

UNIVERSITÀ DEGLI STUDI “ROMA TRE”



Dipartimento di Matematica e Fisica
XXXIII Cycle of the Doctoral School in Physics

Ph.D. Thesis

**Search for Higgs boson pair production
in the single-lepton $WWb\bar{b}$ channel
with the ATLAS detector**

Valerio D'Amico

Supervisor
Prof. B. Di Micco

Ph.D. Coordinator
Prof. G. Degrandi



Contents

Introduction	v
Summary of accomplishments and innovations achievements	vii
1 Theoretical Introduction	1
1 Introduction to the Standard Model of Particle Physics	1
1.1 Weak Interactions and Electroweak Model	3
1.2 Spontaneous Symmetry Breaking and Higgs Mechanism	5
2 Particle production in proton-proton interactions	8
3 The Standard Model Higgs boson	10
3.1 Production processes	10
3.2 Decay modes	11
3.3 Higgs self-coupling	13
3.4 Higgs pair production	13
4 Beyond the Standard Model	15
2 The Large Hadron Collider and the ATLAS Experiment	19
1 The Large Hadron Collider	19
1.1 The accelerator complex	19
1.2 Luminosity and pile-up	21
2 The ATLAS Experiment at the LHC	23
2.1 The ATLAS detector	23
2.1.1 Coordinate system	24
2.1.2 Magnet system	24
2.1.3 Inner Detector	25
2.1.4 Calorimeter system	27
2.1.5 Muon Spectrometer	29
2.1.6 Trigger system	31
2.2 Physics objects definition and reconstruction	32
2.2.1 Tracks and vertices	33
2.2.2 Electrons	34
2.2.3 Muons	35
2.2.4 Taus	37
2.2.5 Hadronic Jets	38
2.2.6 Missing Transverse Energy	41
2.2.7 Overlap Removal	42
3 Higgs boson at the Large Hadron Collider	43
1 Discovery and measurements	43
2 Higgs-pair production: state of the art	47
2.1 ATLAS results	47

2.2	CMS results	52
4	Analysis of the Higgs boson pair production in the $bbWW^*$ decay channel	55
1	Previous measurements	55
2	Single lepton final state overview	58
3	Data and simulated samples	59
3.1	Data samples	59
3.2	Monte Carlo samples	59
3.2.1	Signal samples	60
3.2.2	Background samples	61
4	Object and event selection	62
4.1	Object selection	62
4.2	Event selection	67
4.2.1	Trigger requirements	67
4.2.2	ATLAS derivations	70
4.2.3	Event pre-selection	71
4.2.4	Signal topology selection	72
4.2.5	Higgsness and Topness	74
5	Machine Learning based analysis	76
5.1	Introduction to Machine Learning techniques	77
5.1.1	Boosted Decision Trees	78
5.1.2	Artificial Neural Networks	79
5.2	Deep Neural Network model for the $bbWW^*$ analysis	82
5.2.1	Input features	85
5.2.2	DNN model performances	95
6	Estimation of backgrounds	98
6.1	Definition of the analysis regions	98
6.2	Multijet background estimate	101
6.3	Background normalisation	106
7	Systematic uncertainties	110
7.1	Experimental uncertainties	110
7.2	Theoretical and modelling uncertainties	113
7.2.1	$t\bar{t}$ modelling	115
7.2.2	$W+jets$ modelling	116
8	Statistical analysis	117
8.1	Statistical model	118
8.2	Fit model for the $bbWW^*$ analysis	121
8.3	Background-only fit	123
9	Results	130
9.1	Impact of systematic uncertainties	131
	Conclusions	133
	Bibliography	135
A	Cut-based and BDT truth level studies	145
B	Fake factors for multijet background estimation	149
C	MET trigger scale factors	151
D	Systematic uncertainties list	157

Introduction

The Standard Model of particle physics (SM) is the theory that currently best describes the elementary particles of which the matter is made of and their interactions. Since its introduction in the second half of the 20th century, it has been tested and validated numerous times, giving highly precise predictions of the existing experimental observations and providing predictions for the existence of new particles, leading to several discoveries in the experimental particle physics field. An extremely important success of the SM is the discovery of a new particle compatible with the SM Higgs boson, announced in 2012 by the ATLAS and CMS experiments at the Large Hadron Collider (LHC). The Higgs boson was predicted in 1964 to explain the problem of the mass generation of the SM fundamental particles, and it was not yet observed since then. Its observation has opened a new era in understanding the nature of electroweak symmetry breaking, and now precision measurements of the properties of the Higgs boson are very important to test the validity of the Brout-Englert-Higgs mechanism, the electroweak theory and the SM theory itself. Higgs boson properties are largely studied by both ATLAS and CMS experiments. Among these, the couplings between the Higgs boson and the fundamental particles of the SM are of particular interest as their values are predicted by the SM, given the measured values of the particle's masses and of the vacuum expectation value of the Higgs potential. In the SM the Higgs boson has a self-coupling, the trilinear λ_{HHH} coupling, which rules the interaction of the Higgs boson with itself and controls the shape of the Higgs potential. The value of this particular coupling can be directly accessed by studying the Higgs boson pair production process, which is predicted by the SM and can happen via gluon-gluon fusion (ggF) through top-quark loops, where an Higgs boson is first produced and then splits into an Higgs boson pair. The SM cross section for the Higgs pair production via ggF in LHC pp collision at a centre-of-mass energy of $\sqrt{s} = 13$ TeV is $\sigma(pp \rightarrow HH) = 31.05$ fb. This value is about three orders of magnitude smaller than the single Higgs boson production cross section, making the observation of this process particularly challenging with the statistics collected so far by LHC experiments. Searches for Higgs pair production process have been performed in several decay channels both by ATLAS and CMS using 2015 and 2016 collected data, corresponding to 36.1 and 35.9 fb⁻¹ respectively, at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. So far, the best observed (expected) upper limit on HH cross section has been obtained by the statistical combination of ATLAS searches and is equal to 6.9 (10) times the SM predicted cross section value. The current sensitivity to this process is unlikely to bring to an observation with the available statistics, for which the High Luminosity LHC program is needed. However, analysis of such a rare process is nevertheless important for the development and the improvement of analysis techniques in view of the future optimisation of the searches, and to test the eventual presence of deviations from the SM predictions which may yield a hint of physics beyond the SM (BSM).

