation region requires exactly four leptons sharing the same object definition of the 4ℓ pre-MVA signal region; there has to be two same-flavour $\ell^+\ell^-$ lepton pairs, of which at least one within the Z-boson mass window (namely $|m_{\ell\ell} - 91.2\ \mathrm{GeV}| < 10\ \mathrm{GeV}$). There is good agreement between data and simulation in this region, both for the overall MC normalisation and relevant distributions; Figure 5.6 shows some examples. The expected purity in $ZZ \to 4\ell$ of this validation region is $\sim 98\%$. The study of

[2] See Section 5.12 for further details.

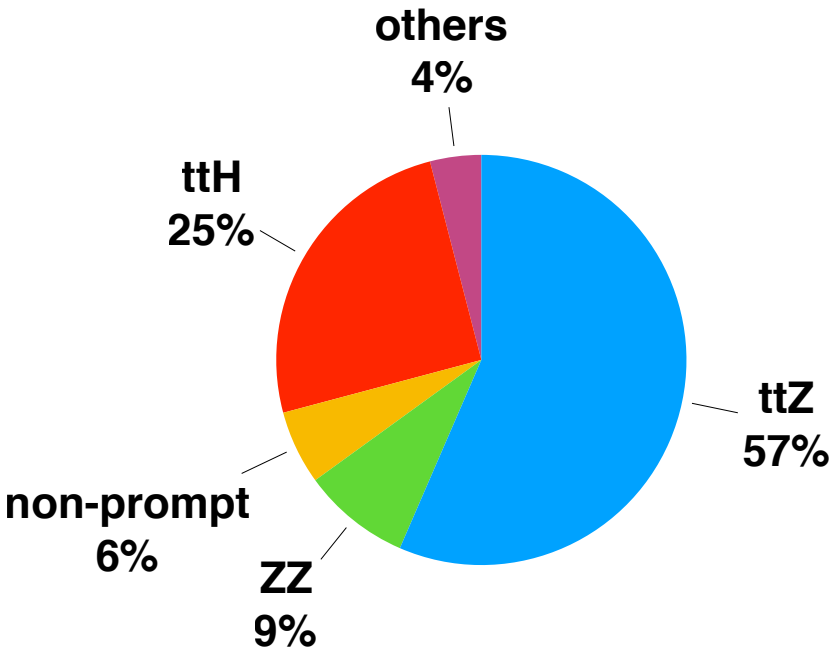


Figure 5.5: The pie chart represents the fraction of processes expected in the 4ℓ pre-MVA signal region. The fraction of $t\bar{t}H$ events with three or less leptons directly coming from the primary vertex is here included as part of $t\bar{t}H$.

the $t\bar{t}Z \rightarrow 4\ell$ validation region is discussed in Section 5.10, together with the modelling of the 4ℓ MVA input observables.

5.6.2 Reducible backgrounds

The reducible background events completely originate from events with one or more non-prompt or fake leptons. Studies relying on the truth record in simulation (see Section 5.7) indicate that, given the chosen lepton definition, most of these events are expected to be due to real electrons and muons produced either by the conversion of a photon in the detector material or the decay of a hadron.

With respect to other multilepton phase spaces, the two- and three-lepton final states in particular, the four-lepton channel benefits from a relatively small contribution of reducible backgrounds. This is due to the fact that the vast majority of eligible processes with a large cross section, especially $t\bar{t}$ and Z + jets, would have to produce at least two non-prompt/fake leptons, in addition to two prompt leptons from the decay of the electroweak bosons; the production cross section of $t\bar{t}W$, which can produce three genuine prompt leptons in the final state, is three orders of magnitude smaller than the $t\bar{t}$ one, and thus it does not contribute significantly to the four-lepton phase space. The estimation of both the normalisation and the systematic uncertainties relative to these backgrounds is illustrated in the next section.

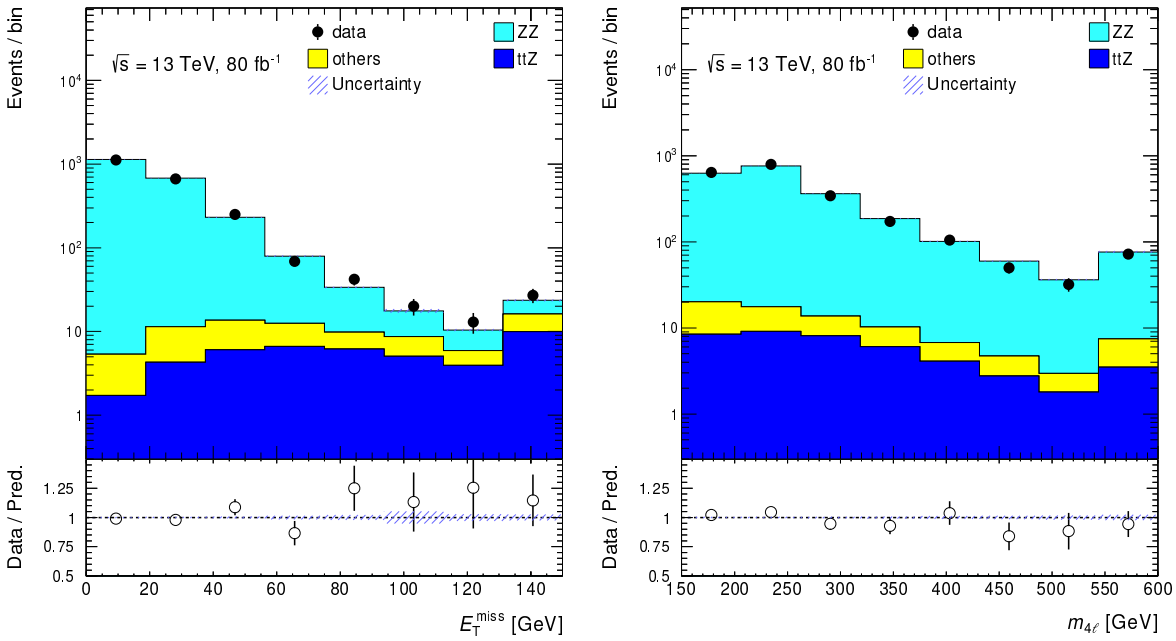


Figure 5.6: E_T^{miss} (left) and invariant mass of the four reconstructed leptons (right) in $ZZ \to 4\ell$ validation region. The lower panel shows the data-to-MC ratio. Only statistical uncertainties are shown, as shaded blue bands.

5.7 The fake scale factor method: data-driven estimation of non-prompt backgrounds

The mechanisms underlying the contribution from processes where one or more lepton candidates are originating either from the semileptonic decay of a hadron or from a jet/photon shower, mimicking electron calorimetric showers, are known to be complex and hard to describe in Monte Carlo simulation. In order to get a reliable estimate of both, the number of expected events with non-prompt and fake leptons and related systematic uncertainties, a new method has been developed and employed in the $t\bar{t}H \to 4\ell$ analysis.

The "*fake scale factor method*" is a semi data-driven technique, whose aim is to correct the contribution from different sources of non-prompt and fake leptons. For this purpose, it relies on the Monte Carlo simulation *truth record* of leptons, i.e. a record stored by the MC generator which allows to trace back the origin of the lepton produced in the simulated event.

According to their truth-record origin, all leptons reconstructed in simulated events can be divided into three families as follows:

- **prompt**, lepton directly matched to the hard-scattering primary vertex;
- **heavy**, non-prompt lepton matched to a bottom or charm hadron;
- **light**, fake lepton coming from the conversion of a real photon or non-prompt lepton from light hadrons.

5.7.1 Fit of fake scale factors

The fake scale factor method is based on a simultaneous six-bin profile likelihood fit to data of normalisation factors scaling the rate of non-prompt and fake leptons source-by-source. Three different normalisation factors, called "Fake Scale Factors" (FSF), are simultaneously fitted to data: heavy $\left(\lambda_e^h\right)$ and light $\left(\lambda_e^\ell\right)$ normalisations for electrons, and one overall normalisation for muons $\left(\lambda_\mu\right)$. Only one FSF is introduced for muons, as the occurrence of non-prompt muons from light sources has an extremely low rate and proves to be negligible in the considered phase space. As the available amount of data for reducible backgrounds is limited in the four-lepton phase space, the fit is performed in the following three-lepton control region:

- exactly three light leptons whose $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and fulfilling the isolation requirements;
- lepton total charge ± 1;
- no same-flavour $\ell^+\ell^-$ pair;
- exactly one or two reconstructed jets.

This 3ℓ control region is extremely pure in $t\bar{t}$ events (see Table 5.3), therefore it is enriched with fake and non-prompt leptons, as a $t\bar{t}$ system can not produce three prompt leptons. All of the three leptons are required to pass the same lepton definition optimised for the 4ℓ signal region, in order to reproduce the exact object definition fulfilled by non-prompt and fake leptons entering the signal-like phase space. In fact, the rate of non-prompt and fake leptons, as well as its reliability in simulation, is known to strictly depend on the lepton definition (requiring leptons to be isolated allows to significantly reduce contributions from reducible backgrounds, as discussed in Section 5.5).

It has to be remarked that, given the lepton definition discussed in Sections 5.4 and 5.5, most of the electrons from light sources are coming from the conversion of a photon, both in the 4ℓ pre-MVA signal and the $t\bar{t}$-enriched 3ℓ control regions. The fraction of light electrons coming from photon conversion is about 99%. This justifies the treatment of light electrons as a single category of leptons in the FSF fit model.

The $t\bar{t}$-enriched 3ℓ control region defined above is characterised by a good data-to-simulation agreement, as demonstrated by Table 5.3 and Figure 5.7.

In order to fit the FSFs, the control region is split into two sub-regions, defined according to the flavour of the leptons: two electrons ($ee\mu$) or one electron ($e\mu\mu$); this split allows to get one region enriched with non-prompt and fake electrons and one region enriched with non-prompt muons. The fit is performed considering three bins of the $E_{\mathrm{T}}^{\mathrm{miss}}$ spectrum in both regions. The $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution allows to partially "disentangle" the two sources of background electrons. In fact, $t\bar{t}$ events featuring a non-prompt lepton, mostly produced in the decay of a heavy-flavour hadron via $b \to \mu\,\nu\,X$ and $b \to c \to \mu\,\nu\,X$ (see Section 4.4), will tend to have larger missing transverse momentum than events with an electron from photon conversion. The normalisation of events with three prompt leptons is known to be well described, thus is assumed to be fixed at SM value. On the other hand, the normalisation of non-prompt and fake electrons from heavy and light sources and non-prompt muons are free to float in all of the six bins. The resulting FSFs, using the 80 fb^{-1} dataset, are:

- $\lambda_e^\ell = 1.37^{+1.31}_{-1.28}$;
- $\lambda_e^h = 0.89 \pm 0.21$;
- $\lambda_\mu = 1.07 \pm 0.05$.

Table 5.3: Expected yields and observed numbers of events in the inclusive $t\bar{t} \to 3\ell$ control region. Reported uncertainties are purely statistical.

$t\bar{t} \to 3\ell$ CR	Exp./obs. events
$t\bar{t}$	$1\,823 \pm 21$
Z + jets	45 ± 11
WZ	56.6 ± 1.5
$t\bar{t}W$	35.21 ± 0.62
$t\bar{t}H$	9.84 ± 0.36
$t\bar{t}Z$	8.07 ± 0.25
ZZ	9.08 ± 0.28
others	19.2 ± 1.9
Total expected	$2\,005 \pm 24$
data	1 974

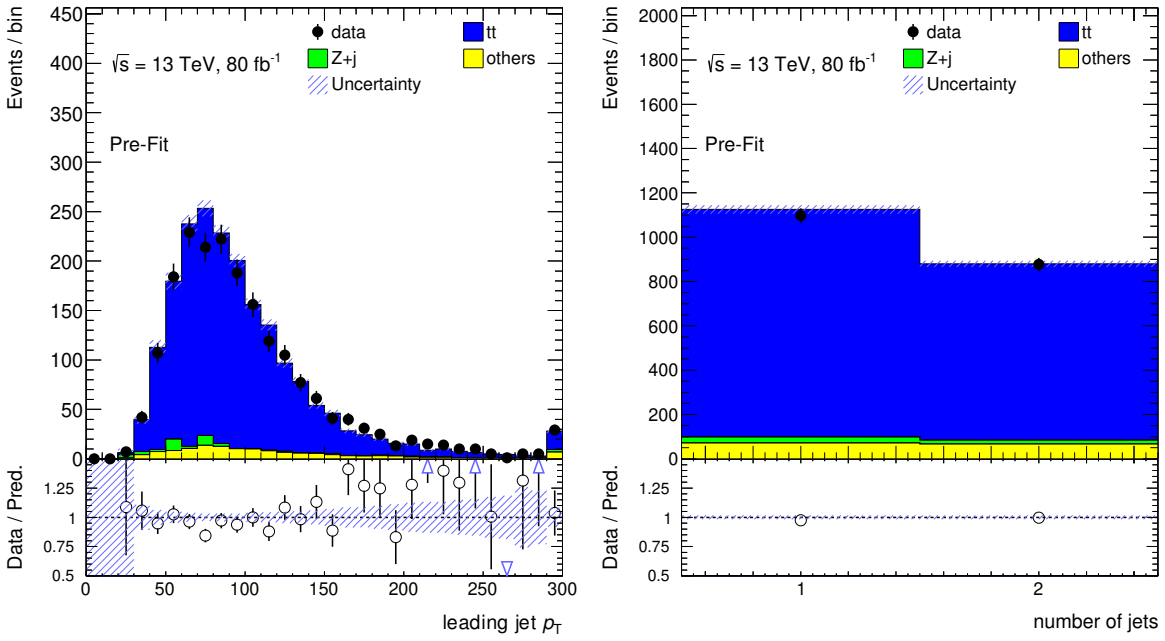


Figure 5.7: Data-to-simulation comparison for the leading jet p_{T} (left) and the number of reconstructed jets (right) in the $t\bar{t}$-enriched 3ℓ control region. The uncertainties, represented by shaded blue bands, are purely statistical.

The quoted uncertainties are only statistical. Figure 5.8 shows a summary of pre- and post-fit yields in the two fitted regions, while Figures 5.9 and 5.10 show pre- and post-fit $E_{\mathrm{T}}^{\mathrm{miss}}$ shapes in each of the two regions. A good post-fit data-to-simulation agreement is observed within statistical uncertainties in all of the six fitted bins.

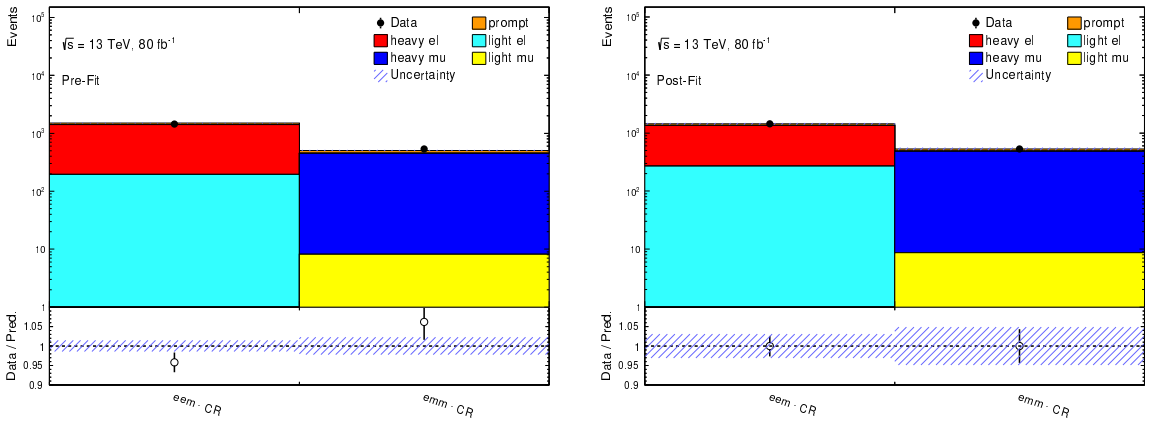


Figure 5.8: Pre-fit (left) and post-fit (right) yields for events with three prompt leptons ("prompt") and events with a fake electron or a fake muon coming from a heavy or light source ("heavy el", "light el", "heavy mu", "light mu"). Heavy and light muons are scaled by the same λ_μ.

The data-to-simulation pre-fit comparison for the $E_{\mathrm{T}}^{\mathrm{miss}}$ is also reported in Figure 5.11, showing $ee\mu$ and $e\mu\mu$ regions separately and with a finer[3] binning. A good agreement between prediction and observed data events is found within statistical uncertainties.

5.7.2 Extraction of systematic uncertainties on data-driven estimation

The correction to fake and non-prompt lepton rates, achieved by fitting their normalisation to data in an inclusive control region with leptons above 10 GeV, may a priori depend on the p_{T} of the non-prompt lepton. Most of the non-prompt and fake contributions, namely more than 80%, entering the 4ℓ pre-MVA signal region is due to leptons whose p_{T} is below 50 GeV; therefore the dependence of FSFs has been studied up to this threshold. Table 5.4 shows fitted FSFs for six different requirements on the p_{T} of same-sign leptons. This choice is due to the fact that in a 3ℓ event the non-prompt/fake lepton is expected to be one of the two same-sign leptons. As the FSFs show a reasonable stability, an overall conservative 40% systematic uncertainty is considered. This covers the whole variation for all of the three FSFs and results in an overall estimated systematic and statistical uncertainty of:

- $\lambda_e^\ell = 1.37 \pm 1.41$;
- $\lambda_e^h = 0.89 \pm 0.41$;
- $\lambda_\mu = 1.07 \pm 0.43$.

[3] As compared to the binning used in the FSF fit configuration.

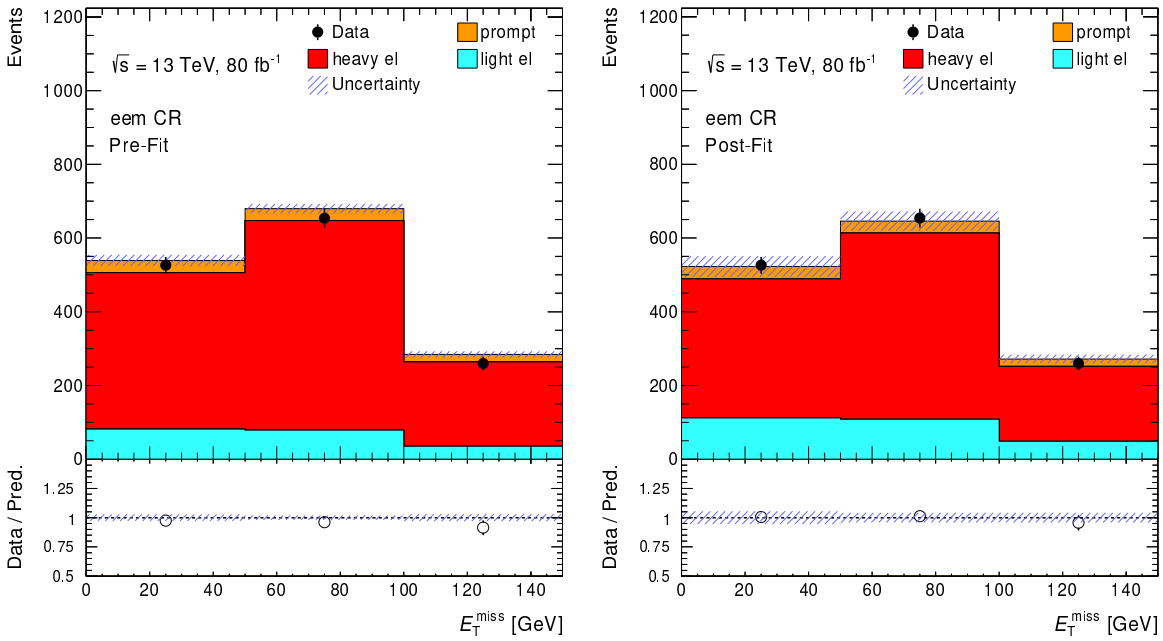


Figure 5.9: Pre-fit (left) and post-fit (right) yields for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $ee\mu$ sub-region. Shown contributions are: events with three prompt leptons ("prompt") and events with a fake/non-prompt electron coming from a heavy or light source ("heavy el", "light el").

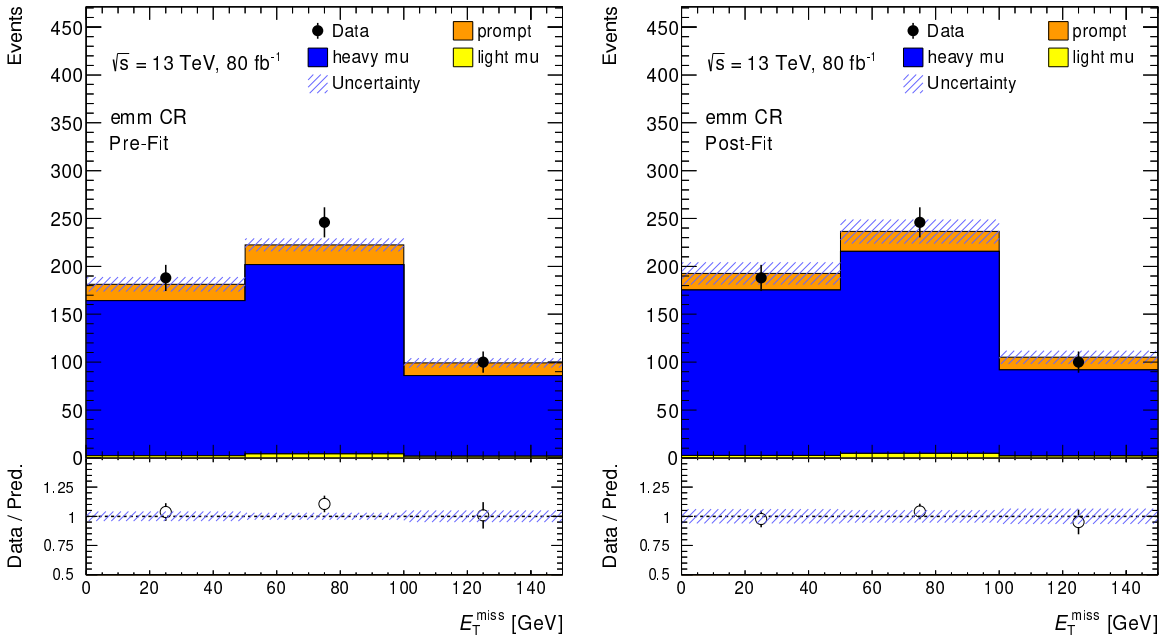


Figure 5.10: Pre-fit (left) and post-fit (right) yields for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $e\mu\mu$ sub-region. Shown contributions are: events with three prompt leptons ("prompt") and events with a fake/non-prompt muon coming from a heavy or light source ("heavy mu", "light mu"). Heavy and light muons are scaled by the same λ_μ.

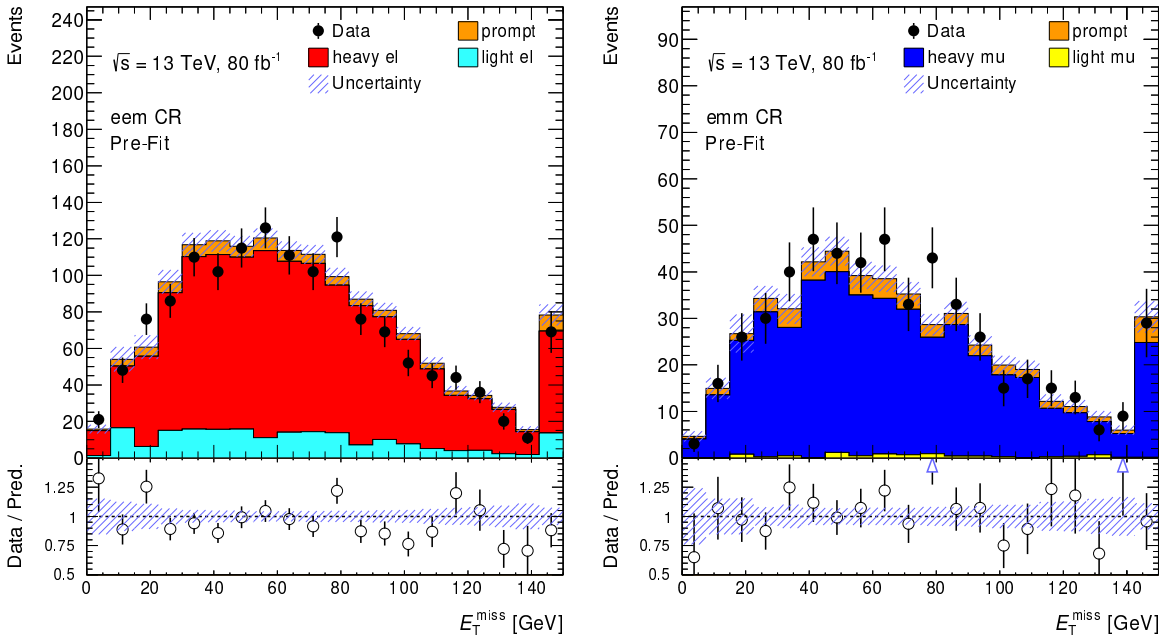


Figure 5.11: Pre-fit shapes for the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in $ee\mu$ (left) and $e\mu\mu$ (right) subregions. Shown contributions are: events with three prompt leptons ("prompt") and events with a fake electron or a fake muon coming from a heavy or light source ("heavy el", "light el", "heavy mu", "light mu").

Table 5.4: The table shows FSFs determined applying different requirements on the p_{T} of same-sign leptons (p_{T}^{SS}) in the $t\bar{t}$-enriched 3ℓ control region. Red labels show the largest variation with respect to the nominal value for each FSF. Shown uncertainties are purely statistical.

p_{T}^{SS} > [GeV]	10	15	20	25	30	40	50
λ_e^h	0.89 ± 0.21	$0.97^{+0.15}_{-0.26}$	$0.95^{+0.18}_{-0.31}$	$0.92^{+0.21}_{-0.31}$	$0.88^{+0.29}_{-0.33}$	$0.98^{+0.22}_{-0.33}$	$1.14^{+0.26}_{-0.34}$
λ_e^ℓ	$1.37^{+1.31}_{-1.28}$	$0.83^{+1.71}_{-0.83}$	$0.96^{+1.99}_{-0.96}$	$1.08^{+1.94}_{-1.08}$	$1.51^{+1.96}_{-1.51}$	$0.93^{+1.78}_{-1.33}$	$0.98^{+2.32}_{-1.62}$
λ_μ	1.07 ± 0.05	1.13 ± 0.06	1.14 ± 0.07	1.16 ± 0.07	1.12 ± 0.08	1.12 ± 0.10	1.15 ± 0.11

5.8 Event reconstruction

The event reconstruction, based on both, reconstructed observables and truth-record origin information in simulation, has been explored and developed in the $t\bar{t}H \rightarrow 4\ell$ analysis to discriminate $t\bar{t}H$ from background processes; in particular the observables are optimised to separate $t\bar{t}H$ events against $t\bar{t}Z$, i.e. the largest background contribution to the Z-enriched signal region. The construction of two discriminants, fed as inputs to the training of a BDT, is discussed in this section.

5.8.1 Template-based discriminator ($\mathcal{P}.\mathcal{M}.\mathcal{E}.$)

The template-based discriminator is built following an approximated Matrix-Element approach, the so-called Pseudo-Matrix Element [174–176] ($\mathcal{P}.\mathcal{M}.\mathcal{E}.$) approach, and results in one of the most discriminating observables of the analysis. The $\mathcal{P}.\mathcal{M}.\mathcal{E}.$ intends to discriminate those processes exhibiting a very signal-like behaviour. The approach aims at identifying signal-like events by partially[4] reconstructing resonances (namely top quarks, Higgs and Z bosons) and exploiting peculiar kinematic properties of the reconstructed and selected objects. The $\mathcal{P}.\mathcal{M}.\mathcal{E}.$ relies on the so-called "truth-matching": the origin of a reconstructed object is determined, at simulation level, by matching it to the closest true particle (fermion or boson before hadronisation and parton showering) within $\Delta R < 0.3$.

The $\mathcal{P}.\mathcal{M}.\mathcal{E}.$ discriminant is defined as:

$$\mathcal{P}.\mathcal{M}.\mathcal{E}.(\mathbf{x}) = \frac{P^{\text{sig}}(\mathbf{x})}{P^{\text{bkg}}(\mathbf{x})}, \tag{5.22}$$

where $P^{\text{sig}}(\mathbf{x})$ and $P^{\text{bkg}}(\mathbf{x})$ represent the probability density functions (p.d.f.s) of a given event under the signal hypothesis ($t\bar{t}H$) and under the background hypothesis ($t\bar{t}Z$), respectively. Both p.d.f.s are functions of $\mathbf{x}$, representing the four-momentum vectors of all final-state particles at the reconstruction level: the leptons (ℓ), jets and b-jets (j and b, respectively) selected in the analysis. In this analysis, $\mathbf{x}$ is extended to include not only the four-momenta of leptons and jets, but also the invariant masses and angular properties; this allows to exploit the different mass, spin and angular distribution of the involved (daughter) resonances (Higgs, Z and W bosons).

$P^{\text{sig}}(\mathbf{x})$ and $P^{\text{bkg}}(\mathbf{x})$ are defined as the product of the normalised p.d.f.s (see the list below) of each reconstructed resonance in the event. The p.d.f.s are determined by employing "templates", which are constructed from simulated signal and background events using the reconstructed lepton and/or jets corresponding to the correct parton-object assignment, determined via truth-matching. These templates are constructed as unit-normalised one-dimensional histograms.

Identifying ℓ_0 and ℓ_1 with the leptons truth-matched to the top quark, and ℓ_2 and ℓ_3 with the leptons truth-matched to either the Higgs or the Z boson, the exploited templates are:

- all lepton and jet four-momenta;
- $m(\ell_2, \ell_3)$ and $\Delta R(\ell_2, \ell_3)$;
- $m(\ell_0, b)$, $m(\ell_1, b)$, $m(\ell_2, b)$ and $m(\ell_3, b)$;
- $\Delta\phi(\ell_2, b)$ and $\Delta\phi(\ell_3, b)$;

[4] The presence of two or more neutrinos in the final state does not allow for full reconstruction of the $t\bar{t}H/t\bar{t}Z$ system.

where ℓ_0 (ℓ_2) and ℓ_1 (ℓ_3) are sorted by p_{T} and b is the true b-jet[5] closest to a lepton. The quantities $m(a,b)$, $\Delta R(a,b)$ and $\Delta\phi(a,b)$ denote the invariant mass of a and b, and the ΔR and $\Delta\phi$ angular distance between a and b, respectively.

For each event, four different signal- ($P^{\mathrm{sig}}(\mathbf{x})$) and background-likeness ($P^{\mathrm{bkg}}(\mathbf{x})$) probabilities are computed, as there are four possible opposite-sign lepton pair combinations. The final discriminant is taken as the base-10 logarithm of the signal-to-background probability ratio for the combination of objects yielding the largest signal-likeness probability: $\log\left(P^{\mathrm{sig}}(\mathbf{x})/P^{\mathrm{bkg}}(\mathbf{x})\right)\Big|_{P^{\mathrm{sig}}_{\mathrm{max}}}$. According to the Neyman-Pearson lemma [177] this ratio is the most powerful discriminant between the signal and background processes. Figure 5.12 shows the separation provided by the $\mathcal{P}.\mathcal{M}.\mathcal{E}.$ discriminator between $t\bar{t}H$ and $t\bar{t}Z$ processes.

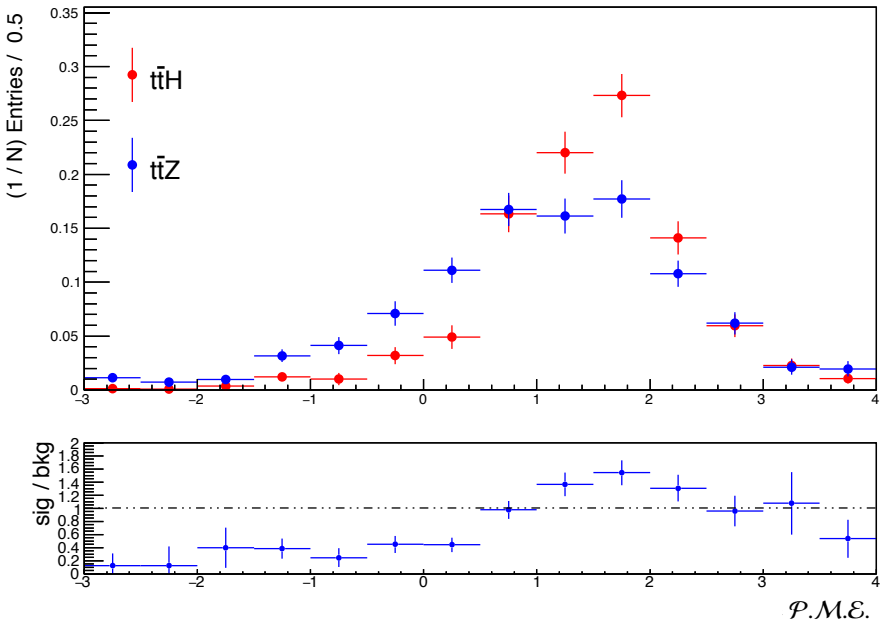


Figure 5.12: Normalised $\mathcal{P}.\mathcal{M}.\mathcal{E}.$ distribution for $t\bar{t}H$ and $t\bar{t}Z$ events in the Z-enriched signal region.

5.8.2 The effective invariant mass ($m_{\ell_2\ell_3 E_{\mathrm{T}}^{\mathrm{miss}}}$)

According to the truth-record information, more than 97% of the $t\bar{t}Z$ events entering the Z-enriched signal region are expected to be given by two leptonically-decaying top quarks and $Z/\gamma^* \to \ell^+\ell^-$ and more than 70% of the $t\bar{t}H$ events are expected to be given by two leptonically-decaying top quarks and $H \to WW^\star \to \ell^+\bar{\nu}\ell^-\nu$; the Feynman diagram in Figure 5.13 shows the largest contribution to the non-resonant $t\bar{t}H \to 4\ell$ final state.

A discriminant observable is built in order to exploit differences in the spatial distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ in $t\bar{t}H(\to WW^\star) \to 4\ell 4\nu$ and $t\bar{t}Z \to 4\ell 2\nu$: the invariant mass of two lepton candidates from the boson (either Higgs or Z boson) together with the $E_{\mathrm{T}}^{\mathrm{miss}}$ ($m_{\ell_2\ell_3 E_{\mathrm{T}}^{\mathrm{miss}}}$). The two lepton candidates are selected out of the four reconstructed leptons by excluding the closest lepton to a b-tagged jet and the closest

[5] The 4ℓ pre-MVA region requires at least one jet to be b-tagged, but simulated $t\bar{t}H$ and $t\bar{t}Z$ events (almost) always have two true b-jets.

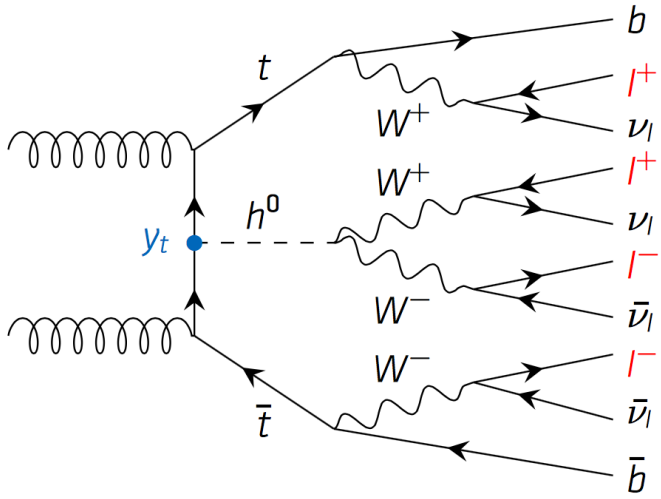


Figure 5.13: The Feynman diagram shows the largest process contributing to the non-resonant $t\bar{t}H \rightarrow 4\ell$. The four prompt leptons are all produced by the leptonic decay of a W boson.

opposite-charged lepton to another jet. In $\sim$ 70% of the cases this association is found to be correct for signal events (according to the truth-matching method). Figure 5.14 shows the $m_{\ell_2\ell_3 E_{\mathrm{T}}^{\mathrm{miss}}}$ shape, comparing $t\bar{t}H$ to $t\bar{t}Z$.

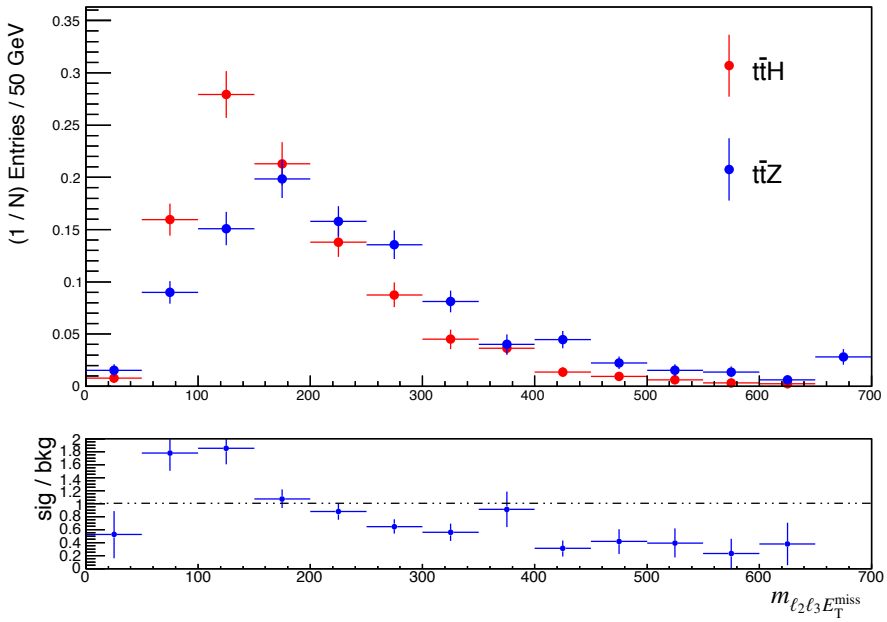


Figure 5.14: Normalised $m_{\ell_2\ell_3 E_{\mathrm{T}}^{\mathrm{miss}}}$ distribution for $t\bar{t}H$ and $t\bar{t}Z$ events in the Z-enriched signal region.

5.9 BDTG training in the Z-enriched signal region

A gradient-boosted decision tree discriminant (BDTG, see Section 4.2.4 for details) is trained in order to reduce the relatively large contribution from $t\bar{t}Z$ in the Z-enriched signal region, i.e. requiring events to have same-flavour $\ell^+\ell^-$ pairs on top of the pre-MVA selection provided in Section 5.5.

The final set of seven discriminating variables exploited as input to the BDTG training is the result of a detailed performance optimisation, based on the integral of the ROC curve for the test evaluation. The variables are:

- template-based discriminator ($\mathcal{P}.\mathcal{M}.\mathcal{E}.$);
- effective invariant mass ($m_{\ell_2\ell_3 E_{\mathrm{T}}^{\mathrm{miss}}}$);
- four-lepton invariant mass ($m_{4\ell}$);
- same-flavour $\ell^+\ell^-$ invariant mass of the best Z-boson candidate ($m_{\ell\ell}^{\text{best } Z}$);
- remaining two-lepton invariant mass ($m_{\ell\ell}^{\text{other pair}}$);
- scalar sum of lepton transverse momenta ($H_{\mathrm{T}}^{\mathrm{lep}}$);
- $E_{\mathrm{T}}^{\mathrm{miss}}$.

Figure 5.15 shows the distributions of the five variables that enter the BDTG training besides the two event-reconstruction observables discussed in Section 5.8. They all provide an effective separation between signal and background.

The BDTG is trained to select $t\bar{t}H$ events against $t\bar{t}Z$ and ZZ in the Z-enriched signal region with the following configuration in the TMVA program:

- number of trees: 400;
- maximum depth: 3;
- minimum node size: 10%;
- number of cuts (granularity): 12.

Simulated events with negative matrix-element weight, present for both signal and background events, are ignored in the training; these events, however, are considered for testing and application purposes.

Figure 5.16 shows the BDTG distribution for signal and background, comparing the MVA response on the training and the testing samples (statistically independent sub-sets of the simulation dataset). No evidence for overtraining is observed. Furthermore, in order to make full use of the available simulated samples and mitigate possible overtraining effects, a cross-training procedure is adopted: half of the simulation sample is used for training a BDTG applied to the remaining half and viceversa. As a consequence no MVA score is evaluated on an event used for the training of the BDTG algorithm itself.

Figure 5.17 shows the comparison between the $t\bar{t}H$ BDTG shape predicted by Powheg-BOX and the same predicted by MadGraph5_aMC@NLO (an alternative $t\bar{t}H$ generator). In Figure 5.18 the $t\bar{t}Z$ BDTG shape predicted by MadGraph5_aMC@NLO+Pythia 8 (the nominal $t\bar{t}Z$ simulation) is compared to the one predicted by the alternative $t\bar{t}Z$ simulation (Sherpa). Predicted shapes are in agreement within statistical uncertainties; this proves that the modelling of relevant shapes for the $t\bar{t}H$ and $t\bar{t}Z$ processes is not sensitive to the choice of event or showering generators. More details about the considered modelling systematic uncertainties are discussed in Section 5.11.

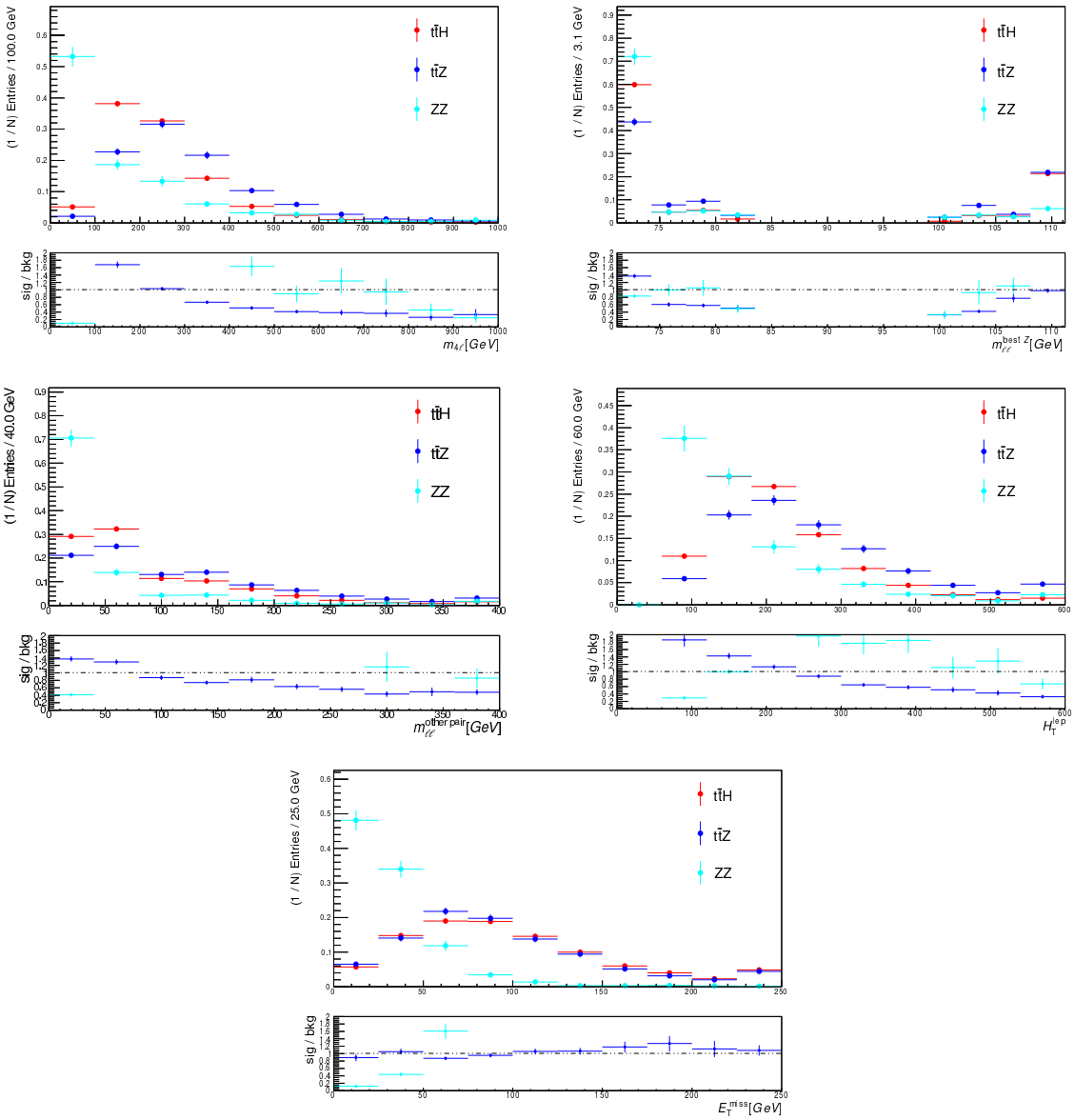


Figure 5.15: Distributions of five BDTG input variables: $m_{4\ell}$, $m_{\ell\ell}^{\text{best } Z}$, $m_{\ell\ell}^{\text{other pair}}$, $H_{\text{T}}^{\text{lep}}$ and $E_{\text{T}}^{\text{miss}}$; the $t\bar{t}H$ (red), $t\bar{t}Z$ (blue) and ZZ (light blue) shapes are compared in the Z-enriched signal region.

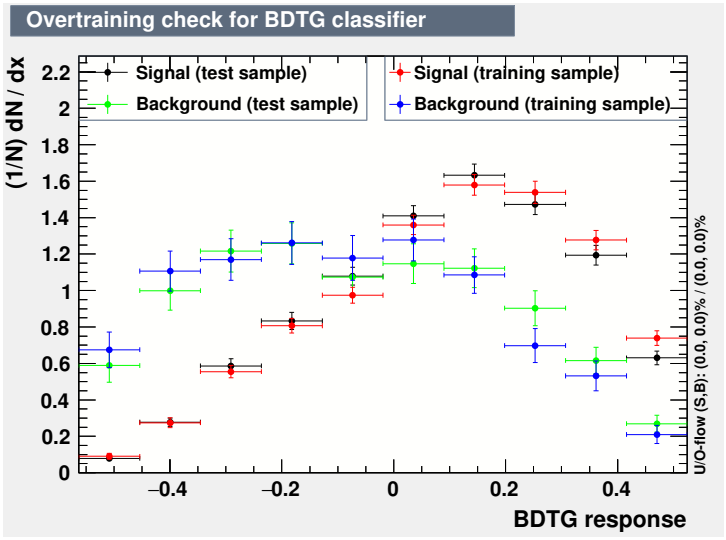


Figure 5.16: Comparison between the BDTG response on test and training samples for signal and all background events entering the Z-enriched signal region.

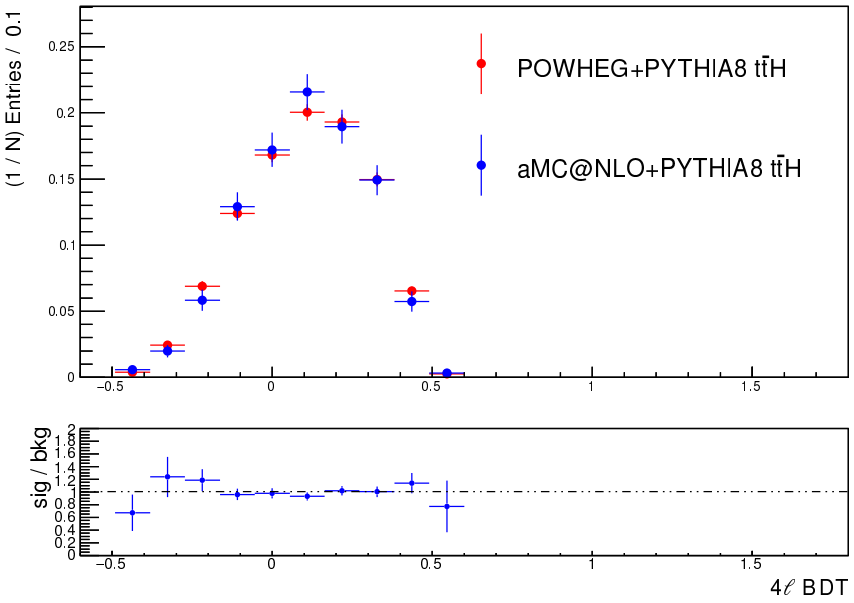


Figure 5.17: Comparison between the $t\bar{t}H$ BDTG response predicted by POWHEG-BOX+PYTHIA 8 (in red) and the same predicted by MADGRAPH5_aMC@NLO+PYTHIA 8 (in blue) in the Z-enriched signal region. The lower panel shows the ratio between the two shapes. Statistical-only uncertainties are shown.

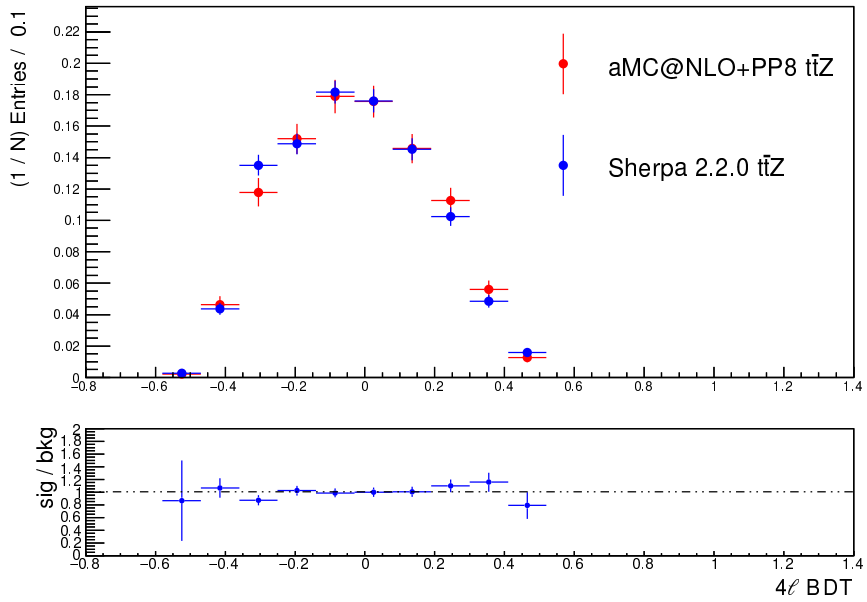


Figure 5.18: Comparison between the $t\bar{t}Z$ BDTG response predicted by MadGraph5_aMC@NLO+Pythia 8 (in red) and the same predicted by Sherpa (in blue) in the Z-enriched signal region. The lower panel shows the ratio between the two shapes. Statistical-only uncertainties are shown.

5.10 Modelling of MVA inputs in the $t\bar{t}Z \to 4\ell$ validation region

This section shows the modelling of the BDTG input variables in the four-lepton $t\bar{t}Z$-enriched validation region defined in Section 5.5. Table 5.5 shows the expected and observed yields in this region. The resulting validation region is pure in $t\bar{t}Z$ events, yielding 76% of the total number of expected events. In order to better compare the shapes of the input variables the $t\bar{t}Z$ normalisation is scaled by 1.10 to match the number of observed data events. The data-to-simulation comparison for the MVA input variables is reported in Figures 5.19 and 5.20. The shapes of the MVA input variables predicted by simulation are in good agreement with observation within statistical uncertainties.

Table 5.5: Expected yields and observed number of events in the $t\bar{t}Z$-enriched validation region. Reported uncertainties are purely statistical.

$t\bar{t}Z$ VR	Exp./obs. events
$t\bar{t}H$	2.444 ± 0.041
$t\bar{t}Z$	57.81 ± 0.68
ZZ	6.33 ± 0.13
others	9.55 ± 0.59
Total expected	76.14 ± 0.91
data	82

The scale factor of 1.10 for the $t\bar{t}Z$ normalisation is obtained subtracting all MC expected yields, but $t\bar{t}Z$, to the number of observed data events and dividing by the $t\bar{t}Z$ expected yield; the aim of this

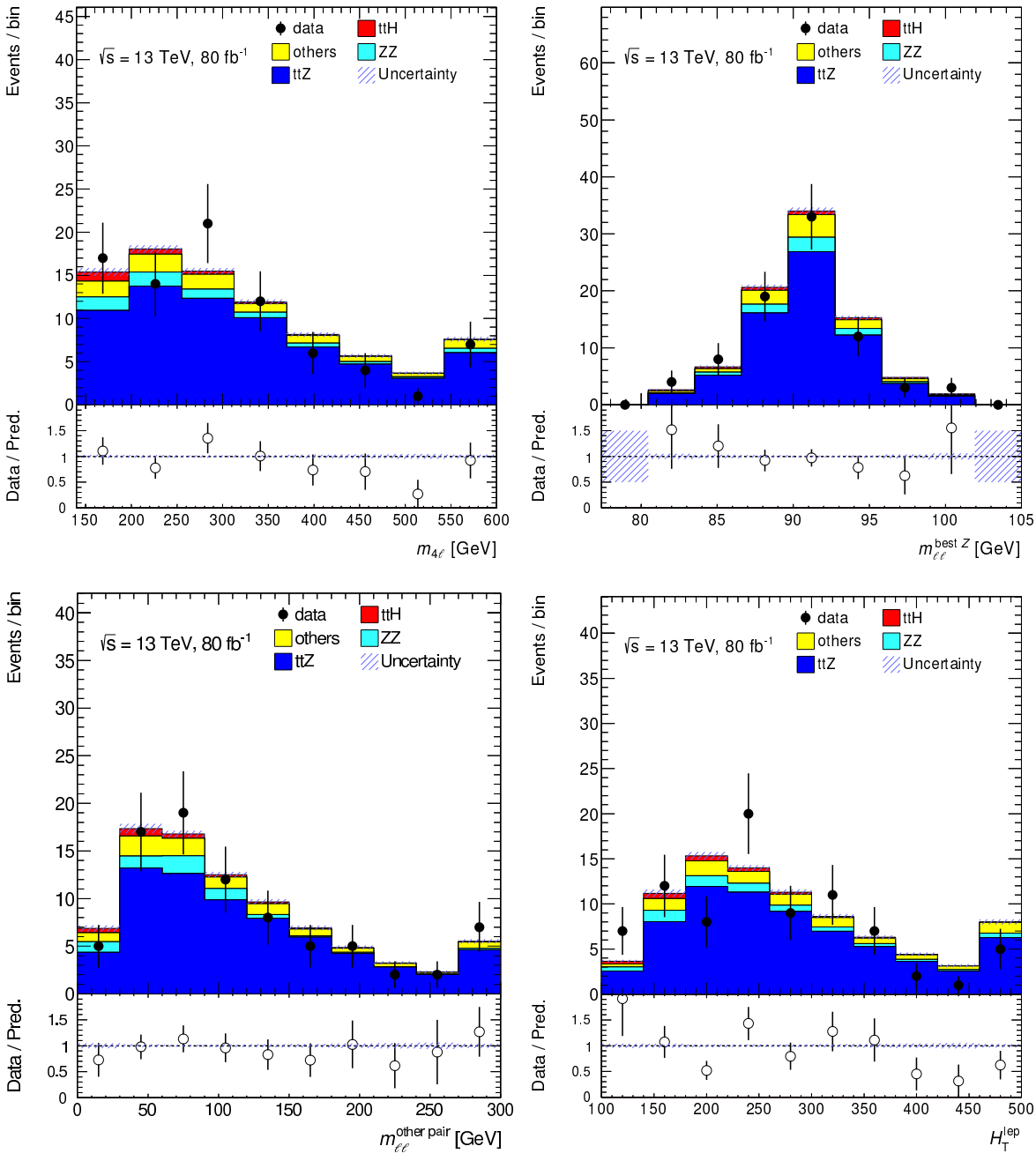


Figure 5.19: Data-to-simulation comparison in the four-lepton $t\bar{t}Z$ control region for four of the input variables to the BDTG training: $m_{4\ell}$, $m_{\ell\ell}^{\text{best } Z}$, $m_{\ell\ell}^{\text{other pair}}$ and $H_{\text{T}}^{\text{lep}}$. The $t\bar{t}Z$ normalisation is scaled by 1.10 to match the number of observed data events. Shown uncertainties are only statistical.

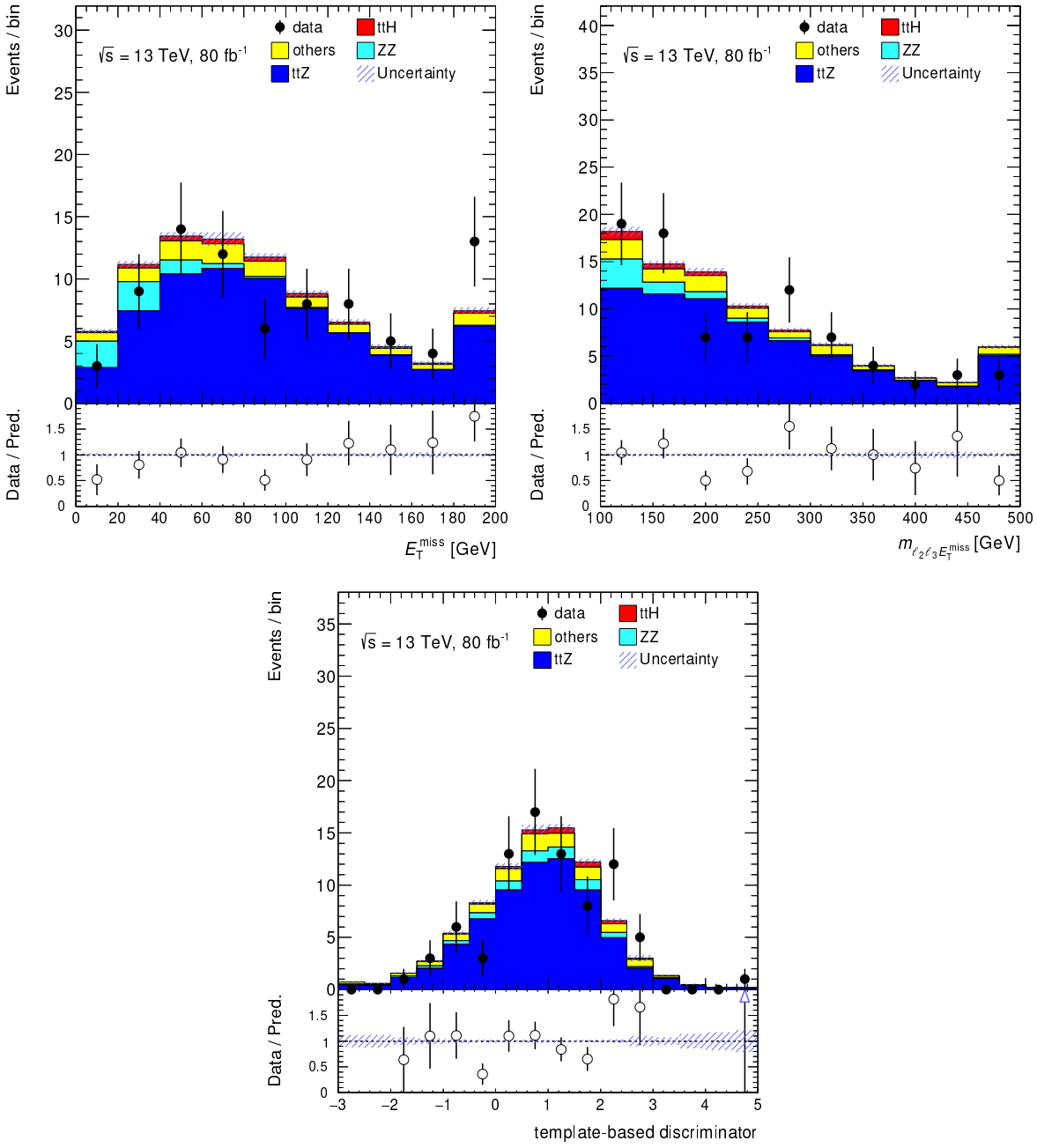


Figure 5.20: Data-to-simulation comparison in the four-lepton $t\bar{t}Z$ control region for three of the input variables to the BDTG training: $E_{\mathrm{T}}^{\mathrm{miss}}$, $m_{\ell_2\ell_3 E_{\mathrm{T}}^{\mathrm{miss}}}$ and $\mathcal{P}.\mathcal{M}.\mathcal{E}.$. The $t\bar{t}Z$ normalisation is scaled by 1.10 to match the number of observed data events. Shown uncertainties are only statistical.

procedure is to show that relevant shapes are well modelled in a $t\bar{t}Z$-dominated region, independently of the overall $t\bar{t}Z$ normalisation in the validation region. Propagating statistical uncertainties leads to a scale factor of 1.10 ± 0.12; the inclusion of QCD scale uncertainties on the $t\bar{t}Z$ cross section ($\sim 10\%$, see Section 5.11) leads to 1.10 ± 0.16. Thus this observation is fully compatible with SM predictions.

5.11 Systematic uncertainties

The systematic uncertainties taken into account for this analysis are summarised in Table 5.6; they can be grouped into six main categories and they are discussed in this section. Each systematic uncertainty is introduced in the fit model as an independent nuisance parameter and impacts the estimated signal and/or background rates.

Uncertainties on the integrated luminosity, trigger efficiency, reconstruction of physics objects and MC models follow an established treatment within the ATLAS collaboration; a total of 112 nuisance parameters is introduced in the fit model. Their impact on the signal strength is summarised in Figure 5.23 in Section 5.13.1.

A *pruning* procedure is applied to this large set of nuisance parameters included in the profile likelihood fit. The pruning consists in the removal of small systematic components, which do affect the result below a given threshold, to speed up the fit procedure and make it more robust. In the $t\bar{t}H \to 4\ell$ analysis, the normalisation components are pruned if their relative impact is found to be smaller than 0.5% on all processes in all regions.

5.11.1 Detector systematic uncertainties

The uncertainty in the combined 2015–2017 integrated luminosity is 2.0%. It is derived following a methodology similar to that detailed in Ref. [178] and using the LUCID-2 detector for the baseline luminosity measurements [179], from calibration of the luminosity scale using x–y beam-separation scans.

The identification and the reconstruction of objects relies on several properties of the objects themselves; these are subject to experimental uncertainties, which are evaluated and combined on an event-by-event basis, depending on the nature and characteristics of the reconstructed objects entering each final state. The experimental systematic uncertainties on single objects are centrally evaluated by the ATLAS performance groups and are used in the analysis either as an overall event re-weighting or as a rescaling of the object energy and momentum. These experimental uncertainties are related to the trigger efficiency, to the reconstruction and identification of light leptons, to the reconstruction and b-tagging of jets and to the reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$.

The uncertainties associated to the light lepton selection originate from the limited knowledge of efficiencies related to trigger (TRIG), reconstruction (RECO), identification (ID) and isolation (ISO) selection [91, 92] and lepton momentum scale and resolution [92, 180]. For muons also the track-to-vertex association (TTVA) efficiency systematic uncertainties, related to the requirements applied on the muon impact parameter components, are included. The reconstruction and identification efficiencies of electrons and muons, as well as the efficiency of the triggers used to select events to be recorded, differ between data and simulation. In order to correct for this effect, scale factors and associated uncertainties are derived using a tag-and-probe method applied to electrons and muons produced in the decay of Z, W bosons and J/ψ resonances. The accuracy of the lepton momentum scale and resolution in simulation is checked through reconstructed distributions of the $Z \to \ell^+\ell^-$ and $J/\psi \to \ell^+\ell^-$ masses. For electrons, also energy-to-momentum ratio studies are performed with $W \to e\nu$ events. Small discrepancies are

Table 5.6: Sources of systematic uncertainty considered in the $t\bar{t}H \rightarrow 4\ell$ analysis. Some of the systematic uncertainties are split into several uncorrelated components, as indicated by the number in the rightmost column.

Systematic uncertainty	Components
Integrated luminosity	1
Pile-up modelling	1
Physics Objects	
Electron	6
Muon	15
Jet energy scale and resolution	32
Jet vertex fraction	1
Jet flavour tagging	17
$E_{\mathrm{T}}^{\mathrm{miss}}$	4
Total (Experimental)	77
Data-driven non-prompt/fake leptons	
Fake scale factors	1
Total (Data-driven reducible background)	1
$t\bar{t}H$ modelling	
Cross section (QCD and PDF variations)	2
Renormalisation and factorisation scales	3
Parton shower and hadronisation model	1
Higgs-boson branching ratio	4
Shower tune	1
$t\bar{t}Z$ modelling	
Cross section (QCD and PDF variations)	2
Renormalisation and factorisation scales	3
Matrix-element MC generator	1
Shower tune	1
Modelling of other background processes	
Cross section	15
Shower tune	1
Total (Signal and background modelling)	34
Total (Overall)	112

observed between data and simulation, and corrections for the lepton energy scale and resolution are applied accordingly. In the case of muons, momentum scale and resolution corrections are only applied to the simulation. Uncertainties on both the momentum scale and resolutions in the muon spectrometer and the tracking systems are considered, and varied independently. To account for different trigger efficiencies in data and simulation, each lepton in the final state is weighted by a scale factor, dependent on the lepton p_{T} and η; related systematic uncertainties are also considered.

The jet energy resolution (JER) and the jet energy scale (JES) uncertainties are estimated combining the information from test-beam data and simulation [181]. The JER has been measured separately for data and simulation using two in-situ techniques [182]. The expected fractional p_{T} resolution for a given jet is estimated as a function of its p_{T} and η. A systematic uncertainty is defined as the quadratic difference between the jet energy resolutions for data and simulation. To estimate the effect of the systematic uncertainty in the analysis, the energy of jets in the simulation is smeared by this residual difference, and the changes in the normalisation and shape of the final discriminant are compared to the default prediction. Since jets in the simulation cannot be un-smeared, by definition the resulting uncertainty on the normalisation and shape of the final discriminant is one-sided. This uncertainty is then symmetrised in the profile likelihood fit. The sources that contribute to the uncertainty in the JES [181, 183] are decomposed into 31 uncorrelated components and treated as independent sources in the analysis. The largest among these uncertainties is due to the flavour composition of jets in the event. Scale factors associated to the JVT association correct for different efficiencies, in data and simulation, in selecting jets from the primary vertex and rejecting pile-up jets; JVT systematic uncertainties range between 1% and 2% per single jet with a p_{T} below 60 GeV. The total per-jet uncertainty varies from 1.0% to 5.5% depending on the p_{T} of the considered jet.

The uncertainties in the b-tagging efficiencies are measured in dedicated calibration analyses [139] and also decomposed into uncorrelated components. These components are uncorrelated since they are obtained by diagonalising the matrix which parametrises the b-tagging efficiencies as a function of the jet p_{T}. The relative size of the b-tagging efficiency uncertainty is approximately 2%, 10% and 30% for b-jets, c-jets and light jets, respectively.

Data and simulation are compared to study the $E_{\mathrm{T}}^{\mathrm{miss}}$ properties [184] in regions enriched with $Z \to \mu\mu, ee$, produced in association with additional jets (Z + jets). Since no neutrinos are expected in these events, this region is used to check the momentum imbalance between all reconstructed objects and additional soft contributions to determine the $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution, as well as their corresponding uncertainties.

5.11.2 Systematic uncertainties on non-prompt/fake backgrounds

The fake scale factors, whose derivation is discussed in Section 5.7.1, are applied to simulated events entering the 4ℓ signal region and containing at least one non-prompt or fake lepton in the final state (non-prompt and fake events). The total variation of single FSFs is given by the sum in quadrature of the statistical uncertainty of the fit, arising from the limited amount of data and simulation in the exploited control regions, and a systematic uncertainty, covering their dependence on the non-prompt/fake lepton p_{T}, as discussed in Section 5.7.2.

Non-prompt/fake events are rescaled according to how many non-prompt/fake leptons they have in the final state, their flavour (electrons or muons) and, in the case of electrons, their source in the simulation truth-record origin (light or heavy). A single nuisance parameter (*overall FSF*) is introduced, incorporating the overall event-by-event variation, arising from the multiplication of lepton-by-lepton corrections to the normalisation of non-prompt/fake events; the up/down systematic variations on single FSFs correspond to the total uncertainties quoted in Section 5.7.2. It has to be remarked that the electron

FSFs (namely λ_e^ℓ and λ_e^h) are anti-correlated:[6] an upward variation of the first determines a downward fluctuation of the second, and viceversa. This is taken into account in the fit by evaluating the $+1\sigma(-1\sigma)$ variation of the overall scale factor with a $+1\sigma(-1\sigma)$ variation in λ_μ, $+1\sigma(-1\sigma)$ in λ_e^ℓ and $-1\sigma(+1\sigma)$ in λ_e^h. The variations in the overall FSF, where λ_e^ℓ and λ_e^h are swapped, is also evaluated, and the final systematic variations in the overall FSF are taken as the envelope of all up/down variations.

From dedicated studies performed on simulation in the 4ℓ signal regions, also $t\bar{t}H$, $t\bar{t}Z$ and ZZ events present a small fraction (between 3 and 7%) of events, where one of the leptons is not a prompt lepton and mostly arising from a semileptonic b-decay. For the sake of consistency, these signal and background events are treated exactly as other non-prompt/fake events, and are thus scaled by the respective overall FSF and varied according to the same related systematic uncertainties. Non-prompt and fake events in the signal simulation are, at the same time, scaled by the signal strength $\mu_{t\bar{t}H}$, as done for all of the $t\bar{t}H$ genuinely-prompt events.

5.11.3 Systematic uncertainties on the $t\bar{t}Z/\gamma^\star$ background

Systematic uncertainties on the $t\bar{t}Z/\gamma^\star$ cross section are evaluated through variations in the factorisation and renormalisation scales (see Section 3.5.2) of the hard process and in the parton shower tune. In particular, cross-section variations arising from both, QCD scale and PDF variations, and corresponding to $^{+9.6}_{-11.3}$% and ±4%, respectively, are included. In addition, the difference between the nominal aMC@NLO and the alternative Sherpa generator is included as an additional systematic uncertainty on the $t\bar{t}Z/\gamma^\star$ normalisation.

5.11.4 Systematic uncertainties on the diboson background

The ZZ normalisation is found to be well described in a phase space dominated by light-flavour jets, as shown in Section 5.6.1. However, ZZ events entering the $t\bar{t}H \to 4\ell$ signal regions are required to have at least one b-tagged jet and are, therefore, enriched with additional heavy-flavour jets. No dedicated measurement of ZZ production in association with heavy-flavour jets is currently available. An overall 50% prior variation of the ZZ normalisation is considered in the fit.

5.11.5 Systematic uncertainties on rare SM backgrounds

The rare SM background processes considered are: $t\bar{t}W^+W^-$, WtZ, VVV, $t\bar{t}t\bar{t}$, $tHjb$ and WtH. Since no measurement of the production cross section for these processes is currently available, an arbitrary and overall 50% normalisation uncertainty is introduced in the fit model for each of these processes, as an additional nuisance parameter.

5.11.6 Systematic uncertainties on $t\bar{t}H$

The systematic uncertainties on the $t\bar{t}H$ signal process come from the theoretical cross-section uncertainty and the shower and fragmentation models. Uncertainties include variations in the QCD factorisation and renormalisation scales $\left(^{+5.8}_{-9.2}\%\right)$, and in the PDFs and strong coupling α_S (±3.6%). The uncertainty from the parton shower and fragmentation models are evaluated comparing the nominal $t\bar{t}H$ (Powheg-BOX+PYTHIA 8) to $t\bar{t}H$ simulated with Powheg-BOX+HERWIG++.

[6] The post-fit correlation yields −92%.

5.12 Optimisation of the fit strategy

For the analysis carried out with 13.2 fb^{-1} [185] an inclusive phase space, corresponding to the 4ℓ pre-MVA region defined above, has been used to extract the $t\bar{t}H$ signal strength. In order to enhance the 4ℓ analysis sensitivity with respect to the inclusive phase space, the 4ℓ pre-MVA region is split into two categories: Z-depleted (no same-flavour $\ell^+\ell^-$ pair) and Z-enriched (at least one same-flavour $\ell^+\ell^-$ pair). The former is very pure in signal by itself. The shape fit to the BDTG, trained in the Z-enriched signal region and described in Section 5.9, has been explored for a four-bin fit including the systematic uncertainties described in Section 5.11. The shape fit does not bring a significant enhancement in expected sensitivity, as compared to the fit to a single bin. Therefore a requirement on the BDTG is applied to increase the signal sensitivity of the combined (i.e. Z-depleted and Z-enriched) $t\bar{t}H \rightarrow 4\ell$ fit. The yields in each of the two categories are fitted as single bins. The split into two regions allows to enhance the 4ℓ channel sensitivity by 15%, compared to the whole channel as a single region.

The requirement on the BDTG in the Z-enriched region is optimised on the overall 4ℓ expected sensitivity, i.e. the expected sensitivity of Z-enriched (including BDTG requirement) and Z-depleted combination. All of the systematic uncertainties described in Section 5.11 are included for this optimisation. Figure 5.21 shows a scan of the 4ℓ expected median significance for the rejection of the background-only hypothesis as a function of a requirement on the MVA score. A fine scan is performed, corresponding to 2400 fits, including systematic uncertainties. The $\sim 1.3\sigma$ plateau starting at -0.6 corresponds to the combined sensitivity of the Z-enriched and Z-depleted signal regions before any additional requirement; the sensitivity increases before a steep fall up to the Z-depleted expected sensitivity (around 0.9σ, all events in the Z-enriched signal region are removed).

The BDTG requirement maximising the expected median significance is at -0.03, corresponding to an expected median significance of 1.39σ. Further details on the resulting Asimov fit can be found in Section 5.13.1.

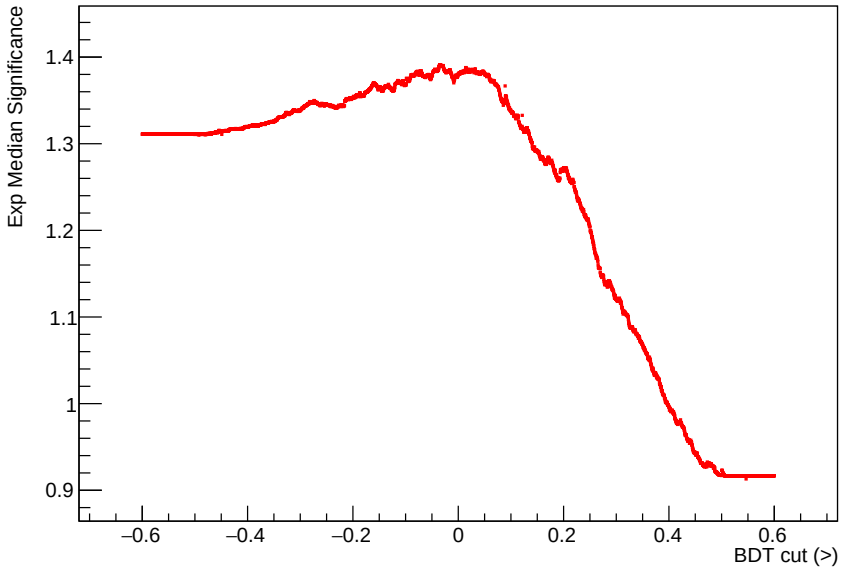


Figure 5.21: Expected median significance for the rejection of the background-only hypothesis as a function of a requirement on the BDTG score in the 4ℓ Z-enriched signal region. The significance is compared for 2400 different profile likelihood fits combining both 4ℓ signal regions.

5.13 Results

Two bins, corresponding to the Z-enriched, including the optimal BDTG requirement discussed in Section 5.12, and the Z-depleted regions, are combined as inputs to a test statistics in order to look for the presence of a $t\bar{t}H \rightarrow 4\ell$ signal. For this purpose, the profile-likelihood-based method, described in Section 5.1, is used.

The profile likelihood fit is performed by assuming the signal-plus-background hypothesis to estimate the signal strength μ, treated as the point of interest and a free parameter of the fit. The production of $t\bar{t}H$ is assumed to be described by the SM. First, a study of the expected performance is described. The unblinded fit results, both for the 4ℓ channel standalone and the latest multilepton combination, follow in Sections 5.13.3 and 5.14.

5.13.1 Signal region before the fit and expected fit results

The final selection, as discussed in Section 5.12, involves a requirement on the BDTG at -0.03 in the Z-enriched signal region. Expected yields in this and in the Z-depleted region are shown in Table 5.7 for both, background and signal processes.

The Asimov fit, described in Section 5.1.7, is exploited in order to study the expected performance of the fit in the 4ℓ channel. As expected, the $t\bar{t}H$ signal strength is centred around one and all nuisance parameters, corresponding to systematic uncertainties, are centred at zero. None of the nuisance parameters is constrained by the fit, as shown in Figure 5.22; this means that the fit model configuration does not have the power to shrink the confidence interval of nuisance parameters with respect to their priors. This is expected, given the small amount of data events in the $t\bar{t}H \rightarrow 4\ell$ channel.

The Asimov fit in the 4ℓ channel, accounting for all of the systematic uncertainties discussed in Section 5.11, yields a best-fit signal strength value of $\mu^{\text{Asimov}}_{t\bar{t}H\rightarrow 4\ell} = 1^{+1.03}_{-0.79} = 1^{+0.96}_{-0.76}$ (stat.) $^{+0.37}_{-0.22}$ (syst.), corresponding to an expected median significance of 1.4σ (1.5σ without systematic uncertainties). Figure 5.23 shows the expected pre- and post-fit impact of the different nuisance parameters on the μ central value. The blue and cyan bars show the $\pm 1\sigma$ impact of the nuisance parameter on the signal strength (shown on the top axis), while the points and associated error bars show the best-fit values of the nuisance parameters and their post-fit uncertainties (shown on the bottom axis). Among the different nuisance parameters, electron identification uncertainties, $t\bar{t}Z$ and $t\bar{t}H$ cross-section variations (both driven by QCD scale uncertainties) are the systematic uncertainties showing the largest expected impact on the $t\bar{t}H$ signal strength. This impact, nevertheless, is small if compared to statistical uncertainties: the expected fit result is dominated by statistical uncertainties.

5.13.2 Asymptotic limit and Monte Carlo toys

The conditions of asymptotic limit discussed in Section 5.1.6 may not apply to the $t\bar{t}H \rightarrow 4\ell$ channel; in fact, the available statistics is relatively small as compared to other analyses. Therefore, the robustness of such hypothesis has been tested by generating 10^4 pseudo-experiments ("MC toys"). As shown in Figure 5.24, the dispersion of the expected $t\bar{t}H$ signal strength, estimated through these Monte Carlo toys, is about 0.87; this value is compatible with the Asimov result discussed in Section 5.13.1, $\mu^{\text{Asimov}}_{t\bar{t}H\rightarrow 4\ell} = 1^{+1.03}_{-0.79} = 1^{+0.96}_{-0.76}$ (stat.) $^{+0.37}_{-0.22}$ (syst.) that has a symmetrised uncertainty of ~ 0.9.

5.13.3 Unblinding and fit to data

The expected numbers of events before the fit and the observed events in the 4ℓ signal regions are shown in Table 5.8. The best-fit signal strength from the fit to data yields $\mu_{t\bar{t}H\rightarrow 4\ell} = 1.02^{+0.99}_{-0.73} = 1.02^{+0.92}_{-0.71}$ (stat.)

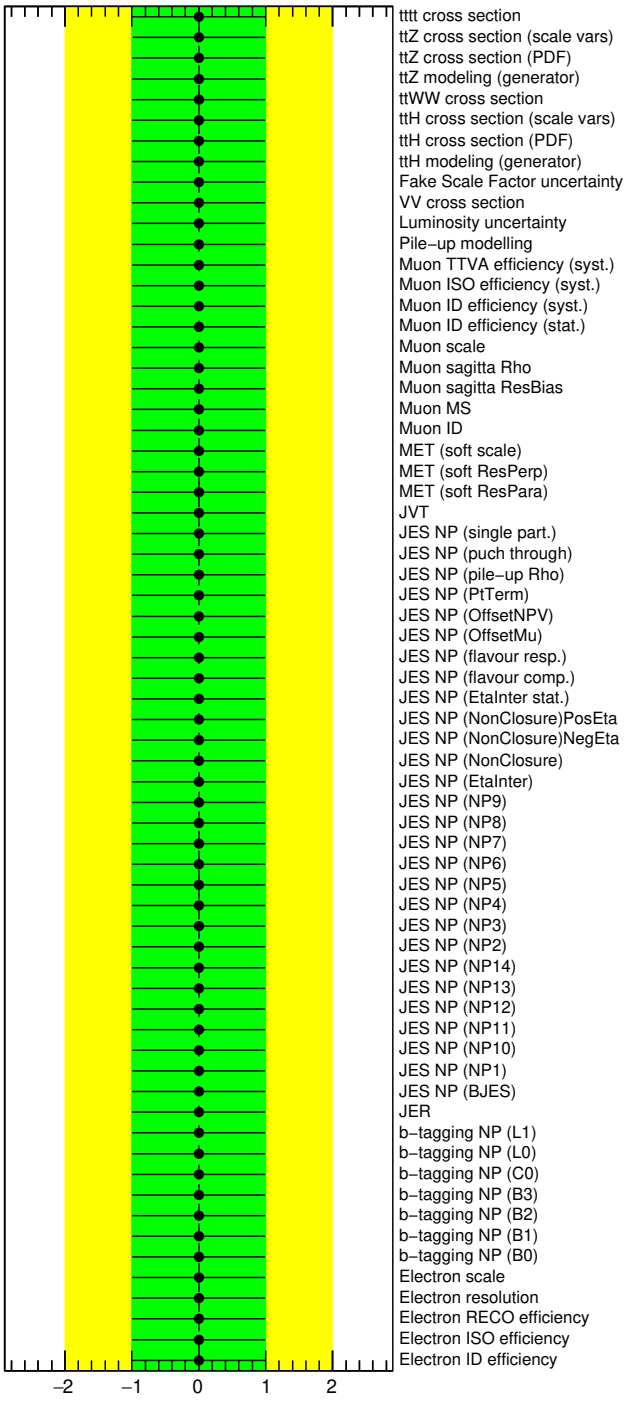


Figure 5.22: Constraints on the nuisance parameters for the 4ℓ fit to the Asimov dataset.

Table 5.7: Expected pre-fit yields in Z-enriched (after the BDTG requirement) and Z-depleted signal regions for $\mathcal{L} = 80\,\text{fb}^{-1}$. Contributions with at least one non-prompt or fake lepton are grouped as "Non-prompt". Expected prompt contributions from rare SM backgrounds are grouped under the label "others". Systematic uncertainties are included.

Process	4ℓ Z-enriched	4ℓ Z-depleted
$t\bar{t}H$	3.06 ± 0.28	0.680 ± 0.061
$t\bar{t}Z/\gamma^*$	4.96 ± 0.78	0.125 ± 0.058
Non-prompt	1.19 ± 0.62	0.100 ± 0.041
ZZ	0.4 ± 0.4	< 0.05
others	0.30 ± 0.15	0.14 ± 0.08
Total expected	9.9 ± 1.2	1.05 ± 0.11

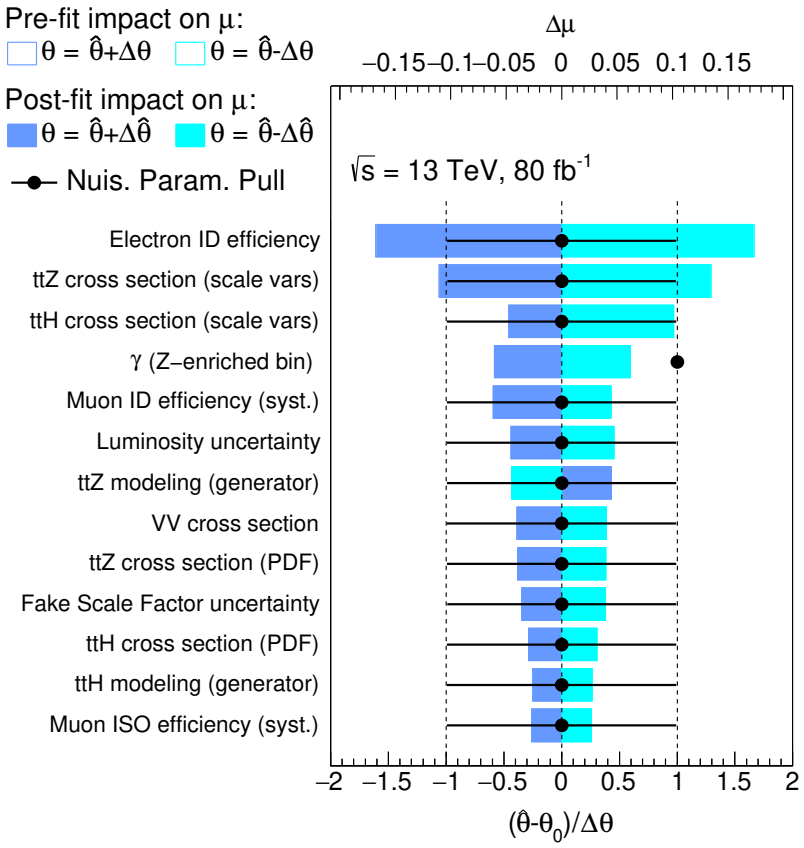


Figure 5.23: Ranking of the impact of nuisance parameters on the μ central value in the 4ℓ Asimov fit; the impact of the thirteen most important systematic uncertainties is shown. In each bin a γ parameter (see Section 5.1.7) is introduced, in order to account for the limited amount of simulated events in some of the samples.

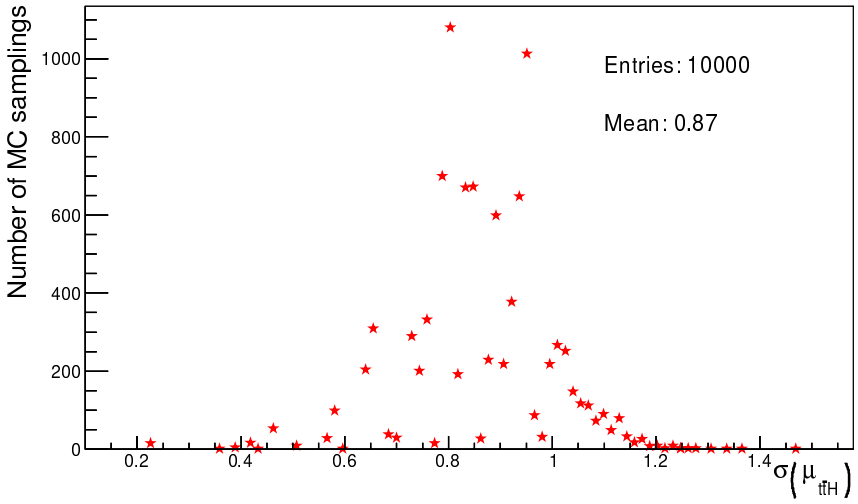


Figure 5.24: Distribution of the signal-strength symmetrised uncertainty, resulting from the fit of 10^4 pseudo-experiments. The compatibility between the mean value of this distribution (0.87) and the expected result from the Asimov fit (0.9) proves the validity of the asymptotic limit hypothesis.

Table 5.8: Number of expected (pre-fit) and observed events in the 4ℓ Z-enriched, including the optimal BDTG requirement, and the Z-depleted signal regions. Expected prompt contributions from rare SM backgrounds are grouped under the label "others". Systematic uncertainties are included.

Region	$t\bar{t}Z/\gamma^*$	Non-prompt	ZZ	others	tot bkg	$t\bar{t}H$	observed
4ℓ Z-enr.	5.0 ± 0.8	1.2 ± 0.6	0.4 ± 0.4	0.30 ± 0.15	7 ± 1	3.1 ± 0.3	8
4ℓ Z-dep.	0.12 ± 0.06	0.10 ± 0.04	< 0.05	0.14 ± 0.08	0.37 ± 0.09	0.68 ± 0.06	2

$^{+0.37}_{-0.17}$ (syst.). The background-only hypothesis is rejected at the 1.5σ level. This result reflects a balance between the observed data in the two 4ℓ regions, with a downward fluctuation in the Z-enriched region and an upward fluctuation in the Z-depleted one. Figures 5.25 and 5.26 show the fitted 4ℓ bins before and after the fit, respectively, comparing observed data events to pre- and post-fit yields from simulation.

The nuisance parameters, as can be seen in Figure 5.27, do not show significant constraints (as expected from the Asimov fit) or shifts from their central value ("pulls"). Figure 5.28 shows the pre- and post-fit impact from the different nuisance parameters on the μ central value for the fit to data; among the different nuisance parameters, electron identification uncertainties, $t\bar{t}Z$ and $t\bar{t}H$ cross-section variations (both driven by QCD scale uncertainties) are the systematic uncertainties showing the largest expected impact on the $t\bar{t}H$ signal strength, as for the fit to the Asimov dataset. Also here the impact is small if compared to statistical uncertainties and the observed result is statistically-dominated.

The observed and expected exclusion limits are obtained on the signal-plus-background hypothesis at 95% CL (see Section 5.1.5). The observed (expected) exclusion limit is $\mu_{\text{obs.}} < 2.89$ ($\mu_{\text{exp.}} < 1.83$).

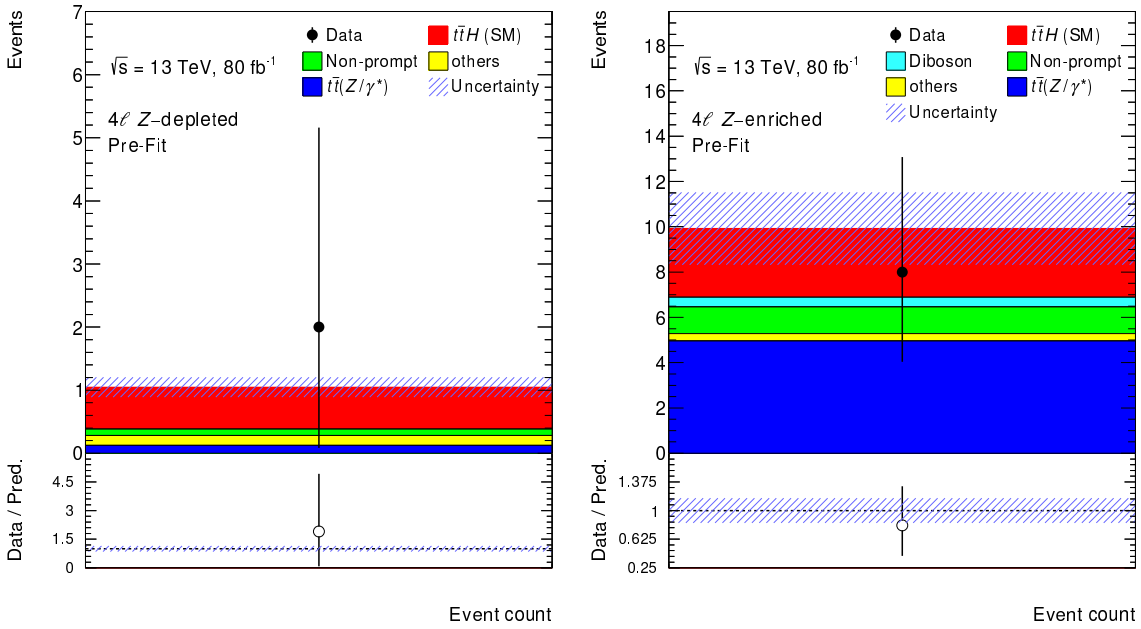


Figure 5.25: Expected yields and observed data events in the 4ℓ Z-depleted (left) and Z-enriched (right) regions. The background pre-fit contributions are shown as filled histograms. The Higgs-boson signal is shown as a filled red histogram. The size of the combined statistical and systematic uncertainty on the sum of the signal and background is indicated by the blue hatched band. The ratio of the data to the sum of signal and background expected yields is shown in the lower panel. Prompt contributions from rare SM backgrounds are grouped under "others".

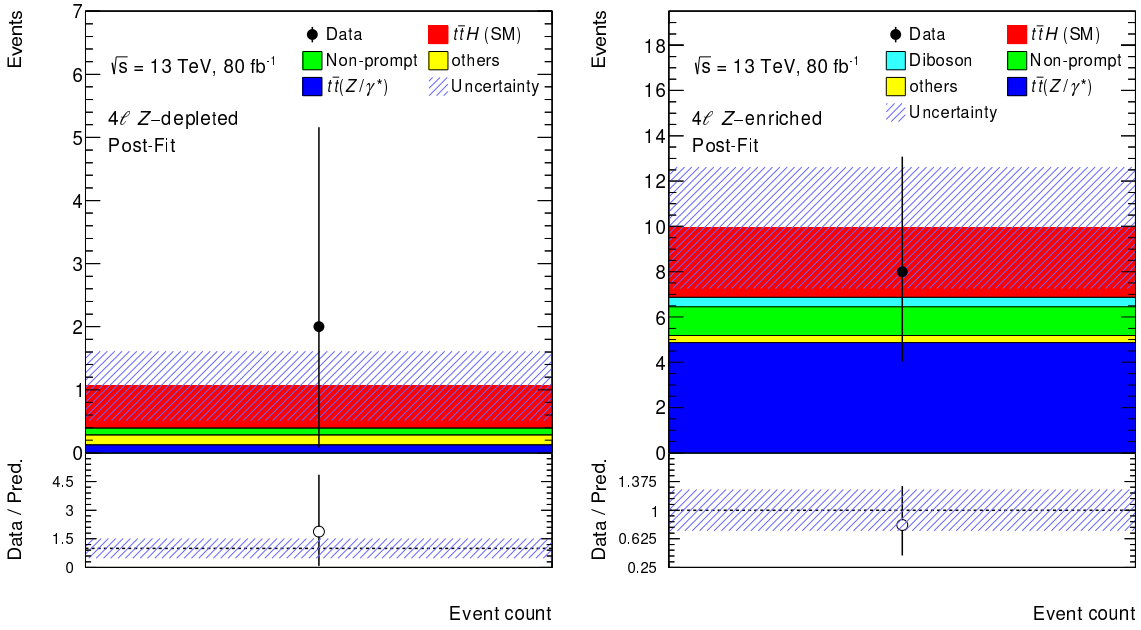


Figure 5.26: Post-fit yields and observed data events in the 4ℓ Z-depleted (left) and Z-enriched (right) regions. The background post-fit contributions are shown as filled histograms. The Higgs-boson signal is shown as a filled red histogram. The size of the combined statistical and systematic uncertainty on the sum of the signal and background is indicated by the blue hatched band. The ratio of the data to the sum of signal and background expected yields is shown in the lower panel. Prompt contributions from rare SM backgrounds are grouped under "others".

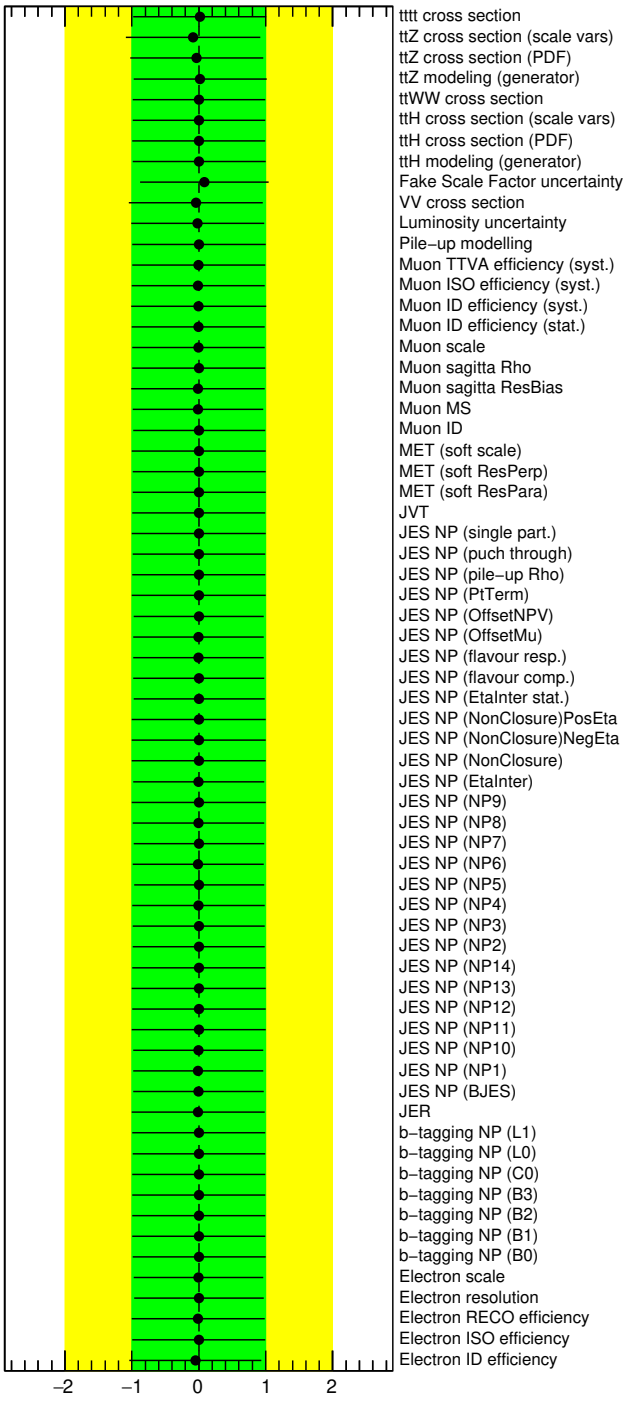


Figure 5.27: Pulls and constraints on the nuisance parameters for the 4ℓ fit to the observed dataset.

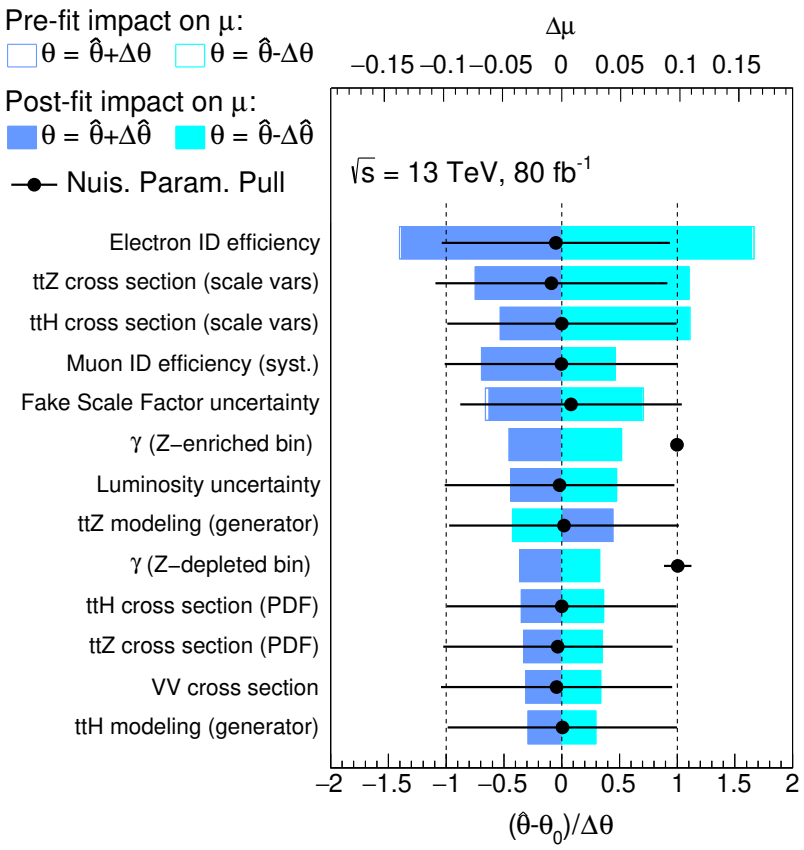


Figure 5.28: Ranking of the impact of nuisance parameters on the μ central value in the fit to observed data events in 4ℓ regions; the impact of the thirteen most important systematic uncertainties is shown. In each bin a γ parameter (see Section 5.1.7) is introduced, in order to account for the limited amount of simulated events in some of the samples.

5.14 ATLAS combination: observation of $t\bar{t}H$ production

The analysis discussed in the previous sections of this chapter is very similar to the $t\bar{t}H \to 4\ell$ analysis I developed and carried out with the 2015 and 2016 ATLAS dataset, corresponding to an integrated luminosity of 36.1 fb^{-1}. That analysis [34] is integral part of the ATLAS combination which provided observation of $t\bar{t}H$ production [36].

The ATLAS combination of the seven multilepton channels introduced in Section 5.2, corresponding to eight signal regions and four control regions, lead to an observed (expected) significance of 4.1 (2.8) standard deviations using 2015 and 2016 data events [34]. The observed (expected) best-fit value of the signal strength is $\mu = 1.6^{+0.3}_{-0.3}(\text{stat.})\,^{+0.4}_{-0.3}(\text{syst.}) = 1.6^{+0.5}_{-0.4}$ $\left(\mu = 1^{+0.3}_{-0.3}(\text{stat.})\,^{+0.3}_{-0.3}(\text{syst.}) = 1^{+0.4}_{-0.4}\right)$. The extrapolation from the fiducial to the inclusive phase space gives a measured cross section of $\sigma_{t\bar{t}H} = 790^{+150}_{-150}(\text{stat.})\,^{+170}_{-150}(\text{syst.}) = 790^{+230}_{-210}$ fb, to be compared to the SM prediction of $\sigma^{\text{SM}}_{t\bar{t}H} = 507^{+35}_{-50}$ fb. Table 5.9 shows expected and observed best-fit values of the signal strength and the associated significance under the SM background-only hypothesis; in Figure 5.29 data, background and signal yields are compared, where the final-discriminant bins in all of the multilepton signal regions are combined into bins of log(S/B), S being the expected signal events and B the fitted background yield.

Table 5.9: Observed and expected best-fit values of the $t\bar{t}H$ signal strength and associated significance with respect to the SM background-only hypothesis. The expected values are shown for the pre-fit background estimates. The observed significance is omitted for the channels where μ is negative [34].