This thesis presents a search for the SM Higgs boson pair production in the single-lepton $WWb\bar{b}$ channel using the full Run 2 data of pp collisions collected at a centre-

of-mass energy of $\sqrt{s} = 13$ TeV with the ATLAS detector and corresponding to an integrated luminosity of 139 fb^{-1} . The $WWb\bar{b}$ channel has the second largest branching fraction of 25%, exploiting the two leading Higgs decay channels $H \rightarrow b\bar{b}$ and $H \rightarrow WW$. This work focuses on the single-lepton final state, which is obtained requiring one W boson to decay hadronically and one to decay into a lepton and a neutrino. The semi-leptonic final state is a compromise between signal efficiency, that is smaller than the full hadronic final state but retains 30% of the $HH \rightarrow b\bar{b}WW$ events, and background reduction of the QCD induced multijet processes which hugely affect the hadronic final state. The dominant background of this analysis is the $t\bar{t}$ background, which, having a final state equal to the signal one, is very difficult to reduce. In addition, smaller background contributions come from W/Z +jets, multijet and single top-quark production processes. This channel has been already studied using 2015 and 2016 data, corresponding to 36.1 fb^{-1} . The result of this previous analysis was an upper limit of 300 times the SM prediction, not comparable with the most sensitive analyses. The aim of this thesis is the development of a new optimised analysis of the single-lepton $WWb\bar{b}$ channel, exploiting machine learning techniques to improve the rejection of the major backgrounds and the sensitivity of the analysis itself.

The thesis has the following structure. In Chapter 1 an introduction to the Standard Model, with the description of the electroweak model and of the Higgs mechanism, as well as the basic concepts of proton-proton collisions, are presented. Moreover, a detailed summary of single Higgs and di-Higgs production and decay processes at the LHC, is reported in Chapter 1 as well as a brief description of the problems of the SM and of possible BSM model predictions for resonant di-Higgs production.

Chapter 2 gives a description of the LHC accelerator complex and the ATLAS detector, giving details on the reconstruction techniques of physics objects.

The state of the art of single Higgs and di-Higgs experimental measurements performed at the LHC is reported in Chapter 3.

The main work of this thesis is described in detail in Chapter 4. First, an introduction to the $WWb\bar{b}$ channel is presented focusing on the single-lepton final state together with a brief summary of the previous di-Higgs searches covering the $WWb\bar{b}$ channel. Then the complete analysis of Higgs boson pair production in the single-lepton $WWb\bar{b}$ channel is described. Data and simulated samples used, as well as the description of the objects and event selection performed in this analysis are reported in detail. Then, an introduction to machine learning techniques and a detailed description of the deep neural network classifier developed and used for a further optimisation of the analysis event selection are presented. Background estimation and systematic uncertainties are then reported before the description of the statistical analysis performed for the extraction of the final result of this work on the cross section of the $HH \rightarrow b\bar{b}WW^* \rightarrow b\bar{b}l\nu qq$ process.

Summary of accomplishments and innovations achievements

The personal contribution to the work described in this thesis is fully contained in Chapter 4. A description of the theoretical framework and of the experimental apparatus needed to introduce the description of this work are, instead, presented in Chapters 1 and 2 respectively, while the state of the art of the di-Higgs searches are presented in Chapter 3.

My work has been devoted to the optimisation and the innovation of the $bbWW^*$ analysis, with respect to its previous iteration, covering many aspects which are here briefly summarised:

- Development and maintenance of the analysis framework.
- Optimisation of the object selection (see Section 4.1).
- Optimisation of the event selection (see Section 4.2).
- Study and introduction of new discriminating variables: Higgsness and Topness (see Section 4.2.5).
- Development of new Machine Learning models for the signal to background discrimination; in particular the selection of the input features, the development and the optimisation of a multi-class Deep Neural Network model (see Section 5).
- Definition of the analysis regions (see Section 6.1).
- Normalisation of the background contribution (see Section 6.3).
- Evaluation of the systematic uncertainties affecting the analysis (see Section 7).
- Optimisation of the statistical analysis, developing the fit model for the comparison of the observed data with the background-only hypothesis in a discovery fit (see Section 8).
- Extraction of the upper limit on the Higgs pair production cross section and evaluation of the impact of the systematic uncertainties (see Section 9).

The studies of the MET triggers (see Appendix C) and the estimation of the multi-jet background contribution (see Section 6.2) have been performed inside the analysis group.

Chapter 1

Theoretical Introduction

The Standard Model of particle physics (SM) is the theory that describes the elementary particles of which the matter is made of and their interactions. This theory has been born and developed in the second half of the 20th century and it has been tested and validated numerous times during this last 70 years. The SM gives highly precise predictions of the existing experimental observations. Moreover, it has been capable of providing predictions for the existence of new particles, leading to new discoveries in the experimental particle physics field.

There are however some experimental observations done in particle physics and astrophysics fields for which the SM is not able to provide an explanation. These observations leads to the possibility of the existence of a physics beyond the SM (BSM).

This chapter gives the basis of the theoretical framework needed for the description of this thesis work. A brief introduction to the SM is given in Section 1 together with the description of the weak interactions and the Electroweak model in Section 1.1 and the spontaneous symmetry breaking mechanism in Section 1.2. Section 2 briefly summarises the particle production in proton-proton interaction to prepare the field for a deeper description of the Higgs boson in Section 3. An overview of the Higgs production and decay modes is presented in Sections 3.1 and 3.2 respectively. Section 3.3 is dedicated to the description of the Higgs self-coupling and Section 3.4 contains the description of the Higgs pairs production modes. Finally an overview of possible BSM models linked to the subject of this thesis is given in Section 4.

1 Introduction to the Standard Model of Particle Physics

The Standard Model is currently the theory that better describes elementary particles and their interactions. The elementary particles are classified in two groups: fermions and bosons. Fermions are half-integer spin particles that follow Fermi-Dirac statistics and satisfy the Pauli exclusion principle, while bosons have integer spin and obey Bose-Einstein statistics. Fermions are classified in leptons and quarks, depending on the force they are subject to: leptons and quarks interact via the electromagnetic and weak force, while only quarks interact with the strong force. For this property they have an additional quantum number with respect to leptons, related to the strong interaction, the colour charge (often referred to red, green and blue). Leptons and quarks are further divided into three doublets, called generations or families, of increasing mass:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \quad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \quad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix} \quad \begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}. \quad (1.1)$$

Each lepton doublet has one negative charged lepton and one neutrino matching the flavour of the corresponding lepton. The electron is the lightest charged lepton, followed by the muon, while the tau is the heaviest. Neutrinos are considered to be massless in the SM scenario, but there are experimental evidences of their oscillation [1, 2], for which it is required a neutrino mass different from zero. Quarks are classified in quark-up like with $+2/3$ electric charge and quark-down like with $-1/3$ electric charge, forming a doublet. According to the Dirac equation all these fermions have an associated anti-particle, having the same mass and opposite quantum numbers. In Table 1.1 are summarized the leptons and the quarks with their charge and their mass.

Table 1.1: Charge Q , in units of electron charge e , and mass, in units of GeV, of leptons (a) and quarks (b). For the neutrinos are reported the mass limits related to the flavour eigenstates reported by [3].

(a)			(b)		
Leptons	Q/e	mass [GeV]	Quarks	Q/e	mass [GeV]
Electron (e)	-1	0.511×10^{-3}	Up (u)	$+2/3$	$2.2^{+0.5}_{-0.3} \times 10^{-3}$
Electron neutrino (ν_e)	0	$< 2 \times 10^{-9}$	Down (d)	$-1/3$	$4.7^{+0.5}_{-0.2} \times 10^{-3}$
Muon (μ)	-1	106×10^{-3}	Charm (c)	$+2/3$	1.27 ± 0.02
Muon neutrino (ν_μ)	0	$< 0.19 \times 10^{-3}$	Strange (s)	$-1/3$	$93^{+11}_{-5} \times 10^{-3}$
Tau (τ)	-1	1.777	Top (t)	$+2/3$	172.9 ± 0.4
Tau neutrino (ν_τ)	0	$< 18.2 \times 10^{-3}$	Bottom (b)	$-1/3$	$4.18^{+0.03}_{-0.02}$

Quarks are not color singlets and for this reason they are not observable alone. Hadrons are color singlet particles composed by quarks. They can be made of three quarks, in this case they are called baryons and have half-integer spin. They can also be made of a quark-antiquark pair and in this case they are called mesons and they have integer-spin. Fermions interact through the exchange of mediators particles, called gauge bosons:

- photon (γ), mediator of the electromagnetic interaction between charged particles;
- $W^\pm$ and Z bosons, mediators of the weak interaction, responsible of processes like nuclear decays;
- gluons (g), mediators of the strong interaction, responsible of the attractive force between the quarks in neutrons and protons, and between neutrons and protons inside the atom's nucleus;
- graviton (G), theorized mediator of the gravitational interaction, not predicted by the SM.

In Table 1.2 are reported the main properties of the gauge bosons.

In addition to all these particles there is the Higgs boson, which is a neutral fundamental scalar particle introduced in the Standard Model in order to generate the masses of the gauge bosons and of all the other elementary particles. It will be better described in Sections 1.2 and 3.

The Standard Model is a Quantum Field Theory (QFT), i.e. a theory having quantum fields as fundamental objects. The fundamental interactions are invariant under local gauge transformations. In a Gauge Theory, the interactions are described by a Lagrangian, which is invariant under some group of transformations, called symmetries.

Table 1.2: Charge Q , in units of electron charge e , spin and mass, in units of GeV, of SM gauge bosons [3].

Gauge boson	Q/e	Spin	mass [GeV]
Photon (γ)	0	1	$< 10^{-27}$
$W^\pm$	± 1	1	80.379 ± 0.012
Z	0	1	91.1876 ± 0.0021
Gluons (g)	0	1	0 (theoretical value)
Graviton (G)	0	2	$< 6 \times 10^{-41}$

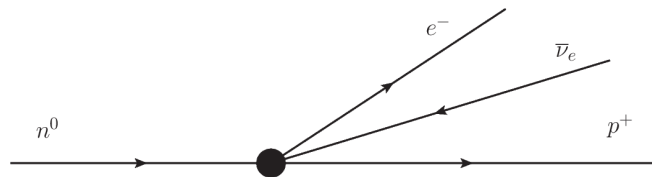
According to the Noether's Theorem [4], for each symmetry, the associated current and charge are conserved. If the symmetry is local, the quantities are conserved at the space-time point where the interaction occurs. In a QFT, in order to get the theory invariant under local transformations, vector boson fields are introduced, interacting with the fundamental fields in a gauge invariant manner.

The SM, in particular, is based on the gauge symmetry $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. The $U(1)$ group has one gauge field and indicates the Quantum Electrodynamics (QED) theory. This theory describes the interaction between electrically charged particles and the mediator of the electromagnetic interaction, i.e. the gauge boson of this group, the photon. The $SU(2)$ group has three gauge bosons and indicates the Weak interactions. In this case the gauge bosons are the $W^\pm$ and Z bosons. The $SU(3)_C$ group has eight gauge bosons that are the gluons of the Quantum Chromodynamics (QCD) theory, the non-abelian theory that describes the strong interactions between quarks and gluons. The "C" pedix stands for the color charge that is exchanged in the strong interactions. $SU(2)_L \otimes U(1)_Y$ indicates the electroweak symmetry group, which unifies electromagnetic and weak interactions in the so-called "electroweak theory". This theory is further described in Section 1.1.

1.1 Weak Interactions and Electroweak Model

The weak interactions have been proposed by Enrico Fermi in 1934 [5] to explain the process of the β decay of the neutron: $n \rightarrow p + e^- + \bar{\nu}_e$. This interaction was characterised by a smaller relative intensity compared to the electromagnetic and strong interactions. This property is visible by looking at the much bigger lifetimes of the particles weakly decaying compared to electromagnetic and strong interaction lifetimes, e.g. the weakly decaying charged pion with lifetime of $O(10^{-8})$ compared to the electromagnetic decaying neutral pion with lifetime of $O(10^{-16})$.

The original idea of Fermi was a point-like vectorial current (V) interaction between the particles taking part in the process, visible in Figure 1.1.

Figure 1.1: Fermi point-like interaction between the four fermions of the neutron β decay.

This theory was a good approximation at low energy, but it did not explain the experi-

mental observation that weak interactions violate parity. For this reason, Fermi's theory was extended introducing to the model an axial (A) term which conserves its sign under parity transformations. In this way the violation of parity comes from the combination of the vectorial and axial currents with the form " $V - A$ " [6, 7].

Weak interactions involve only left-handed chiral particles and right-handed chiral anti-particles. The weak field is invariant under $SU(2)_L$ transformations, where the subscript "L" indicates that only left-handed particles participate to these interactions. There are two types of weak interactions: the charged-current interaction mediated by W^+ or W^- bosons, and the neutral-current mediated by the Z^0 boson.