Channel	Best fit $\mu_{t\bar{t}H}$		Significance	
	Observed	Expected	Observed	Expected
2ℓ OS+$1\tau_{\text{had}}$	$1.7\,^{+1.6}_{-1.5}$ (stat.) $^{+1.4}_{-1.1}$ (syst.)	$1.0\,^{+1.5}_{-1.4}$ (stat.) $^{+1.2}_{-1.1}$ (syst.)	0.9σ	0.5σ
1ℓ+$2\tau_{\text{had}}$	$-0.6\,^{+1.1}_{-0.8}$ (stat.) $^{+1.1}_{-1.3}$ (syst.)	$1.0\,^{+1.1}_{-0.9}$ (stat.) $^{+1.2}_{-1.1}$ (syst.)	-	0.6σ
4ℓ	$-0.5\,^{+1.3}_{-0.8}$ (stat.) $^{+0.2}_{-0.3}$ (syst.)	$1.0\,^{+1.7}_{-1.2}$ (stat.) $^{+0.4}_{-0.2}$ (syst.)	-	0.8σ
3ℓ+$1\tau_{\text{had}}$	$1.6\,^{+1.7}_{-1.3}$ (stat.) $^{+0.6}_{-0.2}$ (syst.)	$1.0\,^{+1.5}_{-1.1}$ (stat.) $^{+0.4}_{-0.2}$ (syst.)	1.3σ	0.9σ
2ℓ SS+$1\tau_{\text{had}}$	$3.5\,^{+1.5}_{-1.2}$ (stat.) $^{+0.9}_{-0.5}$ (syst.)	$1.0\,^{+1.1}_{-0.8}$ (stat.) $^{+0.5}_{-0.3}$ (syst.)	3.4σ	1.1σ
3ℓ	$1.8\,^{+0.6}_{-0.6}$ (stat.) $^{+0.6}_{-0.5}$ (syst.)	$1.0\,^{+0.6}_{-0.5}$ (stat.) $^{+0.5}_{-0.4}$ (syst.)	2.4σ	1.5σ
2ℓ SS	$1.5\,^{+0.4}_{-0.4}$ (stat.) $^{+0.5}_{-0.4}$ (syst.)	$1.0\,^{+0.4}_{-0.4}$ (stat.) $^{+0.4}_{-0.4}$ (syst.)	2.7σ	1.9σ
Combined	$1.6\,^{+0.3}_{-0.3}$ (stat.) $^{+0.4}_{-0.3}$ (syst.)	$1.0\,^{+0.3}_{-0.3}$ (stat.) $^{+0.3}_{-0.3}$ (syst.)	4.1σ	2.8σ

In addition to the multilepton channels, within the ATLAS Collaboration, the search for $t\bar{t}H$ production has been carried out in other Higgs-boson decay modes, namely:

- $H \to b\bar{b}$, in the lepton+jets and dileptonic $t\bar{t}$ decay channels [186];
- $H \to \gamma\gamma$, in leptonic and all-hadronic $t\bar{t}$ decay channels [36]. In addition, dedicated categories sensitive to $tHqb/WtH$ production also have significant $t\bar{t}H$ acceptance and are included;
- $H \to ZZ^\star \to 4\ell$, including all $t\bar{t}$ decay channels [36].

The $H \to b\bar{b}$ and multilepton analyses exploit the same dataset, namely an integrated luminosity of 36.1 fb^{-1}. On the other hand, the $H \to \gamma\gamma$ and $H \to ZZ^\star \to 4\ell$ decay channels are based on the 80 fb^{-1} dataset. An overall combination was also performed with the $t\bar{t}H$ searches based on datasets

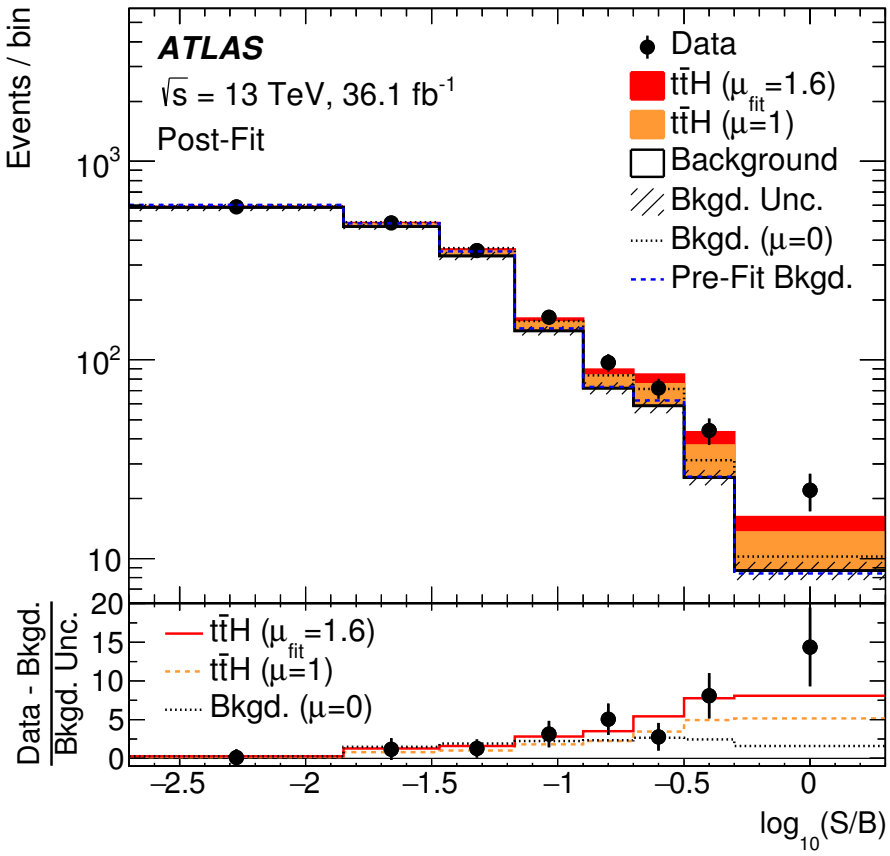


Figure 5.29: Event yields as a function of $\log_{10}$(S/B) for data, background and a Higgs-boson signal for the latest $t\bar{t}H$ multilepton combination. The discriminant bins in all signal regions are combined into bins of $\log_{10}$(S/B), where S is the expected signal yield and B the fitted background yield. The background yields are shown as the fitted values, while the signal yields are shown as both fitted values ($\mu_{t\bar{t}H}$=1.6) and the expectation from the SM ($\mu_{t\bar{t}H}$=1). The total background before the fit is shown as a dashed blue histogram. The pull (residual divided by its uncertainty) of the data with respect to the background-only prediction is shown in the lower panel, where the full red line (dashed yellow line) indicates the pull of the prediction for signal with $\mu_{t\bar{t}H}$=1.6 ($\mu_{t\bar{t}H}$=1) and background with respect to the background-only prediction [34].

corresponding to integrated luminosities of 4.5 fb^{-1} at $\sqrt{s} = 7$ TeV and 20.3 fb^{-1} at $\sqrt{s} = 8$ TeV [187]. The overall combination using 7, 8 and 13 TeV analyses leads to an observed (expected) significance for the exclusion of the background-only hypothesis of 6.3 (5.1) standard deviations [36]. This constitutes observation of the production of $t\bar{t}H$. Based on the analyses performed at 13 TeV, the measured total cross section for $t\bar{t}H$ production is 670 ± 90 (stat.)$^{+110}_{-100}$ (syst.) fb, in agreement with the Standard Model prediction ($\sigma^{\mathrm{SM}}_{t\bar{t}H} = 507^{+35}_{-50}$ fb).

5.15 Four-lepton channel prospects with full Run 2 luminosity

In this section a projection of the $t\bar{t}H \to 4\ell$ analysis to higher integrated luminosity is shown; the benchmark luminosity, 140 fb^{-1}, corresponds to the amount of data collected by ATLAS during Run 2 and good for physics analyses. All systematic uncertainties described in Section 5.11 are assumed to be the same, as for the fit performed at 80 fb^{-1}. Also in this case, with a larger dataset, no constraint is expected yet on any of the considered nuisance parameters.

Table 5.10 shows the expected sensitivity of the $t\bar{t}H \to 4\ell$ analysis, comparing the expected fit results with 80 fb^{-1} to the same performed with 140 fb^{-1} of data. With 1.75 times more data, the uncertainties on the $t\bar{t}H$ signal strength are expected to reduce by about 21%. This is compatible with the observation that the 4ℓ channel is dominated by statistical uncertainties and, therefore, uncertainties on the signal strength approximately scale with the square root of the integrated luminosity. The fit to 140 fb^{-1} of data is expected to allow a rejection of the background-only hypothesis at the 1.7σ level.

Table 5.10: Expected $t\bar{t}H$ and $t\bar{t}Z$ yields in the 4ℓ Z-enriched, including the BDTG requirement, signal region and 4ℓ combined sensitivity at 80 fb^{-1} (first row) and 140 fb^{-1} (second row). Systematic uncertainties are taken into account both for the expected yields and the Asimov fit results.

$\mathcal{L}$ (fb^{-1})	$t\bar{t}H$	$t\bar{t}Z/\gamma^{\star}$	$\mu^{\mathrm{Asimov}}_{t\bar{t}H\to 4\ell}$	Significance
80	3.1 ± 0.3	5.0 ± 0.8	$1^{+1.03}_{-0.79} = 1^{+0.96}_{-0.76}$ (stat.) $^{+0.37}_{-0.22}$ (syst.)	1.4σ (1.5σ stat.)
140	5.4 ± 0.5	9 ± 1	$1^{+0.79}_{-0.64} = 1^{+0.71}_{-0.59}$ (stat.) $^{+0.35}_{-0.25}$ (syst.)	1.7σ (1.8σ stat.)

Statistical uncertainties are expected to decrease as the amount of data collected at the LHC grows; improvements on the analysis and fit strategy, like for instance the introduction of dedicated control regions, will be needed to constrain systematic uncertainties and limit their impact on the signal sensitivity.

Projections at much higher luminosities are not reliable, as a better knowledge of the background and signal modelling may have a primary impact on the overall sensitivity to non-resonant $t\bar{t}H \to 4\ell$ processes.

CHAPTER 6

First evidence for the production of three massive vector bosons

This chapter is devoted to the first search for the production of triboson states with at least one charged and one neutral massive vector boson, i.e. $W^{\pm}W^{\mp}Z$ and $W^{\pm}ZZ$, in pp collisions. Example Feynman diagrams for the processes of interest are shown in Figures 6.1 and 6.2. These processes are studied in both, three- and four-lepton final states. The production of three massive vector bosons (VVV, where $V = W, Z$) in pp collisions at the LHC, as discussed in Section 2.7, is sensitive to SM triple and quartic gauge couplings. It also constitutes a preferential portal connecting the pure EW gauge theory to the Higgs-boson sector, as the Higgs-mediated $VH(\rightarrow VV)$ is one of the mechanisms involved in the production of VVV. Deviations from the SM predictions may point to BSM effects described by anomalous TGC and QGC (see Section 2.6). The full expression for triple and quartic gauge couplings, arising from the pure EW Lagrangian, is derived for the physical gauge bosons in Section 2.2.1.

The first section describes the set of data events and MC simulation samples, as well as the signal definition. An overview of the physics object definition and event selection follows in Sections 6.2 and 6.3, respectively. The modelling of major background processes has been thoroughly studied, as discussed in Section 6.4. The analysis sensitivity is optimised by means of several MVAs trained to discriminate the signal from the largest background sources, especially diboson (see Sections 6.6.1 and 6.6.2). The fit to the Asimov and the observed datasets, accounting for the systematic uncertainties described in Section 6.7, is outlined in Sections 6.8 and 6.9, respectively. The last section describes the statistical combination of the WVZ analysis with the $W^{\pm}W^{\pm}W^{\mp}$ search; this provides the first evidence for the production of three massive vector bosons.

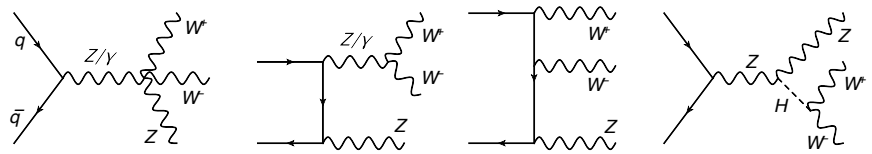


Figure 6.1: Example diagrams for massive triple vector boson production $W^{\pm}W^{\mp}Z$ [188].

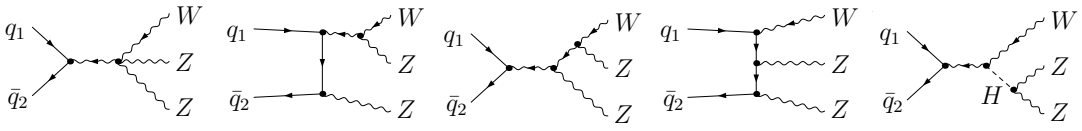


Figure 6.2: Example diagrams for massive triple vector boson production $W^{\pm}ZZ$ [189].

6.1 Data, Monte Carlo simulation and signal definition

This section summarises the set of data events and MC simulation samples used in this analysis, referring to Section 5.3.2 for what concerns MC simulation samples used in both works. The analysis described in this chapter is carried out using the same 80 fb^{-1} dataset exploited for the $t\bar{t}H \to 4\ell$ search, see Section 5.3.1.

In addition to the full leptonically decaying VVV samples described in Section 5.3.2, three dedicated MC simulation samples, i.e. $W^{\pm}W^{\mp}Z \to 3Ljj$, $W^{\pm}ZZ \to 4Ljj$ and $W^{\pm}ZZ \to 3Ljj$, with $L = e, \mu, \tau$, have been generated to access the semileptonic phase space of $W^{\pm}W^{\mp}Z$ and $W^{\pm}ZZ$ processes, where all three W or Z bosons are on-shell[1] and decay into final states with three or four charged leptons. All on-shell VVV samples are generated with consistent parameters in Sherpa [107], with 0 additional partons at NLO in QCD and 1 or 2 additional partons at LO in QCD. These matrix elements are merged with the Sherpa parton shower using the MEPS@NLO prescription [164]. The associated production of $WH \to WZZ^*$ and $ZH \to ZWW^*$ events is inclusively generated with Pythia 8 [140] + EvtGen [168]. Both processes, on-shell WVZ and $WH(ZH) \to WZZ^*(ZWW^*)$, are generated at NLO in QCD [161, 190, 191] and are included in the definition of signal events, see Section 6.8 for more details. Signal events are labelled as WVZ in the following. EW NLO corrections are available for the $W^{\pm}W^{\mp}Z$ and $W^{\pm}ZZ$ processes [188, 189]. However, they are not considered in this work, as they are very small if compared to the sensitivity of this analysis, inducing order 1% corrections.

The same Sherpa MC simulation described in Section 5.3.2 is exploited to model the production of diboson processes, including the loop-induced gg-initiated and the VBS production modes. Alternative diboson samples are generated with Powheg v2.0 [110] and interfaced with Pythia 8 for the parton showering and fragmentation. A generator-level cut of at least 4 GeV is applied to the invariant mass of any same-flavour $\ell^+\ell^-$ pair.

The production of Z + jets is modelled by the same simulation introduced in Section 5.3.2. $Z + \gamma$ production is also modelled using Sherpa, in particular at NLO accuracy for 0 and 1 additional partons, and at LO for 2 and 3 additional partons. The p_{T} of the photon is required to be larger than 7 GeV and the distance between the γ and the Z-boson in the $\eta - \phi$ plane is required to be $\Delta R > 0.1$. In addition, leptons from the Z-boson decay are required to fulfil $m_{\ell^+\ell^-} > 2$ GeV. A removal of events characterised by the presence of matrix-element photons is performed in Z + jets, in order not to double-count their contribution.

The modelling of the production of $t\bar{t}W$, $t\bar{t}Z$, WtZ, $t\bar{t}$ and $t\bar{t}+\gamma$ relies on the same simulation samples outlined in Section 5.3.2. On the other hand, $t\bar{t}H$ processes are simulated by MadGraph5_aMC@NLO [109], instead of Powheg-BOX [110], interfaced with Pythia 8. Rare SM processes involving the production of a Higgs boson, i.e. $tHqb$ and WtH, and those not involving it, i.e. $t\bar{t}W^+W^-$ and $t\bar{t}t\bar{t}$, are modelled with the same configurations discussed in Section 5.3.2.

Matrix-element and parton-shower generators used for simulating signal and background processes are summarised in Table 6.1. As discussed in Section 5.3.2, all Monte Carlo samples are processed and

[1] Produced at the pole mass.

generated in two statistically independent samples (mc16a and mc16d).

Table 6.1: The table shows the configurations used for event generation of signal and background processes. If only one parton distribution function (PDF) is shown, the same one is used for both the matrix-element (ME) and the parton-shower generators; if two are shown, the first is used for the matrix-element calculation and the second for the parton shower. "Tune" refers to the underlying-event tune of the parton shower generator. "MG5_aMC" refers to MadGraph5_aMC@NLO 2.2.1 [109]; "Pythia 8" refers to version 8.2 [140]; "Herwig++" refers to version 2.7 [167]. Samples using Pythia 8 have heavy-flavour hadron decays modeled by EvtGen 1.2.0 [168]. All samples include leading-logarithm photon emission, either modelled by the parton-shower generator or by PHOTOS [169].

Process	ME generator (alternative)	Parton shower (alternative)	PDF set	Tune
$W^{\pm}W^{\mp}Z$	Sherpa [107]	Sherpa	NNPDF 3.0 NNLO [101]	Sherpa default
$W^{\pm}ZZ$	Sherpa	Sherpa	NNPDF 3.0 NNLO	Sherpa default
$W^{\pm}W^{\pm}W^{\mp}$	Sherpa	Sherpa	NNPDF 3.0 NNLO	Sherpa default
ZZZ	Sherpa	Sherpa	NNPDF 3.0 NNLO	Sherpa default
VH	Pythia 8 [140]	Pythia 8	NNPDF 2.3 LO [170]	A14 [144]
VV, $qqVV$	Sherpa	Sherpa	NNPDF 3.0 NNLO	Sherpa default
	(Powheg-BOX [110])	(Pythia 8)	CT10 [141]/CTEQ6L1 [172]	A14
$Z \to \ell^+\ell^-(+\gamma)$	Sherpa	Sherpa	NNPDF 3.0 NLO	Sherpa default
$t\bar{t}Z/\gamma^*$	MG5_aMC [109]	Pythia 8	NNPDF 3.0 NLO /NNPDF 2.3 LO	A14
$tWZ/\gamma^*(\to \ell^+\ell^-)$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}$	Powheg-BOX	Pythia 8	CT10/CTEQ6L1	Perugia2012 [173]
$t\bar{t}\gamma$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}W$	MG5_aMC	Pythia 8	NNPDF 3.0 NLO	A14
$t\bar{t}H$	MG5_aMC	Pythia 8	NNPDF 3.0 NLO	A14
$tHqb$	MG5_aMC	Pythia 8	CT10	A14
tWH	MG5_aMC	Herwig++ [167]	CT10	UE-EE-5 [171]
$t\bar{t}t\bar{t}$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}W^+W^-$	MG5_aMC	Pythia 8	NNPDF 2.3 LO	A14

6.2 Object definition

This section describes the basic definition of physics objects, such as leptons and jets, exploited in order to carry out the analysis described in the following sections of this chapter. Events are required to have been selected by at least one of the ATLAS single-lepton (single-electron or single-muon) triggers. The possible gain in signal acceptance due to the inclusion of dilepton triggers, used for the work outlined in the previous chapter, has been studied. The WVZ enhancement is expected to be very small (0.5% inclusively), while the enhancement in background acceptance is significant (about 7%). Therefore dilepton triggers are not further considered. The difference in WVZ and $t\bar{t}H$ trigger acceptance derives from the fact that WVZ leptons tend to be more energetic. At least one lepton firing one of the single-lepton triggers is required to have at least 1 GeV more than the p_{T} threshold of the corresponding trigger. This cut, applied consistently to simulated and data events, allows to reach the trigger efficiency plateau: typically, below this cut the trigger efficiency grows steeply as a function of the lepton p_{T}.

The electron and muon loose definitions exactly match the one described in Section 5.4, apart from the p_{T} requirement, which is raised to $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ instead of $p_{\mathrm{T}} > 10\,\mathrm{GeV}$. This allows to significantly

reduce contamination from both, reducible and irreducible backgrounds, retaining a good signal efficiency. In particular, the Z + jets contribution in the inclusive three-lepton phase space is reduced by more than a factor 2. The tight lepton selection involves a MVA-based isolation definition and is discussed in Section 6.3.

Jets are reconstructed, selected and b-tagged exactly as previously discussed (see Section 5.4). Any jet, also reconstructed as hadronically-decaying τ-lepton candidate, is treated as a jet.

Ambiguities due to double counting of reconstructed objects are avoided through an overlap removal procedure, performed on the loose objects previously defined and slightly different from the one described in Section 5.4. The following logical order is used: if two electron candidates are overlapping on the 2^{nd} layer cluster of the ECal (see Section 3.2.4) or share the reconstructed track, the electron with the higher p_{T} is kept and the other electron is removed; any CT[2] muon candidate, sharing the ID track with an electron candidate, is removed; any electron candidate sharing the ID track with a combined muon is removed; any jet candidate within $\Delta R = 0.2$ of a reconstructed electron is removed; any electron candidate within $\Delta R = 0.4$ of a jet is removed; a jet candidate is removed if a muon track is within $\Delta R = 0.2$ of it; any muon candidate within $\Delta R = 0.4$ of a jet is removed. Table 6.2 gives an overview of the overlap removal procedure adopted for this analysis. It is important to remark that the differences between the overlap removal procedures employed in the WVZ and $t\bar{t}H$ multilepton analyses are due to the different sought-after signature. In fact, the $t\bar{t}H$ final state tends to be busy, characterised by the presence of multiple jets; therefore an optimisation of the $t\bar{t}H$ overlap removal between jets and leptons has been performed, e.g. the p_{T}-dependent overlap removal between jets and muons reported in Table 5.2. On the other hand, the WVZ analysis is not as sensitive to such requirements.

Table 6.2: Summary of the overlap removal procedure between electron, muon and jet candidates. The row order reflects the logical order adopted.

Keep	**Remove**	**Cone size (ΔR) or track**
electron	electron (lower p_{T})	overlapping 2^{nd} layer cluster or shared track
electron	CT muon	shared track
muon	electron	shared track
electron	jet	0.2
jet	electron	0.4
muon	jet	0.2
jet	muon	0.4

6.3 Event selection

Each event is classified as a three- (3ℓ) or four-lepton (4ℓ) event based on the multiplicity of loose leptons, whose definitions are provided in Section 6.2. Therefore no events are shared between the two channels.

A common tight lepton definition is introduced in order to suppress contributions from events with at least one fake or non-prompt lepton. The same $p_{\text{T}}^{\text{cone 30(20)}}$ isolation cut discussed in Section 5.5 is applied to tight muons (electrons). Furthermore, the tight lepton definition relies on a MVA discriminant, called "*Prompt Lepton Veto*" (PLV), combining isolation and b-tagging related variables to enhance the rejection

[2] See Section 3.4.4.

of fake and non-prompt leptons. It constitutes an improved[3] version of the MVA-based isolation exploited for the latest published $t\bar{t}H$ multilepton search [34]. The main idea is to identify non-prompt leptons using lifetime information associated with a jet matching the selected lepton. This lifetime information is computed using tracks contained within the jet. Typically, the hadron lifetime is determined using the impact parameter of the track reconstructed by the ID which is matched to the reconstructed lepton. Using additional close-by reconstructed charged particle tracks increases the precision of identifying possible displaced decay vertices of bottom or charm hadrons that produced a non-prompt lepton. The MVA also includes information related to the isolation of the lepton. Prompt leptons are defined as background, non-prompt and fake leptons as signal events for the training of a BDT; therefore prompt leptons tend to cluster at low MVA scores. The expected PLV BDT response is shown in Figure 6.3 for the lepton having the smallest p_{T} in the region with exactly three loose leptons (3$^{\mathrm{rd}}$ leading lepton): the PLV distribution is shown for an irreducible background, namely WZ, and a reducible background, i.e. Z + jets.

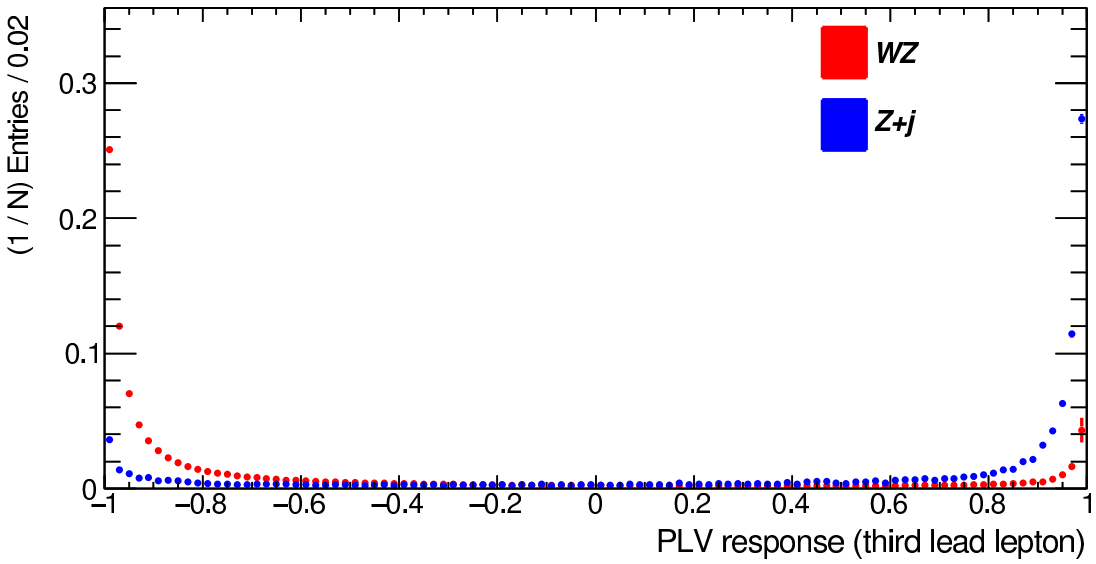


Figure 6.3: Comparison between the expected PLV response on 3$^{\mathrm{rd}}$ leading leptons produced in WZ (in red) and Z + jets (in blue). Statistical-only uncertainties are shown.

The tight lepton definition, as shown in Table 6.3, corresponds to a PLV < -0.7 (-0.5) cut for electrons (muons); on top of the loose definition, described in Section 6.2, tight electrons are also required to pass the *TightLH*[4] working point (as opposed to loose electrons required to fulfil the *LooseLH* identification working point) and the *charge misidentification* BDT cut. There is a small, but non-negligible, probability that electrons or positrons are reconstructed with an incorrect charge. This occurs when an electron (positron) emits a hard bremsstrahlung photon; if the photon subsequently converts to an asymmetric electron-positron pair, and the positron (electron) has high momentum and is reconstructed, the lepton charge can be misidentified. It can also occur when the curvature of a track is poorly estimated, which typically happens at high momentum. The probability for muons to be reconstructed with incorrect charge is small enough that the charge misassignment is negligible. To reject electrons reconstructed

[3] Improved in terms of expected performance on simulation.
[4] See Section 3.4.3

with an incorrect electric charge, a BDT discriminant is built [34], using the following electron cluster and track properties as input: the electron transverse momentum and pseudo-rapidity, the track curvature significance (defined as the ratio between the electric charge and the track momentum divided by the estimated uncertainty in the measurement) and its transverse impact parameter times the electric charge, the cluster width along the azimuthal direction, and the quality of the matching between the track and the cluster, both in terms of energy/momentum and azimuthal position. The charge misidentification allows to achieve a rejection factor of 17 for electrons produced in $Z \to e^+e^-$ events and passing the TightLH identification requirements with a wrong charge assignment, while selecting properly measured electrons with an efficiency of 95%. Correction factors to account for differences in the selection efficiency between data and MC simulation, which are within a few per cent for $|\eta| < 2.4$ but larger in the forward region, $2.4 < |\eta| < 2.47$, are applied to the selected electrons in the MC simulation.

Table 6.3 gives an overview of the lepton selection, both for the loose (L) and the tight (T) definition. The tight lepton definition, as demonstrated in the following sections, allows to significantly reduce the number of expected fake and non-prompt leptons.

Table 6.3: Loose (L), and tight (T) lepton definitions. The tight selection, involves a MVA-based isolation definition.

	e		μ	
	L	T	L	T
$p_{\mathrm{T}}^{\mathrm{cone}}$ isolation	No	Yes	No	Yes
MVA-based isolation (PLV)	No	< -0.7	No	< -0.5
Identification	LooseLH	TightLH	Loose	
Charge misassignment veto BDT	No	Yes	N/A	
Transverse impact parameter significance $\lvert d_0\rvert/\sigma_{d_0}$	< 5		< 3	
Longitudinal impact parameter $\lvert z_0 \sin\theta\rvert$	< 0.5 mm			

6.3.1 The three-lepton channel

The three-lepton (3ℓ) phase space is known to be populated by a non-negligible fake and non-prompt lepton contribution, arising from processes with high cross section such as Z + jets and $t\bar{t}$ with an additional lepton candidate originating from a jet, a hadron or a photon. Table 6.4 shows the expected yields, at 80 fb^{-1}, when selecting events with exactly three reconstructed loose leptons, whose charge sum is $|\sum q_i| = 1$, and at least one same-flavour $\ell^+\ell^-$ pair. This portion of the 3ℓ phase space is dominated by the Z + jets contribution.

The best $Z \to \ell^+\ell^-$ candidate, already introduced in Chapter 5, is defined as the same-flavour $\ell^+\ell^-$ pair whose invariant mass is the closest to the Z-boson mass. The requirement of the tight lepton quality for the lepton not belonging to the best $Z \to \ell^+\ell^-$ candidate significantly suppresses contributions from non-prompt and fake leptons, as also shown in Table 6.4. Z + jets and $t\bar{t}$ contributions are reduced by a factor 13 and 2.5, respectively.

A veto on b-tagged jets is applied in the 3ℓ regions, as it allows to significantly reduce top-quark-like contributions (such as $t\bar{t}$, $t\bar{t}Z$ and tZ), retaining 93% of the signal. Only 10% of signal events is expected

Table 6.4: Expected yields in the three-loose-lepton ("3 loose ℓ") region, the three-lepton region where the tight lepton definition is applied to the lepton not belonging to the best $Z \to \ell^+\ell^-$ candidate ("2 loose 1 tight") and the three-lepton pre-selection ("3ℓ pre-sel") region with an integrated luminosity of 80 fb^{-1}. When requiring three loose leptons the $t\bar{t}$ and Z + jets are among the largest expected contributions. Statistical-only uncertainties are shown.

Process	3 loose ℓ	2 loose 1 tight	3ℓ pre-sel
WVZ	257.1 ± 0.7	221.7 ± 0.6	158.8 ± 0.5
ZZ	7 360 ± 30	4 280 ± 20	1 154 ± 8
WZ	22 800 ± 300	18 460 ± 30	7 770 ± 20
Z + jets	52 900 ± 700	4 200 ± 200	690 ± 50
$Z + \gamma$	8 400 ± 200	1 420 ± 60	140 ± 20
$t\bar{t}$	21 320 ± 60	9 360 ± 40	540 ± 20
$t\bar{t}Z$	863 ± 2	691 ± 2	109.7 ± 0.8
tZ	317 ± 2	253 ± 2	70.8 ± 0.9
tWZ	150 ± 2	126 ± 2	36.3 ± 0.9
others	113 ± 8	71.8 ± 7	51 ± 7
Total expected	114 500 ± 800	39 100 ± 200	10 720 ± 60

not to have jets, mostly corresponding to $W^\pm ZZ \to 3\ell 3\nu$; on the other hand, almost 50% of the WZ contribution is expected not to have jets in the final state, therefore events entering the 3ℓ *pre-selection region* are required to have one or more reconstructed jets. It has to be remarked that, considering the accuracy of WZ Monte Carlo simulation and that any genuine jet in a $WZ \to 3\ell$ event must originate from additional partons, predictions of the jet multiplicity will need to be corrected (see Section 6.5).

In order to remove leptons from quarkonia decays, a lower cut on the invariant mass of all possible same-flavour $\ell^+\ell^-$ pairs is applied at 12 GeV. Furthermore, since in both, the $W^\pm W^\mp Z$ and $W^\pm ZZ$ 3ℓ final states, at least one of the Z bosons is expected to decay to two charged leptons, the best $Z \to \ell^+\ell^-$ candidate is required to be within 10 GeV of the Z-boson mass. This allows to further suppress top-quark processes, where no on-shell Z-boson is produced (e.g. $t\bar{t}$ and $t\bar{t}W$). The final 3ℓ pre-selection region reads as follows:

- exactly 3 loose leptons are selected and the lepton not belonging to the best $Z \to \ell^+\ell^-$ candidate fulfils the tight requirements;
- the invariant mass of any same-flavour $\ell^+\ell^-$ pair is above 12 GeV;
- at least one same-flavour $\ell^+\ell^-$ pair whose invariant mass is within 10 GeV of the Z-boson mass;
- one or more reconstructed jets are selected and none of them is b-tagged.

The expected yields in the 3ℓ pre-selection region are shown in Table 6.4.

6.3.2 The four-lepton channel

The four-lepton (4ℓ) channel, as discussed for the $t\bar{t}H \to 4\ell$ analysis in Chapter 5, benefits from a relatively small reducible background contribution, i.e. involving non-prompt and fake leptons, with respect to other multilepton regions (for instance, the three-lepton final state, discussed in the previous

section). Nevertheless, as shown in Table 6.5, the relative contribution from fake and non-prompt processes proves not to be negligible when requiring four loose leptons, with total electric charge zero. This contribution can be significantly reduced by requiring the two subleading leptons, more likely to be fake/non-prompt, to pass the tight lepton requirements; the achieved suppression of reducible backgrounds is also shown in Table 6.5. The efficiency of this lepton selection on Z + jets events is expected to be about 1%.

A different efficiency for such tight selection is observed in WVZ and ZZ events (68% and 59%, respectively), which are both expected to produce four prompt leptons in the final state; the difference between the two is due to the significant dependence of the PLV efficiency on the prompt lepton p_{T}, as shown in Figure 6.4, and the large difference in the lepton p_{T} spectra for the two processes (WVZ leptons being considerably harder), as shown in Figure 6.5.

Table 6.5: Expected yields in the four-loose-lepton ("4 loose ℓ"), the four-lepton region where the tight lepton definition is applied to the two subleading leptons ("2 loose 2 tight") and the four-lepton pre-selection ("4ℓ pre-sel") region with an integrated luminosity of 80 fb^{-1}. The Z + jets contribution is not negligible when requiring four loose lepton. The $t\bar{t}$ contribution is taken into account and shown under "others". The quoted uncertainties are only statistical.

Process	4 loose ℓ	2 loose 2 tight	4ℓ pre-sel
WVZ	24.3 ± 0.1	16.5 ± 0.1	14.9 ± 0.1
ZZ	$2\,120 \pm 10$	$1\,247 \pm 5$	$1\,149 \pm 5$
WZ	14.1 ± 0.7	2.5 ± 0.2	2.1 ± 0.2
Z + jets	91 ± 20	1 ± 1	0.2 ± 0.1
$Z + \gamma$	22 ± 7	< 1	< 1
$t\bar{t}Z$	91.9 ± 0.7	53.5 ± 0.6	9.8 ± 0.2
tZ	2.2 ± 0.2	0.19 ± 0.04	0.11 ± 0.03
tWZ	17.4 ± 0.6	10.7 ± 0.5	3.7 ± 0.3
others	202 ± 6	9.0 ± 0.6	3.3 ± 0.3
Total expected	$2\,580 \pm 30$	$1\,340 \pm 5$	$1\,184 \pm 5$

For the same arguments discussed in Section 6.3.1, vetoes on leptonic decays of quarkonia resonances and b-tagged jets are applied. Also in the 4ℓ channel at least one of the Z bosons from the WVZ states is expected to decay to two charged leptons, therefore at least one $Z \to \ell^+\ell^-$ candidate is required to be selected. In summary, the 4ℓ *pre-selection region* is defined as follows:

- exactly 4 loose leptons are selected;
- the 3$^{\mathrm{rd}}$ and 4$^{\mathrm{th}}$ leading leptons pass the tight lepton requirements;
- the invariant mass of any same-flavour $\ell^+\ell^-$ pair is above 12 GeV;
- at least one same-flavour $\ell^+\ell^-$ pair is present with invariant mass within 10 GeV of the Z-boson mass;
- no reconstructed and selected jet is b-tagged.

The expected yields in the 4ℓ pre-selection region are shown in Table 6.5.

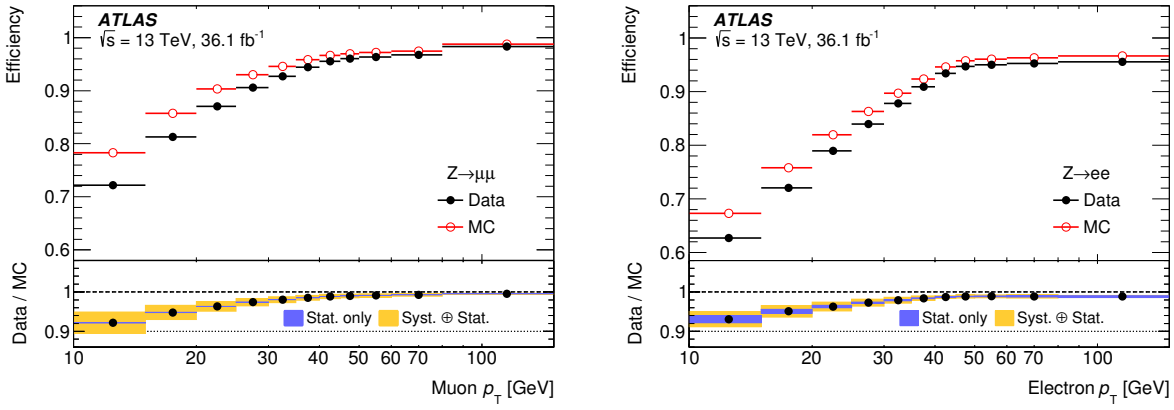
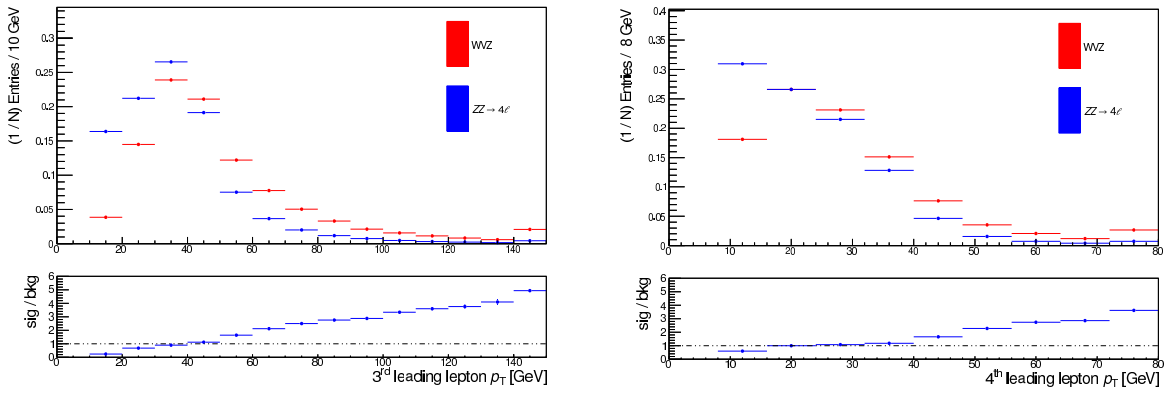


Figure 6.4: Efficiency to select prompt (left) muons and (right) electrons in $Z \to \ell^+\ell^-$ events with the previous version of the PLV, as a function of the lepton p_{T}. The measurements in data (simulation) are shown as full black (open red) circles. The bottom panel displays the data-to-simulation ratio, with the blue (yellow) band representing the statistical (total) uncertainty. The ratio corresponds to the scale factor that is applied to correct the simulation [34].



(a) (b)

Figure 6.5: Shape comparison for the (a) 3$^{\mathrm{rd}}$ and (b) 4$^{\mathrm{th}}$ leading lepton p_{T} for the WVZ (red) and ZZ (blue) processes, requiring four loose reconstructed leptons with total charge zero. The lower panels show the ratio between the WVZ and ZZ shapes. Statistical-only uncertainties are shown.

6.4 Background modelling

Background processes affecting the WVZ search can be split into reducible and irreducible backgrounds, introduced in Section 5.6. This section shows the validation of the most important background processes in this analysis, including both, irreducible backgrounds, in particular WZ in 3ℓ and ZZ in 4ℓ, and reducible backgrounds, dominated by Z + jets/γ in the 3ℓ channel. Among the irreducible backgrounds also $t\bar{t}Z$ plays an important role, therefore the validation of the $t\bar{t}Z \to 3\ell$ control region, included in the final WVZ fit,[5] is shown. As for the validation of the largest irreducible background, i.e. Z + jets/γ, a dedicated validation region is defined, see Section 6.4.2. It has to be remarked that the reweighting procedure, discussed later in Section 6.5, is not applied to the distributions shown in this section.

6.4.1 Irreducible backgrounds

The three-lepton pre-MVA region

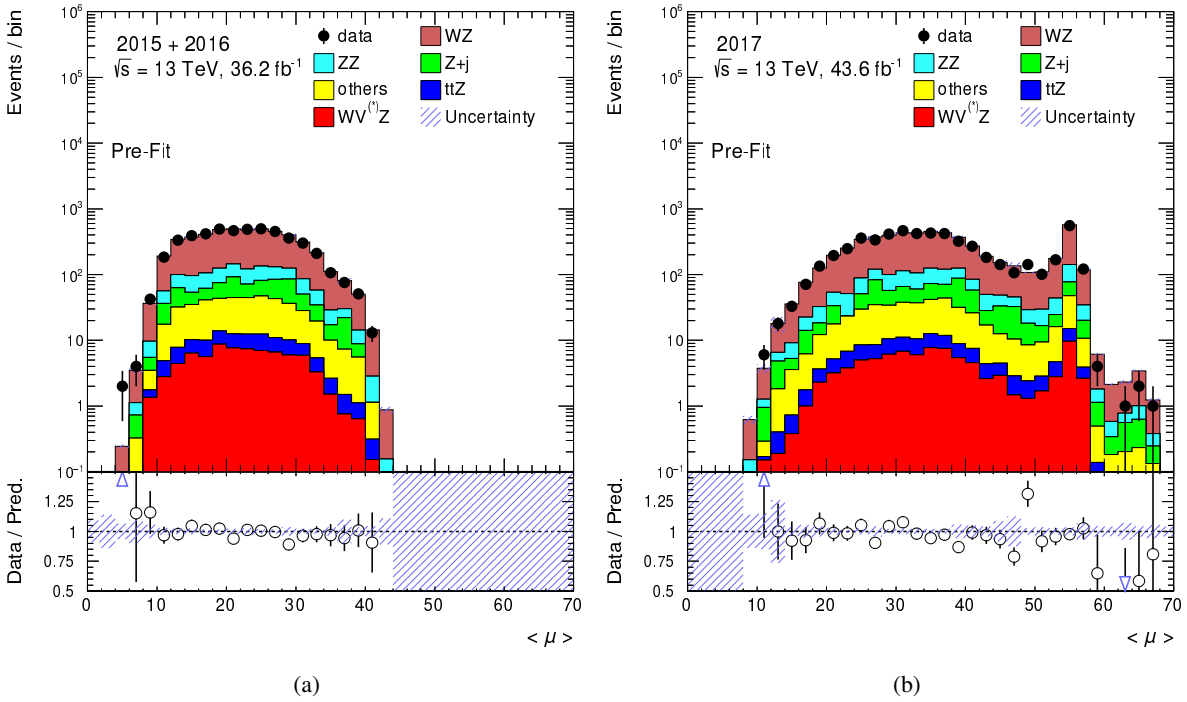
The 3ℓ pre-selection region, as shown in Table 6.4, is characterised by the non-negligible contamination from reducible backgrounds. This contribution can be reduced by requiring events to also have a scalar sum of all selected leptons' and jets' p_{T} (H_{T}) larger than 200 GeV. More details about the definition of this region, called 3ℓ *pre-MVA*, can be found in Section 6.6.1. The modelling of irreducible backgrounds has been probed in the 3ℓ pre-MVA region. Expected and observed yields in this region are reported in Table 6.6, the purity in WZ (ZZ) events is 75(9)%. ZZ processes enter this region mostly with events where four leptons are produced, and one of the leptons is not reconstructed/selected, and events where one of the Z bosons decays to two leptons, while the other one to a leptonically decaying tau lepton and a τ_{had}.

Table 6.6: Expected and observed yields in the 3ℓ pre-MVA region with an integrated luminosity of 80 fb^{-1}. Only statistical uncertainties are shown.

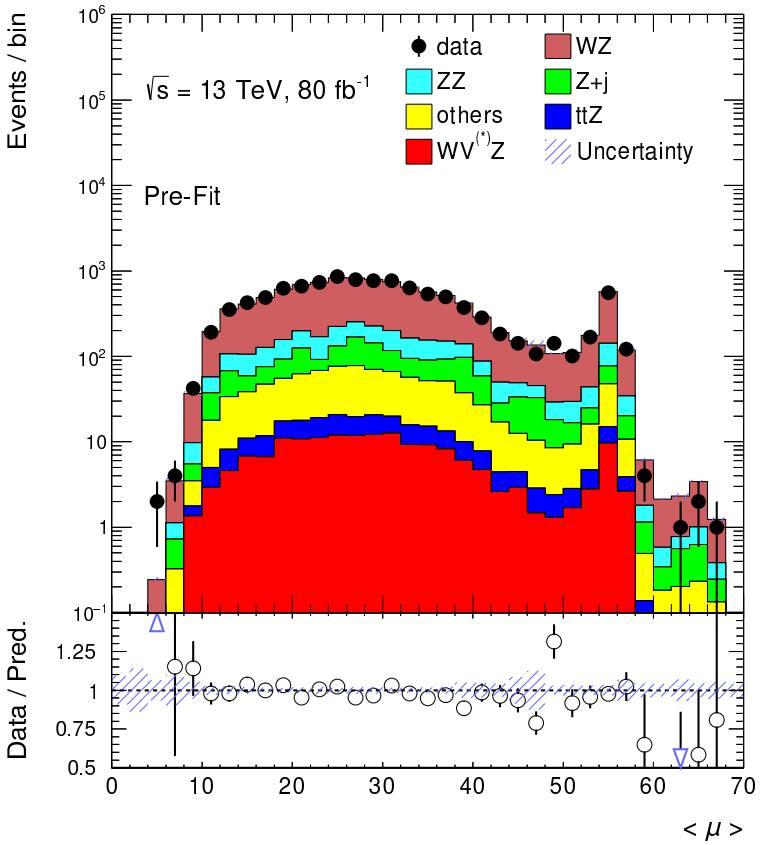
3ℓ pre-MVA	Exp./obs. yields
WVZ	137.6 ± 0.5
ZZ	706 ± 6
WZ	$5\,860 \pm 10$
Z + jets	330 ± 30
$Z + \gamma$	84 ± 10
$t\bar{t}Z$	107.5 ± 0.8
tZ	60.3 ± 0.8
tWZ	35.2 ± 0.9
others	442 ± 9
Total expected	$7\,770 \pm 40$
data	7 454

Figures 6.6 and 6.7 show some kinematic distributions involving leptons and jets, separately for the 2015+2016 and the 2017, as well as for the inclusive 80 fb^{-1} datasets. More distributions, showing data-to-simulation comparison in this region, can be found in Appendix A. An overall good agreement is observed within statistical uncertainties.

[5] See Section 6.8 for further details.