The Electroweak theory has been formulated in late 60's by Weinberg [8] and Salam [9] to unify the electromagnetic and weak interactions. This theory is based on the $SU(2)_L \otimes U(1)_Y$ symmetry group to describe the lepton and quark interactions with photons and weak mediators.

The generators of $SU(2)$ and $U(1)$ are the weak isospin T and weak hypercharge Y_W . The electric charge arises as a linear combination of Y and the T_3 component of weak isospin, that are invariant under gauge transformations:

$$Q = T_3 + \frac{1}{2}Y_W \quad (1.2)$$

The gauge bosons of the electroweak interaction arise from these generators. They are organized in the weak isospin triplet W_1, W_2 , and W_3 , and weak hypercharge singlet B , and corresponds to the $\gamma, W^\pm$ and Z bosons. The propagation of the weak boson fields definitions are the following:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.3)$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon^{abc}W_\mu^b W_\nu^c \quad (1.4)$$

where g is the electroweak coupling and ϵ^{abc} is the Levi-Civita tensor.

The Lagrangian of the electroweak interaction consists of the kinetic part of both $SU(2)$ and $U(1)$ and the interaction term between matter and fields:

$$\mathcal{L}_{EW} = \sum_{j=1}^3 i\bar{\psi}_j(x)\gamma^\mu D_\mu \psi_j(x) - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu} \quad (1.5)$$

where ψ is the fermionic field, γ^μ are the Dirac gamma matrices and D_μ is the covariant derivative defined as:

$$D_\mu = \partial_\mu + ig'\frac{Y}{2}B_\mu(x) + ig\frac{\sigma_a}{2}W_\mu^a(x) \quad (1.6)$$

where σ_a are the Pauli spin matrices.

The observable vectors fields are obtained with the following relations:

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2) \quad (1.7)$$

$$A_\mu = B_\mu \cos \theta_W + W_\mu^3 \sin \theta_W \quad (1.8)$$

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^3 \cos \theta_W \quad (1.9)$$

The B and W^3 fields are rotated by the Weinberg angle θ_W to generate the electromagnetic vector potential and the neutral weak vector potential.

At this stage the theory describes all the gauge bosons as massless. Fermionic masses are also forbidden, otherwise they would produce an explicit breaking of the gauge symmetry. The mass terms can be put by hand, but this would make the theory not any-more renormalizable. The experimental evidences of massive gauge bosons requires the masses of the particles to be present also in the theory. This discrepancy is resolved by the Higgs mechanism that introduces mass terms in the Standard Model Lagrangian, that is described in Section 1.2.

1.2 Spontaneous Symmetry Breaking and Higgs Mechanism

The problem of the mass generation in the Standard Model was solved by Higgs, Brout and Englert in 1964 [10, 11]. This new model proposed a spontaneous breaking of the $SU(2)_L \otimes U(1)_Y$ symmetry by the introduction of a scalar field in the electroweak Lagrangian. This mechanism is known as "Higgs mechanism" and provides the generation of the masses of the fundamental particles without breaking the gauge symmetry of the theory. The new scalar field is described by a complex scalar $SU(2)_L$ spin-0 doublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.10)$$

The corresponding Higgs Lagrangian is then added to the electroweak Lagrangian:

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) \quad (1.11)$$

where $V(\phi^\dagger \phi)$ is the Higgs potential:

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \quad \lambda > 0 \quad (1.12)$$

The shape of the potential in Equation 1.12 depends on the sign of μ^2 :

$$\mu^2 \geq 0: \quad \frac{\partial V}{\partial \phi} = 0 \rightarrow |\phi_{min}| = 0 \quad (1.13)$$

$$\mu^2 < 0: \quad \frac{\partial V}{\partial \phi} = 0 \rightarrow |\phi_{min}| = \sqrt{\frac{-\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} \quad (1.14)$$

where v is the vacuum expectation value $\langle 0 | \phi | 0 \rangle$.

In the case of $\mu^2 \geq 0$ the potential has a single point of minimum, or ground state. The vacuum expectation value (vev) vanishes and there is no spontaneous symmetry breaking. In the case of $\mu^2 < 0$ the minimum is degenerate along a circumference, as shown in Figure 1.2, and there are an infinite number of ground states with a non-zero vacuum expectation value.

The electroweak symmetry is then spontaneously broken by selecting one of these vacuum states. Since no positive charged permanent field is observed, the ground state ϕ_0 takes the form:

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.15)$$

According to the Goldstone theorem [13], the spontaneous breaking of a global symmetry must be accompanied by the appearance of new massless scalar particles, called

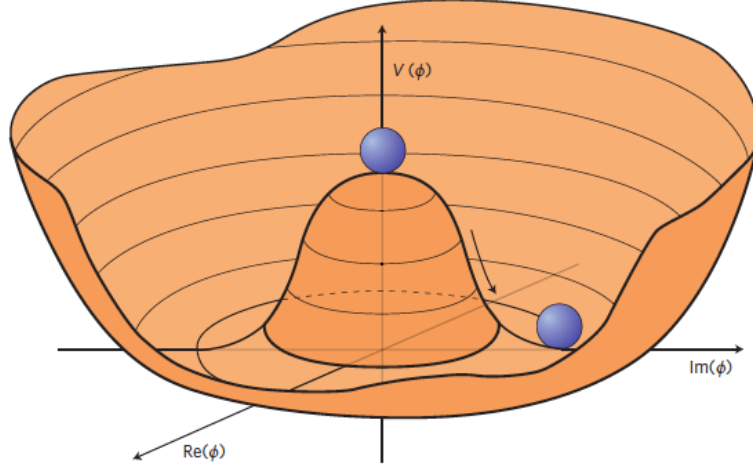


Figure 1.2: Illustration of the Higgs potential shape in the $\text{Re}(\phi), \text{Im}(\phi)$ space for the values of $\mu^2 < 0$ [12].

Nambu-Goldstone bosons. Expanding the ϕ field by a perturbation around the vacuum expectation value we obtain:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_1 + i\theta_2 \\ v + h(x) + i\theta_3 \end{pmatrix} \quad (1.16)$$

where the $\theta_{1,2,3}$ are the three Goldstone bosons generated in the spontaneous symmetry breaking, and the $H(x)$ is the scalar field that represents the physical massive Higgs boson. By a SU(2) transformation the three θ components can be removed, removing three Goldstone bosons. Thus, the final ϕ field becomes:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (1.17)$$

The corresponding Higgs Lagrangian expressed by use of this new field is then:

$$\begin{aligned} \mathcal{L}_{Higgs} = & \frac{1}{2} \partial_\mu H \partial^\mu H + \frac{g^2}{4} (v + H)^2 \left(W_\mu^\dagger W^\mu + \frac{1}{2 \cos^2 \theta_W} Z_\mu Z^\mu \right) - \mu^2 H^2 + \\ & - \lambda_{HHH} v H^3 - \frac{\lambda_{HHHH}}{4} H^4 \end{aligned} \quad (1.18)$$

The result of the Higgs mechanism is that $W^\pm$ and Z vector bosons have gained a mass, while the photon is left massless. Moreover, a new scalar neutral boson H has appeared in the theory. The first three terms in the Equation 1.18 represent the kinetic and the mass terms of the W and Z bosons, and the interaction between them and the Higgs boson. The last two terms describe the trilinear and quartic couplings of the Higgs scalar field, also known as self-interactions or self-couplings of the Higgs boson.

From the Higgs Lagrangian we can extract some interesting parameters of this model:

$$m_{W^\pm} = \frac{gv}{2} \quad (1.19)$$

$$m_Z = \frac{gv}{2 \cos^2 \theta_W} \quad (1.20)$$

$$v = (\sqrt{2} G_F)^{\frac{1}{2}} \sim 246 \text{ GeV} \quad (1.21)$$

$$\lambda_{HHH} = \frac{3m_H^2}{v} \quad (1.22)$$

$$\lambda_{HHHH} = \frac{3m_H^2}{v^2} \quad (1.23)$$

$$m_H = \sqrt{-2\mu^2} = \sqrt{-2\lambda v} \quad (1.24)$$

The self-couplings is related to the mass of the Higgs boson, that is a free parameter of the theory.

The masses of the fermions arises by adding mass terms via Yukawa couplings to the Lagrangian:

$$\mathcal{L}_{Yukawa} = - \sum_{f=\ell,q} y_f [\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi \psi_L] \quad (1.25)$$

where y_f are the Yukawa couplings matrices for leptons and quarks. After the spontaneous symmetry breaking, i.e. substituting the ϕ field with the one in Equation 1.17, the Yukawa Lagrangian becomes:

$$\mathcal{L}_{Yukawa}^{leptons} = - \frac{y_\ell}{\sqrt{2}} (v \bar{\ell}_L \ell_R + \bar{\ell}_L \ell_R H) + h.c. \quad (1.26)$$

$$\mathcal{L}_{Yukawa}^{quarks} = - \frac{1}{\sqrt{2}} v (y_d \bar{d}_L d_R + \tilde{y}_u \bar{u}_L u_R) - \frac{1}{\sqrt{2}} (y_d \bar{d}_L d_R + \tilde{y}_u \bar{u}_L u_R) H + h.c. \quad (1.27)$$

where, since for the quarks there are also right-handed down-type quarks, there is the need of the conjugate component of the doublet that give rise to the $\tilde{y}$ couplings.

The fermion masses are given by the following relation with the Yukawa coupling:

$$m_f = \frac{y_f v}{\sqrt{2}} \quad (1.28)$$

Inverting the Equation 1.28, it can be seen that the Higgs coupling with the fermions is proportional to the fermion mass.

The Yukawa couplings are free parameters of the theory and therefore also the fermion masses are not predicted by the theory. The neutrinos, instead, remain massless, but their masses can be added to the theory with a minimal extension of the SM.

The y_f matrices can be diagonalised in order to get the eigenvalues of the fermion masses. For the leptons this transformation has no effect due to the absence of right-handed neutrinos. While for the quarks the rotation to the mass eigenstate basis provides a mixing among the quark flavours. This mixing is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \quad (1.29)$$

that can be parametrised by three mixing angles, that control the mixing among each generation pair, and one complex phase, responsible for CP-violating phenomena of the electroweak interactions.

As a result of the mixing between the flavour eigenstates, the charged-current $W^\pm$ interactions can provide the change of the flavour of the interacting quark. Moreover, because of CKM matrix coupling depends on the quarks involved in the transition, there are no more equal couplings for all weak interactions, and the universality of the weak interactions is preserved requiring that V is a unitary matrix.

2 Particle production in proton-proton interactions

As described in Section 1, hadrons, e.g. the protons, are composite particles made of constituents called "partons". These constituents are the quarks and gluons, that behave as free particles because of the QCD asymptotic freedom regime [14, 15]. The structure of hadrons and the way they interact in high energy physics are described by the "parton model", introduced by Richard Feynman in the late 60's [16].

The constituents of the proton are three "valence" quarks, two up quarks and one down quark, uud . The quarks are bound by exchanging gluons, that can temporarily split into quark-antiquark pairs, called "sea" quarks. In proton-proton collisions the fundamental interaction occurs between partons. Both valence quarks, sea quarks and gluons take part in this interaction and they contribute carrying only a fraction x_i of the proton momentum. The center-of-mass energy of the partonic interaction $\sqrt{\hat{s}}$ is therefore smaller than the proton-proton center-of-mass energy s :

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s} \quad (1.30)$$

where x_1 and x_2 are the momentum fractions of the interacting partons from the two protons. The probability of finding a parton with a particular momentum fraction x_i inside the proton is described by the parton distribution functions, usually referred to PDFs. They depend on the parton type and on the momentum transfer scale Q^2 of the collision, as visible in Figure 1.3. The PDFs can be evolved from a Q^2 scale to any other Q^2 scale by numerically solving the "DGLAP" equations [17–19], as done in the calculations shown in Figure 1.3.

Gluons dominate the distribution at low x . As a result, at low partonic $\sqrt{\hat{s}}$ processes initiated by a gluon-gluon interaction have a much larger cross section at high energy hadron interaction, compared to the processes initiated by quark-quark interactions. In particular, in the Large Hadron Collider, the collisions are then mainly gluon-gluon fusions, i.e. hard scattering of two gluons.

The interactions in proton-proton collision can be split into the subsequent sub-processes that occur during the interaction, as described by the factorisation theorem:

- the initial state, with the proton partons taking part in the interaction and the "spectators" partons which do not contribute to the interaction;
- the hard scattering process between the interacting partons, that is described by the matrix element (ME) of the interaction equation;
- the parton shower (PS), i.e. the cascades of radiation produced from QCD processes and interactions, like emissions or splitting of gluons and quark-antiquark pair production;
- the hadronisation, i.e. the formation of visible color neutral hadrons from the quarks and gluons produced in the parton shower.

These steps are schematically shown in Figure 1.4.