(a) (b)



(c)

Figure 6.6: Data/simulation comparison in the 3ℓ pre-MVA region for the average pile-up profile comparing (a) data collected in 2015 and 2016 ($36.2\,\text{fb}^{-1}$) to mc16a, (b) data collected in 2017 ($43.6\,\text{fb}^{-1}$) to mc16d and (c) the full $80\,\text{fb}^{-1}$ dataset to mc16a and mc16d. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

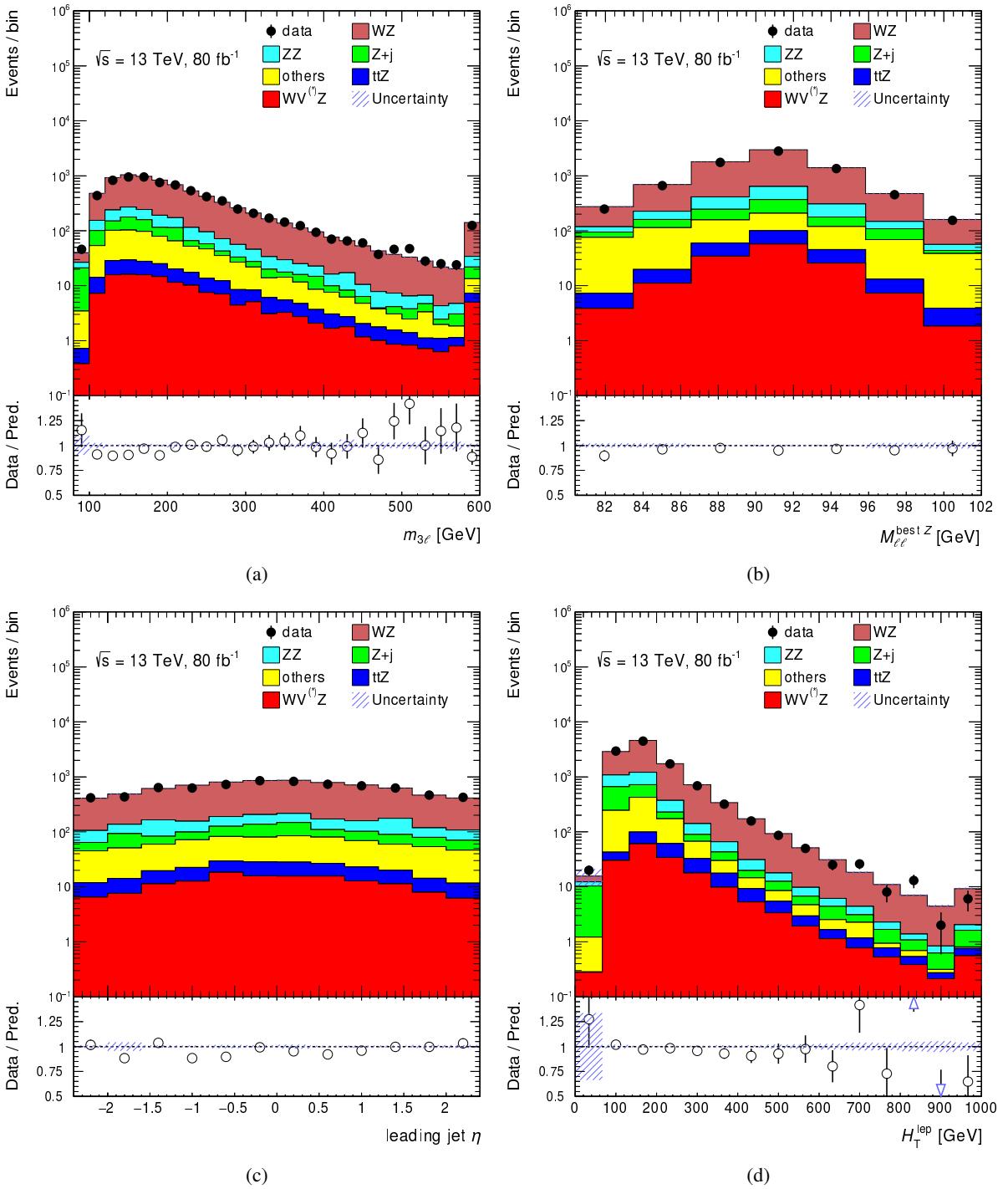


Figure 6.7: Data/simulation comparison in the 3ℓ pre-MVA region for the (a) three-lepton invariant mass, (b) best $Z \to \ell^+\ell^-$ candidate invariant mass, (c) leading jet η and (d) scalar sum of lepton p_T (H_T^{lep}). The full 80 fb^{-1} dataset is compared to the complete set of simulation samples. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

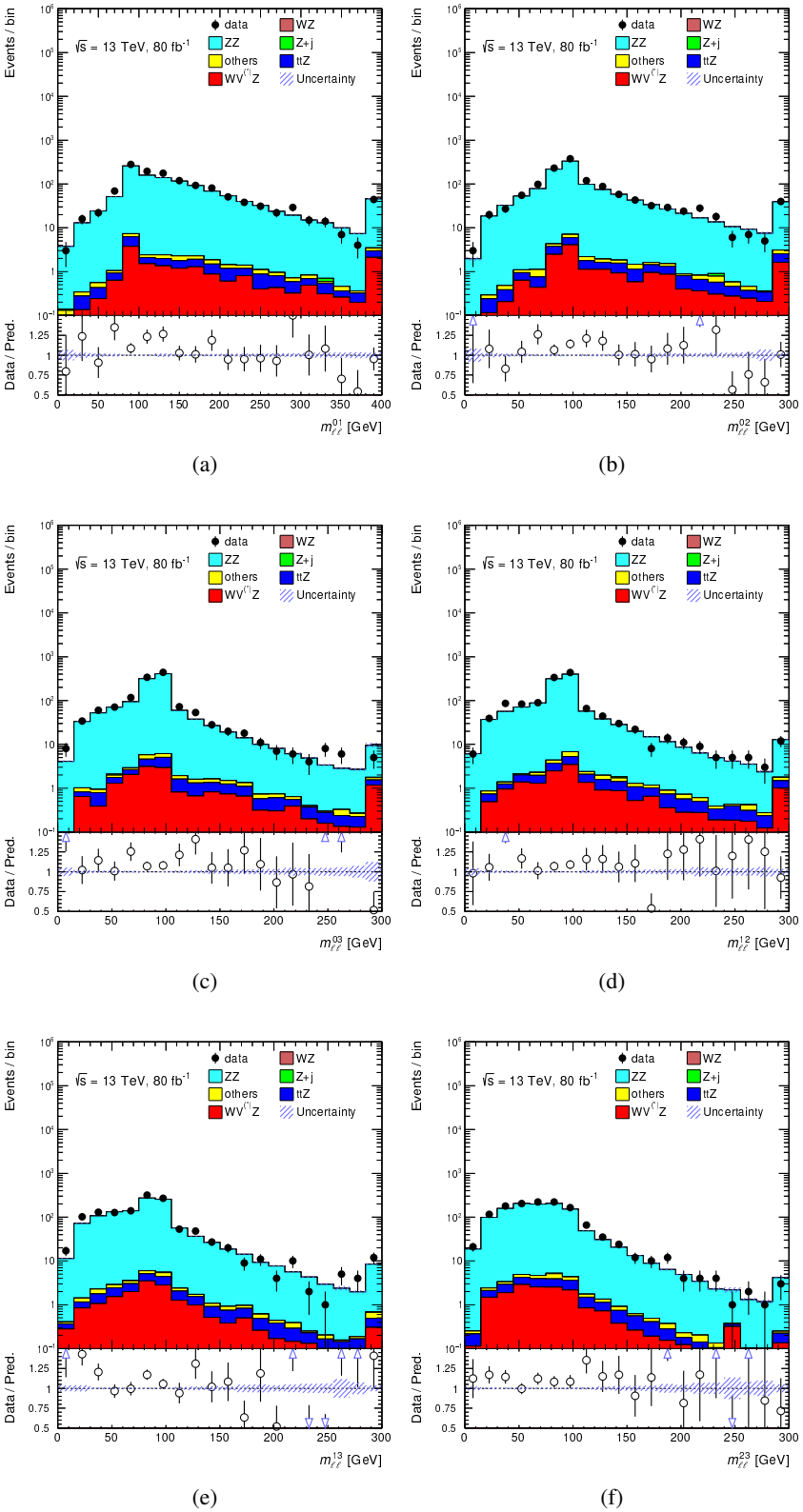
The four-lepton pre-selection region

The modelling of the ZZ process has been studied in the 4ℓ pre-selection region, defined in Section 6.3.2. Expected and observed yields in this region are reported in Table 6.7. The region is very pure in ZZ events, yielding 97% of the total MC simulation prediction. A small excess in data, with respect to predictions, is observed. This is also reflected in the observed fit results, discussed later in Section 6.9.

Table 6.7: Expected and observed yields in the 4ℓ pre-selection region with an integrated luminosity of $80\,\text{fb}^{-1}$. Only statistical uncertainties are shown.

4ℓ pre-selection	Exp./obs. yields
WVZ	14.9 ± 0.1
ZZ	$1\,149 \pm 5$
WZ	2.1 ± 0.2
Z + jets	0.2 ± 0.1
$Z + \gamma$	0 ± 0
$t\bar{t}Z$	9.8 ± 0.2
tZ	0.11 ± 0.03
tWZ	3.7 ± 0.3
others	3.3 ± 0.3
Total expected	$1\,184 \pm 5$
data	1 308

In Figures 6.8 and 6.9 kinematic distributions involving leptons are shown, comparing the inclusive $80\,\text{fb}^{-1}$ dataset to simulation. $m_{\ell\ell}^{xy}$ indicates the invariant mass of the x^{th} and y^{th} leptons, sorted by decreasing transverse momentum. An overall good agreement in the description of the relevant shapes is observed.



(a) (b) (c) (d) (e) (f)

Figure 6.8: Data/simulation comparison in the 4ℓ pre-selection region for the invariant mass of all possible lepton pairs, comparing the full $80\,\text{fb}^{-1}$ dataset to mc16a and mc16d. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

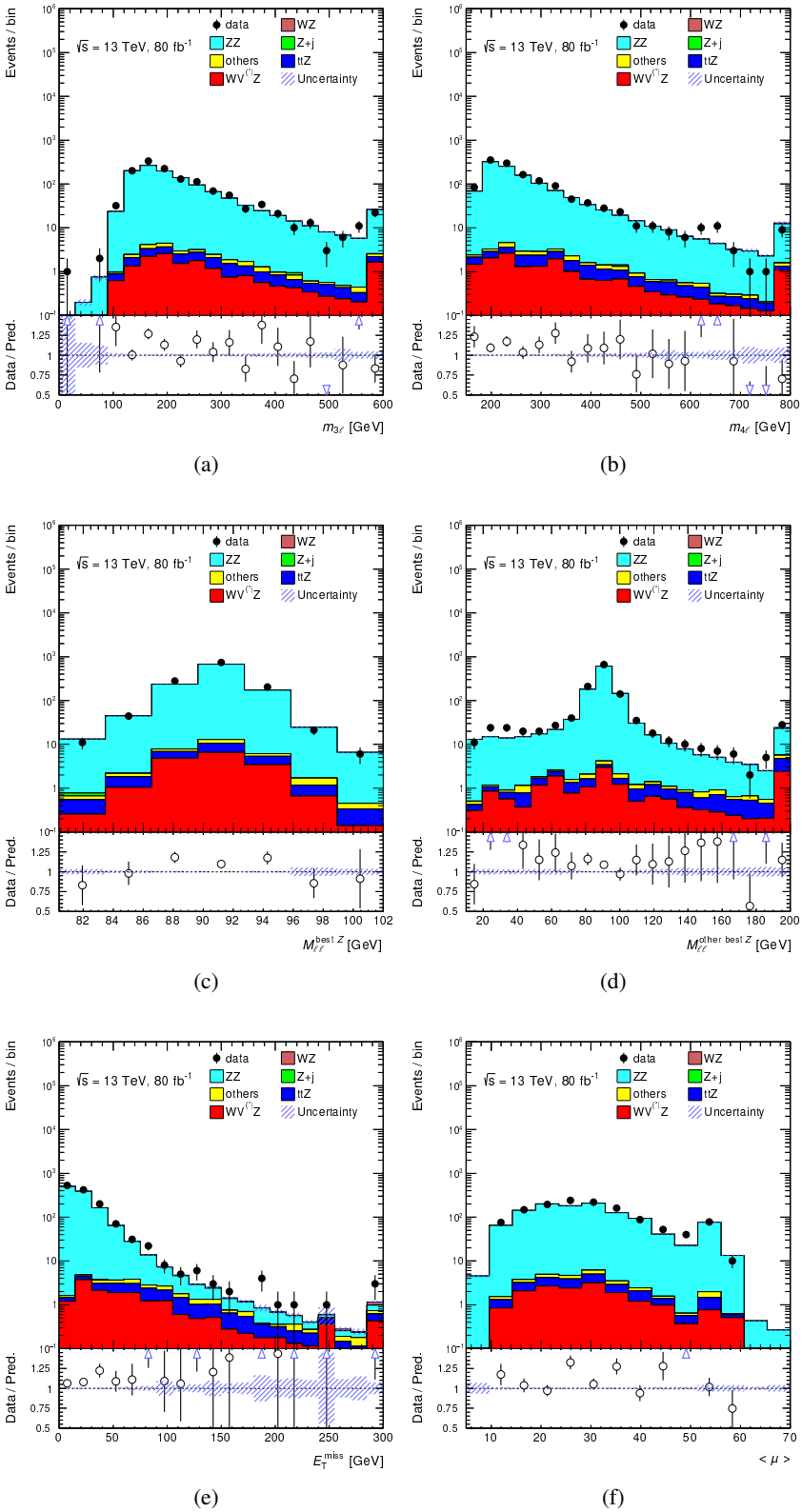


Figure 6.9: Data/simulation comparison in the 4ℓ pre-selection region for the invariant mass of (a) the three leading leptons and (b) all four leptons, (c) best Z-candidate and (d) second best lepton pair invariant masses, (e) $E_{\mathrm{T}}^{\mathrm{miss}}$ and (f) average pile-up profile, comparing the full 80 fb^{-1} dataset to mc16a and mc16d. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

The three-lepton $t\bar{t}Z$ control region

The production of $t\bar{t}Z$ is an important irreducible background in both, the 3ℓ and 4ℓ channels, as shown in Sections 6.3.1 and 6.3.2, respectively. The $t\bar{t}Z$ modelling is shown in a 3ℓ phase space.[6] The $t\bar{t}Z \to 3\ell$ control region is defined as follows:

- exactly 3 loose leptons are selected and the lepton not belonging to the best $Z \to \ell^+\ell^-$ candidate fulfils the tight requirements;
- the invariant mass of any same-flavour $\ell^+\ell^-$ pair is above 12 GeV;
- at least one same-flavour $\ell^+\ell^-$ pair with invariant mass within 10 GeV of the Z-boson mass;
- four or more reconstructed jets, of which at least 2 b-tagged, are selected.

Expected and observed yields are shown in Table 6.8: this region is expected to have a 73% purity in $t\bar{t}Z$ events.

Table 6.8: Expected and observed yields in the $t\bar{t}Z \to 3\ell$ control region with an integrated luminosity of 80 fb^{-1}. Only statistical uncertainties are shown.

$t\bar{t}Z \to 3\ell$ CR	Exp./obs. yields
WVZ	0.49 ± 0.03
ZZ	0.90 ± 0.05
WZ	6.1 ± 0.2
Z + jets	0.5 ± 0.1
$Z + \gamma$	0.4 ± 0.4
$t\bar{t}Z$	114.7 ± 0.9
tZ	8.5 ± 0.3
tWZ	10.2 ± 0.5
others	16 ± 2
Total expected	157 ± 2
data	170

Figure 6.10 shows the data-to-simulation comparison for some kinematic distributions; more distributions can be found in Appendix A. A good description of the relevant shapes is observed. This region is included in the WVZ fit combination, discussed in Section 6.8, in order to constrain $t\bar{t}Z$ systematic uncertainties (see Section 6.7). Given the requirements on the number of b-tagged jets, no event is shared between this region and either of the 3ℓ or the 4ℓ pre-selection region.

[6] As opposed to the $t\bar{t}Z \to 4\ell$ validation region explored in Section 5.10.

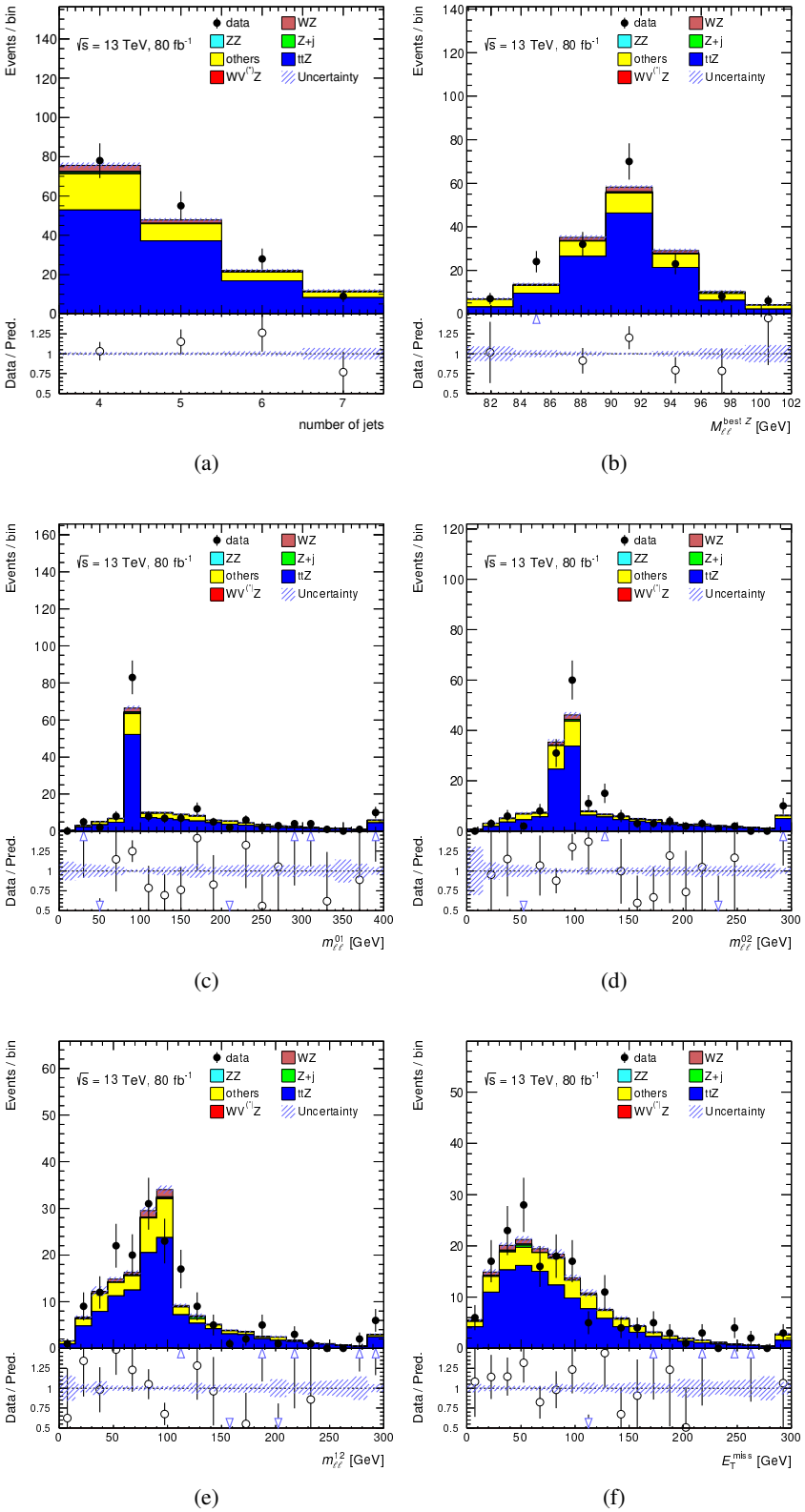


Figure 6.10: Data/simulation comparison in the $t\bar{t}Z \to 3\ell$ control region, comparing the full 80 fb^{-1} dataset to mc16a and mc16d, for: (a) the number of jets, (b) the best Z-candidate invariant mass, (c) $m_{\ell\ell}^{01}$, (d) $m_{\ell\ell}^{02}$, (e) $m_{\ell\ell}^{12}$ and (f) $E_{\mathrm{T}}^{\mathrm{miss}}$. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

6.4.2 Reducible backgrounds

The three-lepton Z + jets validation region

The largest contamination from reducible background sources is provided by Z + jets/γ, as shown in Table 6.4. The modelling of Z + jets/γ has been studied in a dedicated 3ℓ validation region; the selection reads:

- the third leading lepton fulfils $10 < p_{\mathrm{T}} < 15$ GeV;
- exactly 3 loose leptons and the lepton not belonging to the best $Z \to \ell^+\ell^-$ candidate fulfils the tight requirements;
- the invariant mass of any same-flavour $\ell^+\ell^-$ pair is above 12 GeV;
- at least one same-flavour $\ell^+\ell^-$ pair with invariant mass within 10 GeV of the Z-boson mass;
- the three-lepton invariant mass is smaller than 150 GeV;
- exactly one reconstructed non-b-tagged jet is present.

The leptons used in the analysis, as discussed in Section 6.2, are required to fulfil $p_{\mathrm{T}} > 15$ GeV. In order to allow for a good acceptance and purity for fake and non-prompt leptons, the softest lepton for each event entering this region is required to fulfil $10 < p_{\mathrm{T}} < 15$ GeV. There is no overlap, therefore, between this region and the regions defined before. The requirement on the three-lepton invariant mass allows to further enrich this phase space with Z + jets/γ events, otherwise still dominated by the WZ processes. Expected and observed yields in this region are reported in Table 6.9: about 55% of the events are expected to be due to the production of three leptons in a Z + jets/γ system.

Table 6.9: Expected and observed yields in the 3ℓ Z + jets validation region with an integrated luminosity of 80 fb^{-1}; a good agreement between data and expectation is observed within statistical uncertainties.

3ℓ Z + jets VR	Exp./obs. yields
WVZ	1.14 ± 0.05
ZZ	92 ± 7
WZ	182 ± 4
Z + jets	370 ± 50
$Z + \gamma$	30 ± 10
$t\bar{t}Z$	0.27 ± 0.04
tZ	1.5 ± 0.1
tWZ	0.17 ± 0.07
others	38 ± 3
Total expected	720 ± 50
data	743

In Figure 6.11, some kinematic distributions, including the scalar sum of the p_{T} of all selected leptons ($H_{\mathrm{T}}^{\mathrm{lep}}$) and of all selected jets ($H_{\mathrm{T}}^{\mathrm{had}}$), are shown for the inclusive 80 fb^{-1} dataset. More distributions are available in Appendix A. An overall good agreement is observed within statistical uncertainties. According to the MC truth-record origin, 82% of the non-prompt and fake leptons entering this region

are coming from the decay of a heavy-flavour hadron, while the rest is fake electrons from the conversion of a real photon. This composition is very similar to the one expected in the 3ℓ pre-selection region.

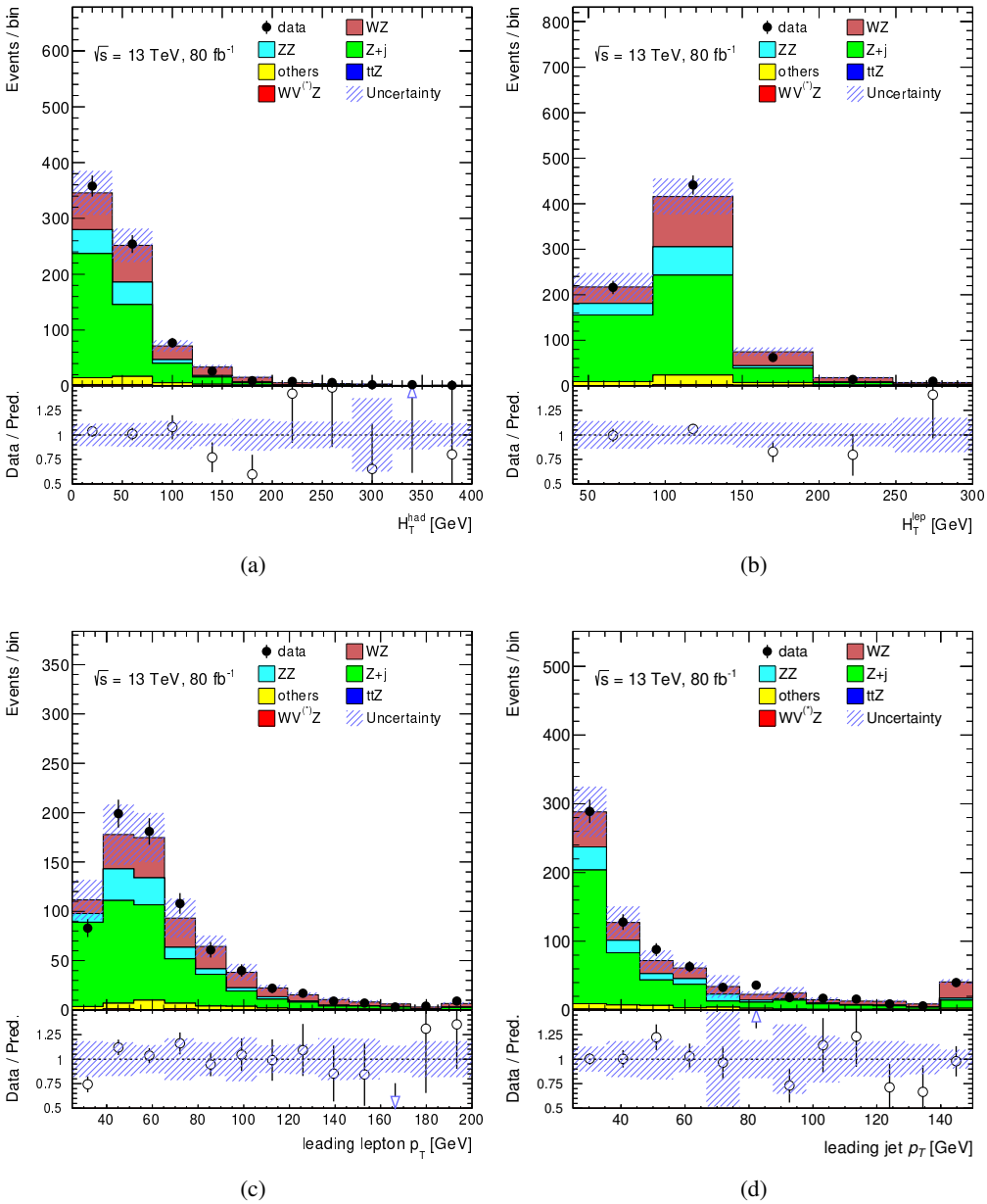


Figure 6.11: Data/simulation comparison in the 3ℓ Z + jets validation region for (a) $H_{\mathrm{T}}^{\mathrm{had}}$, (b) $H_{\mathrm{T}}^{\mathrm{lep}}$, the leading (c) lepton and (d) jet p_{T}, comparing the full $80\,\mathrm{fb}^{-1}$ dataset to mc16a and mc16d. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

6.5 Jet multiplicity-based reweighting

In both, the 3ℓ pre-MVA and the 4ℓ pre-selection regions a good agreement is observed between data and simulation, as discussed in Section 6.4.1. However, the distribution of the number of reconstructed jets is poorly described in both regions, as shown in Figure 6.12.

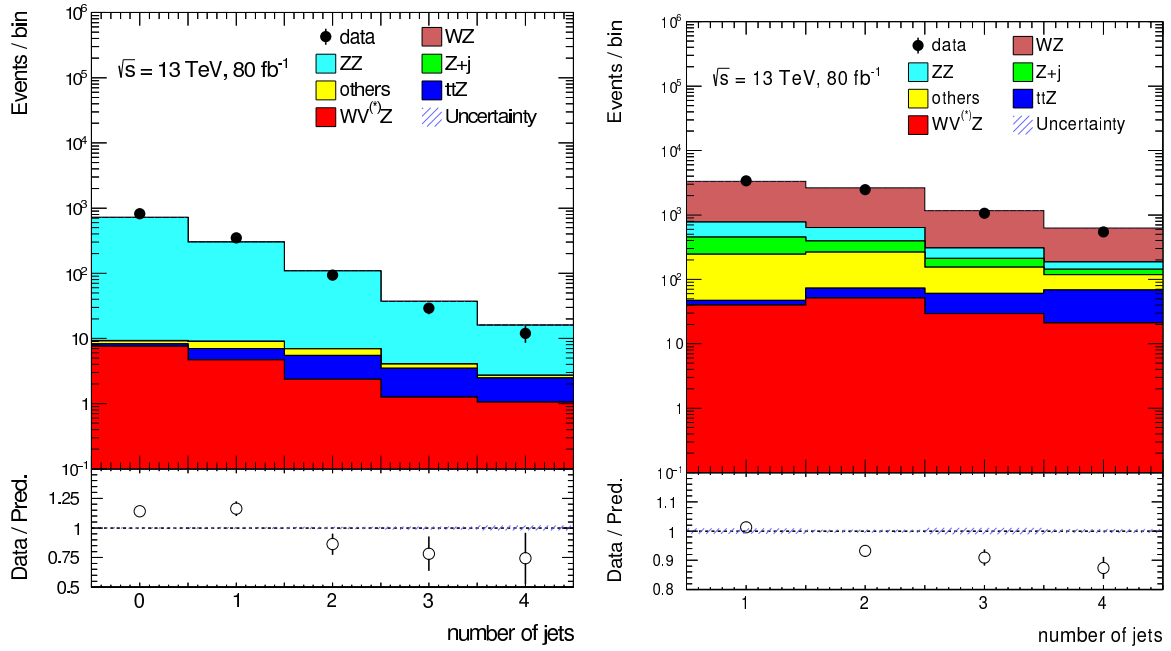


Figure 6.12: Comparison between the expected and the observed number of reconstructed jets in the (a) 4ℓ pre-selection and (b) 3ℓ pre-MVA region. The lower panel shows the ratio between data and simulation. Only statistical uncertainties are shown.

The disagreement observed in the jet multiplicity, in the context of this analysis, is compatible with the one observed by previous ATLAS measurements at 36.1 fb^{-1}. In particular, the same trend in the number of reconstructed jets has been observed in the latest WZ [192] and ZZ [193] cross-section measurements, where the same MC generator is exploited to simulate the VV processes. It has to be noted that Figure 6.12 does not include systematic effects, an important argument for the discussion of the theoretical uncertainties in Section 6.7.

This section outlines a reweighting procedure, employed throughout the analysis in order to improve the agreement between data and simulation. The assumption, tested comparing the modelling before and after the procedure at issue, is that the poor description of the distributions in Figure 6.12 is mainly due to the WZ and ZZ backgrounds. In fact, these processes clearly dominate all of the bins of the distribution. It's important to highlight, that the diboson samples are generated at NLO accuracy only for 0 and 1 additional partons and at LO accuracy for 2 and 3 additional partons; therefore some deviations might be expected at high jet multiplicity, where the simulation accuracy becomes less and less reliable (or, equivalently, more and more affected by systematic uncertainties).

Two principles are underlying the reweighting procedure:

- it is *shape-only*, i.e. WZ (ZZ) events are reweighted to match the shape of the jet multiplicity

in data in 3ℓ (4ℓ), but their overall normalisation is unchanged with respect to the simulation prediction;

- it is *combined* in the two channels: since the ZZ contamination is not negligible in the 3ℓ channel, the WZ reweighting is performed on top of the ZZ reweighting, determined in the 4ℓ channel and applied to the 3ℓ channel.

These two principles allow to preserve the VV normalisation prediction and avoid inconsistencies among the two channels, when performing their statistical combination. A scale factor is extracted in each bin i of the jet-multiplicity distribution:

$$K_{VZ}^{\text{rew},i} = \frac{f_{\text{data}}^{i} - f_{\text{MC}-VZ}^{i}}{f_{VZ}^{i}}, \tag{6.1}$$

where data and MC simulation distributions are normalised to their own total yield (shape-only reweighting), f_{data}^{i} is the number of observed events, $f_{\text{MC}-VZ}^{i}$ is the expected yield from all simulated processes but the involved diboson process, whose number of expected events is labelled as f_{VZ}^{i}.

As mentioned above, the reweighting scale factors are extracted in the 4ℓ channel for ZZ and, afterwards, for WZ in 3ℓ. In the following, the reweighting procedure is discussed in the 4ℓ and the 3ℓ channel.

6.5.1 Jet multiplicity-based reweighting in the four-lepton channel

In the 4ℓ pre-selection region a good agreement is observed between data and simulation, as discussed in Section 6.4.1, except for the number of reconstructed jets. The assumption of the reweighting procedure is that the poor description of the distribution in Figure 6.12 is mainly due to ZZ. This process is the dominant contribution in all bins of the distribution at issue.

The scale factors, given by Eq. 6.1 (where $VZ = ZZ$), are extracted in three bins: 0 or 1 jets, exactly 2 jets and 3 or more jets. They are then applied bin-by-bin to the ZZ simulation in order to match the shape observed in data. Table 6.10 shows a comparison between expected and observed yields in these regions and the corresponding scale factors. The jet multiplicity distribution, after applying the reweighting procedure, is shown in Figure 6.13. As shown in Section 6.4.1, the ZZ overall normalisation is 10% above expectation.

Table 6.10: Expected and observed yields in the 4ℓ pre-selection region, separately for: 0 or 1 jets, exactly 2 jets and 3 or more jets; the second-to-last row shows the purity in ZZ, while the last row shows the extracted scale factors in each region. The quoted uncertainties are only statistical.

4ℓ pre-selection	≤ 1 jets	2 jets	≥ 3 jets
WVZ	9.31 ± 0.09	2.95 ± 0.05	2.66 ± 0.04
ZZ	$1\,002 \pm 5$	102 ± 1	46.2 ± 0.5
Total expected	$1\,022 \pm 5$	109 ± 1	53.2 ± 0.6
data	1 173	94	41
ZZ purity (%)	98	94	87
K_{ZZ}^{rew}	1.04	0.70	0.64

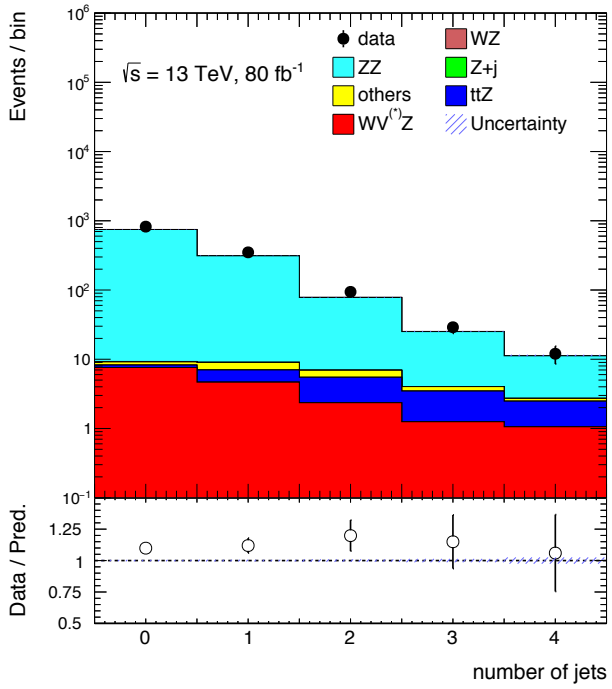


Figure 6.13: Comparison between the reweighted simulation and the observed number of reconstructed jets in the 4ℓ pre-selection region. The lower panel shows the ratio between data and simulation. Statistical-only uncertainties are shown.

6.5.2 Jet multiplicity-based reweighting in the three-lepton channel

The 3ℓ pre-MVA region data-to-simulation agreement was discussed in Section 6.4.1. As for the 4ℓ pre-selection region, also the 3ℓ pre-MVA region is characterised by a discrepancy between the expected and the observed number of reconstructed jets (see Figure 6.12). As previously discussed in this section, the scale factors extracted for the reweighting of ZZ in 4ℓ (see Table 6.10) are applied to the ZZ yields in 3ℓ. Table 6.11 shows a comparison between the expected and observed yields in three 3ℓ regions and the resulting WZ scale factors (from Eq. 6.1, where $VZ = WZ$). The WZ scale factors are also extracted in three bins: exactly 1 jet ($3\ell 1j$), exactly 2 jets ($3\ell 2j$) and 3 or more jets ($3\ell 3j$). The jet multiplicity distribution, after applying the ZZ and WZ reweighting procedures, is shown in Figure 6.14.

In order to test the improvement in the diboson modelling, a global χ^2 test is performed for jet-related and non-jet-related distributions, separately. This test, performed on several distributions, shows that the description of jet-related shapes is improved (the reduced χ^2 goes from 1.24 to 0.92) and no deterioration in the modelling of non-jet-related observables is observed (the reduced χ^2 goes from 1.23 to 1.22) after the reweighting. All input distributions to the χ^2 test can be found in Appendix B.

The scale factors reported in Tables 6.10 and 6.11 are applied to the WZ and ZZ simulated events before the training of MVAs, see Sections 6.6.1 and 6.6.2, and the fits, as discussed in Sections 6.8 and 6.9.

Table 6.11: Expected and observed yields in the $3\ell 1j$, $3\ell 2j$ and $3\ell 3j$ regions at 80 fb^{-1}; the second-to-last row shows the purity in WZ, while the last row shows the extracted scale factors in each region. The ZZ expected yields are scaled by the scale factors reported in Table 6.10. The quoted uncertainties are only statistical.

3ℓ pre-MVA	$3\ell 1j$	$3\ell 2j$	$3\ell 3j$
WVZ	37.3 ± 0.2	50.9 ± 0.3	49.6 ± 0.3
ZZ	335 ± 5	172 ± 3	89 ± 2
WZ	$2\,560 \pm 10$	$2\,007 \pm 5$	$1\,299 \pm 3$
Total expected	$3\,340 \pm 30$	$2\,640 \pm 10$	$1\,790 \pm 10$
data	3 379	2 466	1 609
WZ purity (%)	77	76	73
K_{WZ}^{rew}	1.05	0.98	0.93

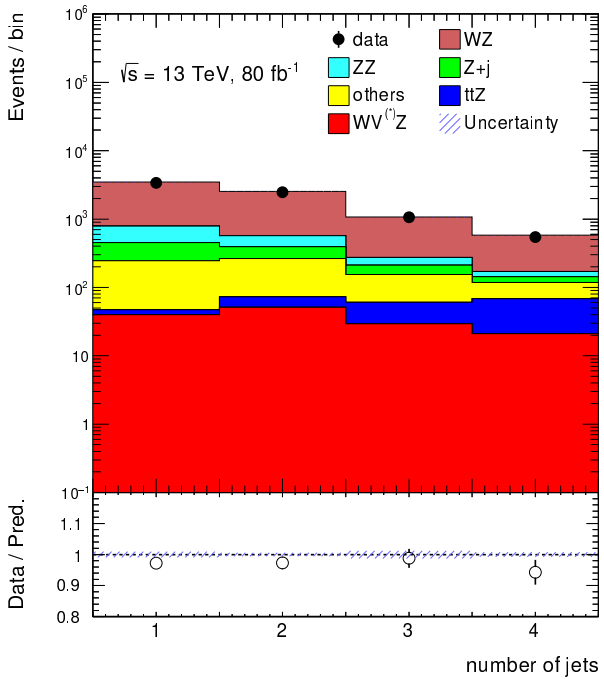


Figure 6.14: Comparison between the reweighted simulation and the observed number of reconstructed jets in the 3ℓ pre-MVA region. The lower panel shows the ratio between data and simulation. Statistical-only uncertainties are shown.

6.6 Signal selection

This section is devoted to the description of the signal selection, involving multivariate techniques in both, the three- (Section 6.6.1) and the four-lepton (Section 6.6.2) channels.

6.6.1 Signal selection in the three-lepton channel

The 3ℓ pre-selection region, as shown in Table 6.4, shows a non-negligible contamination expected to come from reducible backgrounds, in particular $Z \to e^+e^-/\mu^+\mu^-$ plus a fake/non-prompt lepton. This contribution can be further reduced before employing multivariate techniques, as these events tend to be much softer than signal events. Figure 6.15 shows the predicted shape of the scalar sum of all selected leptons' and jets' p_{T} (H_{T}) in the 3ℓ pre-selection region for WVZ, Z + jets and WZ. A lower cut on H_{T} is applied at 200 GeV to define the 3ℓ *pre-MVA region*. Table 6.12 shows the expected yields in this region, where the Z + jets contribution has been reduced by more than a factor of 2 with respect to the 3ℓ pre-selection region. Similarly to the Z + jets validation region (see Section 6.4.2) most, namely 74%, of the fake and non-prompt leptons entering the 3ℓ pre-MVA region is expected to originate in heavy-flavour hadron decays. The rest is associated to fake electrons from real photon conversion.

The kinematic properties of the WVZ and WZ processes show a strong dependence on the multiplicity of the reconstructed jets in the final state. Also, the signal purity is expected to be higher at larger jet multiplicities. Therefore, the 3ℓ pre-MVA region is split into three *signal regions* according to the number of jets:

- exactly 1 reconstructed jet ($3\ell 1j$);
- exactly 2 reconstructed jets ($3\ell 2j$);
- 3 or more reconstructed jets ($3\ell 3j$).

Expected yields, after applying the jet multiplicity-based reweighting, are reported in Table 6.12, for each of the three 3ℓ signal regions.

Table 6.12: Expected yields in the $3\ell 1j$, $3\ell 2j$, $3\ell 3j$ and 3ℓ pre-MVA regions at 80 fb^{-1}. The jet multiplicity-based reweighting is applied to the WZ and ZZ expected yields. The uncertainties shown are only statistical.

3ℓ pre-MVA	$3\ell 1j$	$3\ell 2j$	$3\ell 3j$	Total
WVZ	37.3 ± 0.2	50.9 ± 0.3	49.6 ± 0.3	137.6 ± 0.5
ZZ	335 ± 5	172 ± 3	89 ± 2	596 ± 6
WZ	$2\,690 \pm 10$	$1\,967 \pm 5$	$1\,208 \pm 3$	$5\,860 \pm 10$
Z + jets	170 ± 30	96 ± 9	70 ± 10	330 ± 30
$Z + \gamma$	42 ± 8	32 ± 7	10 ± 4	84 ± 10
$t\bar{t}Z$	7.3 ± 0.2	21.6 ± 0.3	78.6 ± 0.7	107.5 ± 0.8
tZ	23.7 ± 0.5	24.5 ± 0.5	12.1 ± 0.4	60.3 ± 0.8
tWZ	4.2 ± 0.3	11.3 ± 0.5	19.7 ± 0.7	35.2 ± 0.9
others	173 ± 6	156 ± 6	112 ± 5	442 ± 9
Total expected	$3\,460 \pm 30$	$2\,600 \pm 10$	$1\,700 \pm 10$	$7\,770 \pm 40$

Several discriminating observables have been built and studied in each of the three 3ℓ regions, in order to discriminate the signal against the overwhelming contribution from WZ. These observables involve

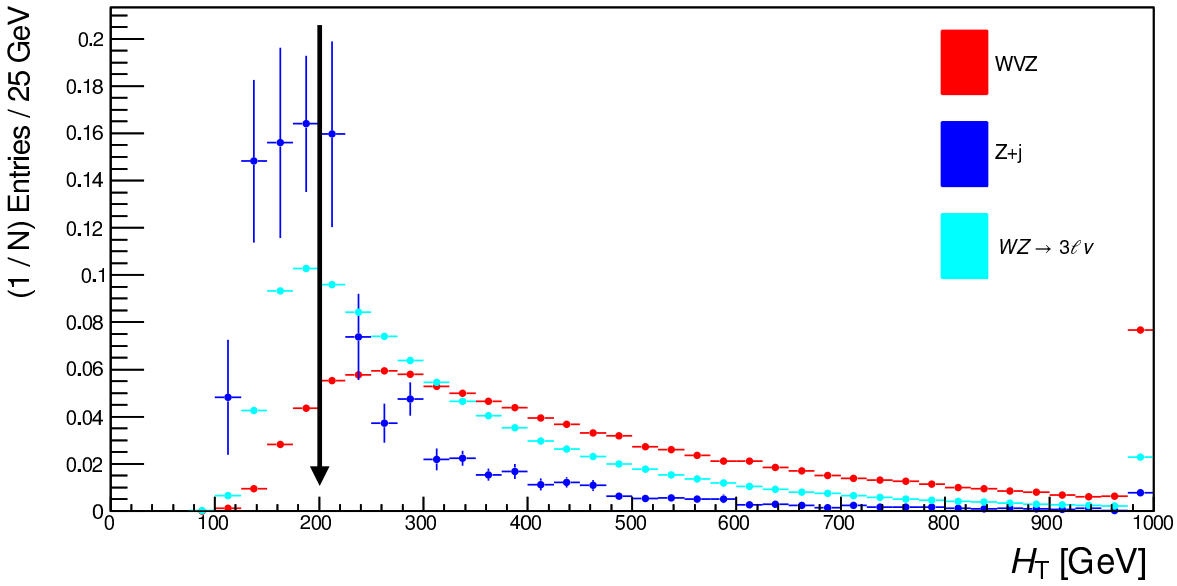
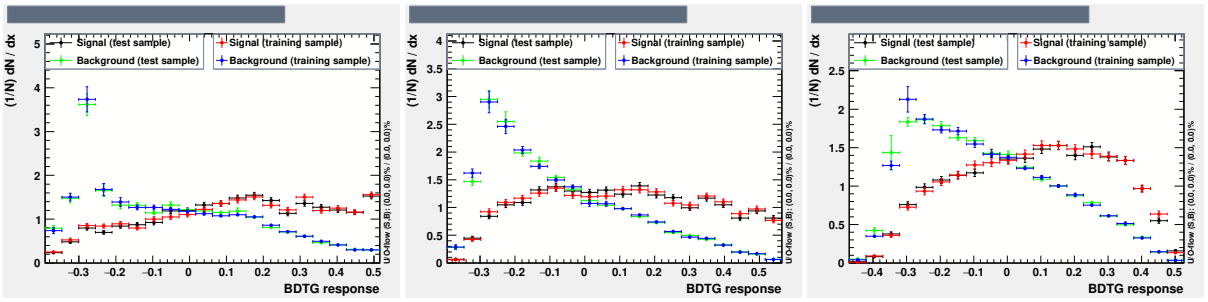


Figure 6.15: Comparison between the expected H_T shapes for WVZ, Z + jets and WZ in the 3ℓ pre-selection region. The black arrow shows the $H_T > 200$ GeV requirement applied to events entering the 3ℓ pre-MVA region. Statistical-only uncertainties are shown.

kinematic properties of the reconstructed candidates in the WVZ system, namely $W \to \ell\nu$, $V \to qq'$ and $Z \to \ell^+\ell^-$. The final set of inputs, selected according to the optimal performance,[7] used for the the training of three gradient-boosted decision trees (one per region), is reported in Table 6.13. The reweighting procedure, outlined in Section 6.5, is applied to the training samples before performing the training of the MVAs. Figure 6.16 shows the discriminating power of the final MVA responses in each of the three 3ℓ signal regions. The three BDT shapes are then combined into a profile likelihood shape fit, as discussed later in Section 6.8.



(a) (b) (c)

Figure 6.16: MVA responses in the (a) $3\ell 1j$, (b) $3\ell 2j$ and (c) $3\ell 3j$ regions. Only statistical uncertainties are shown.

[7] Evaluated as the maximum ROC integral on the testing sample.

Table 6.13: List of discriminating variables fed as inputs to the 3ℓ MVA trainings.

Input Variable	$3\ell 1j$	$3\ell 2j$	$3\ell 3j$
$m_{3\ell}$	×	×	×
$m_{\ell\ell}^{01}$	×	×	
$m_{\ell\ell}^{02}$	×	×	
$m_{\ell\ell}^{12}$	×	×	
$p_{\mathrm{T}}^{\ell_0}$	×	×	
$p_{\mathrm{T}}^{\ell_1}$	×	×	×
$p_{\mathrm{T}}^{\ell_2}$	×	×	×
total lepton charge	×	×	×
$H_{\mathrm{T}}^{\mathrm{lep}}$	×	×	×
H_{T}	×	×	
leading jet p_{T}	×	×	
invariant mass of all leptons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$	×		×
invariant mass of best $Z \to \ell^+\ell^-$ and leading jet	×		
$E_{\mathrm{T}}^{\mathrm{miss}}$		×	×
subleading jet p_{T}		×	×
m_{jj}^{01}		×	
$m_{\mathrm{T}}^{W\to\ell\nu}$		×	
$H_{\mathrm{T}}^{\mathrm{had}}$			×
n. of reconstructed jets			×
$m_{jj}^{\text{best } W}$			×
smallest m_{jj}			×

6.6.2 Signal selection in the four-lepton channel

The four-lepton phase space is dominated, as shown in Table 6.5, by the ZZ contribution. Relatively pure regions in $W^{\pm}W^{\mp}Z$ and ZZ can be obtained by simply splitting the 4ℓ pre-selection region according to the flavour of the lepton pair not being the best $Z \to \ell^+\ell^-$ candidate[8] (at least one is required, see Section 6.3.2). Requiring this, called "*second best lepton pair*", to be of different flavour (4ℓ DF region), i.e. $e^+\mu^-$ or $e^-\mu^+$, allows to select a region very sensitive to the production of $W^{\pm}W^{\mp}Z \to 4\ell 2\nu$, where ZZ mostly enters with a Z-boson decaying to two leptonically decaying tau leptons. Figure 6.17 shows the MC truth-record decay of the ZZ system in the DF signal region; more than 96% of the events are expected to be ZZ events with the second best lepton pair decaying to two leptonically decaying taus. On the other hand, when the second best lepton pair is same flavour (4ℓ SF region), i.e. e^+e^- or $\mu^-\mu^+$, most of $W^{\pm}ZZ \to 4\ell 2q$ and half of $W^{\pm}W^{\mp}Z \to 4\ell 2\nu$ events are selected together with a large fraction of ZZ events.

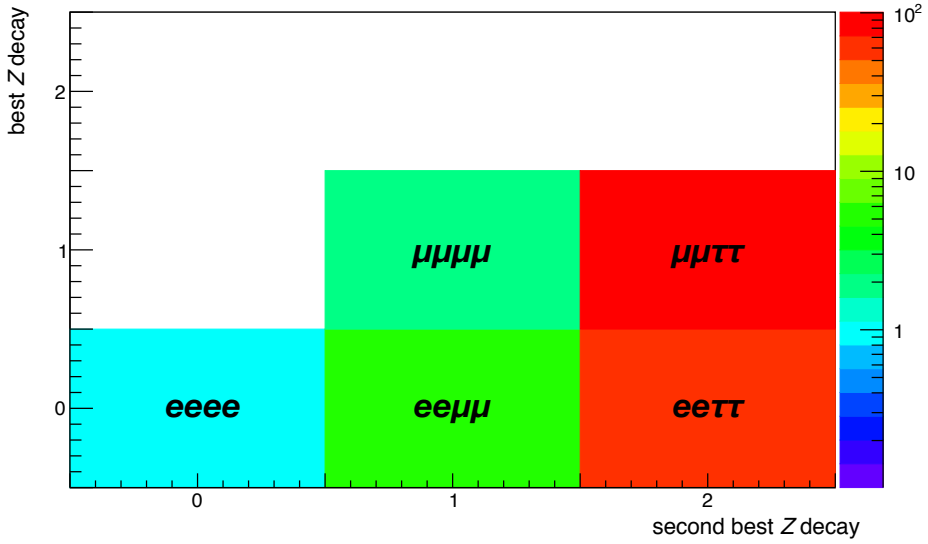


Figure 6.17: MC truth-record decays of ZZ events falling in the 4ℓ DF region. The Z-boson decays are defined as: $0 = ee$, $1 = \mu\mu$, $2 = \tau\tau$. The overall normalisation of the ZZ simulation is arbitrary.

The split of a phase space into a purer and a less pure region allows to enhance the analysis sensitivity, as discussed in Section 5.12 for the split into Z-enriched and Z-depleted signal regions in the context of the $t\bar{t}H \to 4\ell$ analysis. This is the case for the SF region, where $W^{\pm}W^{\mp}Z$ events show a flat invariant mass of the second best lepton pair, while $W^{\pm}ZZ$ features a peak at the Z-boson mass. Figure 6.18 shows the $W^{\pm}W^{\mp}Z$, $W^{\pm}ZZ$ and ZZ shapes of the invariant mass of the second best lepton pair in the 4ℓ SF region. It is split into on-shell SF, i.e. the second best lepton pair invariant mass ($m_{\ell\ell}^{\text{second best pair}}$) within 10 GeV of the Z-boson mass, and off-shell SF regions, i.e. $\left|m_{\ell\ell}^{\text{second best pair}} - m_Z\right| > 10\,\text{GeV}$. Dividing the 4ℓ SF region into an on-shell and an off-shell region brings a $\sim 5\%$ improvement in the 4ℓ statistical-only expected significance.

Similarly to the three-lepton channel case, a dedicated gradient-boosted decision tree is trained in each of the three 4ℓ *signal regions*: DF, on-shell SF and off-shell SF. In Table 6.14 the full set of discriminating

[8] When more than one same-flavour $\ell^+\ell^-$ pair with invariant mass within 10 GeV of the Z-boson mass is selected, the closest one to 91.2 GeV is taken as the best $Z \to \ell^+\ell^-$ candidate.