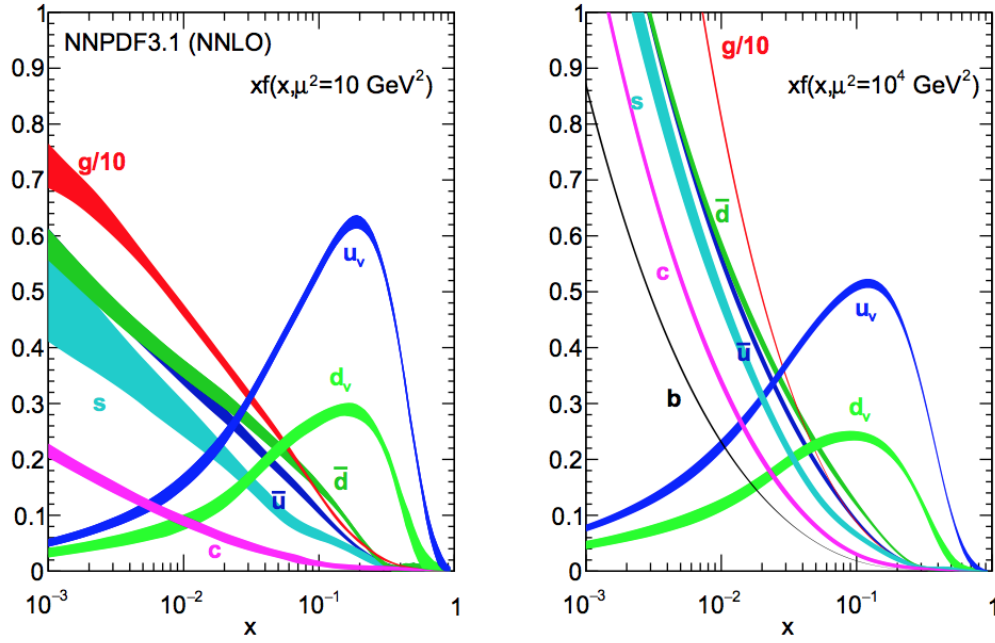


Figure 1.3: Parton distribution functions for protons for Q^2 scales (μ^2 in the plots) of 10 GeV^2 (left) and 10^4 GeV^2 (right) [20].

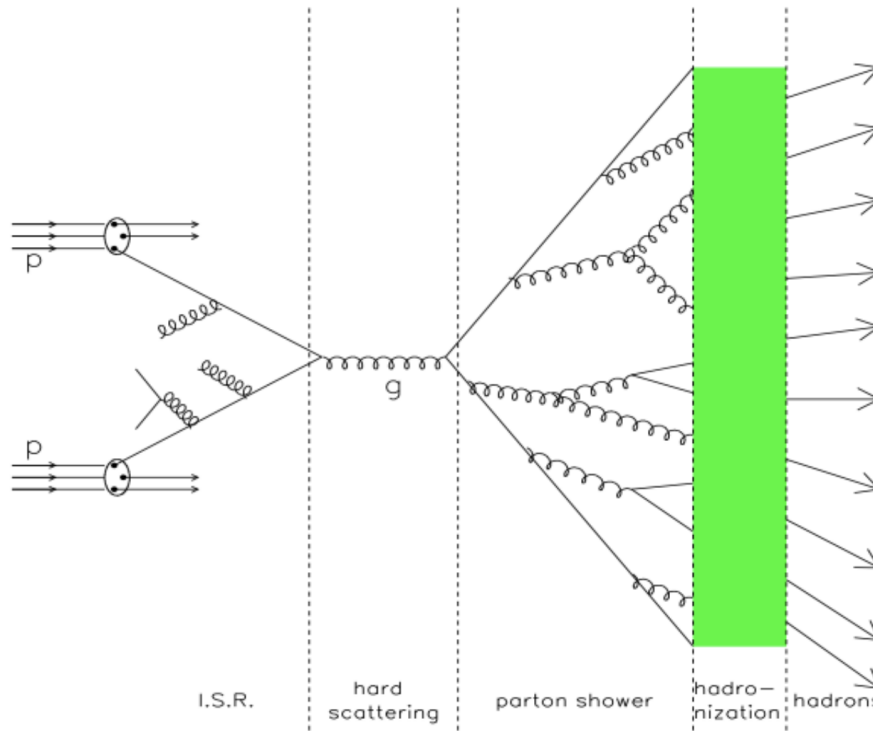


Figure 1.4: Schematic illustration of the different steps of proton-proton collisions.

The spectator partons that do not take part in the hard scattering process, continue their paths recombining themselves and hadronising. This secondary process is known as the "underlying event" of the collision.

Particle interactions are commonly described by the use of the Feynman diagrams, like the one represented in Figure 1.4. Both in the initial state and in the final state there is the possibility of having an additional emission of particles, like photons and

gluons. These processes are often referred to as initial state radiation and final state radiation, and represent the "real" corrections to the SM processes, since these emitted particles are visible in the event. In higher order of perturbation theory the diagrams can also contain "virtual" corrections represented as loops of particles, where there is an emission and an absorption of an elementary particle, like a boson or a quark. All these phenomena are known as next-to-leading order (NLO) corrections, and are of primary importance in the precise measurements of the Standard Model properties.

3 The Standard Model Higgs boson

The Higgs boson existence, as already described in Section 1.2, was predicted since the Higgs mechanism formulation in 1964. After many years from its prediction it was the only particle of the Standard Model not yet observed. Huge efforts have been made in the search for this particle at LEP [21] and at Tevatron [22]. Finally, the observation of a particle compatible with the SM Higgs boson with a mass of about 125 GeV was made by ATLAS and CMS in July 2012 [23, 24]. After the discovery, much progress has been made in the understanding and precision measurements and calculations of the electroweak symmetry breaking process. The properties of this new particle need to be studied in order to determine if they are in agreement with the predictions for the SM Higgs boson. The current reference value for the mass of the Higgs boson comes from the combination of the ATLAS and CMS results in the four leptons final state [25], and it is:

$$m_H = 125.09 \pm 0.24 \text{ GeV} \quad (1.31)$$

An overview of the main properties of the Higgs boson will be shown in the following sections. Starting with a description of the production and the decay modes for a single Higgs boson production, and then concentrating on the Higgs self-coupling and the processes of Higgs boson pair production that can make accessible the measure of this particular coupling. The description of the state of the art of the Higgs boson experimental measurements will be given instead in Chapter 3.