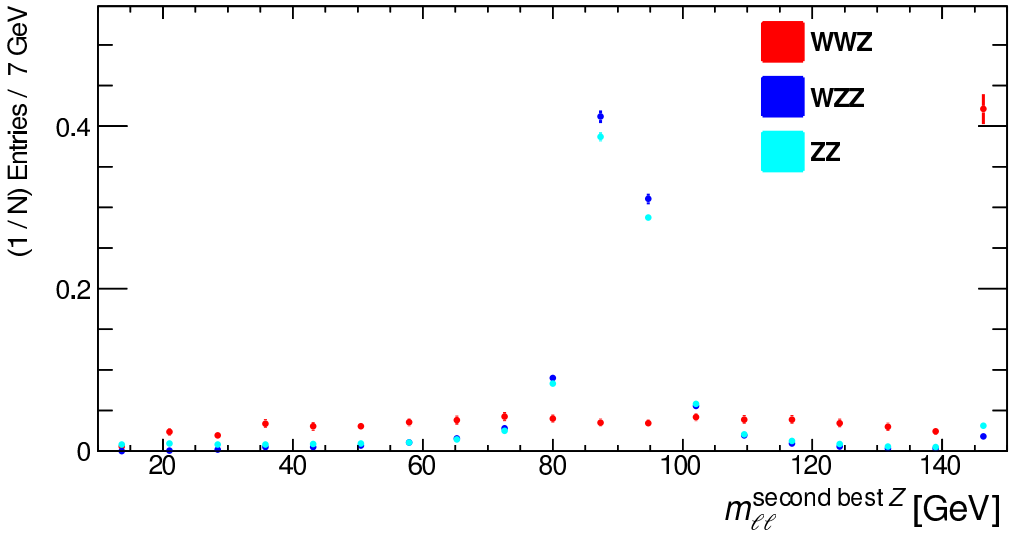


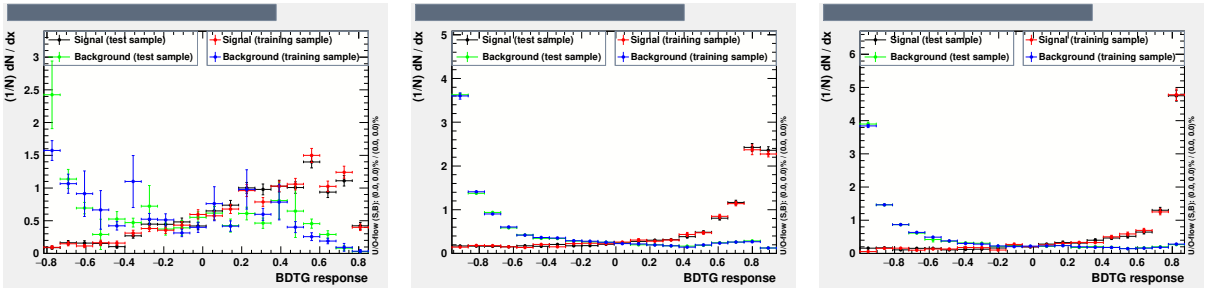
Figure 6.18: Shape of the second best lepton pair invariant mass for the $W^{\pm}W^{\mp}Z$, $W^{\pm}ZZ$ and ZZ processes in the 4ℓ SF region. Statistical-only uncertainties are shown.

inputs to the BDT trainings in the 4ℓ channel is reported, while Figure 6.19 shows the final discriminating MVA responses. The reweighting procedure, outlined in Section 6.5 and affecting ZZ simulated events, is applied before training the MVAs. Similarly to the 3ℓ channel, the resulting BDT shapes in the three 4ℓ regions are exploited as distinct inputs to be combined through a profile likelihood shape fit, as discussed in Section 6.8.

Table 6.14: List of discriminating variables fed as inputs to the 4ℓ MVA trainings.

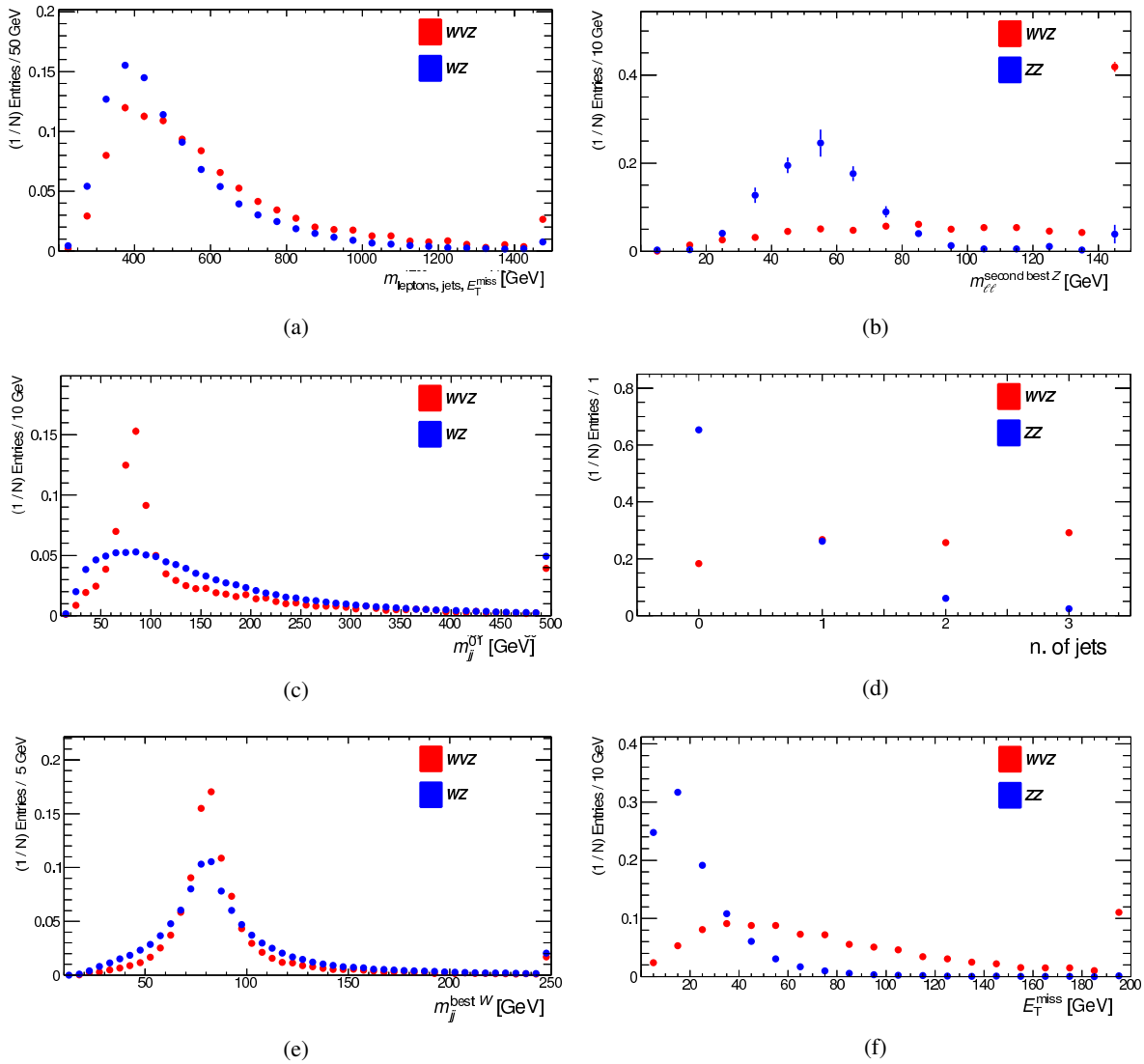
Input Variable	DF	on-shell SF	off-shell SF
number of reconstructed jets	×	×	×
$m_{4\ell}$	×	×	×
$E_{\text{T}}^{\text{miss}}$	×	×	×
$H_{\text{T}}^{\text{lep}}$	×		
$H_{\text{T}}^{\text{had}}$	×		
$m_{\ell\ell}^{\text{second best pair}}$	×	×	×
$m_{\ell\ell}^{\text{best } Z}$		×	×
H_{T}		×	×

The expected shape of the most discriminating input variable in each of the six signal regions (three 3ℓ and three 4ℓ) MVA training is shown in Figure 6.20, comparing the signal to the largest background: WZ in the 3ℓ and ZZ in the 4ℓ regions. The modelling of the MVA input variables has been extensively studied in the six signal regions. Figure 6.21 shows the data-to-simulation comparison for the most discriminating input variable in each of the signal regions. The comparison is provided for all other MVA input variables in Appendix C.



(a) (b) (c)

Figure 6.19: MVA responses in the 4ℓ (a) DF, (b) on-shell SF and (c) off-shell SF regions. Only statistical uncertainties are shown.



(a) (b) (c) (d) (e) (f)

Figure 6.20: Top-ranked MVA input variables in the (a) $3\ell 1j$, (c) $3\ell 2j$, (e) $3\ell 3j$ and the 4ℓ (b) DF, (d) on-shell SF and (f) off-shell SF regions. The signal is compared to the WZ (ZZ) background process in the 3ℓ (4ℓ) regions. Only statistical uncertainties are shown.

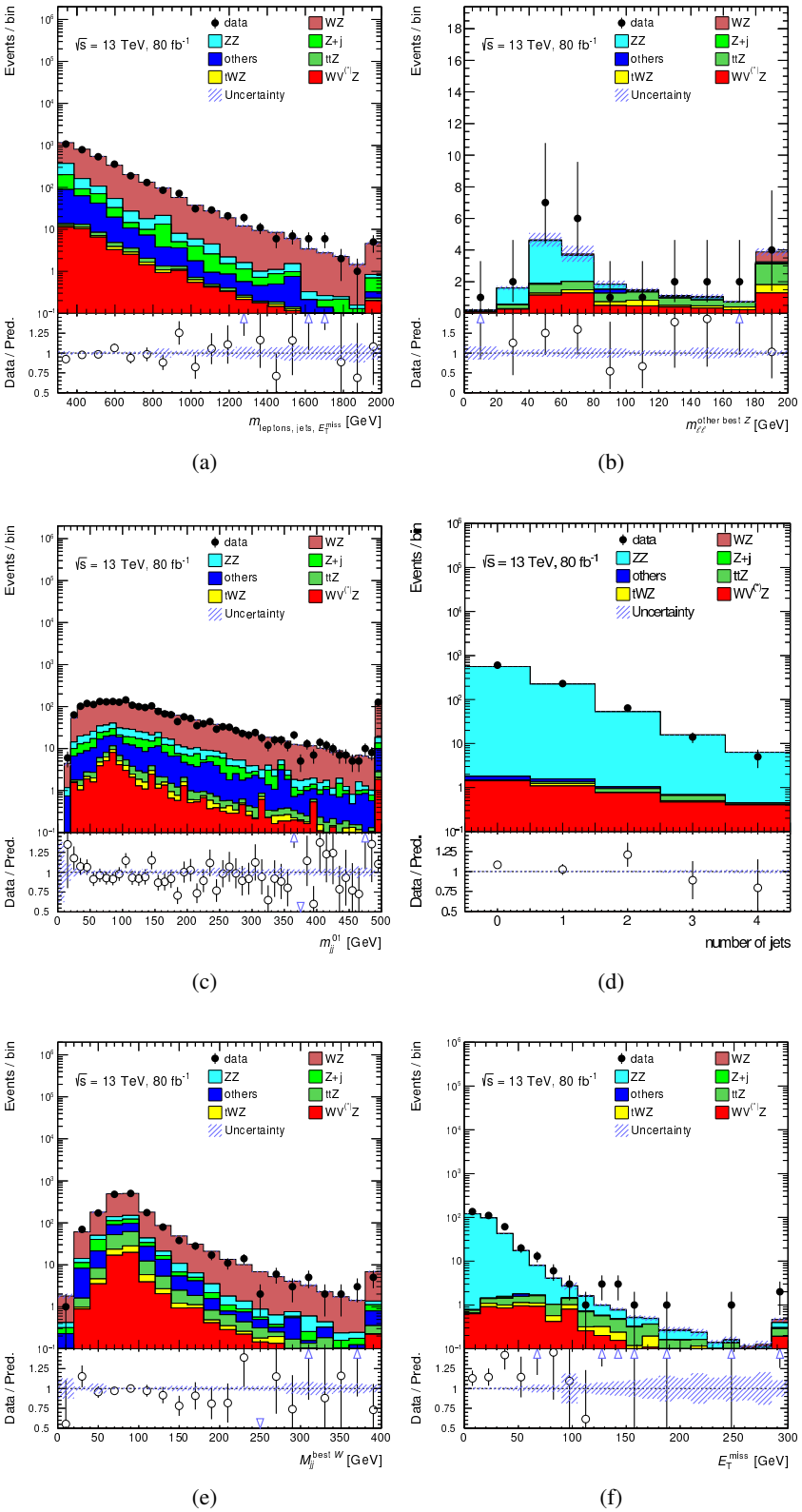


Figure 6.21: Data-to-simulation comparison for the top-ranked MVA input variables in the (a) $3\ell 1j$, (c) $3\ell 2j$, (e) $3\ell 3j$ and the 4ℓ (b) DF, (d) on-shell SF and (f) off-shell SF regions. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

6.7 Systematic uncertainties

The systematic uncertainties taken into account for this analysis are presented in this section, covering experimental systematic uncertainties in Section 6.7.1 and signal and background modelling systematic uncertainties in Sections 6.7.2–6.7.4. All systematic uncertainties are summarised in Table 6.15. Each component of a systematic uncertainty is introduced in the fit model (see Section 5.1.7) as an independent nuisance parameter and impacts the estimated signal and/or background rates. All nuisance parameters are correlated among all of the fitted regions.

In order to avoid statistical fluctuations in the systematic variations two procedures are applied: *averaging* and *pruning*. The first step in the systematic averaging is the *symmetrisation* of the systematic uncertainty. In this analysis a *one-sided* symmetrisation is employed: systematic uncertainties for which the 1σ variation is available in one direction only, by convention the up variation, are complemented by a down variation as the opposite of the up variation, around the nominal prediction. The second step in the averaging procedure is the *smoothing*, which averages systematic uncertainties across bins. It is meant to remove fluctuations in the systematic model, in particular for uncertainties derived from the comparison of simulations with a limited amount of events. The *ROOT smoothing* is directly based on the smoothing function of one-dimensional histograms, i.e. TH1::Smooth [79]. It averages the bin contents based on the neighbouring bin information and the histogram integral.

A pruning procedure is applied to the large set of nuisance parameters included in the profile likelihood fit, as was done in the $t\bar{t}H \to 4\ell$ analysis (Section 5.11). Small systematic uncertainties, not affecting the final result, can be removed before performing the fit. In the WVZ profile likelihood fit, normalisation and shape components of each single nuisance parameter are pruned separately; a normalisation/shape variation, corresponding to a given nuisance parameter, is removed if its impact is found to be smaller than 0.5% on all processes in all bins of all fitted regions.

6.7.1 Detector systematic uncertainties

Experimental systematic uncertainties are considered, that are related to trigger efficiency, lepton reconstruction and identification, jet calibration, b-tagging and global event activity. They are evaluated by the ATLAS "performance groups" and are used in the WVZ analysis either as an overall event reweighting or as a rescaling of the object energy and momentum. Most of the detector systematic uncertainties follow the treatment already discussed in Section 5.11.1; this treatment applies to: luminosity uncertainty, uncertainty in the efficiency of lepton trigger (TRIG), identification (ID, both for LooseLH and TightLH electrons), reconstruction (RECO) and muon track-to-vertex association (TTVA), lepton momentum scale and resolution corrections, jet energy resolution (JER) and jet energy scale (JES), b-tagging efficiencies and $E_{\mathrm{T}}^{\mathrm{miss}}$ (MET, scale and resolution).

Lepton isolation (ISO) systematic uncertainties, associated with the tight lepton definition exploited in this analysis, are measured using the tag and probe method with $Z \to \ell^+\ell^-$ events. The scale factors are 0.92 for $10 < p_{\mathrm{T}} < 15$ GeV muons and 0.97 for electrons, and averaging at 0.98 to 0.99 for higher p_{T} leptons.[9] Calibration scale factors for the tight electrons are derived as a function of η and p_{T}. A one-dimensional parametrisation is, instead, employed for tight muons as a function of the muon p_{T}. The calibration of the isolation working point for muons is derived from the 2017 dataset, and is validated on the 2015 and 2016 dataset. More details, including an expanded description of the method exploited for the retrieval of related systematic uncertainties, can be found in Ref. [34].

[9] Only $p_{\mathrm{T}} > 15$ GeV leptons are considered in this analysis.

Table 6.15: Sources of systematic uncertainty considered in the WVZ analysis. "N" denotes uncertainties affecting only the normalisation for the relevant processes, whereas "S" indicates uncertainties which are considered to affect only the shape of the normalised distributions. "SN" denotes uncertainties affecting both, shape and normalisation. Some of the systematic uncertainties are split into several uncorrelated components, as indicated by the number in the central column.

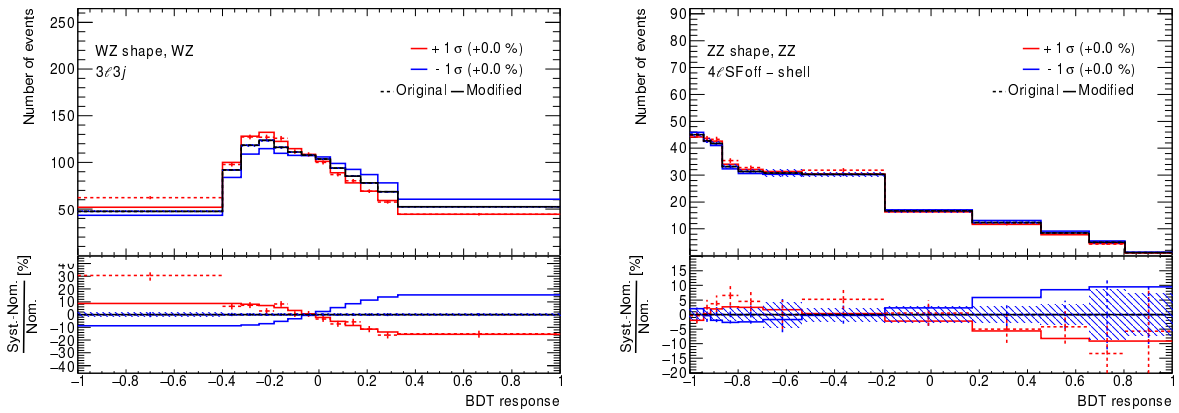
Systematic uncertainty	Components	Type
Integrated luminosity	1	N
Pile-up modelling	1	SN
Physics objects		SN
Electron	6	SN
Muon	15	SN
Jet energy scale and resolution	32	SN
Jet vertex fraction	1	SN
Jet flavour tagging	17	SN
$E_{\mathrm{T}}^{\mathrm{miss}}$	4	SN
Total (Experimental)	77	
Diboson modelling		
Cross section	2	N
NLO generator	2	S
Scale and PDF (acceptance)	6	SN
Modelling of other background processes		
$t\bar{t}Z$ scale and PDF	2	SN
WtZ cross section	1	N
tZ cross section	1	N
Z + jets cross section	1	N
VH (background) cross section	1	N
others cross section	8	N
Total (Background modelling)	24	
Signal modelling		
Scale and PDF	3	SN
Total (Signal modelling)	3	
Total (Overall)	104	

6.7.2 Background cross-section systematic uncertainties

A nuisance parameter corresponding to a normalisation variation is introduced for each of the relevant background processes. It is implemented as a symmetric variation of the cross-section prediction. For WZ and ZZ, two independent normalisation systematic uncertainties with a 20% 1σ prior are assumed or, alternatively, their normalisation is allowed to float freely.[10] As will be shown in Section 6.8, the fit results are independent of this choice and of the prior choice, since the fit model has the power to constrain the diboson normalisations at the 5% level. The Z + jets/γ normalisation is varied with a 40% prior; this is a conservative systematic variation, as proven by the background modelling studies presented in Section 6.4.2. An uncertainty prior of 10% is assigned to the WtZ background cross section, resulting from different prescriptions, diagram removal and diagram subtraction [165], for removing the WtZ interference with the $t\bar{t}$ and $t\bar{t}Z$ processes. For the tZ background, an overall normalisation uncertainty of 30% is assigned, motivated by the measurements of this process presented in Refs. [194, 195]. Uncertainties of 20%, based on calculations from Ref. [26], are assigned to the normalisation of the fraction of VH events, not belonging to the signal definition. All remaining background processes are granted a 50% prior normalisation uncertainty for their cross-section prediction.

6.7.3 Diboson generator modelling uncertainties

The modelling of diboson processes relies on the predictions provided by the Sherpa generator. Since WZ and ZZ are the largest backgrounds in the 3ℓ and 4ℓ channels, respectively, a dedicated modelling systematic uncertainty is introduced to account for differences in the diboson generator modelling. Predictions from the alternative diboson samples introduced in Section 6.1, i.e. Powheg v2.0+Pythia 8, are compared to the nominal simulation to build a shape-only variation in the fitted MVA distributions. For this purpose, the Powheg v2.0+Pythia 8 yield prediction is corrected to match the Sherpa predictions. This procedure allows to avoid a double counting of uncertainties in the normalisation, already described in Section 6.7.2, of the WZ and ZZ components in the fit. Figure 6.22 shows the WZ (ZZ) generator-modelling systematic variations in the $3\ell 1j$ (4ℓ off-shell SF) signal region.



(a) (b)

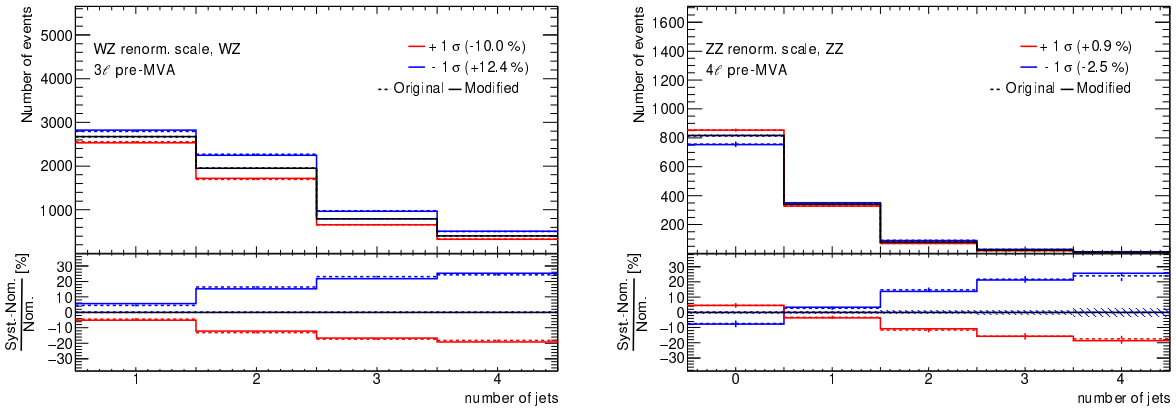
Figure 6.22: Shape-only systematic variations from Powheg v2.0+Pythia 8 in the predicted MVA distributions for (a) WZ in the $3\ell 3j$ and (b) ZZ in the 4ℓ off-shell SF regions. The $\pm 1\sigma$ variations shown within brackets represent the normalisation variation in the inclusive region.

[10] The normalisation is distributed according to a uniform distribution.

6.7.4 Scale and PDF systematic uncertainties

Signal modelling uncertainties are treated as correlated among the different signal processes and fitted regions. Both, cross-section and acceptance variations due to the scale and PDF choice are considered. Two independent variations are considered in order to account for renormalisation (μ_R, "renorm. scale") and factorisation (μ_F, "fact. scale") scale uncertainties. Both scale variations are evaluated as three-point variations: $\mu_{R,F}$ are varied by factors 2.0 (up) and 0.5 (down).

For the main background processes, i.e. WZ, ZZ and $t\bar{t}Z$, scale and PDF variations are also considered as two independent three-point variations. These variations are treated as acceptance-only for the diboson processes; this means that the cross section is renormalised to the nominal prediction, in order not to double count systematic uncertainties in the VV normalisation (see Section 6.7.2). For the $t\bar{t}Z$ processes, the scale and PDF systematic uncertainties account for variations both in the cross section and the acceptance. All of the aforementioned scale and PDF systematic uncertainties introduce variations in the predicted MVA shapes as well. Renormalisation scale variations affecting the WZ (ZZ) jet multiplicity in the 3ℓ (4ℓ) pre-MVA region are shown in Figure 6.23. Interestingly, these acceptance (and shape) variations are of the same order of magnitude of, and thus covering, the deviations discussed in Section 6.5.



(a) (b)

Figure 6.23: Renormalisation scale variations in the predicted jet-multiplicity distribution for (a) WZ in the 3ℓ pre-MVA and (b) ZZ in the 4ℓ pre-selection regions. The $\pm 1\sigma$ variations shown within brackets represent the normalisation variation in the inclusive region.

6.8 Fit model and expected results

This section illustrates the expected fit results in the 3ℓ and 4ℓ channels, as well as the WVZ overall combination; BDT shapes are fitted to the Asimov dataset. The signal strength is defined as:

$$\mu_{WVZ} = \frac{\sigma_{WVZ}}{\sigma_{WVZ}^{\text{SM}}}, \tag{6.2}$$

where σ_{WVZ} corresponds to the measured cross section for the signal WVZ and σ_{WVZ}^{SM} is the corresponding prediction from SM. The relevant part of the VH processes, $WH(\rightarrow WW^*, ZZ^*)$ and $ZH(\rightarrow WW^*)$, is considered as part of the signal definition and scaled by the signal strength in Eq. 6.2. Therefore the production of on-shell $W^{\pm}W^{\mp}Z$ and $W^{\pm}ZZ$, and Higgs-mediated WVV^* are scaled by the same signal strength. For the WVZ combination a profile likelihood fit is performed, combining all six signal regions and the $t\bar{t}Z$ control region in order to extract the signal strength, μ_{WVZ}, defined in Eq. 6.2. The fit exploits a dedicated BDT shape as discriminant in each of the six signal regions and yields (a single bin) in the $t\bar{t}Z$ control region (see definition in Section 6.4.1). The inclusion of this region allows to constrain uncertainties on the $t\bar{t}Z$ normalisation, a major systematic uncertainty in the four-lepton channel. The total number of bins used in the fit is 65. The jet multiplicity-based reweighting procedure, outlined in Section 6.5, is applied to the WZ and ZZ pre-fit predictions. The systematic model presented in Section 6.7 is adopted to perform all of the fits presented in the following.

Asimov fit in the three-lepton channel

The Asimov fit in the 3ℓ channel involves the shape fit of the three BDT distributions, whose training is discussed in Section 6.6.1. Pre-fit distributions are shown in Figure 6.24. The Asimov fit results in a signal strength of $\mu_{WVZ} = 1^{+1.03}_{-1.01}$ and an expected median significance for the rejection of the background-only hypothesis of 0.99σ (2.0σ excluding systematic uncertainties). The three-lepton fit model allows to constrain the WZ shape, scale and cross-section variations; this is expected, given the large amount of data in the 3ℓ region and the magnitude of these systematic uncertainties (see Sections 6.7.2–6.7.4). The top-ranked nuisance parameters, according to their post-fit impact on the central value of μ_{WVZ}, are the WZ renormalisation-scale, shape and cross-section variations.

Asimov fit in the four-lepton channel

The fit in the 4ℓ channel constitutes a profile likelihood shape fit in the three 4ℓ regions: the DF, the on-shell and the off-shell SF regions; the MVA, whose training is illustrated in Section 6.6.2, distribution is fitted in each region. Pre-fit distributions are shown in Figure 6.24. The Asimov fit yields a signal strength of $\mu_{WVZ} = 1^{+0.71}_{-0.61}$ and an expected median significance for the rejection of the background-only hypothesis of 1.75σ (1.9σ without systematic uncertainties). The four-lepton fit allows to constrain systematic uncertainties in the ZZ cross section and scale-variation acceptance. The top-ranked nuisance parameters, according to their post-fit impact on the central value of μ_{WVZ}, are the ZZ renormalisation scale and the variations in the $t\bar{t}Z$ and WtZ cross sections.

Table 6.16 shows an overview of the expected sensitivity in each 3ℓ and 4ℓ region separately.

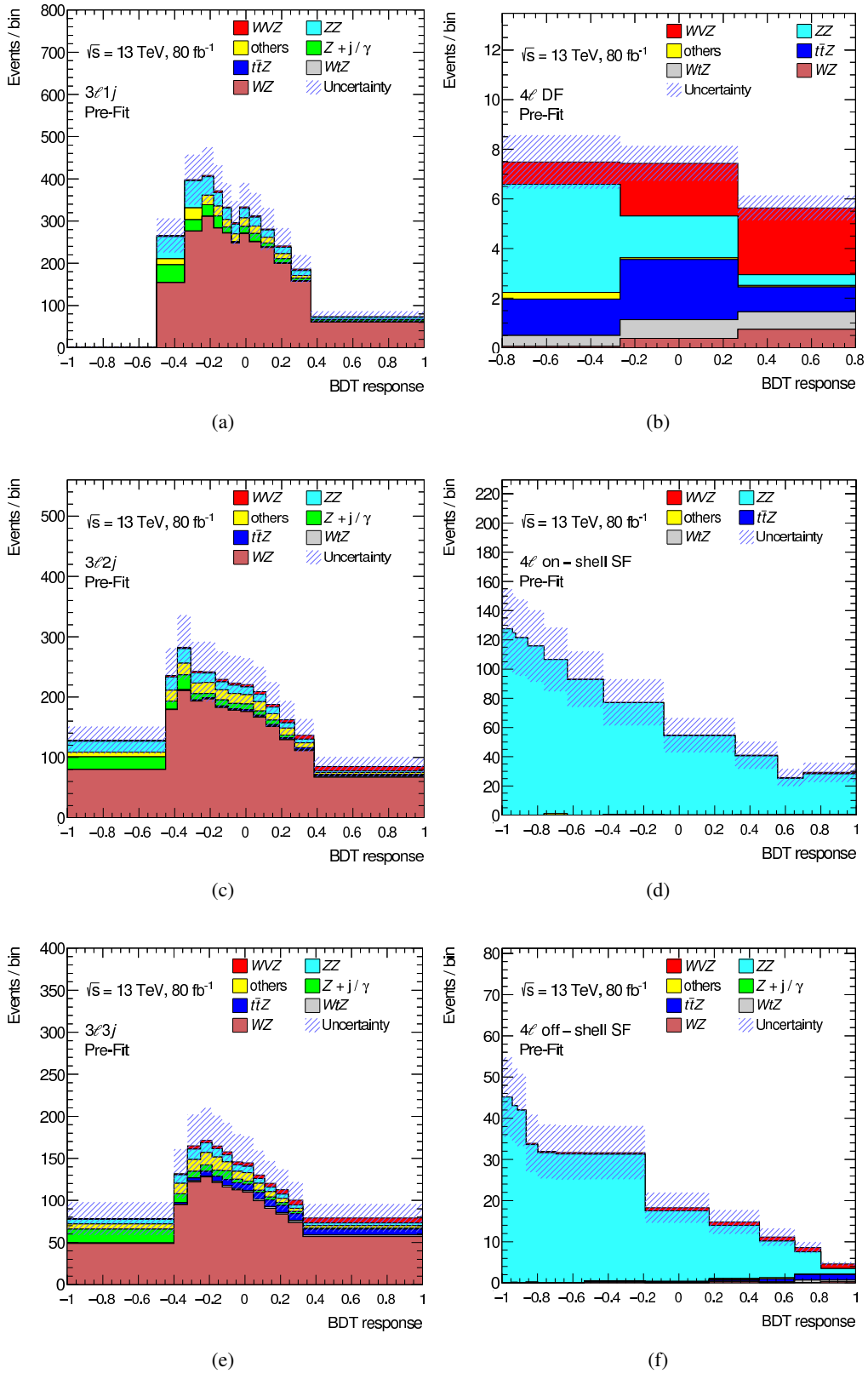


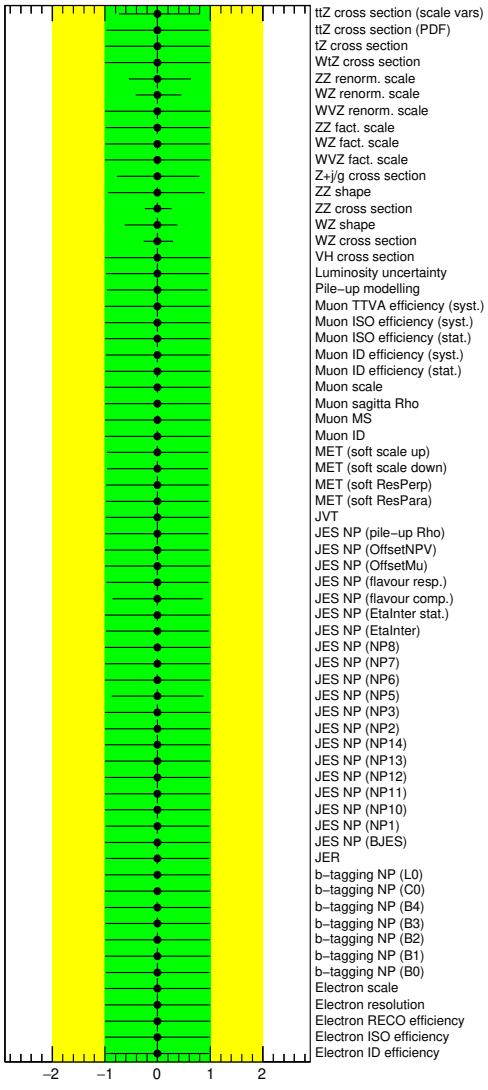
Figure 6.24: Pre-fit distributions in the (a) $3\ell 1j$, (b) 4ℓ DF, (c) $3\ell 2j$, (d) 4ℓ on-shell SF, (e) $3\ell 3j$ and (f) 4ℓ off-shell SF regions. Both, statistical and systematic uncertainties are shown, as shaded blue bands.

Table 6.16: Expected sensitivity from Asimov fits in each 3ℓ and 4ℓ region, and the 3ℓ and 4ℓ combinations. The corresponding expected median significances ("exp. median sig.") are shown, as well as the statistical-only fit results for the 3ℓ and 4ℓ combinations, included within brackets in the rightmost column.

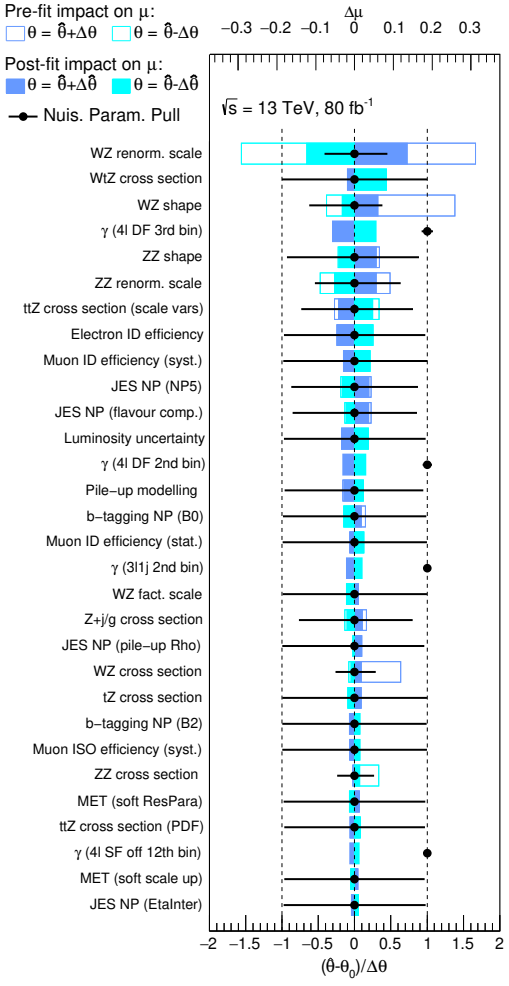
3ℓ region	$3\ell 1j$	$3\ell 2j$	$3\ell 3j$	3ℓ comb.
σ_μ	$^{+4.9}_{-4.7}$	$^{+1.5}_{-1.4}$	$^{+2.4}_{-1.9}$	$^{+1.03}_{-1.01}$ $\left(^{+0.51}_{-0.50}\right)$
exp. median sig.	0.22σ	0.69σ	0.51σ	0.99σ (2.0σ)
4ℓ region	DF	SF on-shell	SF off-shell	4ℓ comb.
σ_μ	$^{+0.82}_{-0.67}$	$^{+6.6}_{-6.6}$	$^{+1.6}_{-1.4}$	$^{+0.71}_{-0.61}$ $\left(^{+0.66}_{-0.57}\right)$
exp. median sig.	1.59σ	0.16σ	0.69σ	1.75σ (1.9σ)

6.8.1 Signal regions before the fit and expected fit results

The full Asimov likelihood fit combination of the shapes shown in Figure 6.24 and the yields reported in Table 6.8 is performed to evaluate the expected signal sensitivity. The WVZ combined fit yields $\mu_{WVZ} = 1^{+0.59}_{-0.53} = 1^{+0.39}_{-0.38}\,(\text{stat.})\,^{+0.45}_{-0.37}\,(\text{syst.})$, corresponding to an expected median significance for the rejection of the null hypothesis of 2.0σ (2.8σ excluding systematic uncertainties). Figure 6.25 shows the resulting nuisance parameters constraints and ranking, according to their post-fit impact on the μ_{WVZ} central value. The fit combination allows to constrain the WZ and ZZ normalisations, acceptance and shape systematic uncertainties, as a natural result of the combination of the 3ℓ and 4ℓ channels. Figure 6.26 shows the expected best-fit signal strength from all of the six single-region fits and from their statistical combination, together with the $t\bar{t}Z \to 3\ell$ control region.



(a)



(b)

Figure 6.25: The figure shows (a) the constraints on the nuisance parameters and (b) the corresponding ranking in terms of their post-fit impact on μ_{WVZ} for the combined Asimov fit.

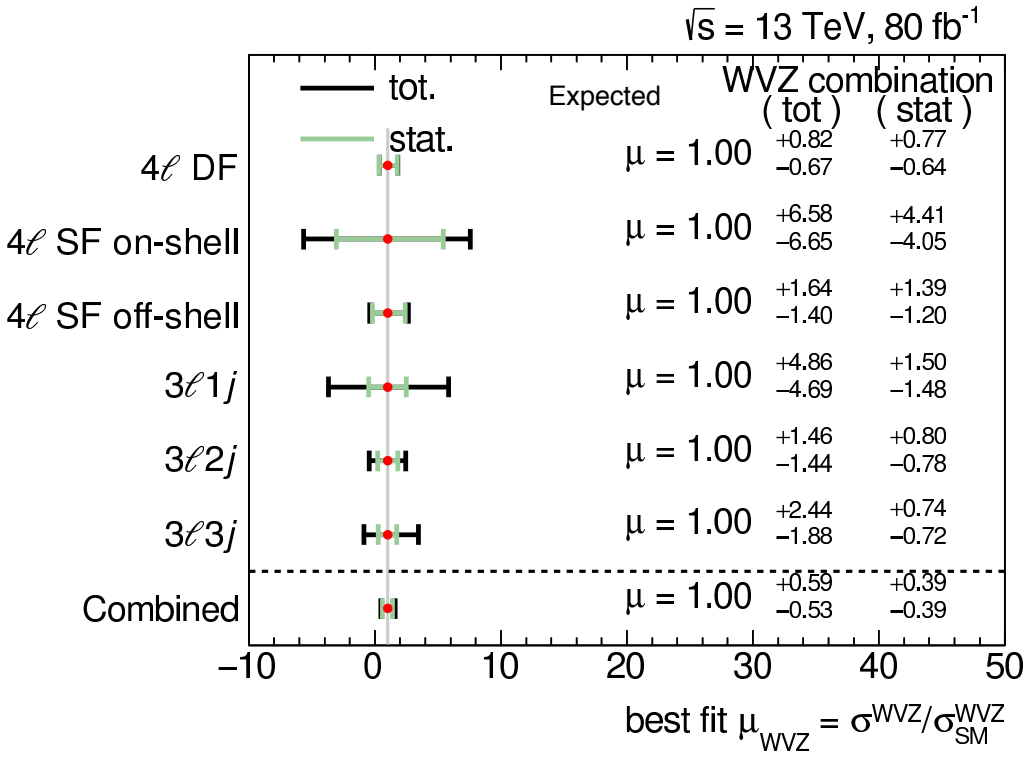


Figure 6.26: The figure shows the Asimov best-fit signal strength from the fit of each single signal region and their combination, together with the $t\bar{t}Z \to 3\ell$ control region. Statistical-only and statistical-plus-systematic fit results are shown for both, the single-region and the combined fits.

6.9 Unblinding and fit to the observed dataset

The observed fit results under the signal-plus-background hypothesis are presented in this section for the individual channels as well as for the WVZ combination.

6.9.1 Pre-fit plots and yields

Table 6.17 and Figure 6.27 show a comparison of the pre-fit background and signal yields to data in the six signal and one control regions. Figure 6.28 shows the pre-fit distributions in the six signal regions fed to the profile likelihood fit.

Table 6.17: Pre-fit background, signal and observed yields in the seven analysis regions with 80 fb^{-1} of data at $\sqrt{s} = 13$ TeV. Uncertainties on the background expectations due to systematic effects and limited amount of simulated events are shown.

	4ℓ DF	4ℓ SF on	4ℓ SF off	$3\ell 1j$	$3\ell 2j$	$3\ell 3j$	$t\bar{t}Z \to 3\ell$ CR
WVZ	5.71 ± 0.17	2.96 ± 0.10	6.23 ± 0.17	37.3 ± 1.7	50.9 ± 1.3	49.6 ± 3.6	0.493 ± 0.032
ZZ	6.5 ± 1.4	910 ± 190	302 ± 61	335 ± 68	177 ± 54	96 ± 26	0.876 ± 0.057
WZ	1.18 ± 0.13	–	1.09 ± 0.14	$2\,710 \pm 560$	$2\,020 \pm 490$	$1\,260 \pm 360$	6.14 ± 0.38
Z + jets	–	–	0.17 ± 0.12	161 ± 70	93 ± 38	71 ± 31	0.46 ± 0.18
$Z + \gamma$	–	–	–	42 ± 19	32 ± 14	9.7 ± 4.2	0.40 ± 0.56
$t\bar{t}Z$	4.89 ± 0.63	0.527 ± 0.088	4.28 ± 0.57	7.2 ± 1.4	21.4 ± 3.3	78.0 ± 9.6	115 ± 14
tZ	0.090 ± 0.029	–	–	24 ± 12	24 ± 12	12.0 ± 6.0	8.52 ± 0.60
tWZ	1.89 ± 0.42	0.22 ± 0.10	1.57 ± 0.37	4.14 ± 0.94	11.2 ± 2.3	19.6 ± 3.9	10.19 ± 0.79
others	0.30 ± 0.18	1.80 ± 0.94	1.24 ± 0.67	173 ± 88	156 ± 80	112 ± 56	15.6 ± 8.3
Total	20.6 ± 1.8	920 ± 190	317 ± 62	$3\,500 \pm 570$	$2\,580 \pm 500$	$1\,700 \pm 370$	157 ± 15
data	28	912	360	3 351	2 438	1 572	170

6.9.2 Best-fit μ_{WVZ} and significance

The observed best-fit value of μ_{WVZ} is $1.48^{+0.60}_{-0.57} = 1.48^{+0.50}_{-0.47}$ (stat.) $^{+0.33}_{-0.32}$ (syst.), where all seven regions are simultaneously fitted. Figure 6.29 shows the best-fit value of μ_{WVZ} for each individual channel and the combination of all channels, including the $t\bar{t}Z \to 3\ell$ single-bin control region. Observed and expected fit results are also shown in Table 6.18, together with the corresponding significances. The observed significance with respect to the no-WVZ hypothesis is 2.9σ. The observed and expected exclusion limits are obtained on the signal-plus-background hypothesis at 95% CLs. The observed (expected) exclusion limit is $\mu_{\text{obs.}} < 2.5$ ($\mu_{\text{exp.}} < 1.0$).

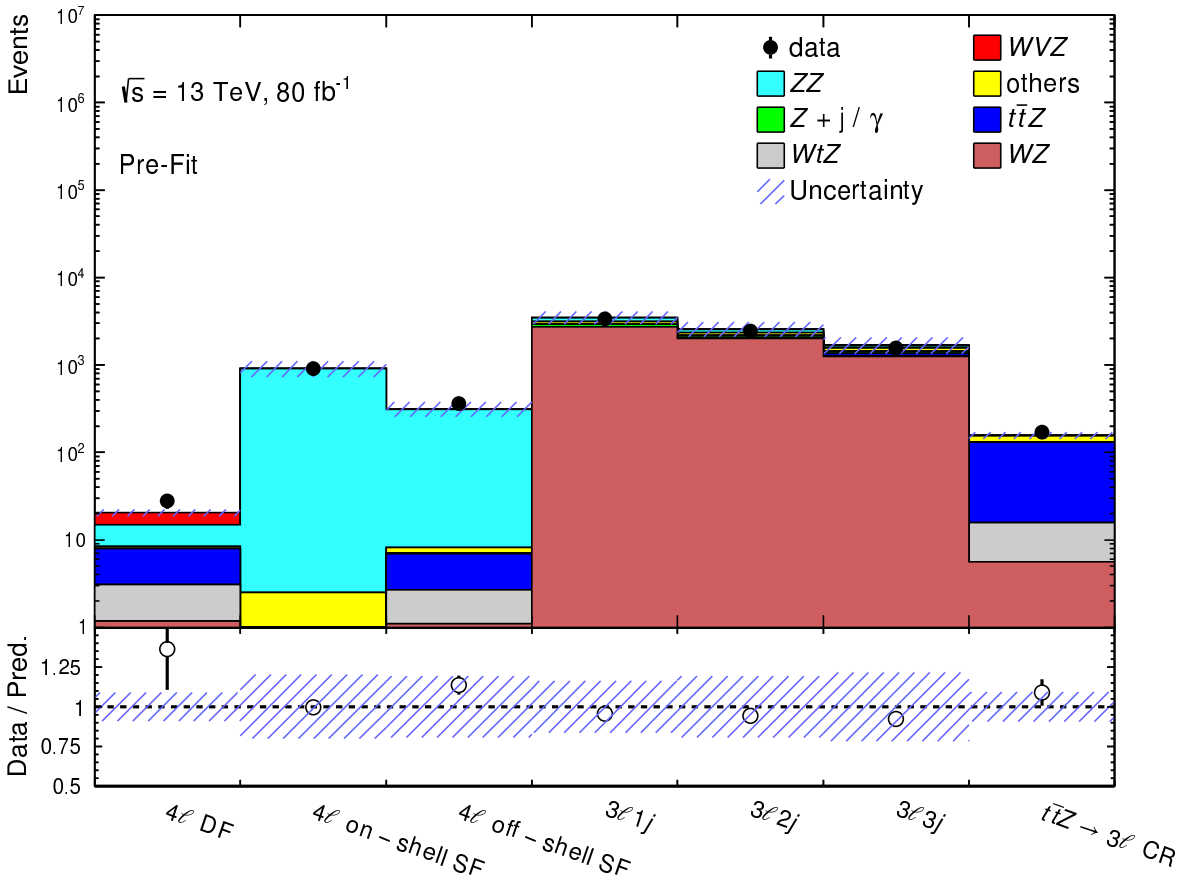


Figure 6.27: Comparison between prediction and observed data before the fit in the six signal and one control regions entering the profile likelihood fit. The lower panel shows the data-to-simulation ratio. The systematic uncertainties on the predicted yields are indicated by the shaded blue bands.

Table 6.18: Observed and expected best-fit values of the signal strength μ_{WVZ} and associated significance with respect to the SM background-only hypothesis. The observed significance is indicated with a – for the channels where μ_{WVZ} is negative.

Channel	Obs. best-fit μ	Exp. best-fit μ	Obs. significance	Exp. significance
4ℓ DF	$1.84^{+1.00}_{-0.84}$	$1^{+0.82}_{-0.67}$	2.50σ	1.59σ
4ℓ on-shell SF	$-3.2^{+6.6}_{-6.7}$	1 ± 6.6	–	0.16σ
4ℓ off-shell SF	$3.9^{+2.1}_{-1.9}$	$1^{+1.6}_{-1.4}$	2.23σ	0.69σ
$3\ell 1j$	$-2.9^{+4.7}_{-4.6}$	$1^{+4.9}_{-4.7}$	–	0.22σ
$3\ell 2j$	-0.3 ± 1.4	$1^{+1.5}_{-1.4}$	–	0.69σ
$3\ell 3j$	$2.5^{+2.0}_{-2.2}$	$1^{+2.4}_{-1.9}$	1.13σ	0.51σ
Combined	$1.48^{+0.60}_{-0.57}$	$1^{+0.59}_{-0.53}$	2.86σ	2.00σ

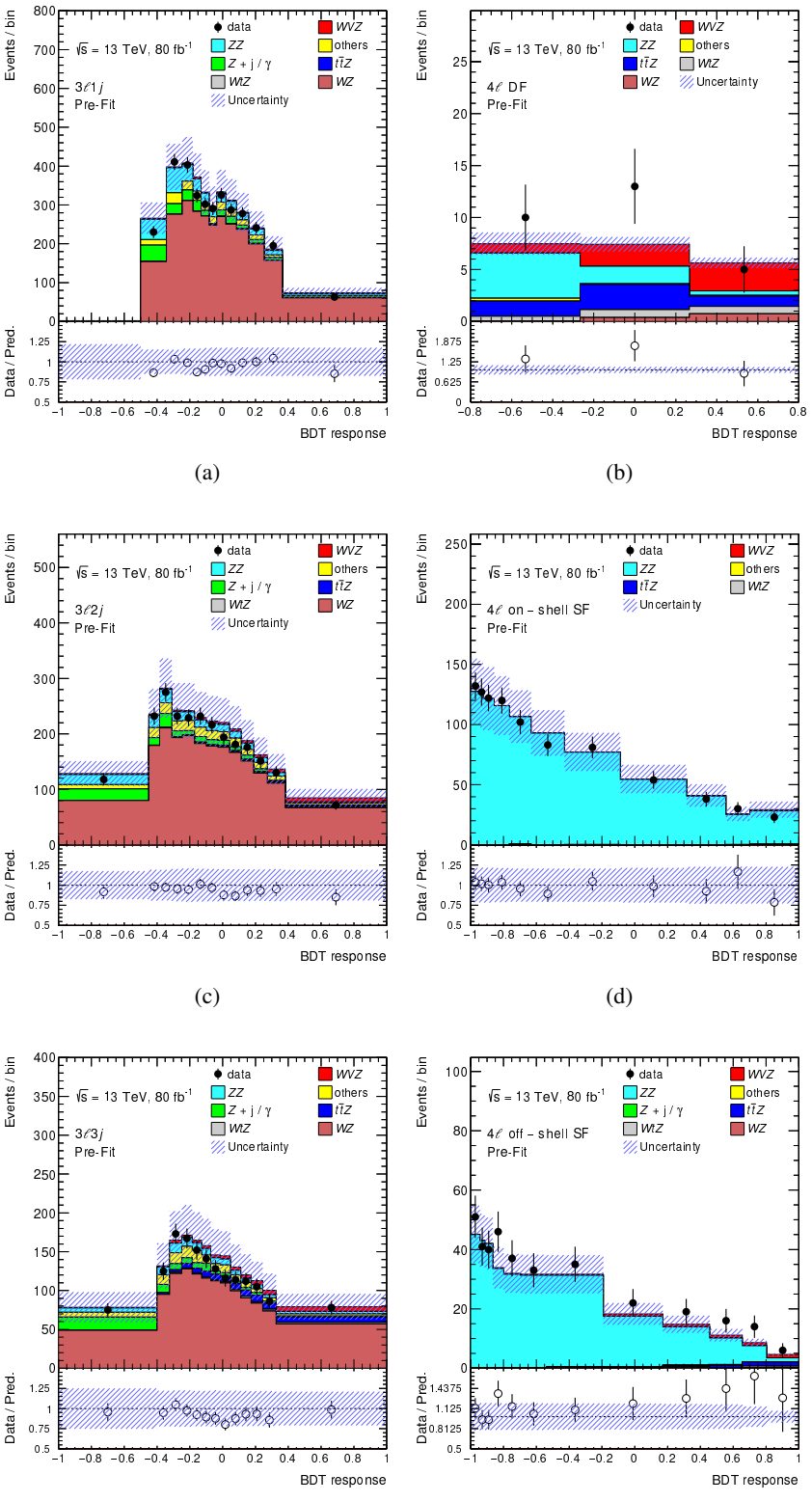


Figure 6.28: Pre-fit distributions in the (a) $3\ell 1j$, (b) 4ℓ DF, (c) $3\ell 2j$, (d) 4ℓ on-shell SF, (e) $3\ell 3j$ and (f) 4ℓ off-shell SF regions exploited by the combined fit to data. The lower panel shows the data-to-simulation ratio. Both, statistical and systematic uncertainties are shown, as shaded blue bands.