3.1 Production processes

The Feynman diagrams of the main production mechanisms of the Standard Model Higgs bosons are shown in Figure 1.8. They are the gluon-gluon fusion (ggF), the vector boson fusion (VBF), the associated production with a vector boson (VH , also referred to as *Higgs-strahlung*) and the associated production with a top quark pair (ttH). The cross sections of these processes depends both on the center-of-mass energy of the collision and on the combination of partons that take part in the initial interaction. In Figure 1.6a and 1.6b are shown the cross sections for each production mode as a function of the Higgs boson mass and as a function of the center-of-mass energy $\sqrt{s}$ respectively.

The gluon-gluon fusion is the most probable production mode of a Higgs boson. This is due to the big quantity of gluon-gluon interactions that happens in the proton-proton collisions at the energies of the Large Hadron Collider. In this process two gluons interact via a loop of quarks and produce an Higgs boson exiting from the third vertex of the triangle loop. Since the coupling between the Higgs boson and the quarks is dependent to the mass of the quark, the loop is composed mainly by the heaviest quark type, that is the top quark.

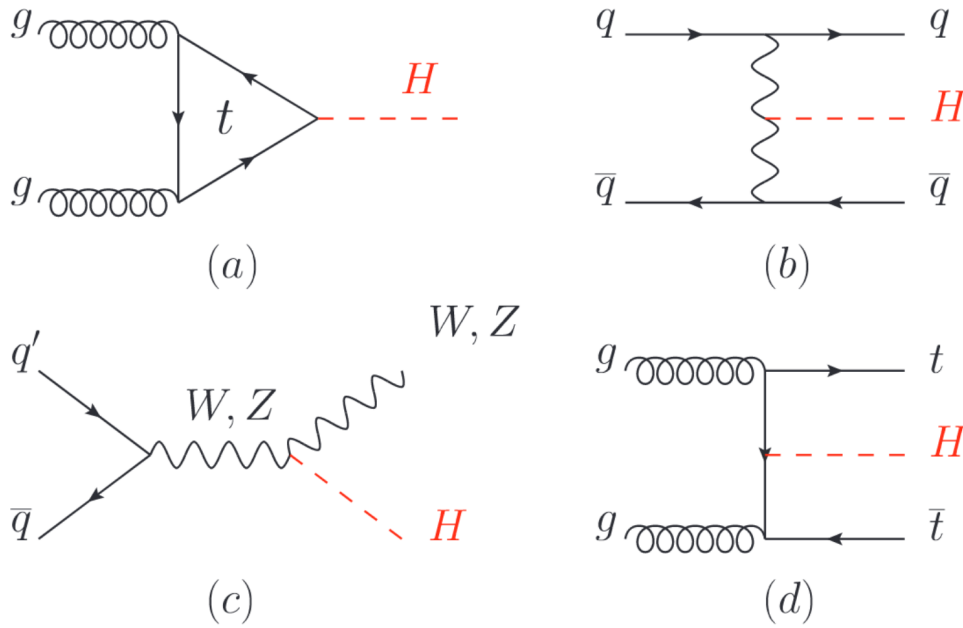


Figure 1.5: Main Leading Order Feynman diagrams contributing to the Higgs production: in (a) gluon-gluon fusion (ggF), (b) Vector-boson fusion (VBF), (c) Higgs-strahlung (or associated production with a gauge boson, VH), (d) associated production with a pair of top quarks (ttH) [3].

The vector boson fusion is the second largest production mode. In this process a quark and an anti-quark from the initial protons interact through the exchange of a virtual W or Z boson. The Higgs boson is then emitted by the exchanged vector boson. In the final state there are, thus, the Higgs boson decay products and two energetic jets in the forward regions of the detector, coming from the hadronisation of the quarks.

The Higgs-strahlung (VH) has the third largest cross section for the Higgs boson production. In this process the Higgs is emitted from a $W^\pm$ or a Z boson that is produced by the annihilation of a quark and an anti-quark. In the final state there are then the Higgs boson and the vector boson decay products.

Last, the associated production with top quarks has the smallest cross section among these production modes. The process shows in the final state two top quarks, that can be tagged by their decay, and the Higgs boson decay products.

3.2 Decay modes

The branching ratio (BR), or branching fraction, of a final state is defined as the fraction of times that a particle decays into this particular final state, among all the possible final states. The Higgs boson has several decay modes with different branching fractions, depending on the mass and on the couplings of the Higgs boson. The couplings of the Higgs boson with vector bosons are proportional to the mass squared of the vector boson, while the ones with the fermions are proportional to the fermion mass. In principle, then, the preferred decays should involve vector bosons. However, since these processes are not kinematically allowed due to the Higgs mass, these processes are suppressed and the decays with fermions are preferred. In Figure 1.7 are shown the branching ratios of the allowed Higgs boson final states as a function of m_H . It is visible that for Higgs boson masses higher than the kinematic threshold of WW production, the decays would be mostly in pairs of vector bosons. The value of the Higgs boson mass, instead, being in this particular region below this threshold, allows differ-

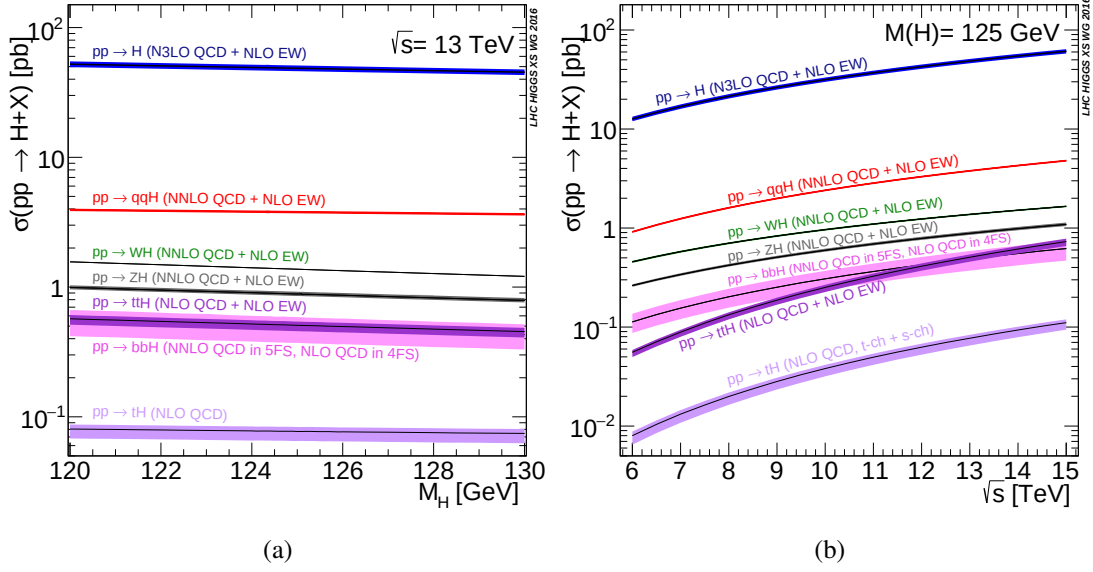


Figure 1.6: Standard Model Higgs boson production cross sections at $\sqrt{s} = 13$ TeV as a function of Higgs boson mass (a) and as a function of the centre-of-mass-energies $\sqrt{s}$ (b). The VBF process is indicated here as qqH [26].

ent decays with non negligible decay probabilities, making possible the measurements of its couplings with several particles.