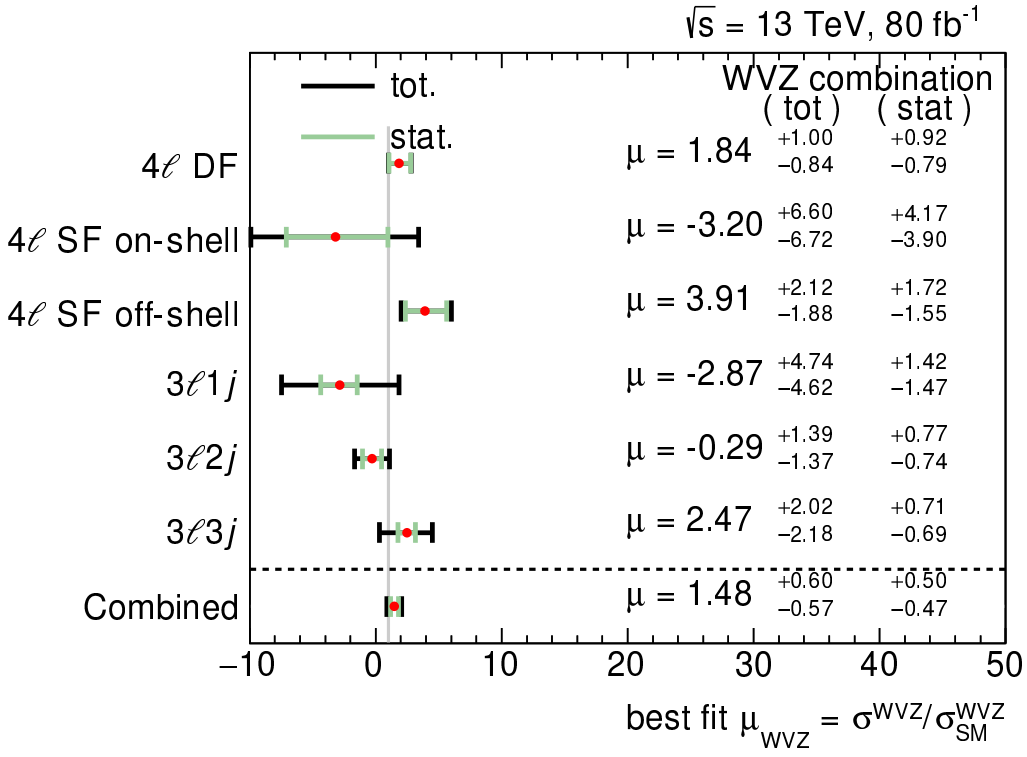


Figure 6.29: The figure shows the observed best-fit signal strength from the fit of each single signal region and their combination, together with the $t\bar{t}Z \rightarrow 3\ell$ control region. Statistical-only and statistical-plus-systematic fit results are shown for both, the single-region and the combined fits.

6.9.3 Post-fit yields and distributions

Table 6.19 and Figure 6.30 show a comparison of the post-fit background and signal yields to data in the six signal and one control regions. Figure 6.31 shows the post-fit distributions in the six signal regions, inputs to the profile likelihood fit. Post-fit distributions show a good agreement between simulation and observation.

Table 6.19: Post-fit background, signal and observed yields in the seven analysis regions with 80 fb^{-1} of data at $\sqrt{s} = 13$ TeV. Uncertainties on the background expectations due to systematic effects and limited amount of simulated events are shown.

	4ℓ DF	4ℓ SF on	4ℓ SF off	$3\ell 1j$	$3\ell 2j$	$3\ell 3j$	$t\bar{t}Z \to 3\ell$ CR
WVZ	8.4 ± 2.6	4.4 ± 1.7	9.3 ± 2.6	55 ± 14	75 ± 21	74 ± 26	0.753 ± 0.032
ZZ	6.69 ± 0.45	927 ± 28	319.3 ± 9.9	339 ± 12	180 ± 23	92 ± 12	0.876 ± 0.056
WZ	1.21 ± 0.13	–	1.12 ± 0.14	$2\,530 \pm 81$	$1\,765 \pm 61$	$1\,056 \pm 48$	5.70 ± 0.36
Z + jets	–	–	0.18 ± 0.12	139 ± 57	80 ± 30	61 ± 25	0.47 ± 0.18
$Z + \gamma$	–	–	–	42.2 ± 8.5	32.5 ± 7.0	9.6 ± 3.5	0.40 ± 0.56
$t\bar{t}Z$	5.14 ± 0.51	0.554 ± 0.077	4.53 ± 0.46	7.7 ± 1.0	22.7 ± 2.5	82.1 ± 7.5	121.7 ± 9.2
tZ	0.091 ± 0.029	–	–	26 ± 13	27 ± 13	13.3 ± 6.4	8.56 ± 0.57
tWZ	1.91 ± 0.42	0.23 ± 0.11	1.59 ± 0.36	4.22 ± 0.90	11.3 ± 2.2	19.8 ± 3.8	10.27 ± 0.77
others	0.30 ± 0.18	1.80 ± 0.94	1.24 ± 0.67	173 ± 88	156 ± 80	112 ± 56	15.6 ± 8.3
Total	23.8 ± 3.4	942 ± 28	329 ± 10	$3\,373 \pm 68$	$2\,426 \pm 41$	$1\,585 \pm 40$	163.5 ± 9.9
data	28	912	360	3 351	2 438	1 572	170

6.9.4 Nuisance parameters

The impact of the most important groups of systematic uncertainties on the measured value of μ_{WVZ} is shown in Table 6.20. The systematic uncertainties with the largest impact are "Theory", "Instrumental" and "Generators"; the first group gathers PDF, scale and normalisation variations, the second detector uncertainties and the third accounts for diboson shape variations from Powheg v2.0+Pythia 8. The small difference between the sum in quadrature of the individual groups and the total uncertainty is due to rounding effects and small correlations between the individual groups. Figure 6.32 shows both, the nuisance parameter pulls and constraints and the ranking of nuisance parameters for the combined fit. No strong (above 1σ) pull is observed.

Table 6.20: Summary of the effects of the most important groups of systematic uncertainties in μ. Due to rounding effects and correlations between the different sources of uncertainties, the total systematic uncertainty can be different from the sum in quadrature of the individual sources.

Uncertainty source	$\Delta\mu$	
Theory	+0.21	− 0.18
Instrumental	+0.15	− 0.15
Generators	+0.11	− 0.10
Gammas	+0.09	− 0.08
Total systematic uncertainty	+0.33	− 0.32

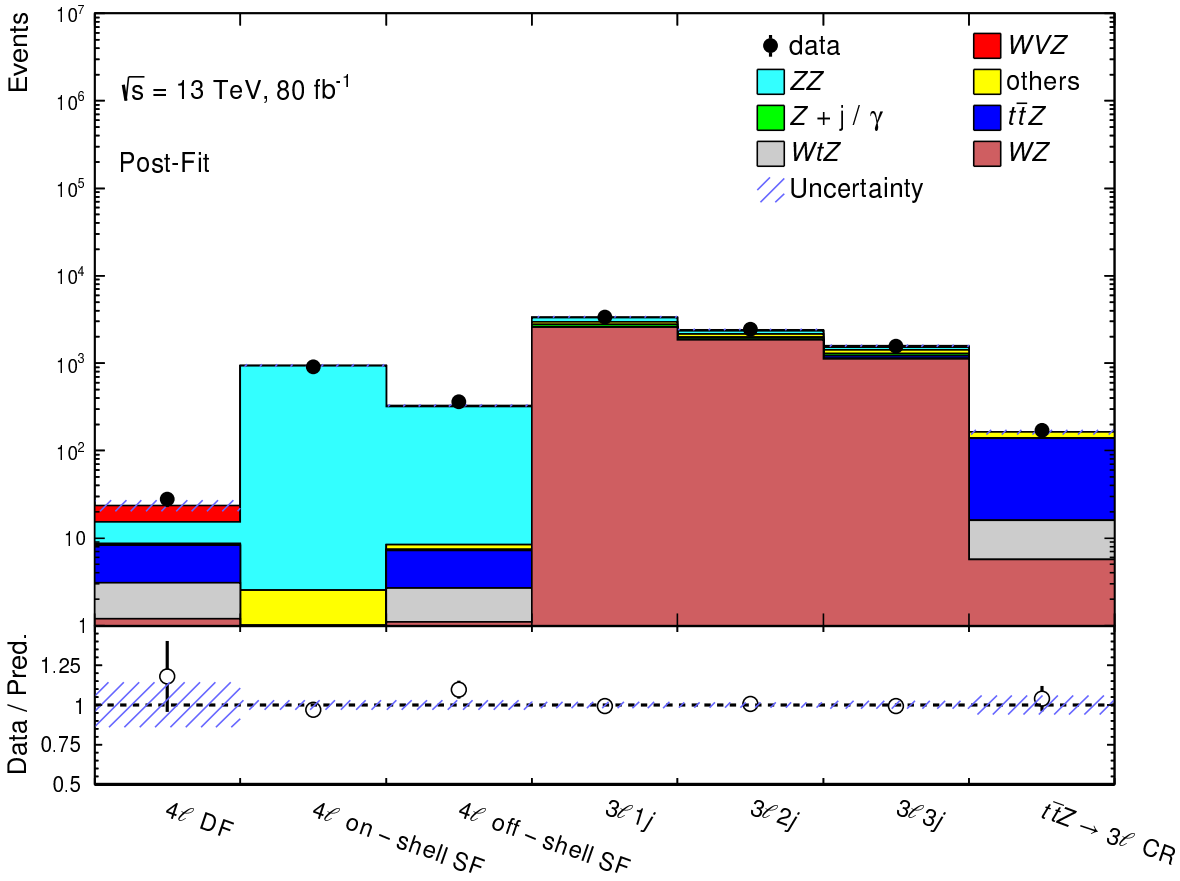


Figure 6.30: Comparison between prediction and observed data after the fit in the six signal and one control regions entering the profile likelihood fit. The lower panel shows the data-to-simulation ratio. The systematic uncertainties on the predicted yields are indicated by the shaded blue bands.

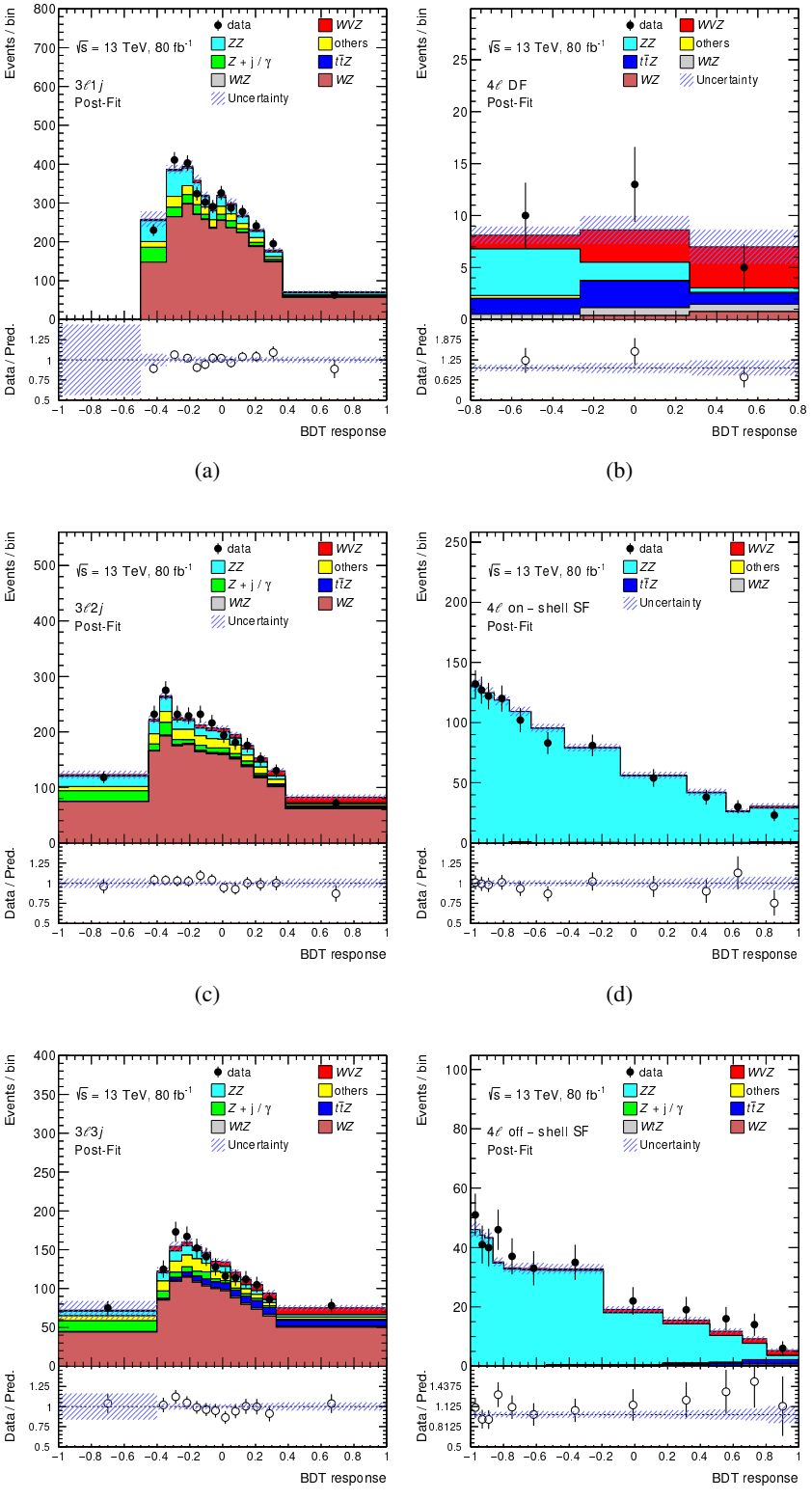
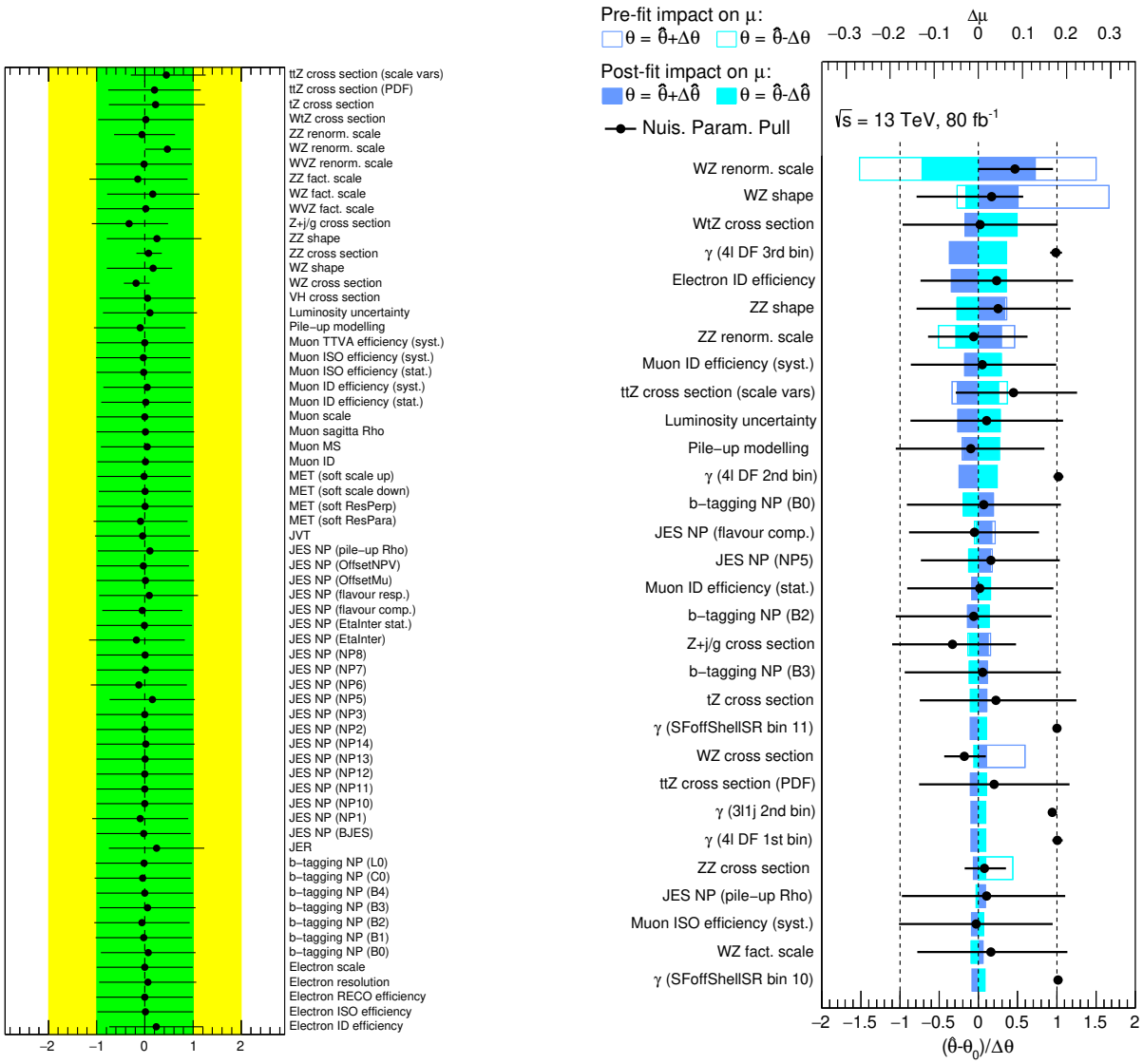


Figure 6.31: Post-fit distributions in the (a) $3\ell 1j$, (b) 4ℓ DF, (c) $3\ell 2j$, (d) 4ℓ on-shell SF, (e) $3\ell 3j$ and (f) 4ℓ off-shell SF regions corresponding to the combined fit to data in 3ℓ and 4ℓ regions. The lower panel shows the data-to-simulation ratio. Both, statistical and systematic uncertainties are shown, as shaded blue bands.



(a) (b)

Figure 6.32: The figure shows (a) the constraints and pulls on the nuisance parameters, and (b) the corresponding ranking in terms of their post-fit impact on μ_{WVZ} for the combined fit to data.

6.10 Combination with $W^{\pm}W^{\pm}W^{\mp}$: evidence for the production of VVV

This section is devoted to the statistical combination of the WVZ analysis, discussed in the previous sections of this chapter, with the $W^{\pm}W^{\pm}W^{\mp}$ analysis, also carried out at $\sqrt{s} = 13\,\text{TeV}$ and with the same $80\,\text{fb}^{-1}$ of data collected by the ATLAS detector. A set of Feynman diagrams very similar to those shown in Figures 6.1 and 6.2 dominates the production of the $W^{\pm}W^{\pm}W^{\mp}$ processes.

6.10.1 The $W^{\pm}W^{\pm}W^{\mp}$ analysis

The tight lepton definition, very similar to the one described in Section 6.3, is exploited to suppress fake and non-prompt contributions in the $W^{\pm}W^{\pm}W^{\mp}$ two same-sign-lepton ($W^{\pm}W^{\pm}W^{\mp} \to \ell^{\pm}\nu\ell^{\pm}\nu jj$) and three-lepton ($W^{\pm}W^{\pm}W^{\mp} \to \ell^{\pm}\nu\ell^{\pm}\nu\ell^{\mp}\nu$) channels. These two channels are hereafter referred to as $\ell\nu\ell\nu jj$ and $\ell\nu\ell\nu\ell\nu$, respectively. The experimental signature of the $\ell\nu\ell\nu jj$ channel is the presence of two same-sign leptons, $E_{\text{T}}^{\text{miss}}$ and two jets with an invariant mass close to the W-boson mass. The signature of the $\ell\nu\ell\nu\ell\nu$ channel is the presence of three leptons and significant $E_{\text{T}}^{\text{miss}}$. Electrons and muons are required to fulfil the tight lepton definitions provided in Table 6.3 and, additionally, to have $p_{\text{T}} > 20\,\text{GeV}$ (as compared to 15 GeV in the WVZ channels). For the veto of b-jets a different MV2c10 working point, namely 85% b-jet efficiency (as compared to 70% in the WVZ analyisis), is used, as it allows to better reject $t\bar{t}$ events in both, the $\ell\nu\ell\nu jj$ and $\ell\nu\ell\nu\ell\nu$ channels.

Selected $\ell\nu\ell\nu jj$ candidate events are required to have exactly two tight leptons with the same electric charge, at least two jets and no identified b-jets. The leading lepton is required to fulfil $p_{\text{T}} > 27$ GeV. The $\ell\nu\ell\nu jj$ channel is split into four signal regions, based on the lepton flavour: ee, $e\mu$, μe, and $\mu\mu$. $e\mu$ (μe) indicates events, where the leading lepton is an electron (a muon). The invariant mass of the dilepton system is required to fulfil $40 < m_{\ell\ell} < 400\,\text{GeV}$ in all four final states and, in the ee channel, it must also fulfil $|m_{ee} - 90\,\text{GeV}| > 10\,\text{GeV}$. The leading (sub-leading) jet must have $p_{\text{T}} > 30\,(20)$ GeV. The dijet system is required to have $m_{jj} < 300\,\text{GeV}$ and $|\Delta\eta_{jj}| < 1.5$, where m_{jj} is the dijet invariant mass and $\Delta\eta_{jj}$ is the pseudorapidity separation between the two jets. $E_{\text{T}}^{\text{miss}}$ lower threshold is set to 55 GeV only in the ee final state, since a smaller contamination from the Z + jets background is expected in the other three regions.

Selected $\ell\nu\ell\nu\ell\nu$ candidate events are required to have exactly three tight leptons and no identified b-jets. The leading lepton is also required to have $p_{\text{T}} > 27$ GeV. To reduce the contribution from the WZ processes, events are required to have zero same-flavour $\ell^{+}\ell^{-}$ lepton pairs. This requirement results in two possible final states: $\mu^{\pm}e^{\mp}e^{\mp}$ and $e^{\pm}\mu^{\mp}\mu^{\mp}$. The WVZ and $W^{\pm}W^{\pm}W^{\mp}$ signal regions are not overlapping.

A WZ-dominated control region is defined by selecting events with exactly three tight leptons and at least one same-flavour $\ell^{+}\ell^{-}$ lepton pair. Events in this region are also required to have zero b-jets, $E_{\text{T}}^{\text{miss}} > 55\,\text{GeV}$ and the three-lepton invariant mass satisfying $m_{3\ell} > 110\,\text{GeV}$. This control region is used for the standalone $W^{\pm}W^{\pm}W^{\mp}$ measurement, but is fully overlapping with the WVZ three-lepton channel and is, therefore, not included in the VVV combination discussed in Section 6.10.2.

The $W^{\pm}W^{\pm}W^{\mp}$ analysis is affected by both, the *charge-flip* and non-prompt/fake processes. The charge-flip background originates from processes where the charge of at least one prompt electron (the muon charge-flip rate is negligible) is misidentified. The estimation of the charge-flip background relies on the measurement of the charge-flip rate as a function of the electron p_{T} and η in a two-lepton same-sign region enriched with $Z \to e^{+}e^{-}$ events [196]. In the $\ell\nu\ell\nu jj$ channel, the charge-flip background is estimated by applying the electron charge-flip rate to data events, selected using all signal criteria, except requiring the two leptons to be opposite-sign. In the $\ell\nu\ell\nu\ell\nu$ channel, this background is estimated by using these rates to reweight the simulation prediciton of the WZ and ZZ processes, based on the probability for

opposite-sign events of this kind to migrate into the signal region, i.e. with zero same-flavour $\ell^+\ell^-$ lepton pairs. Contributions from SM processes (dominated by the $t\bar{t}$ process) that produce at least one non-prompt or fake lepton are estimated using a data-driven technique [196]. Events containing one (two) tight lepton and one non-tight lepton are scaled by a "*fake factor*" to predict the non-prompt/fake background contribution in the $\ell\nu\ell\nu jj$ ($\ell\nu\ell\nu\ell\nu$) channel. Fake factors are derived in a $t\bar{t}$-enriched three-lepton region. Uncertainties on data-driven background estimations mainly come from statistical and systematic uncertainties on the charge-flip rate and the fake factor. Experimental, signal and prompt background modelling systematic uncertainties are implemented as described in Section 6.7.

The $W^{\pm}W^{\pm}W^{\mp}$ profile likelihood fit model combines four m_{jj} distributions, one in each of the $\ell\nu\ell\nu jj$ signal regions, and the yields in the $\ell\nu\ell\nu\ell\nu$ channel. The WZ control region is divided into 4 channels (eee, $ee\mu$, $e\mu\mu$, $\mu\mu\mu$) and binned in the $m_{3\ell}$ distribution. The four $\ell\nu\ell\nu jj$, one $\ell\nu\ell\nu\ell\nu$ and four WZ regions are fitted simultaneously. The $W^{\pm}W^{\pm}W^{\mp}$ expected significance for the rejection of the null hypothesis yields 2.5σ, where the $\ell\nu\ell\nu jj$ ($\ell\nu\ell\nu\ell\nu$) channel, combined with the WZ control region, is expected to reject it at the 1.7σ (2.0) level.

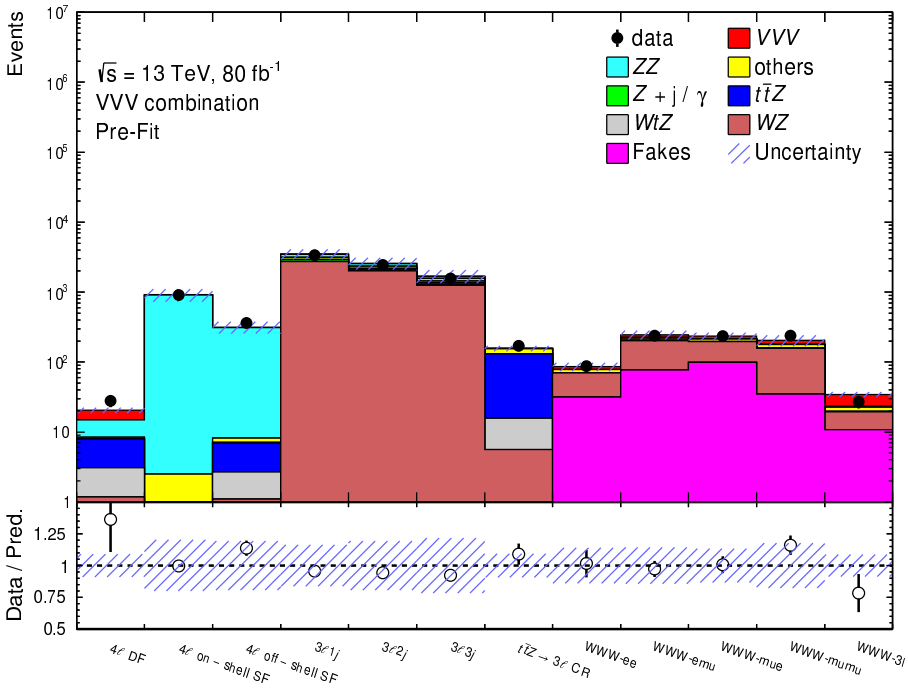
6.10.2 The VVV combination

The $\ell\nu\ell\nu jj$ and $\ell\nu\ell\nu\ell\nu$ channels, corresponding to five signal regions, are combined with the seven WVZ analysis regions already discussed in Section 6.9. All nuisance parameters, those provided in Table 6.15 and the WWW data-driven ones introduced in the previous section, are correlated among all of the fitted regions. In total 186 bins are fitted in the VVV combined fit.

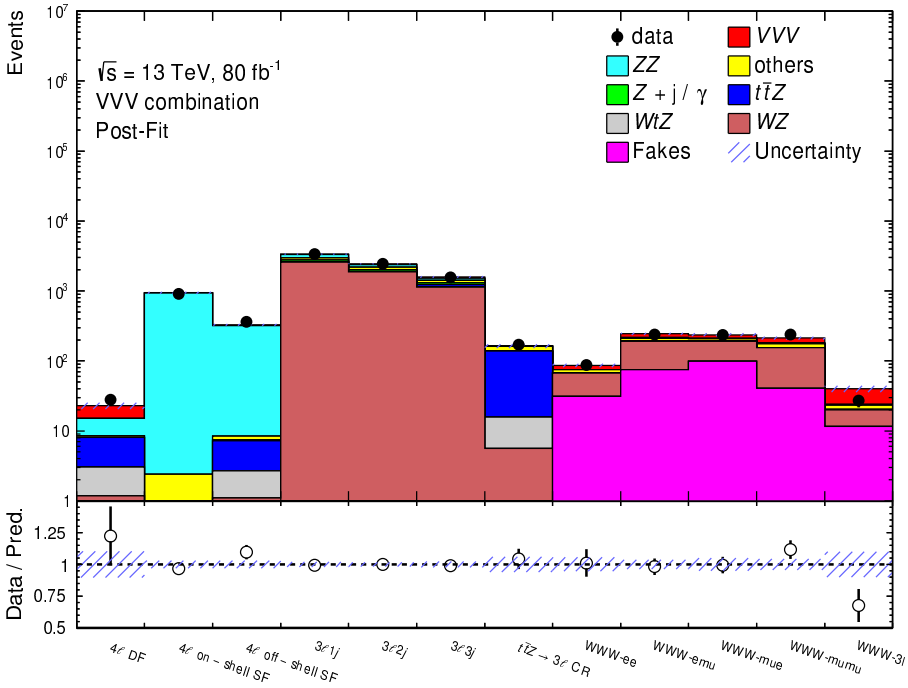
The observed (expected) best-fit signal strength, scaling both, the WVZ and WWW processes, yields $\mu_{VVV} = 1.39^{+0.38}_{-0.36} = 1.39^{+0.30}_{-0.28}$ (stat.) $^{+0.24}_{-0.23}$ (syst.) $\left(1^{+0.35}_{-0.33} = 1^{+0.28}_{-0.27}\text{ (stat.) }^{+0.21}_{-0.19}\text{ (syst.)}\right)$. The no-$VVV$ hypothesis is excluded at the 4.0σ level, with an expectation of 3.1σ in the case of SM signal processes. This constitutes first *evidence* for the production of three massive vector bosons.

Figure 6.33 shows pre- and post-fit yields in all of the twelve regions combined in the VVV fit. Figure 6.34 summarises post-fit event yields as a function of $\log_{10}$(S/B), where S (B) indicates the signal (background) expected yield in the corresponding bin, for all 186 bins of the distributions used in the combined fit. The value of $\log_{10}$(S/B) is calculated according to the post-fit yields in each bin of the fitted distributions. The total number of background and signal events is displayed in bins of $\log_{10}$(S/B). A clear accumulation of signal-like events can be observed for large values of $\log_{10}$(S/B).

Pulls and constraints on the nuisance parameters related to experimental and prompt background systematics are compatible with those already shown in Figure 6.32. Additionally, $< 1\sigma$ pulls on the WWW non-prompt/fake nuisance parameters are observed and allow predictions to fit data in regions dominated by these processes. Table 6.21 shows the impact of the most important groups of systematic uncertainties on the measured value of μ_{VVV}. The systematic uncertainties with the largest impact are "Data-driven", "Theory" and "Instrumental". The first group represents uncertainties in the data-driven background estimations affecting the WWW channels, whereas the second and third group gather scale and normalisation variations, and detector uncertainties, respectively.



(a)



(b)

Figure 6.33: Comparison between prediction and observed data, (a) before and (b) after the fit, in the twelve regions entering the combined *VVV* profile likelihood fit. Contributions due to charge-flip and fake/non-prompt processes entering the *WWW* channels are labelled as "Fakes". The lower panel shows the data-to-simulation ratio. The overall uncertainties are indicated by the shaded blue bands.

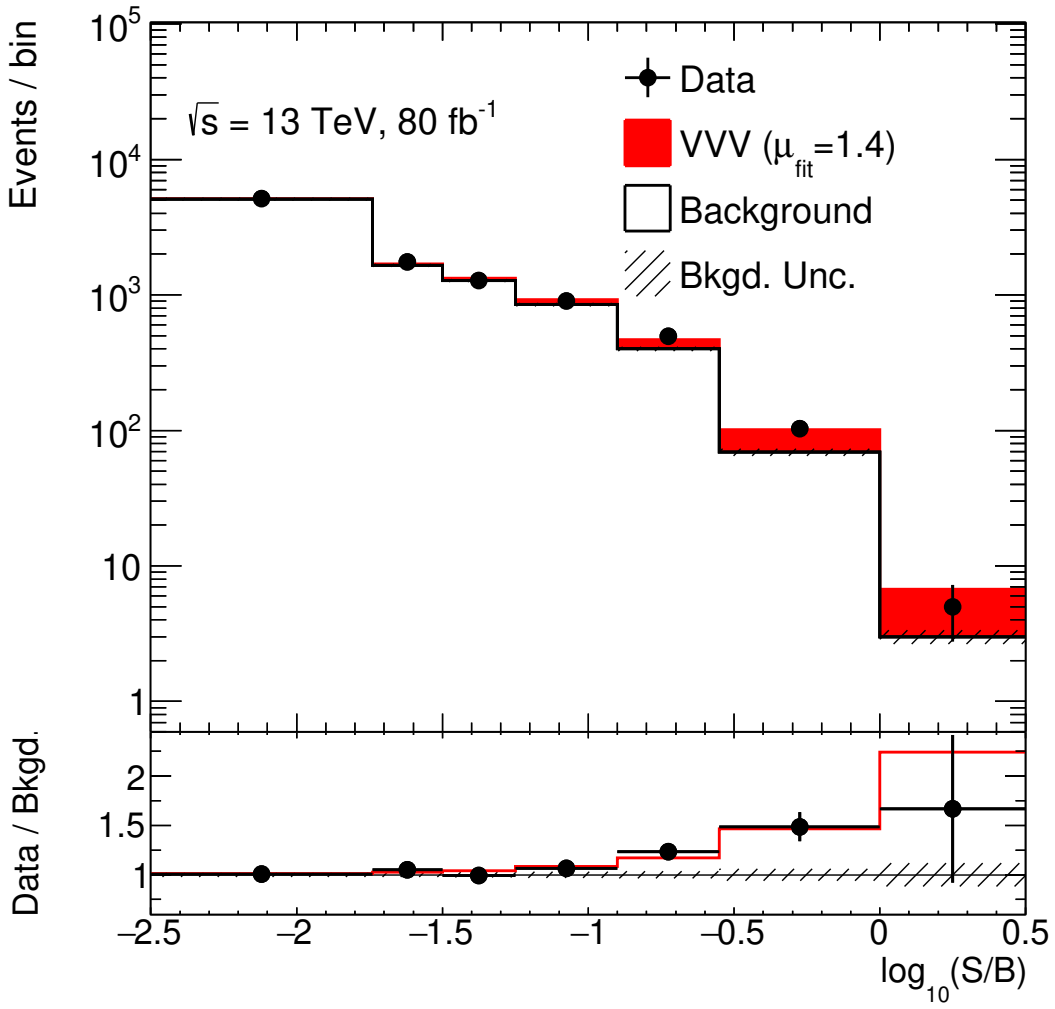


Figure 6.34: Event yields as a function of $\log_{10}$ (S/B), where S (B) indicates the signal (background) expected yield in the corresponding bin. Events in all of the twelve fitted regions are included and the predicted background yield is obtained from the global signal-plus-background fit. The VVV signal is shown for the best-fit (μ = 1.4) value.

Table 6.21: Summary of the effects of the most important groups of systematic uncertainties on μ_{VVV}.

Uncertainty source	$\Delta\mu$
Data-driven	+0.15 − 0.15
Theory	+0.13 − 0.12
Instrumental	+0.11 − 0.08
Gammas	+0.07 − 0.06
Generators	+0.04 − 0.04
Total systematic uncertainty	+0.24 − 0.23

CHAPTER 7

Conclusions and outlook

One of the main goals of the LHC was the discovery of the Higgs boson. After the discovery of a new particle with a mass of 125 GeV compatible with the Standard Model Higgs boson, a large effort has been dedicated to the measurement of its properties. The top quark is the heaviest known elementary particle and, thus, is expected to have the largest Yukawa coupling to the Higgs boson. The associated production of a Higgs boson with a pair of top quarks (referred to as $t\bar{t}H$) allows a direct measurement of this coupling at the LHC. No evidence of $t\bar{t}H$ production was found with the $20\,\text{fb}^{-1}$ of Run 1 data at $\sqrt{s} = 8\,\text{TeV}$.

The excellent operation of the LHC machine and ATLAS detector in 2015, 2016 and 2017 allowed to record $80\,\text{fb}^{-1}$ of data useful for physics analyses at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. This thesis presents two searches ($t\bar{t}H$ and WVZ) in multilepton, namely three- and four-lepton, final states with the $80\,\text{fb}^{-1}$ Run 2 data collected by the ATLAS detector.

Two versions of the $t\bar{t}H$ search in the non-resonant four-lepton final state have been performed, using 36.1 and $80\,\text{fb}^{-1}$. The analysis takes advantage of two signal regions and a gradient boosted decision tree in order to optimise the signal sensitivity. The combined best-fit value of the $t\bar{t}H$ signal strength for the measurement at $80\,\text{fb}^{-1}$ yields $\mu_{t\bar{t}H\to4\ell} = 1.0^{+1.0}_{-0.7}$, where the result is dominated by statistical uncertainties. The background-only hypothesis is rejected at the 1.5σ level. The latest ATLAS $t\bar{t}H$ combination, involving several channels, has provided observation of Standard Model $t\bar{t}H$ production and, thus, of the top-quark Yukawa coupling.

The absence of clear deviations from the Standard Model predictions requires thorough investigations in all unexplored Standard Model sectors. The first search for WVZ production in three- and four-lepton final states has been performed. Six signal regions are considered for this final state: three 3ℓ ($3\ell1j$, $3\ell2j$ and $3\ell3j$) and three 4ℓ (DF, on-shell and off-shell SF) channels; in addition a $t\bar{t}Z \to 3\ell$ single-bin control region is fitted. The combined best-fit value of the WVZ signal strength yields $\mu_{WVZ} = 1.5 \pm 0.6$. The background-only hypothesis is rejected at the 2.9σ level. The statistical combination with the ATLAS search for $W^\pm W^\pm W^\mp$ provides the first evidence for the Standard Model production of three massive vector bosons (VVV).

The combination of multiple channels and final states has allowed to reach both, the observation of $t\bar{t}H$ and evidence for VVV production, important steps towards a tighter and tighter probe of the Standard Model.

Bibliography

[1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[2] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) 367.

[3] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[4] S. L. Glashow, J. Iliopoulos and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285.

[5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[6] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156.

[7] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321.

[8] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. **D98** (2018) 030001.

[9] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1918) 235.

[10] W. Heisenberg, *Über den Bau der Atomkerne. I*, Z. Physik (1932) 77.

[11] C. N. Yang and R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance*, Phys. Rev. **96** (1954) 191.

[12] R. Utiyama, *Invariant Theoretical Interpretation of Interaction*, Phys. Rev. **101** (1956) 1597.

[13] J. S. Schwinger, *A Theory of the Fundamental Interactions*, Annals Phys. **2** (1957) 407.

[14] R. P. Feynman and M. Gell-Mann, *Theory of Fermi interaction*, Phys. Rev. **109** (1958) 193.

[15] E. C. G. Sudarshan and R. E. Marshak, *Chirality invariance and the universal Fermi interaction*, Phys. Rev. **109** (1958) 1860.

[16] J. J. Sakurai, *Mass Reversal and Weak Interactions*, Nuovo Cim. **7** (1958) 649.

[17] S. A. Bludman, *On the universal Fermi interaction*, Nuovo Cim. **9** (1958) 433.

[18] F. J. Hasert et al., *Search for Elastic ν_μ Electron Scattering*, Phys. Lett. **B46** (1973) 121.

[19] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Phys. Lett. **13** (1964) 168.

[20] G. 't Hooft, *Renormalization of Massless Yang-Mills Fields*, Nucl. Phys. **B33** (1971) 173.

[21] A. Pich, *The Standard model of electroweak interactions*, (2008), arXiv: 0705.4264 [hep-ph].

[22] J. Ellis, *Topics in Higgs Physics*, (2017), arXiv: 1702.05436 [hep-ph].

[23] R. Barate et al., *Search for the standard model Higgs boson at LEP*, Phys. Lett. **B565** (2003) 61, arXiv: hep-ex/0306033.

[24] The TEVNPH Working Group, *Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data*, (2012), arXiv: 1203.3774 [hep-ex].

[25] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[26] D. de Florian et al.,
Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector, (2016), arXiv: 1610.07922 [hep-ph].

[27] ATLAS and CMS Collaborations, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*,
Phys. Rev. Lett. **114** (2015) 191803, arXiv: 1503.07589 [hep-ex].

[28] ATLAS and CMS Collaborations,
Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV,
JHEP **08** (2016) 045, arXiv: 1606.02266 [hep-ex].

[29] ATLAS Collaboration, *Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector*, Phys. Lett. **B749** (2015) 519,
arXiv: 1506.05988 [hep-ex].

[30] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*,
Eur. Phys. J. **C75** (2015) 349, arXiv: 1503.05066 [hep-ex].

[31] ATLAS Collaboration, *Search for $H \to \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector*, Phys. Lett. **B740** (2015) 222,
arXiv: 1409.3122 [hep-ex].

[32] CMS Collaboration,
Search for the associated production of the Higgs boson with a top-quark pair,
JHEP **09** (2014) 087, [Erratum: JHEP10(2014)106], arXiv: 1408.1682 [hep-ex].

[33] CMS Collaboration, *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method*,
Eur. Phys. J. **C75** (2015) 251, arXiv: 1502.02485 [hep-ex].

[34] ATLAS Collaboration, *Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector*, Phys. Rev. **D97** (2018) 072003,
arXiv: 1712.08891 [hep-ex].

[35] CMS Collaboration, *Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV*,
JHEP **08** (2018) 066, arXiv: 1803.05485 [hep-ex].

[36] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. **B784** (2018) 173,
arXiv: 1806.00425 [hep-ex].

[37] CMS Collaboration, *Observation of $t\bar{t}H$ production*, Phys. Rev. Lett. **120** (2018) 231801,
arXiv: 1804.02610 [hep-ex].

[38] S. R. Coleman and E. J. Weinberg,
Radiative Corrections as the Origin of Spontaneous Symmetry Breaking,
Phys. Rev. **D7** (1973) 1888.

[39] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, J. Exp. Theor. Phys. **120** (2015) 335, arXiv: 1411.1923 [hep-ph].

[40] G. Altarelli and G. Isidori, *Lower limit on the Higgs mass in the standard model: An Update*, Phys. Lett. **B337** (1994) 141.

[41] N. Cabibbo et al., *Bounds on the fermions and Higgs boson masses in grand unified theories*, Nucl. Phys. **B158** (1979) 295.

[42] C. Ford et al., *The Effective potential and the renormalization group*, Nucl. Phys. **B395** (1993) 17, arXiv: hep-lat/9210033.

[43] M. F. Zoller, *Vacuum stability in the SM and the three-loop β-function for the Higgs self-interaction*, Subnucl. Ser. **50** (2014) 557, arXiv: 1209.5609 [hep-ph].

[44] H. D. Politzer and S. Wolfram, *Bounds on Particle Masses in the Weinberg-Salam Model*, Phys. Lett. **82B** (1979) 242, [Erratum: Phys. Lett.83B(1979)421].

[45] P. Q. Hung, *Vacuum Instability and New Constraints on Fermion Masses*, Phys. Rev. Lett. **42** (1979) 873.

[46] V. Branchina, E. Messina and M. Sher, *Lifetime of the electroweak vacuum and sensitivity to Planck scale physics*, Phys. Rev. **D91** (2015) 013003, arXiv: 1408.5302 [hep-ph].

[47] D. Buttazzo et al., *Investigating the near-criticality of the Higgs boson*, JHEP **12** (2013) 089, arXiv: 1307.3536 [hep-ph].

[48] J. Elias-Miro et al., *Higgs mass implications on the stability of the electroweak vacuum*, Phys. Lett. **B709** (2012) 222, arXiv: 1112.3022 [hep-ph].

[49] C.-S. Chen and Y. Tang, *Vacuum stability, neutrinos, and dark matter*, JHEP **04** (2012) 019, arXiv: 1202.5717 [hep-ph].

[50] S. Baek et al., *Vacuum structure and stability of a singlet fermion dark matter model with a singlet scalar messenger*, JHEP **11** (2012) 116, arXiv: 1209.4163 [hep-ph].

[51] C. Coriano, L. Delle Rose and C. Marzo, *Vacuum Stability in U(1)-Prime Extensions of the Standard Model with TeV Scale Right Handed Neutrinos*, Phys. Lett. **B738** (2014) 13, arXiv: 1407.8539 [hep-ph].

[52] F. L. Bezrukov and M. Shaposhnikov, *The Standard Model Higgs boson as the inflaton*, Phys. Lett. **B659** (2008) 703, arXiv: 0710.3755 [hep-th].

[53] F. Bezrukov, J. Rubio and M. Shaposhnikov, *Living beyond the edge: Higgs inflation and vacuum metastability*, Phys. Rev. **D92** (2015) 083512, arXiv: 1412.3811 [hep-ph].

[54] S. Weinberg, *Phenomenological Lagrangians*, Physica **A96** (1979) 327.

[55] M. Baak et al., "Working Group Report: Precision Study of Electroweak Interactions", *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013, arXiv: 1310.6708 [hep-ph].

[56] G. Gounaris et al., "Triple gauge boson couplings", *AGS / RHIC Users Annual Meeting Upton, New York, June 15-16, 1995*, arXiv: hep-ph/9601233.

[57] CMS Collaboration, *Limits on anomalous triple and quartic gauge couplings*, URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMPaTGC.

[58] ATLAS Collaboration, *Measurement of the production cross section of three isolated photons in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector*, Phys. Lett. **B781** (2018) 55, arXiv: 1712.07291 [hep-ex].

[59] ATLAS Collaboration, *Measurements of $Z\gamma$ and $Z\gamma\gamma$ production in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, Phys. Rev. **D93** (2016) 112002, arXiv: 1604.05232 [hep-ex].

[60] CMS Collaboration, *Measurements of the $pp \to W\gamma\gamma$ and $pp \to Z\gamma\gamma$ cross sections and limits on anomalous quartic gauge couplings at $\sqrt{s}$ = 8 TeV*, JHEP **10** (2017) 072, arXiv: 1704.00366 [hep-ex].

[61] ATLAS Collaboration, *Evidence of $W\gamma\gamma$ Production in pp Collisions at $\sqrt{s}$ = 8 TeV and Limits on Anomalous Quartic Gauge Couplings with the ATLAS Detector*, Phys. Rev. Lett. **115** (2015) 031802, arXiv: 1503.03243 [hep-ex].

[62] CMS Collaboration, *Search for $WW\gamma$ and $WZ\gamma$ production and constraints on anomalous quartic gauge couplings in pp collisions at $\sqrt{s}$ = 8 TeV*, Phys. Rev. **D90** (2014) 032008, arXiv: 1404.4619 [hep-ex].

[63] ATLAS Collaboration, *Search for triboson $W^{\pm}W^{\pm}W^{\mp}$ production in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 141, arXiv: 1610.05088 [hep-ex].

[64] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*, ATLAS-TDR-14, 1999, URL: https://cds.cern.ch/record/391176.

[65] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 2*, ATLAS-TDR-15, 1999, URL: https://cds.cern.ch/record/391177.

[66] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[67] W. Herr and B. Muratori, *Concept of luminosity*, Yellow Report CERN 2006-002, 2006, URL: http://cds.cern.ch/record/941318.

[68] A. Achilli et al., *Total and inelastic cross sections at LHC at $\sqrt{s}$ = 7 TeV and beyond*, Phys. Rev. D **84** (2011) 094009.

[69] L. Evans, *The Large Hadron Collider*, New Journal of Physics **9** (2007) 335.

[70] ATLAS Collaboration, *Luminosity Public Results Run2*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[71] ATLAS Collaboration, *The ATLAS Data Acquisition and High Level Trigger system*, JINST **11** (2016) P06008.

[72] M. Schott and M. Dunford, *Review of single vector boson production in pp collisions at $\sqrt{s}$ = 7 TeV*, Eur. Phys. J. **C74** (2014) 2916, arXiv: 1405.1160 [hep-ex].

[73] ATLAS Collaboration, *The design and performance of the ATLAS Inner Detector trigger for Run 2 LHC Collisions at $\sqrt{s}$ = 13 TeV*, JINST **762** (2016) 012029.

[74] C. Marcelloni De Oliveira, *IBL installation into the inner detector of the ATLAS Experiment side C*, 2014, URL: https://cds.cern.ch/record/1702006.

[75] A. Aloisio et al., *The trigger chambers of the ATLAS muon spectrometer: Production and tests*, Nucl. Instrum. Meth. **A535** (2004) 265.

[76] ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report*, ATLAS-TDR-10, 1997, URL: https://cds.cern.ch/record/331068.

[77] H. P. Beck, *Triggering at high luminosity colliders*, New Journal of Physics **9** (2007) 334.

[78] S. Artz et al., *Upgrade of the ATLAS Central Trigger for LHC Run-2*, JINST **10** (2015) C02030.

[79] R. Brun and F. Rademakers, *ROOT — An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.

[80] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, Eur. Phys. J. **C77** (2017) 673, arXiv: 1704.07983 [hep-ex].

[81] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444.

[82] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, JINST **9** (2014) P09009, arXiv: 1406.7690 [hep-ex].

[83] G. Piacquadio, K. Prokofiev and A. Wildauer, *Primary vertex reconstruction in the ATLAS experiment at LHC*, J. Phys. Conf. Ser. **119** (2008) 032033.

[84] G. Borissov et al., *ATLAS strategy for primary vertex reconstruction during Run-2 of the LHC*, J. Phys. Conf. Ser. **664** (2015) 072041.

[85] F. Meloni, *Primary vertex reconstruction with the ATLAS detector*, JINST **11** (2016) C12060.

[86] J. Miner, *Comparison of Jet reconstruction Algorithms*, URL: http://www.int.washington.edu/REU/2006/MinerJetREU2006.pdf.

[87] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph].

[88] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: https://cds.cern.ch/record/1700870.

[89] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: http://cds.cern.ch/record/1099735.

[90] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012, URL: https://cds.cern.ch/record/1449796.

[91] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024, 2016, URL: https://cds.cern.ch/record/2157687.

[92] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. **C76** (2016) 292, arXiv: 1603.05598 [hep-ex].

[93] J. Illingworth and J. Kittler, *A survey of the Hough transform*, Computer Vision, Graphics, and Image Processing **44** (1988) 87.

[94] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s}$= 13 TeV*, ATL-PHYS-PUB-2015-027, 2015, URL: http://cds.cern.ch/record/2037904.

[95] M. A. Dobbs et al., "Les Houches guidebook to Monte Carlo generators for hadron collider physics", *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 26-June 3, 2003*, 2004 411, arXiv: hep-ph/0403045.

[96] M. Klein, *Structure functions in deep inelastic lepton nucleon scattering*, Int. J. Mod. Phys. **A15S1** (2000) 467, arXiv: hep-ex/0001059.

[97] J. Rojo et al., *The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II*, J. Phys. G **42** (2015) 103103.

[98] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438, [Yad. Fiz.15,781(1972)].

[99] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641, [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[100] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298.

[101] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph].

[102] S. Dulat et al., *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. **D93** (2016) 033006, arXiv: 1506.07443 [hep-ph].

[103] A. D. Martin et al., *Heavy-quark mass dependence in global PDF analyses and 3- and 4-flavour parton distributions*, Eur. Phys. J. **C70** (2010) 51, arXiv: 1007.2624 [hep-ph].

[104] J. C. Collins, D. E. Soper and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1, arXiv: hep-ph/0409313.

[105] T. Plehn, "LHC Phenomenology for Physics Hunters", *Proceedings of Theoretical Advanced Study Institute in Elementary Particle Physics on The dawn of the LHC era (TASI 2008): Boulder, USA, June 2-27, 2008*, 2010 125, arXiv: 0810.2281 [hep-ph].

[106] A. Buckley et al., *General-purpose event generators for LHC physics*, Phys. Rept. **504** (2011) 145, arXiv: 1101.2599 [hep-ph].

[107] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph].

[108] J. Alwall et al., *MadGraph/MadEvent v4: The New Web Generation*, JHEP **09** (2007) 028, arXiv: 0706.2334 [hep-ph].

[109] 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 [hep-ph].

[110] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146.

[111] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph].

[112] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph].

[113] S. Catani et al., *QCD Matrix Elements + Parton Showers*, JHEP **11** (2001) 063.

[114] F. Krauss, *Matrix elements and parton showers in hadronic interactions*, JHEP **08** (2002) 015, arXiv: hep-ph/0205283.

[115] J. Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur. Phys. J. **C53** (2008) 473, arXiv: 0706.2569 [hep-ph].

[116] V. V. Sudakov, *Vertex parts at very high-energies in quantum electrodynamics*, Sov. Phys. JETP **3** (1956) 65.

[117] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31.

[118] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[119] ATLAS collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823, arXiv: 1005.4568 [physics.ins-det].

[120] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, ATL-SOFT-PROC-2012-065, 2012, URL: http://cdsweb.cern.ch/record/1458503/.

[121] T. Kuhr, *Flavor physics at the Tevatron*, Springer Tracts Mod. Phys. **249** (2013) 1.

[122] A. Lenz, *Lifetimes and heavy quark expansion*, Int. J. Mod. Phys. **A30** (2015) 1543005, arXiv: 1405.3601 [hep-ph].

[123] Yu. M. Shabelski, A. G. Shuvaev, and I. V. Surnin, *Heavy quark production in k_t factorization approach at LHC energies*, Int. J. Mod. Phys. **A33** (2018) 1850003.

[124] ALICE Collaboration, *Beauty production in pp collisions at $\sqrt{s} = 2.76$ TeV measured via semi-electronic decays*, Phys. Lett. **B738** (2014) 97.

[125] ATLAS Collaboration, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, JHEP **01** (2015) 069, arXiv: 1409.6212 [hep-ex].

[126] ATLAS Collaboration, *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*, Phys. Lett. **B786** (2018) 59, arXiv: 1808.08238 [hep-ex].

[127] CMS Collaboration, *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. **121** (2018) 121801, arXiv: 1808.08242 [hep-ex].

[128] A. Höcker et al., *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv: physics/0703039.

[129] G. Cowan, *Statistical Data Analysis*, Oxford Science Publications, 1998.

[130] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.

[131] Y. Mansour and D. McAllester, *Boosting Using Branching Programs*, Journal of Computer and System Sciences **64** (2002) 103.

[132] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*, John Wiley & Sons, 2004.

[133] ATLAS Collaboration, *Commissioning of the ATLAS high performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102, 2011, URL: https://cds.cern.ch/record/1369219.

[134] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281.

[135] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, 2012.

[136] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, URL: https://cds.cern.ch/record/2270366.

[137] Theano Development Team, *Theano: A Python framework for fast computation of mathematical expressions*, (2016), arXiv: 1605.02688 [cs].

[138] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: https://cds.cern.ch/record/2160731.

[139] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex].

[140] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph].

[141] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) 074024, arXiv: 1007.2241 [hep-ph].

[142] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. **D89** (2014) 033009, arXiv: 1302.6246 [hep-ph].

[143] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175.

[144] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419.

[145] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189, arXiv: 0901.0002 [hep-ph].

[146] A. Sciandra, *The Soft Muon Tagger for the identification of b jets in ATLAS*, CERN-THESIS-2015-143, Roma III, 2015, URL: http://cds.cern.ch/record/2054489.

[147] M. Karson, *Handbook of Methods of Applied Statistics. Volume I: Techniques of Computation Descriptive Methods, and Statistical Inference. Volume II: Planning of Surveys and Experiments. I. M. Chakravarti, R. G. Laha, and J. Roy, New York, John Wiley; 1967*, JASA **63** (1968) 1047.

[148] ATLAS Collaboration, *Event Displays from Run 2 physics analyses*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayRun2Physics.

[149] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, [Erratum: Eur. Phys. J. C73 (2013) 2501], arXiv: 1007.1727 [physics.data-an].

[150] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society **54** (1943) 426.

[151] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Ann. Math. Statist. **9** (1938) 60.

[152] ATLAS Collaboration, 2018, URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/GoodRunListsForAnalysisRun2.

[153] W. Beenakker et al., *Higgs radiation off top quarks at the Tevatron and the LHC*, Phys. Rev. Lett. **87** (2001) 201805, arXiv: hep-ph/0107081.

[154] W. Beenakker et al., *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl. Phys. **B653** (2003) 151, arXiv: hep-ph/0211352.

[155] S. Dawson et al., *Associated top quark Higgs boson production at the LHC*, Phys. Rev. **D67** (2003) 071503, arXiv: hep-ph/0211438.

[156] S. Dawson et al., *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, Phys. Rev. D **68** (2003) 034022, arXiv: hep-ph/0305087.

[157] Y. Zhang et al., *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider*, Phys. Lett. **B738** (2014) 1, arXiv: 1407.1110 [hep-ph].

[158] S. Frixione et al., *Weak corrections to Higgs hadroproduction in association with a top-quark pair*, JHEP **09** (2014) 065, arXiv: 1407.0823 [hep-ph].

[159] S. Frixione et al., *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, JHEP **06** (2015) 184, arXiv: 1504.03446 [hep-ph].

[160] J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, (2013), ed. by S. Heinemeyer et al., arXiv: 1307.1347 [hep-ph].

[161] ATLAS and CMS Collaborations, 2017, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV.

[162] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph].

[163] K. Melnikov, M. Schulze and A. Scharf, *QCD corrections to top quark pair production in association with a photon at hadron colliders*, Phys. Rev. **D83** (2011) 074013, arXiv: 1102.1967 [hep-ph].

[164] S. Höche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph].

[165] F. Demartin et al., *tWH associated production at the LHC*, Eur. Phys. J. **C77** (2017) 34, arXiv: 1607.05862 [hep-ph].

[166] T. Sjöstrand et al., *High-energy-physics event generation with Pythia 6.1*, Comput. Phys. Commun. **135** (2001) 238, arXiv: hep-ph/0010017.

[167] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: 0803.0883 [hep-ph].

[168] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. Phys. Res. A **462** (2001) 152.

[169] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv: hep-ph/0506026.

[170] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) 244, arXiv: 1207.1303 [hep-ph].

[171] M. H. Seymour and A. Siodmok, *Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data*, JHEP **10** (2013) 113, arXiv: 1307.5015 [hep-ph].

[172] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. **D78** (2008) 013004, arXiv: 0802.0007 [hep-ph].

[173] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018, arXiv: 1005.3457 [hep-ph].

[174] ATLAS Collaboration, *Search for flavour-changing neutral current top quark decays $t \to Hq$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **12** (2015) 061, arXiv: 1509.06047 [hep-ex].

[175] ATLAS Collaboration, *Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **05** (2016) 160, arXiv: 1604.03812 [hep-ex].

[176] N. Bruscino, *A gateway to new physics: direct measurement of the top Yukawa coupling to the Higgs boson*, CERN-THESIS-2017-084, Bonn, 2017, URL: https://cds.cern.ch/record/2273695.

[177] J. Neyman and E. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Phil. Trans. R. Soc. Lond. **231** (1933) 289.

[178] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) 653, arXiv: 1608.03953 [hep-ex].

[179] ATLAS Collaboration, *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.

[180] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74** (2014) 3071, arXiv: 1407.5063 [hep-ex].

[181] ATLAS Collaboration,
Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s}$ = 7 TeV,
Eur. Phys. J. C **73** (2013) 2304, arXiv: 1112.6426 [hep-ex].

[182] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s}$ = 13 TeV*, (2015), URL: http://cds.cern.ch/record/2037613.

[183] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*,
Phys. Rev. **D96** (2017) 072002, arXiv: 1703.09665 [hep-ex].

[184] ATLAS Collaboration,
E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC p-p collisions, (2018),
URL: http://cds.cern.ch/record/2625233.

[185] ATLAS Collaboration, *Search for the Associated Production of a Higgs Boson and a Top Quark Pair in Multilepton Final States with the ATLAS Detector*, (2016),
URL: https://cds.cern.ch/record/2206153.

[186] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*,
Phys. Rev. **D97** (2018) 072016, arXiv: 1712.08895 [hep-ex].

[187] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}$ = 7 and 8 TeV in the ATLAS experiment*,
Eur. Phys. J. **C76** (2016) 6, arXiv: 1507.04548 [hep-ex].

[188] D. T. Nhung, L. D. Ninh and M. M. Weber, *NLO corrections to WWZ production at the LHC*,
JHEP **12** (2013) 096, arXiv: 1307.7403 [hep-ph].

[189] Y.-B. Shen et al.,
NLO QCD + NLO EW corrections to WZZ productions with leptonic decays at the LHC,
JHEP **10** (2015) 186, arXiv: 1507.03693 [hep-ph], Erratum: JHEP **10** (2016) 156.

[190] V. Hankele and D. Zeppenfeld,
QCD corrections to hadronic WWZ production with leptonic decays,
Phys. Lett. **B661** (2008) 103, arXiv: 0712.3544 [hep-ph].

[191] F. Campanario et al.,
QCD corrections to charged triple vector boson production with leptonic decay,
Phys. Rev. **D78** (2008) 094012, arXiv: 0809.0790 [hep-ph].

[192] ATLAS Collaboration, *Measurement of $W^{\pm}Z$ production cross sections and gauge boson polarisation in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, ATLAS-CONF-2018-034, 2018, URL: http://cds.cern.ch/record/2630187.

[193] ATLAS Collaboration, *$ZZ \to \ell^+\ell^-\ell'^+\ell'^-$ cross-section measurements and search for anomalous triple gauge couplings in 13 TeV pp collisions with the ATLAS detector*,
Phys. Rev. **D97** (2018) 032005, arXiv: 1709.07703 [hep-ex].

[194] ATLAS Collaboration, *Measurement of the production cross-section of a single top quark in association with a Z boson in proton–proton collisions at 13 TeV with the ATLAS detector*, ATLAS-CONF-2017-052, 2017, URL: http://cds.cern.ch/record/2273868.

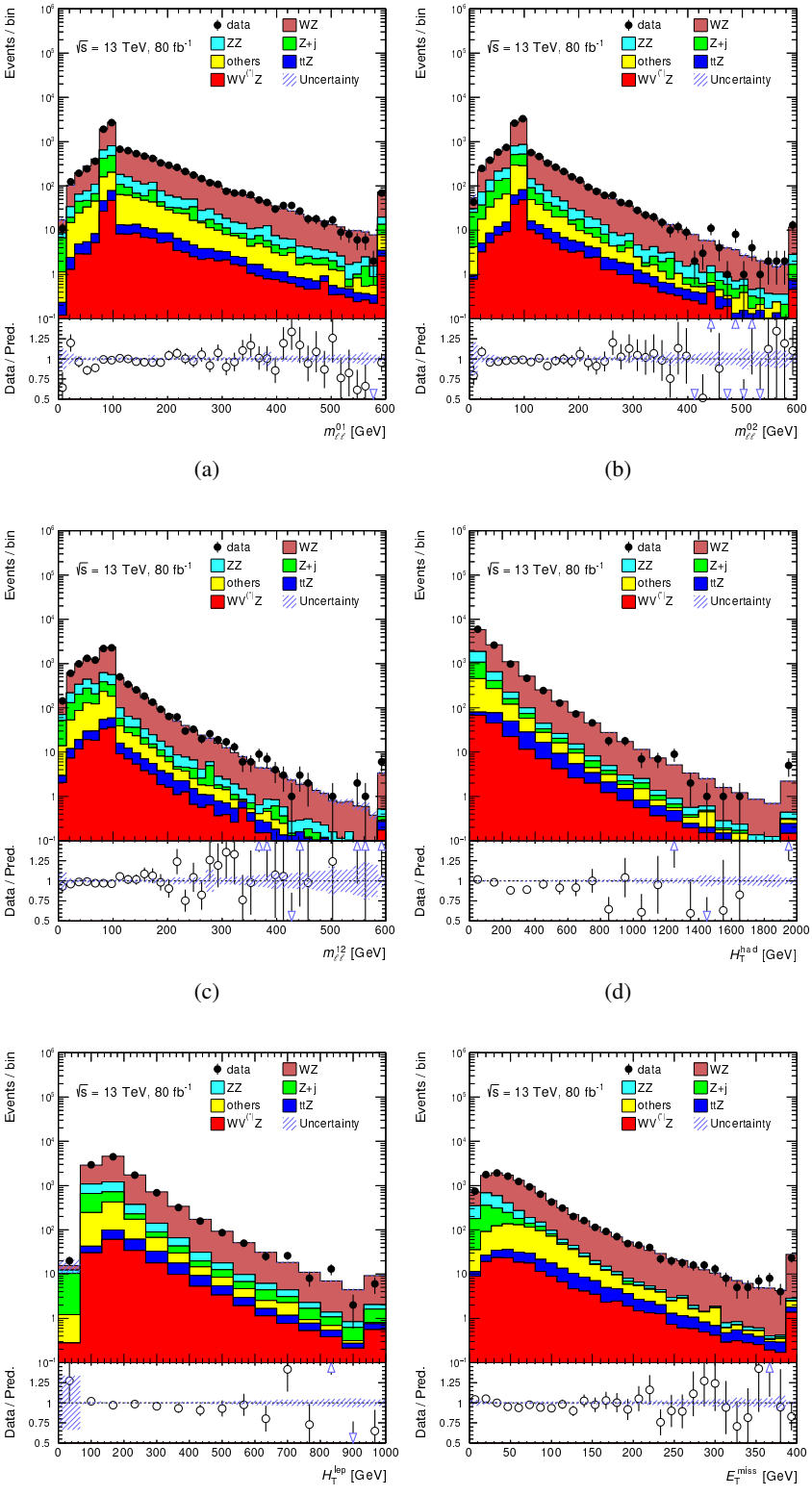
[195] CMS Collaboration, *Measurement of the associated production of a single top quark and a Z boson in pp collisions at $\sqrt{s}$ = TeV*, Phys. Lett. **B779** (2018) 358,
arXiv: 1712.02825 [hep-ex].

[196] ATLAS Collaboration, *Measurement of $W^{\pm}W^{\pm}$ vector-boson scattering and limits on anomalous quartic gauge couplings with the ATLAS detector*, Phys. Rev. **D96** (2017) 012007, arXiv: 1611.02428 [hep-ex].

APPENDIX A

WVZ background validation plots

The validation of irreducible and reducible backgrounds is discussed in Section 6.4. More distributions in the 3ℓ pre-MVA, $t\bar{t}Z$ control and Z + jets validation regions are shown in this appendix section.



(a) (b) (c) (d) (e) (f)

Figure A.1: Data/simulation comparison in the 3ℓ pre-MVA region for (a) $m_{\ell\ell}^{01}$, (b) $m_{\ell\ell}^{02}$, (c) $m_{\ell\ell}^{12}$, (d) $H_{\mathrm{T}}^{\mathrm{had}}$, (e) $H_{\mathrm{T}}^{\mathrm{lep}}$ and (f) $E_{\mathrm{T}}^{\mathrm{miss}}$. The full $80\,\mathrm{fb}^{-1}$ dataset is compared to the complete set of simulation samples. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

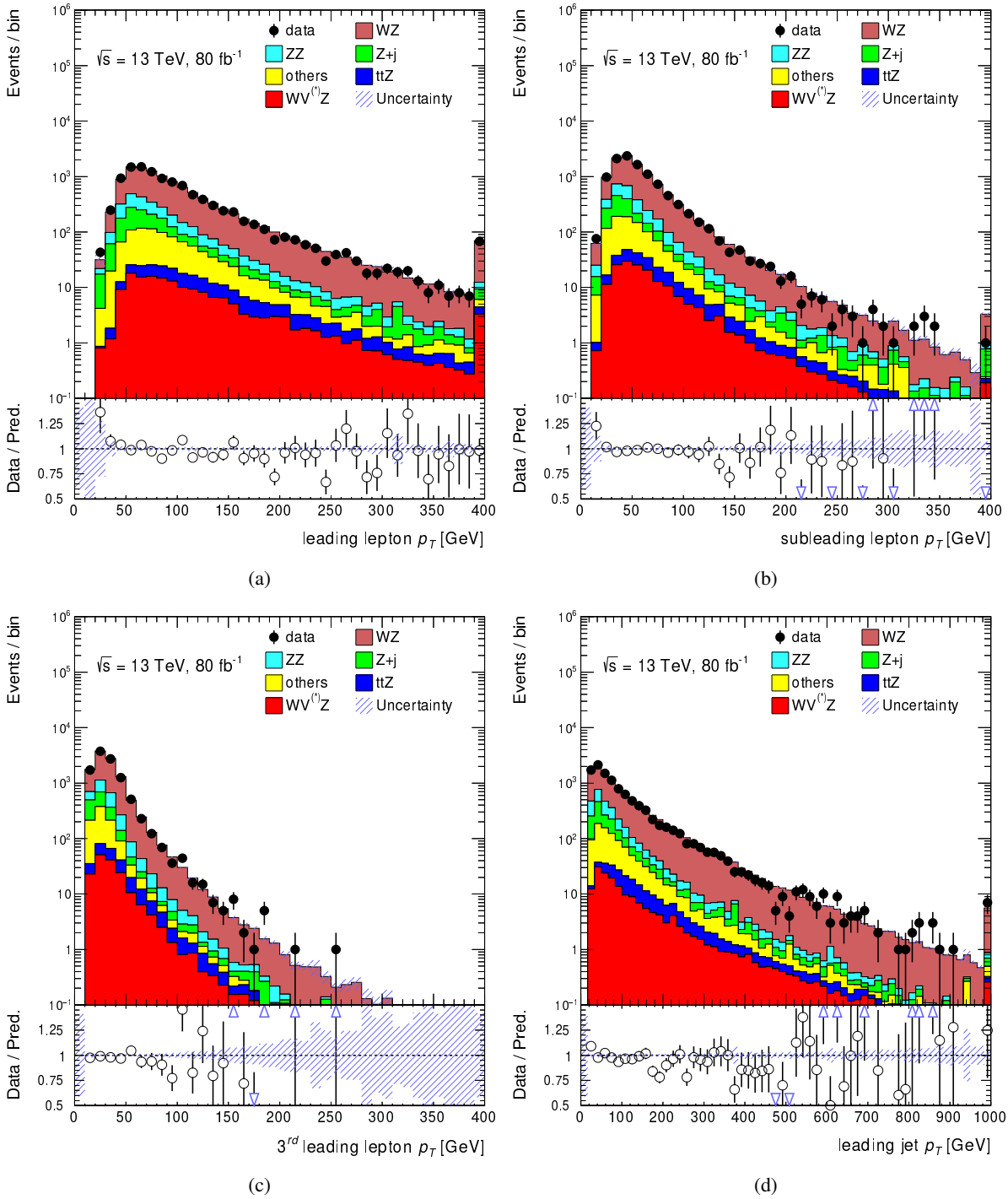


Figure A.2: Data/simulation comparison in the 3ℓ pre-MVA region for the (a) leading lepton p_{T}, (b) sub-leading lepton p_{T}, (c) third leading lepton p_{T} and (d) leading jet p_{T}. The full 80 fb^{-1} dataset is compared to the complete set of simulation samples. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

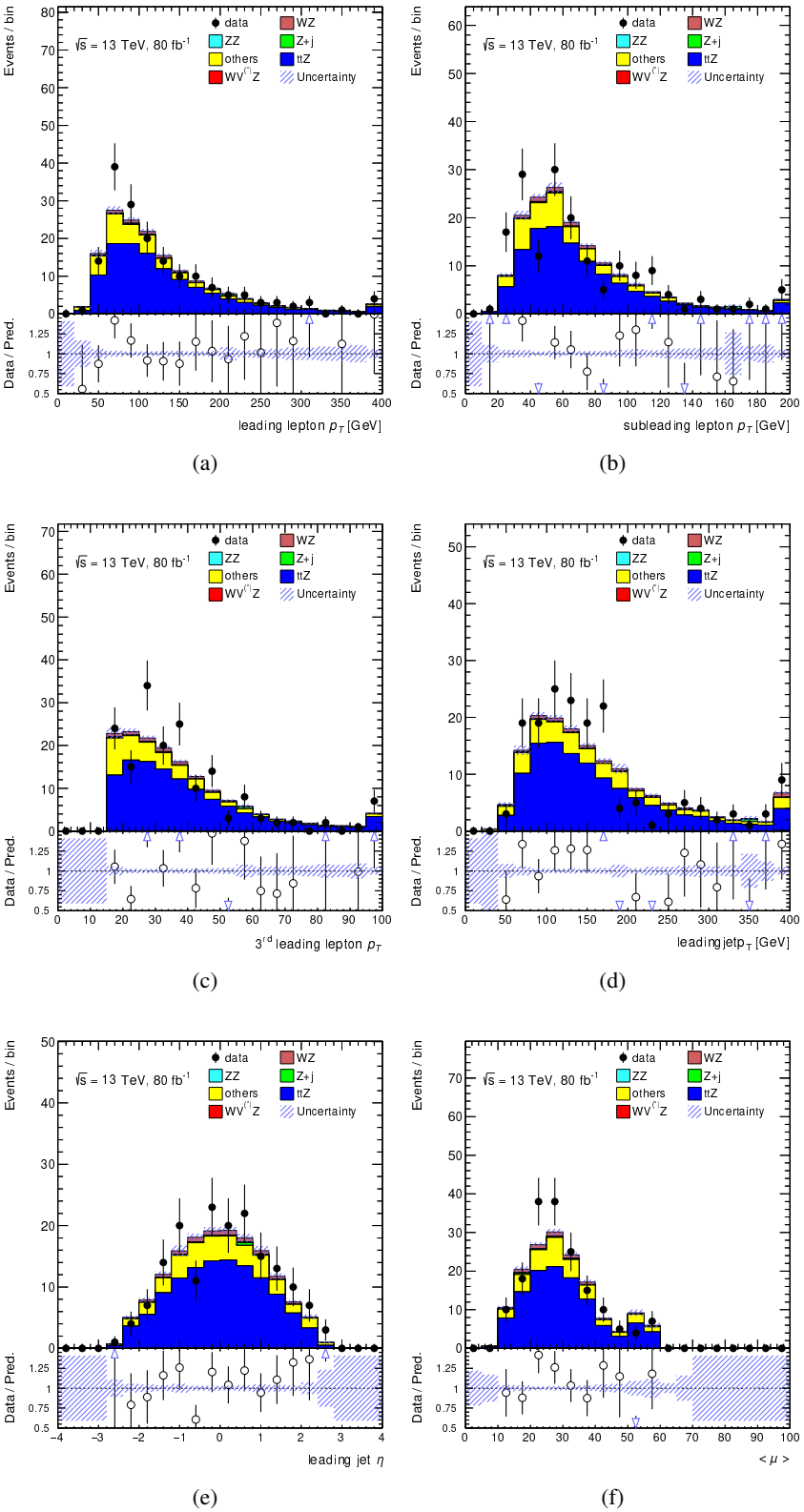


Figure A.3: Data/simulation comparison in the $t\bar{t}Z \to 3\ell$ control region, comparing the full 80 fb^{-1} dataset to mc16a and mc16d, for: (a) the leading lepton p_{T}, (b) the subleading lepton p_{T}, (c) the 3$^{\mathrm{rd}}$ leading lepton p_{T}, (d) the leading jet p_{T}, (e) the leading jet η and (f) the $< \mu >$. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

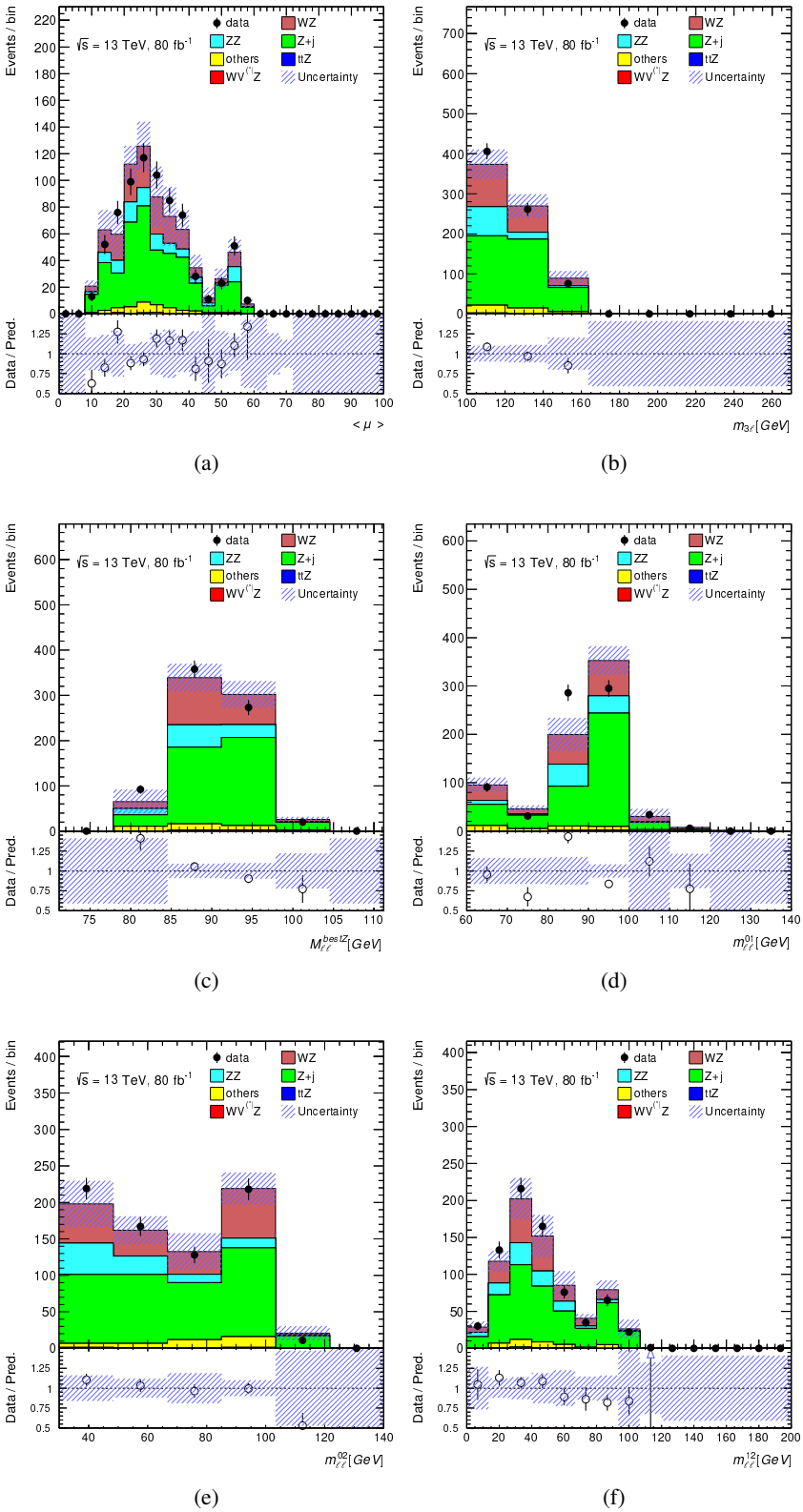
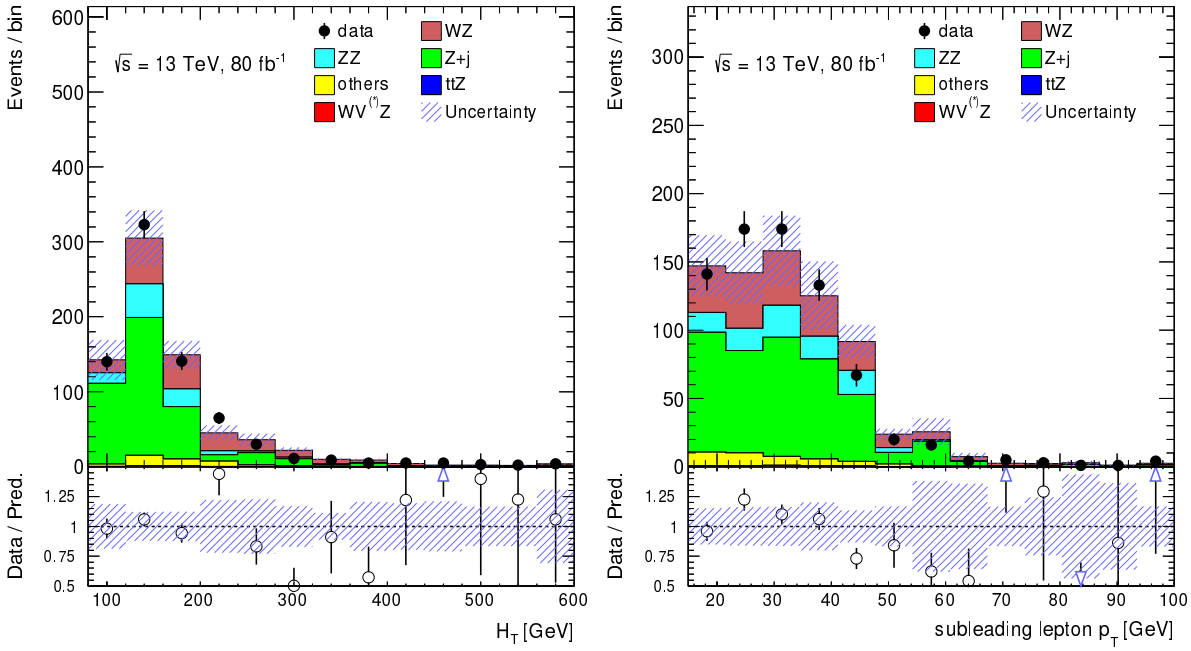
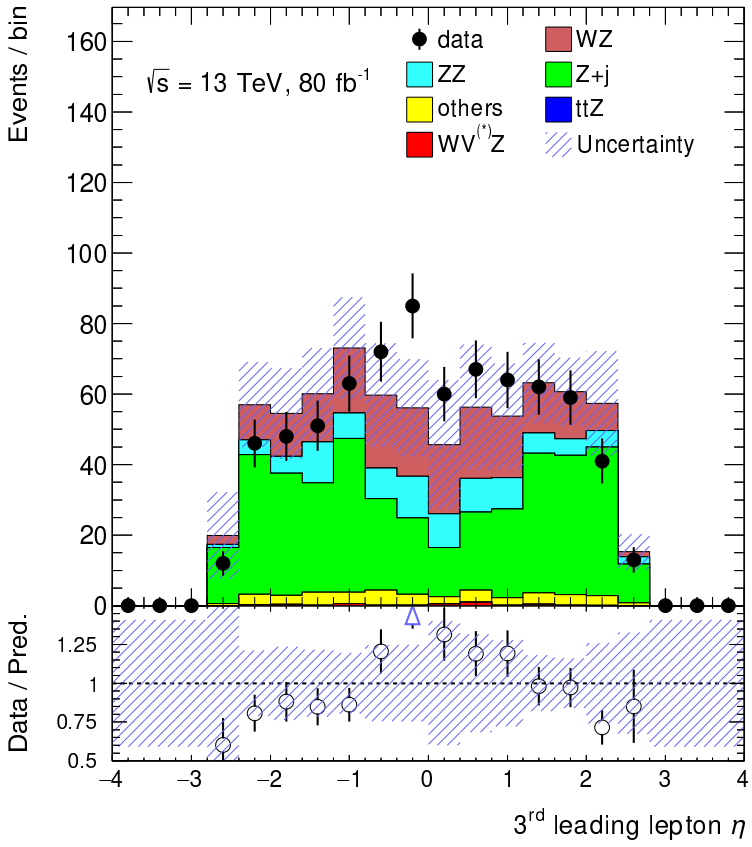


Figure A.4: Data/simulation comparison in the 3ℓ Z + jets validation region for the invariant mass of all possible lepton pairs, comparing the full 80 fb^{-1} dataset to mc16a and mc16d. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.



(a) (b)



(c)

Figure A.5: Data/simulation comparison in the 3ℓ Z + jets validation region for (a) H_{T}, (b) the subleading lepton p_{T} and (c) the third leading lepton η, comparing the full $80\,\mathrm{fb}^{-1}$ dataset to mc16a and mc16d. The lower panel shows the data-to-simulation ratio. Only statistical uncertainties are shown.

APPENDIX B

χ^2 test for jet multiplicity-based reweighting

In this appendix section the full set of distributions, used as inputs to a global χ^2 test to prove the validity of the jet multiplicity-based reweighting, is shown. All χ^2 and numbers of degrees of freedom (NDF) for single distributions have been summed up and the ratio is taken as a reduced χ^2 (correlations between different distributions are neglected).

Two separate tests are performed comparing distributions before and after reweighting in the 3ℓ pre-MVA region: one for jet-related observables and one for non-jet-related quantities.

Jet-related observables

Figure B.1 shows all jet-related distributions before and after reweighting. Jet-related observables lead to 123 NDF and a global χ^2 of 135.5 and 125.5 before and after reweighting, respectively. A significant improvement in the description of jet-related kinematical properties of the events is observed.

Non-jet-related observables

Figures B.2, B.3 and B.4 show all non-jet-related distributions before and after reweighting. They have 286 NDF and a global χ^2 of 300.1 and 296.2 before and after reweighting, respectively. No degradation in the description of observables unrelated to jets is observed.

Impact of reweighting on other observables

Figure B.5 shows the impact of the jet multiplicity-based reweighting on other kinematical distributions, comparing data to simulation before and after reweighting.

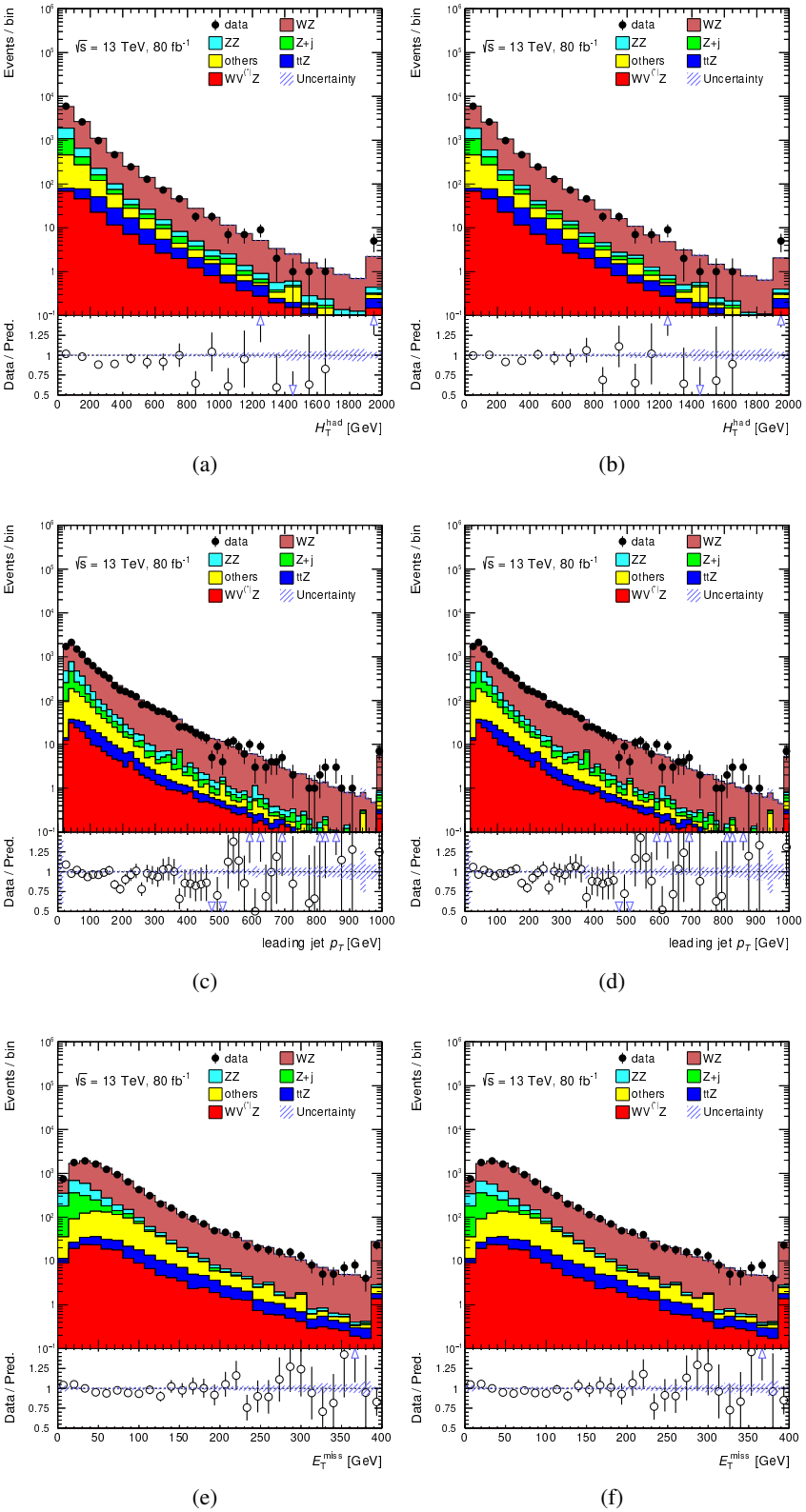
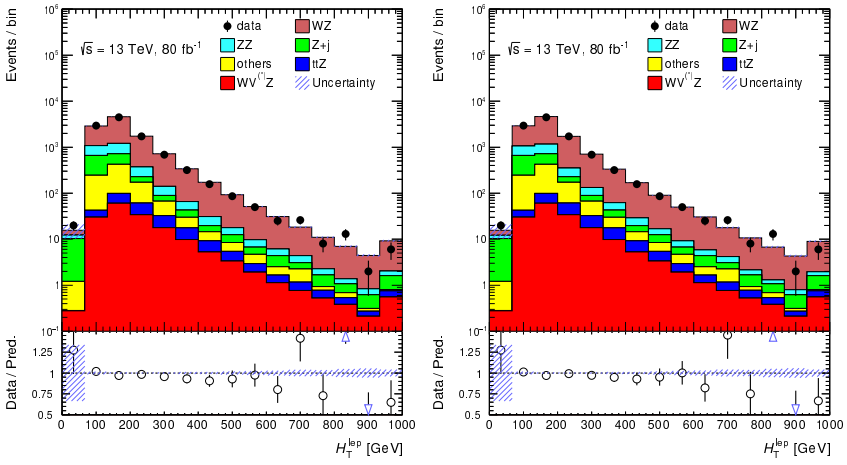
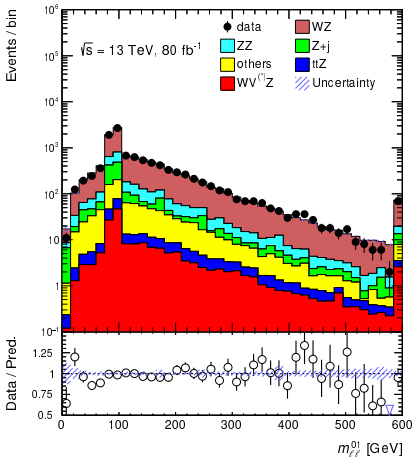


Figure B.1: Data/MC comparison before (left column) and after (right column) reweighting for: (a)(b) $H_{\mathrm{T}}^{\mathrm{had}}$, (c)(d) the leading jet p_{T} and (e)(f) $E_{\mathrm{T}}^{\mathrm{miss}}$. Only statistical uncertainties are shown.

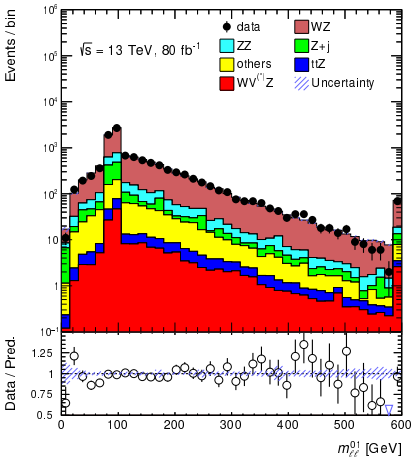


(a)

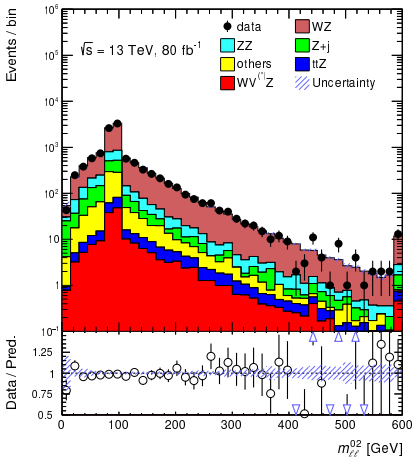
(b)



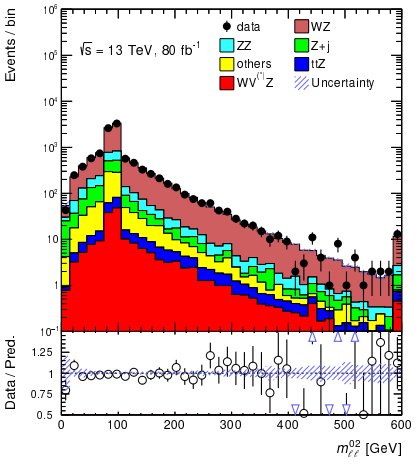
(c)



(d)



(e)



(f)

Figure B.2: Data/MC comparison before (left column) and after (right column) reweighting for: (a)(b) $H_{\mathrm{T}}^{\mathrm{lep}}$, (c)(d) $m_{\ell\ell}^{01}$ and (e)(f) $m_{\ell\ell}^{02}$. Only statistical uncertainties are shown.

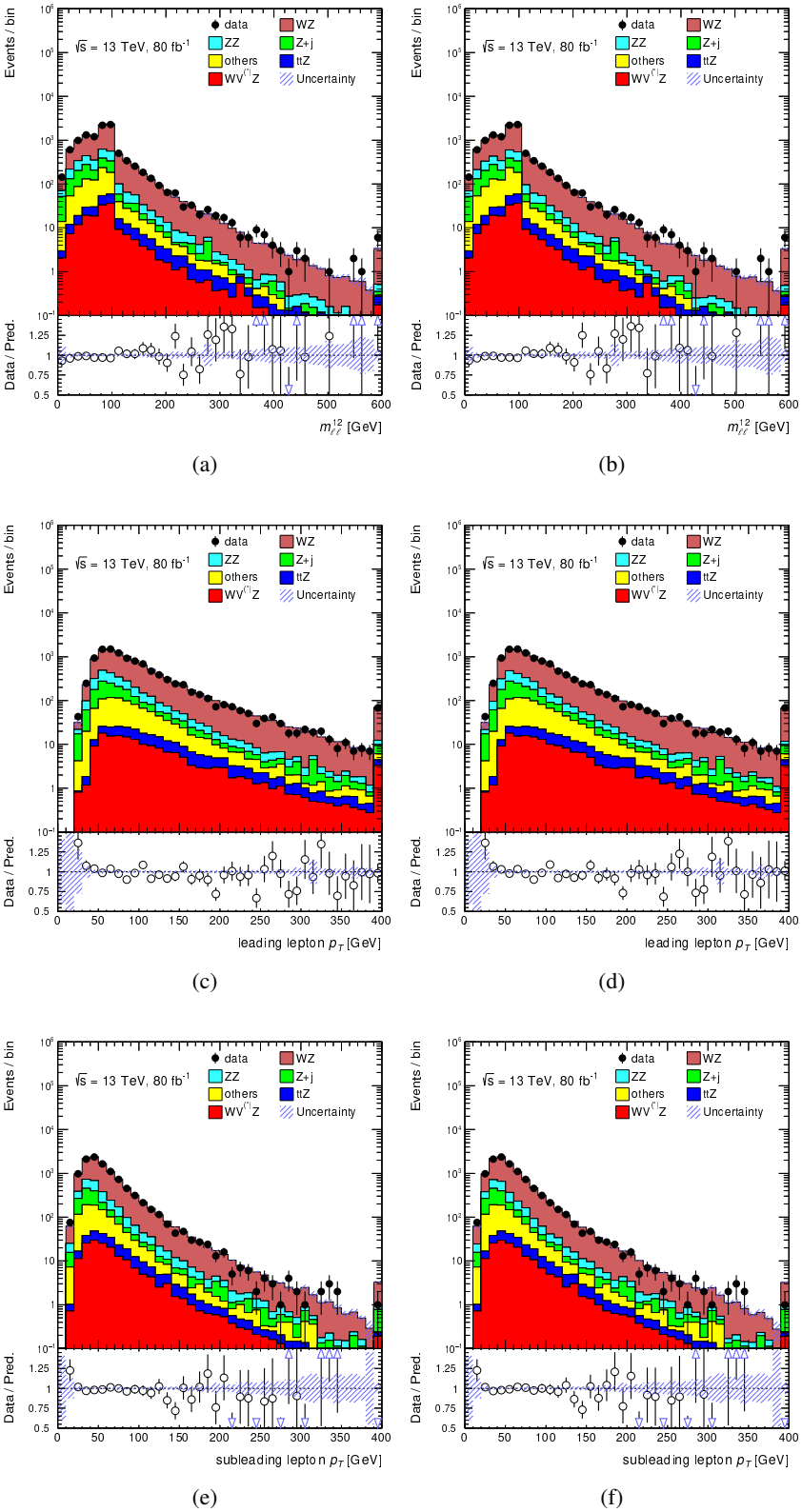


Figure B.3: Data/MC comparison before (left column) and after (right column) reweighting for: (a)(b) $m_{\ell\ell}^{12}$, (c)(d) the leading lepton p_{T} and (e)(f) the subleading lepton p_{T}. Only statistical uncertainties are shown.

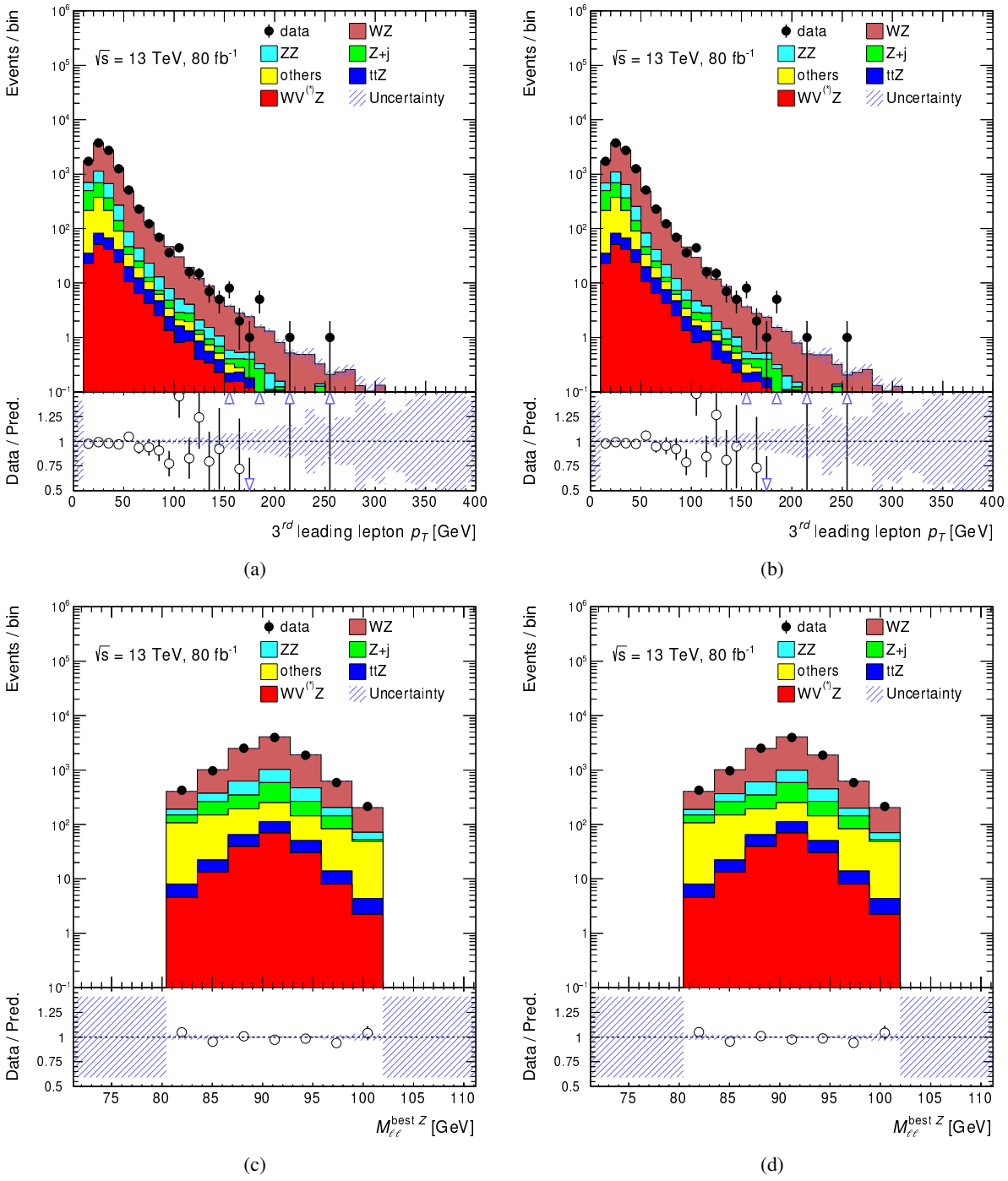


Figure B.4: Data/MC comparison before (left column) and after (right column) reweighting for: (a)(b) the third leading lepton and (c)(d) invariant mass of the best $Z \to \ell\ell$ candidate. Only statistical uncertainties are shown.

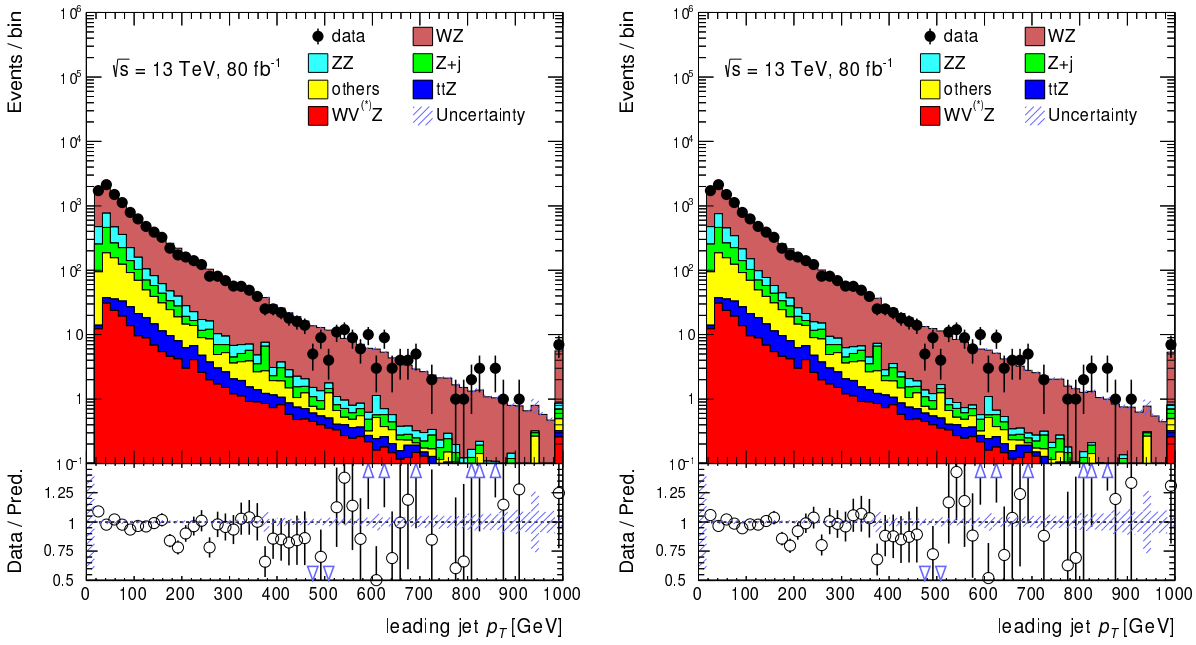


Figure B.5: Data/MC comparison (a) before and (b) after reweighting for the leading jet p_{T}. Only statistical uncertainties are shown.

APPENDIX C

Validation of WVZ MVA inputs

This appendix section is devoted to the modelling of the full set of distributions used as inputs to the WVZ MVAs, discussed in Sections 6.6.1 and 6.6.2.

MVA inputs in 3ℓ signal regions

Figures C.1–C.3 show the data-to-simulation comparison for all of the MVA inputs in the $3\ell 1j$ region, after applying the jet-multiplicity-based reweighting.

Figures C.4–C.6 show the data-to-simulation comparison for all of the MVA inputs in the $3\ell 2j$ region, after applying the jet-multiplicity-based reweighting.

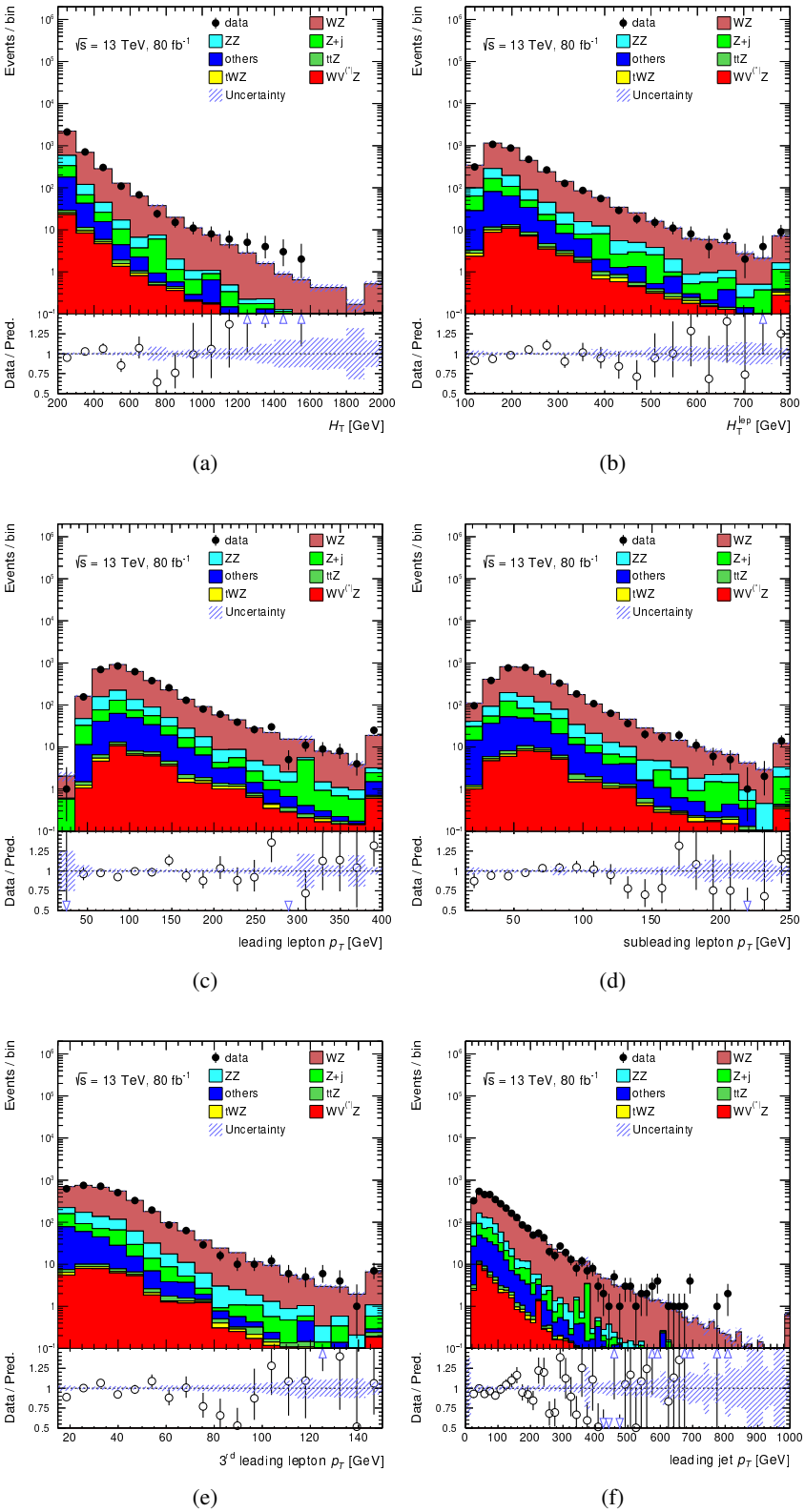
Figures C.7 and C.8 show data-to-simulation comparison distributions for all of the MVA inputs in the $3\ell 3j$ region, after performing the jet-multiplicity-based reweighting.

MVA inputs in 4ℓ signal regions

Figure C.9 shows the data-to-simulation comparison for the MVA inputs in the 4ℓ off-shell SF signal region, after performing the jet-multiplicity-based reweighting.

Figure C.10 shows the data-to-simulation comparison for the MVA inputs in 4ℓ on-shell SF signal region, after performing the jet-multiplicity-based reweighting.

Figure C.11 shows the data-to-simulation comparison for the MVA inputs in 4ℓ DF signal region, after performing the jet-multiplicity-based reweighting.



(a) (b)

(c) (d)

(e) (f)

Figure C.1: Data/MC comparison in the $3\ell 1j$ region for: (a) H_{T}, (b) $H_{\mathrm{T}}^{\mathrm{lep}}$, (c) the leading lepton p_{T}, (d) the subleading lepton p_{T}, (e) the third leading lepton p_{T} and (f) the leading jet p_{T}. Only statistical uncertainties are shown.

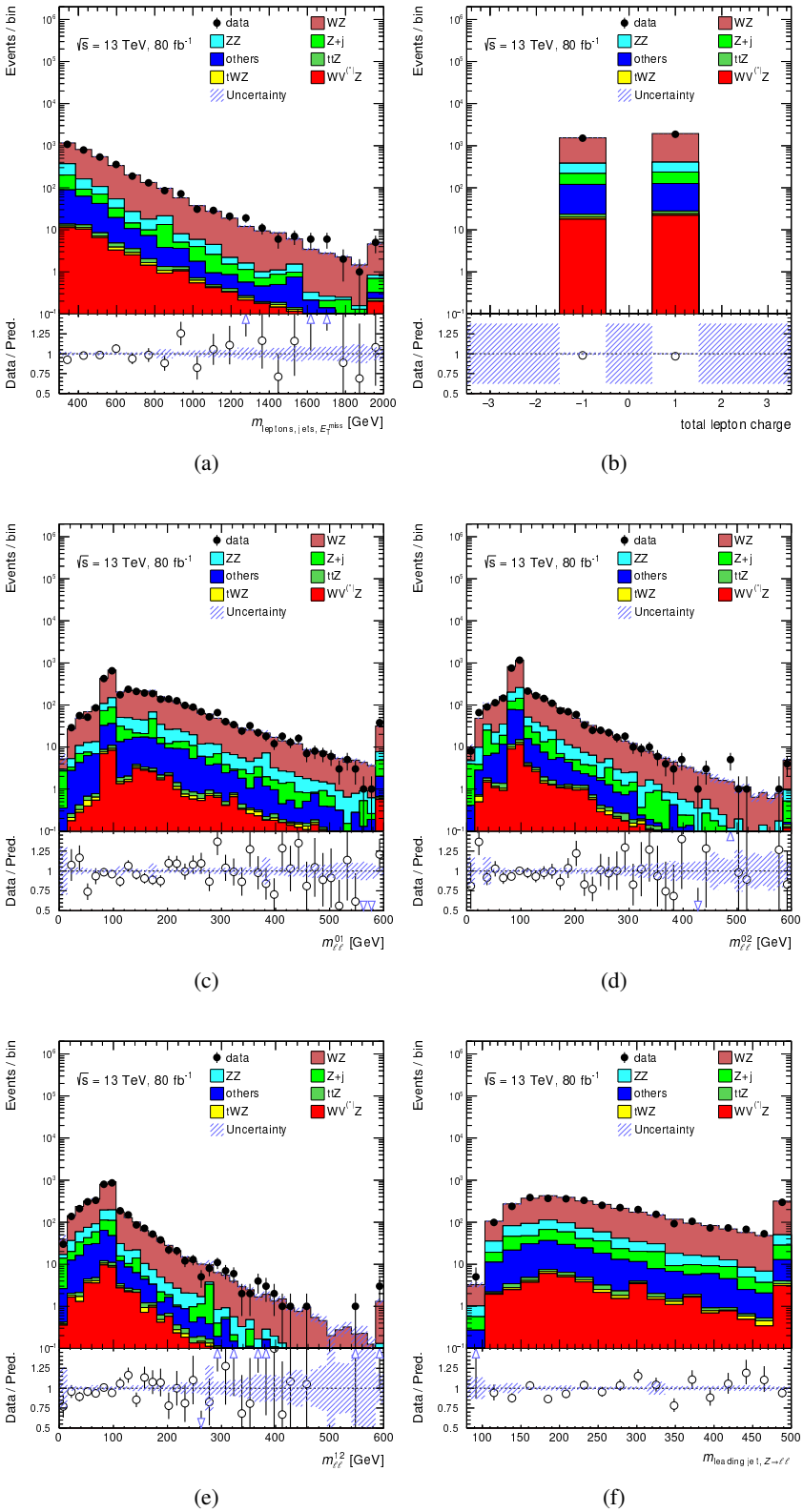


Figure C.2: Data/MC comparison in the $3\ell 1j$ region for: (a) the invariant mass of leptons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$, (b) the total lepton charge, (c) $m_{\ell\ell}^{01}$, (d) $m_{\ell\ell}^{02}$, (e) $m_{\ell\ell}^{12}$ and (f) the invariant mass of best $Z \to \ell\ell$ candidate and the leading jet. Only statistical uncertainties are shown.

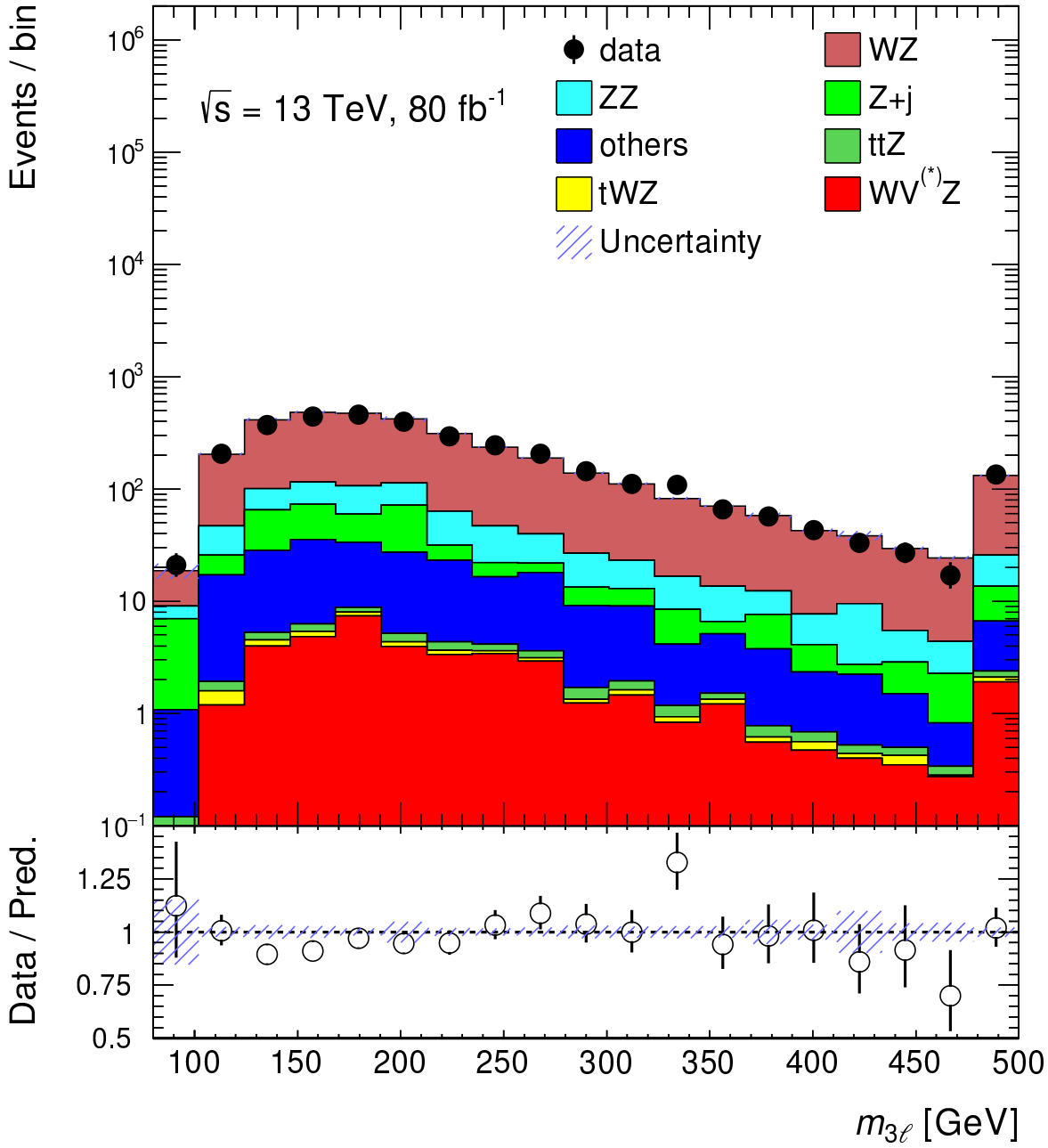


Figure C.3: Data/MC comparison in the $3\ell 1j$ region for the three-lepton invariant mass. Only statistical uncertainties are shown.

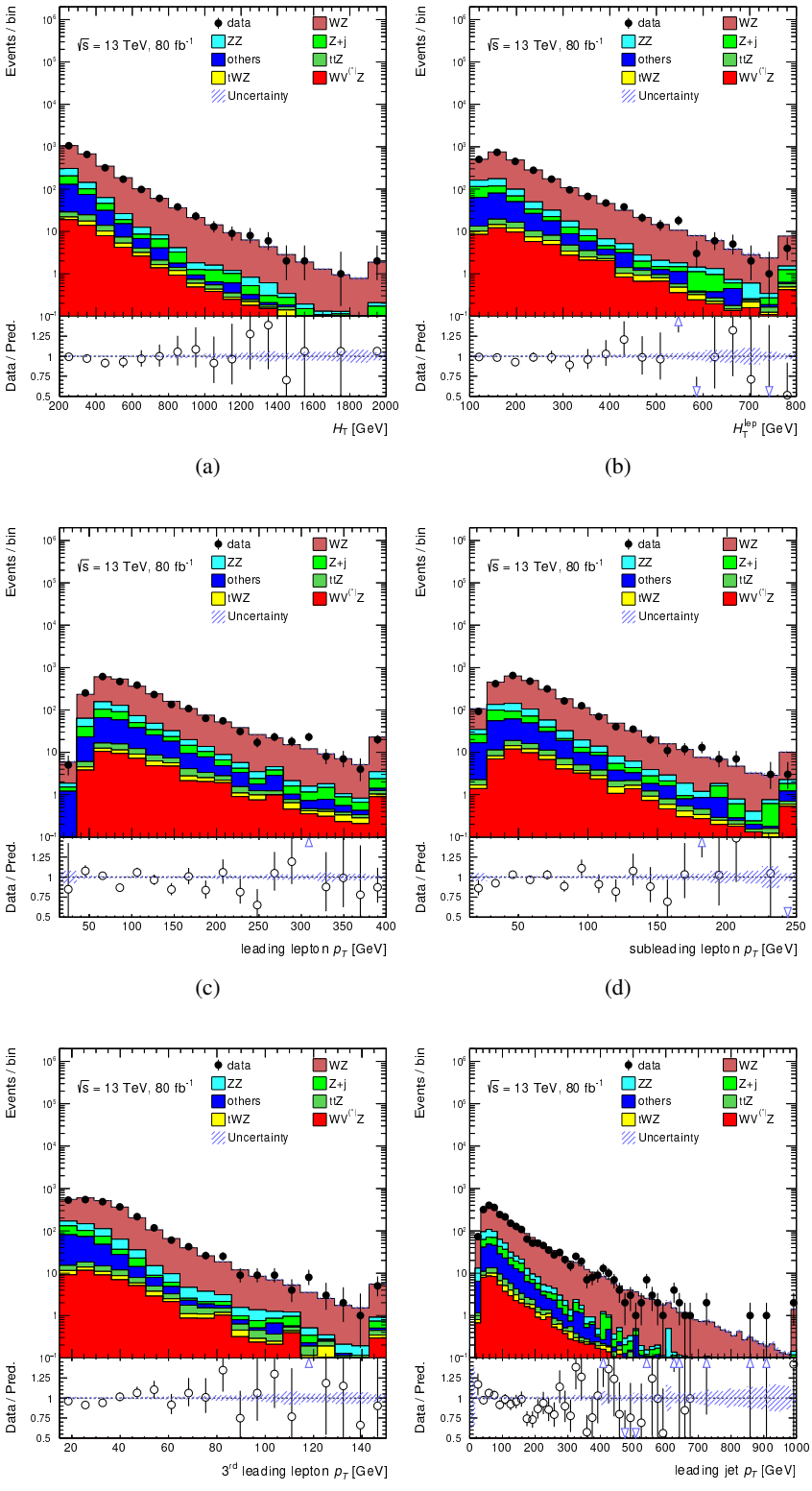
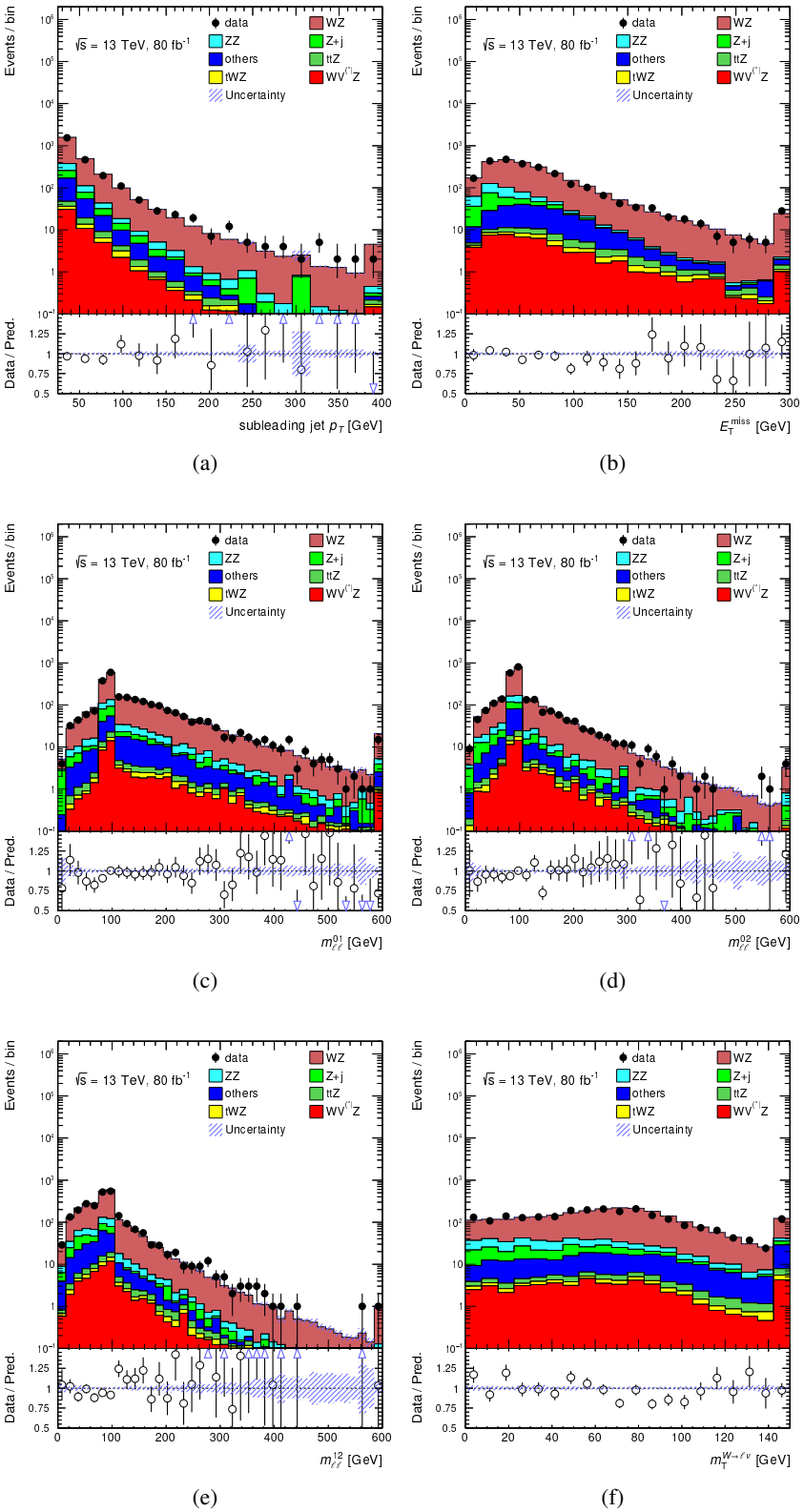


Figure C.4: Data/MC comparison in the $3\ell 2j$ region for: (a) H_{T}, (b) $H_{\mathrm{T}}^{\mathrm{lep}}$, (c) the leading lepton p_{T}, (d) the subleading lepton p_{T}, (e) the third leading lepton p_{T} and (f) the leading jet p_{T}. Only statistical uncertainties are shown.

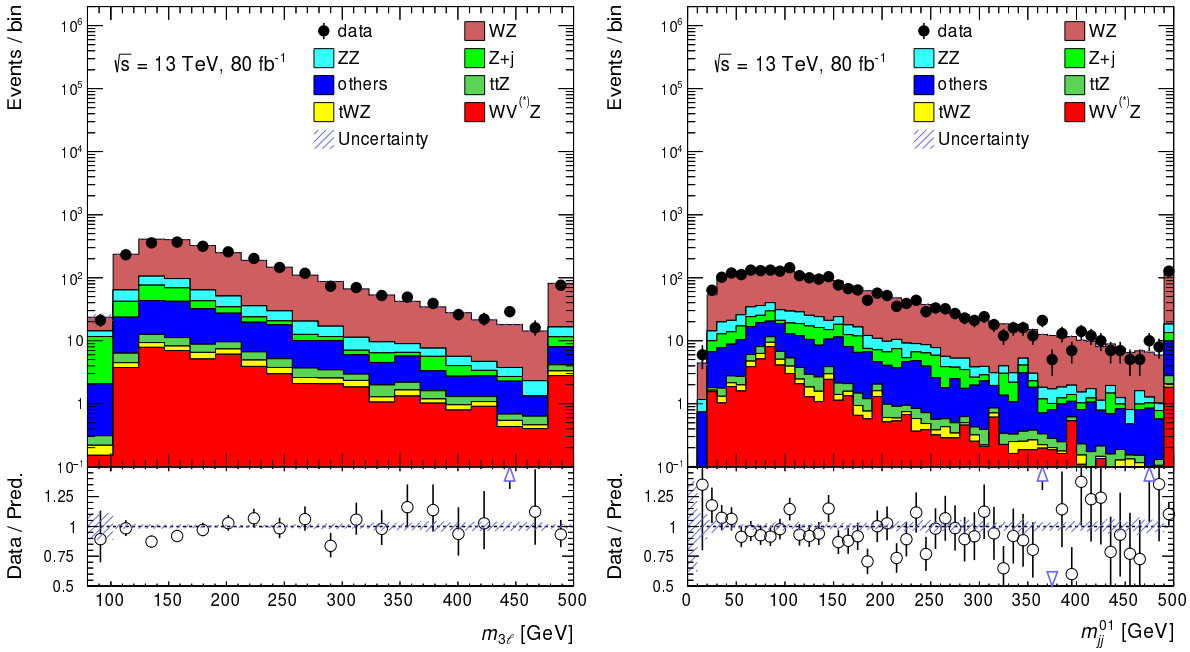


(a) (b)

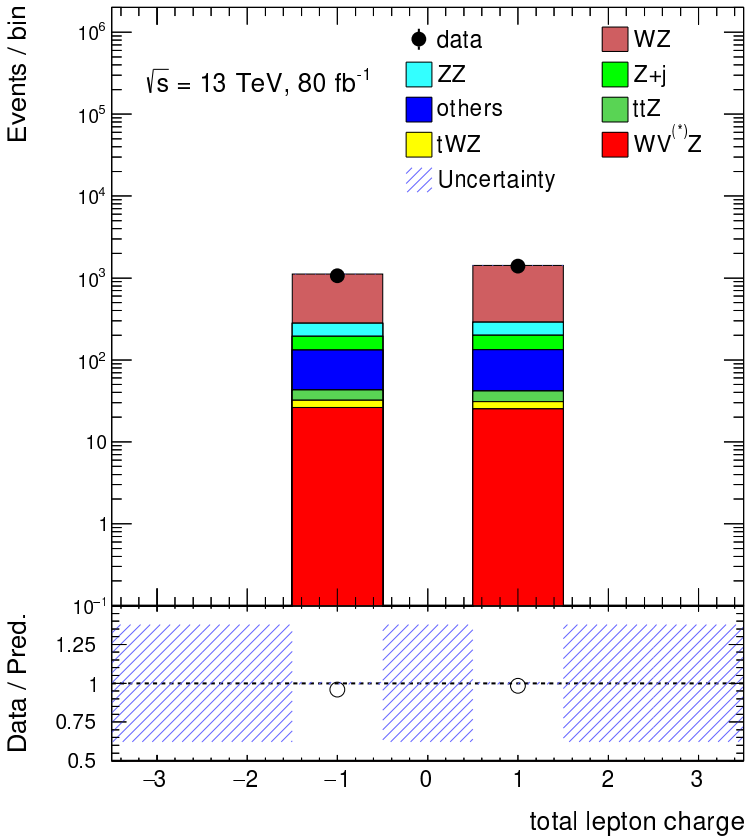
(c) (d)

(e) (f)

Figure C.5: Data/MC comparison in the $3\ell 2j$ region for: (a) the subleading jet p_{T}, (b) $E_{\mathrm{T}}^{\mathrm{miss}}$, (c) $m_{\ell\ell}^{01}$, (d) $m_{\ell\ell}^{02}$, (e) $m_{\ell\ell}^{12}$ and (f) the transverse mass of the best $W \to \ell\nu$ candidate. Only statistical uncertainties are shown.

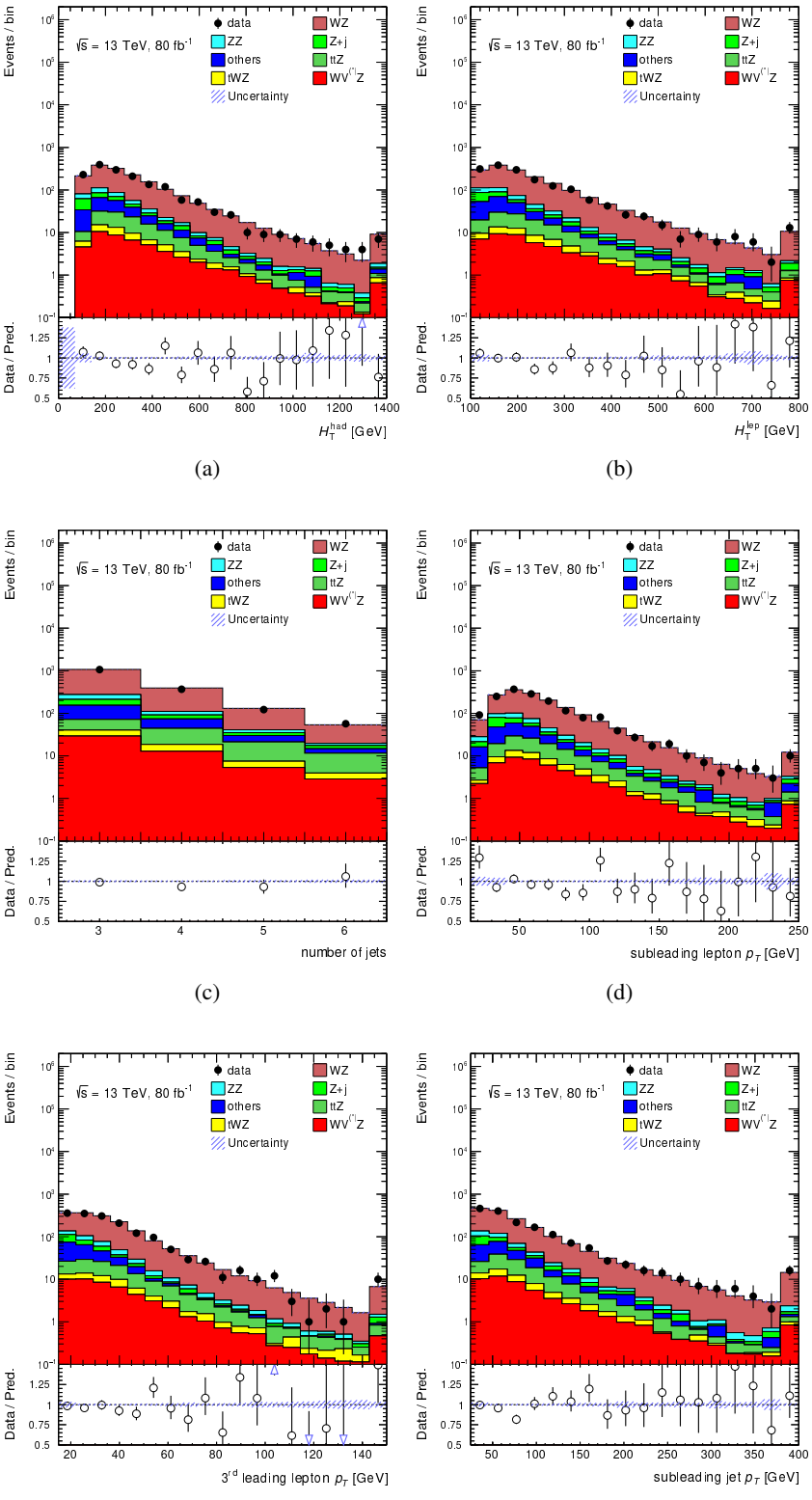


(a) (b)



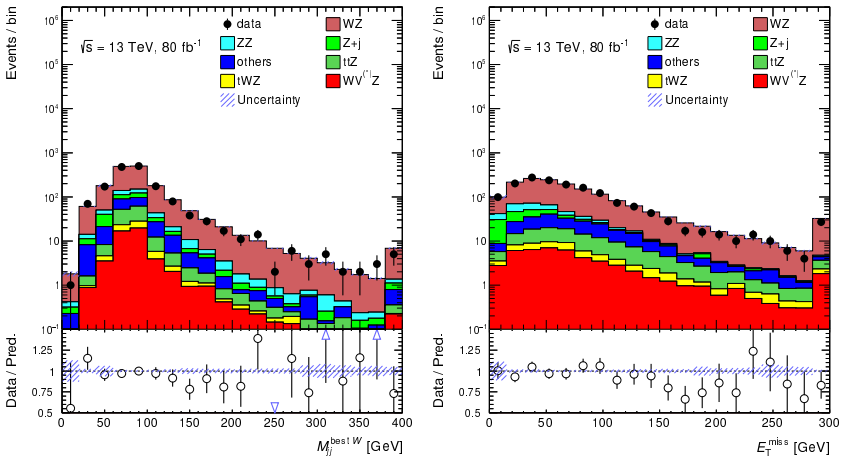
(c)

Figure C.6: Data/MC comparison in the $3\ell 2j$ region for the (a) three-lepton invariant mass, (b) two-jet invariant mass and (c) total lepton charge. Only statistical uncertainties are shown.



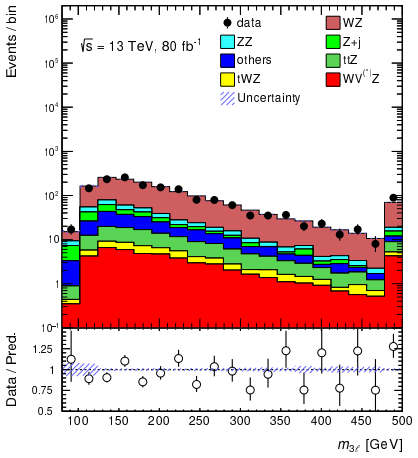
(a) (b) (c) (d) (e) (f)

Figure C.7: Data/MC comparison in the $3\ell 3j$ region for: (a) $H_{\mathrm{T}}^{\mathrm{had}}$, (b) $H_{\mathrm{T}}^{\mathrm{lep}}$, (c) the number of reconstructed jets, (d) the subleading lepton p_{T}, (e) the third leading lepton p_{T} and (f) the subleading jet p_{T}. Only statistical uncertainties are shown.

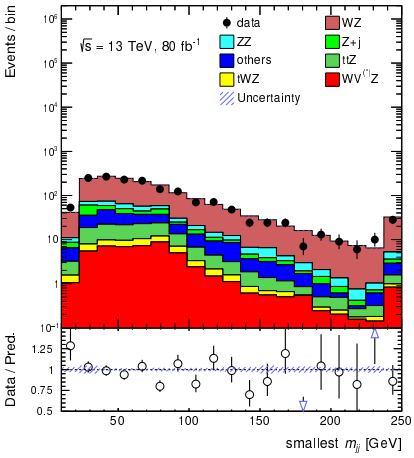


(a)

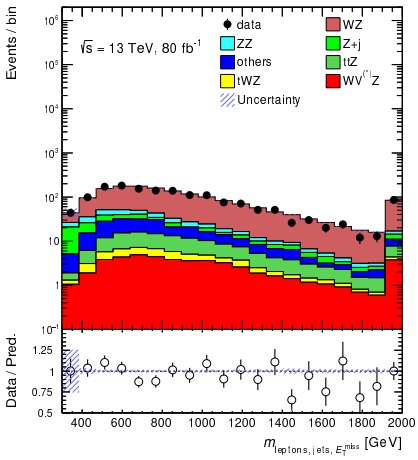
(b)



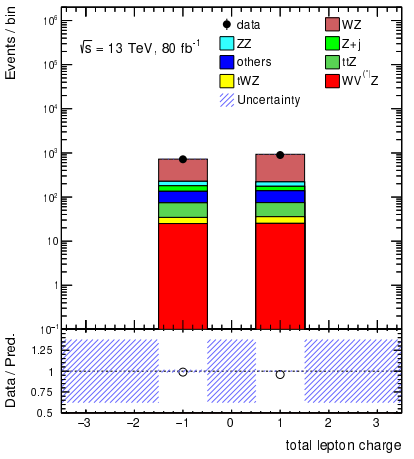
(c)



(d)



(e)



(f)

Figure C.8: Data/MC comparison in the $3\ell 3j$ region for: (a) the $W \to jj$-candidate invariant mass, (b) $E_{\text{T}}^{\text{miss}}$, (c) the three-lepton invariant mass, (d) the smallest two-jet invariant mass, (e) the invariant mass of leptons, jets and $E_{\text{T}}^{\text{miss}}$ and (f) the total lepton charge. Only statistical uncertainties are shown.

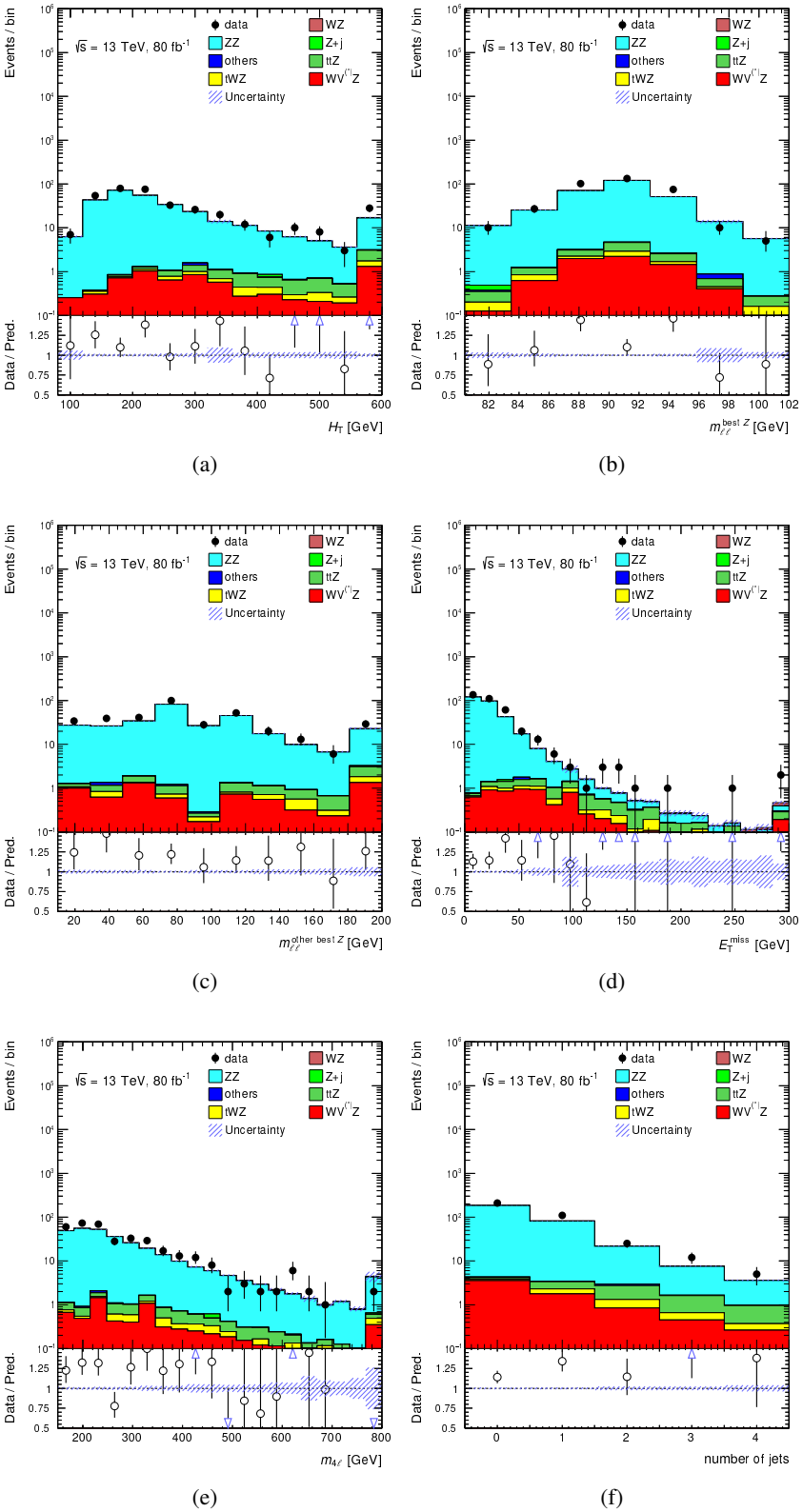
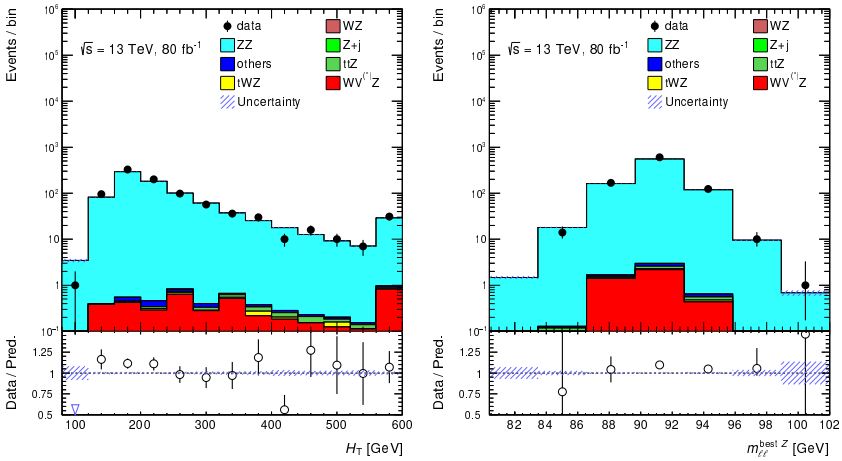
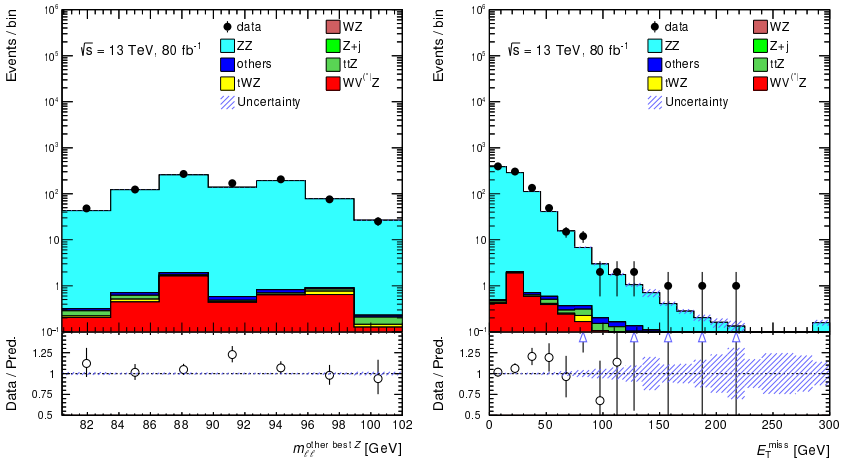


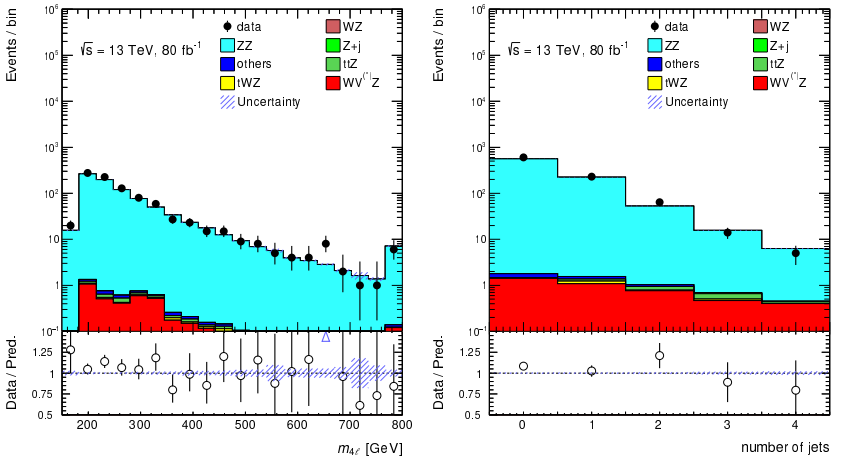
Figure C.9: Data/MC comparison in the off-shell SF signal region for: (a) H_{T}, (b) the best $Z \to \ell\ell$-candidate mass, (c) the second best pair invariant mass, (d) $E_{\mathrm{T}}^{\mathrm{miss}}$, (e) the four-lepton mass and (f) the number of reconstructed jets. Only statistical uncertainties are shown.



(a) (b)



(c) (d)



(e) (f)

Figure C.10: Data/MC comparison in the on-shell SF signal region for: (a) H_{T}, (b) the best $Z \to \ell\ell$-candidate mass, (c) the second best pair invariant mass, (d) $E_{\mathrm{T}}^{\mathrm{miss}}$, (e) the four-lepton mass and (f) the number of reconstructed jets. Only statistical uncertainties are shown.

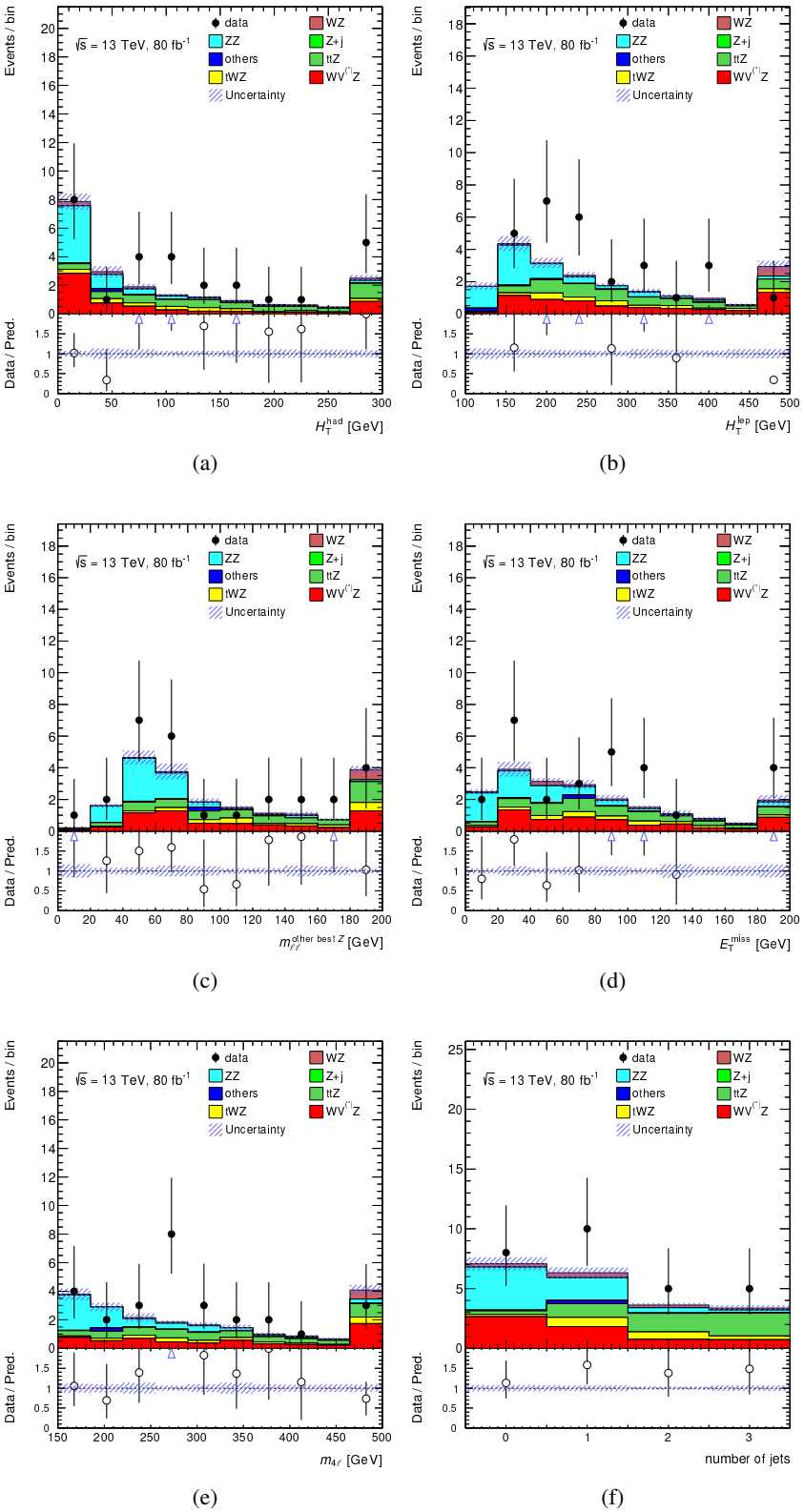


Figure C.11: Data/MC comparison in the DF signal region for: (a) $H_{\mathrm{T}}^{\mathrm{had}}$, (b) $H_{\mathrm{T}}^{\mathrm{lep}}$, (c) the second best pair invariant mass, (d) $E_{\mathrm{T}}^{\mathrm{miss}}$, (e) the four-lepton mass and (f) the number of reconstructed jets. Only statistical uncertainties are shown.

Acknowledgements

I want to express my enormous gratitude to my supervisor PD Dr. Markus Cristinziani, for his precious and careful guidance during these three rich, fruitful and exciting years. His help and support have been key elements to achieve this goal. Thank you for all the long discussions that made this work possible. I would also like to thank the European Research Council (ERC) for their financial support and the council of Bonn-Cologne Graduate School of Physics and Astronomy (BCGS) for offering me a two-year PhD excellence ("honors branch") program.

I sincerely thank Dr. Vadim Kostyukhin, for his essential and constant support in these years. Working and collaborating with him has been fundamental for the completion of the works presented in this dissertation.

Special thanks to Prof. Dr. Klaus Desch, for his precious support during these years.

I want to thank Prof. Dr. Norbert Wermes for hosting our group and for the big opportunity I have had to work in the welcoming and stimulating atmosphere of the ATLAS Bonn group.

Thanks to Alessandra Betti, Dr. Nello Bruscino, Dr. Julien Caudron, Matei Climescu, Dr. Mazuza Ghneimat, Carlo Alberto Gottardo, Sebastian Heer, Dr. David Hohn, Dr. Liza Mijović, Omer Oğul Öncel, Nikolaus Owtscharenko, Keshava Prasad Gubbi, Arshia Ruina and Dr. Kaven Henry Yau Wong for their valuable suggestions and friendliness during these years. I learnt a lot from each of you.

Thanks to my parents and my brother, without your support all of this would not have been possible. From you I learnt to follow my own aspirations, no matter how impervious is the road to reach my goals.

Thank you Eleonora, always present in these years to support me. Thanks for your unconditional love and your everlasting presence in difficult times. My words will never be enough to express my gratitude and affection.