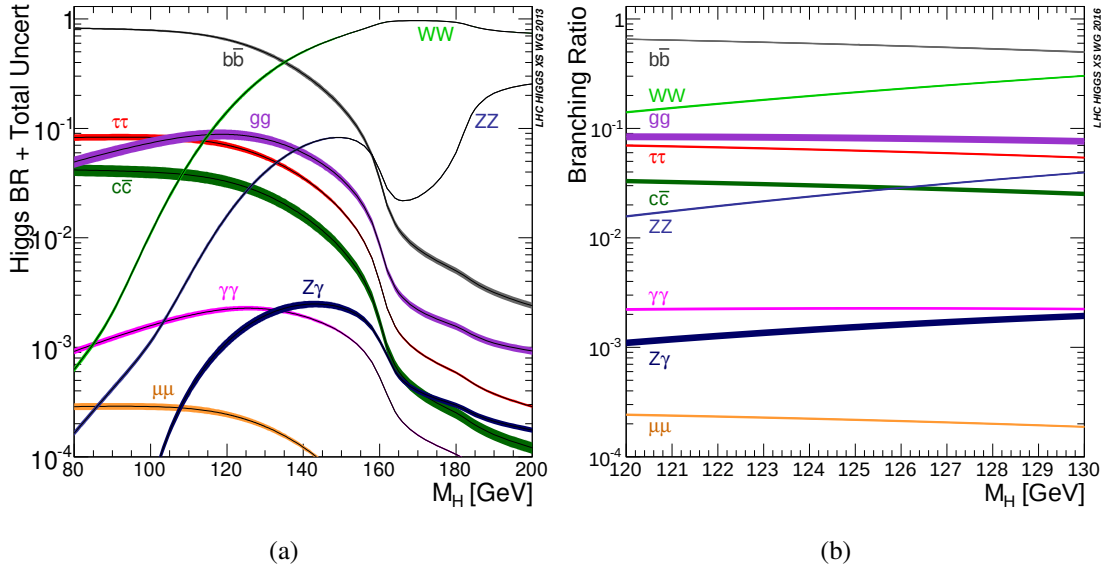


Figure 1.7: The decay branching ratios for the main decays of the SM Higgs boson: (a) for different Higgs boson masses [27], and (b) for the Higgs boson masses near $m_H = 125$ GeV [26].

In Table 1.3 are reported the branching fractions for a SM Higgs boson with a mass of $m_H = 125.09$ GeV. More than half of the times an Higgs boson decays into a pair of bottom quarks. The second largest BR belongs to the W^+W^- final state, that accounts for $\sim 21\%$ of the total decays. All the other final states have lower branching fractions, but some of them have a cleaner final state, that can provide a more precise measurement of the Higgs boson mass compared to the two leading decay channels. These channels are in particular $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ$, the second of which can provide a very clean four lepton final state. In $H \rightarrow \gamma\gamma$, since the Higgs does not couple with massless particles, the decay process happens via a quark loop, in the same way as the

gluon-gluon fusion production process.

Decay channel	Branching Ratio
$H \rightarrow b\bar{b}$	5.809×10^{-1}
$H \rightarrow W^+W^-$	2.152×10^{-1}
$H \rightarrow gg$	8.180×10^{-2}
$H \rightarrow \tau^+\tau^-$	6.256×10^{-2}
$H \rightarrow c\bar{c}$	2.884×10^{-2}
$H \rightarrow ZZ$	2.641×10^{-2}
$H \rightarrow \gamma\gamma$	2.270×10^{-3}
$H \rightarrow Z\gamma$	1.541×10^{-3}
$H \rightarrow \mu^+\mu^-$	2.171×10^{-4}

Table 1.3: Branching ratios for a Standard Model Higgs boson with $m_H = 125.09$ GeV [26].

3.3 Higgs self-coupling

As seen in Section 1.2 the self-couplings of the Higgs boson arise from the expansion of the Higgs potential around the ground state. These couplings, already introduced in Equation 1.22 and 1.23, are:

$$\lambda_{HHH} = \lambda_3 = \frac{3m_H^2}{v}, \quad \lambda_{HHHH} = \lambda_4 = \frac{3m_H^2}{v^2} \quad (1.32)$$

The self-couplings determine the shape of the Higgs potential, which is connected to the phase transition of the early universe from the unbroken to the broken electroweak symmetry. Large deviations of the trilinear and quartic couplings are possible in BSM scenarios. For this reason a measurement of these particular couplings are important both in the validation of the Standard Model predictions and in the possible searches for new physics beyond the Standard Model. The trilinear Higgs self-coupling can be directly accessed in Higgs pair production processes, that will be described in Section 3.4, and also in an indirect way via loop corrections to single-Higgs production processes. The quartic self-coupling appears in triple-Higgs production processes; however, due to the very small cross section of these processes, which are suppressed by a factor v compared to the trilinear self-coupling, this parameter is not accessible at the Large Hadron Colliders [28]. In the following the trilinear coupling will be referred using also the κ -framework notation, i.e. the notation in which the parameters are normalised to their SM predicted value, as κ_λ :

$$\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}} \quad (1.33)$$

3.4 Higgs pair production

At hadron colliders, Higgs boson pairs are dominantly produced with the gluon-gluon fusion mechanism, mainly mediated by top quark loops, in a similar way to the single Higgs ggF production process. There are also other production modes with sub-leading cross section compared to the gluon-gluon fusion's one. These processes are the vector boson fusion (VBF), the double Higgs-strahlung (VHH) and the double Higgs bremsstrahlung off top quarks ($t\bar{t}HH$ and $tjHH$). In Figure 1.8 are illustrated the Feynman diagrams that contribute to the Higgs boson pair production, while in Table 1.4 are summarised the cross sections for the main production modes.

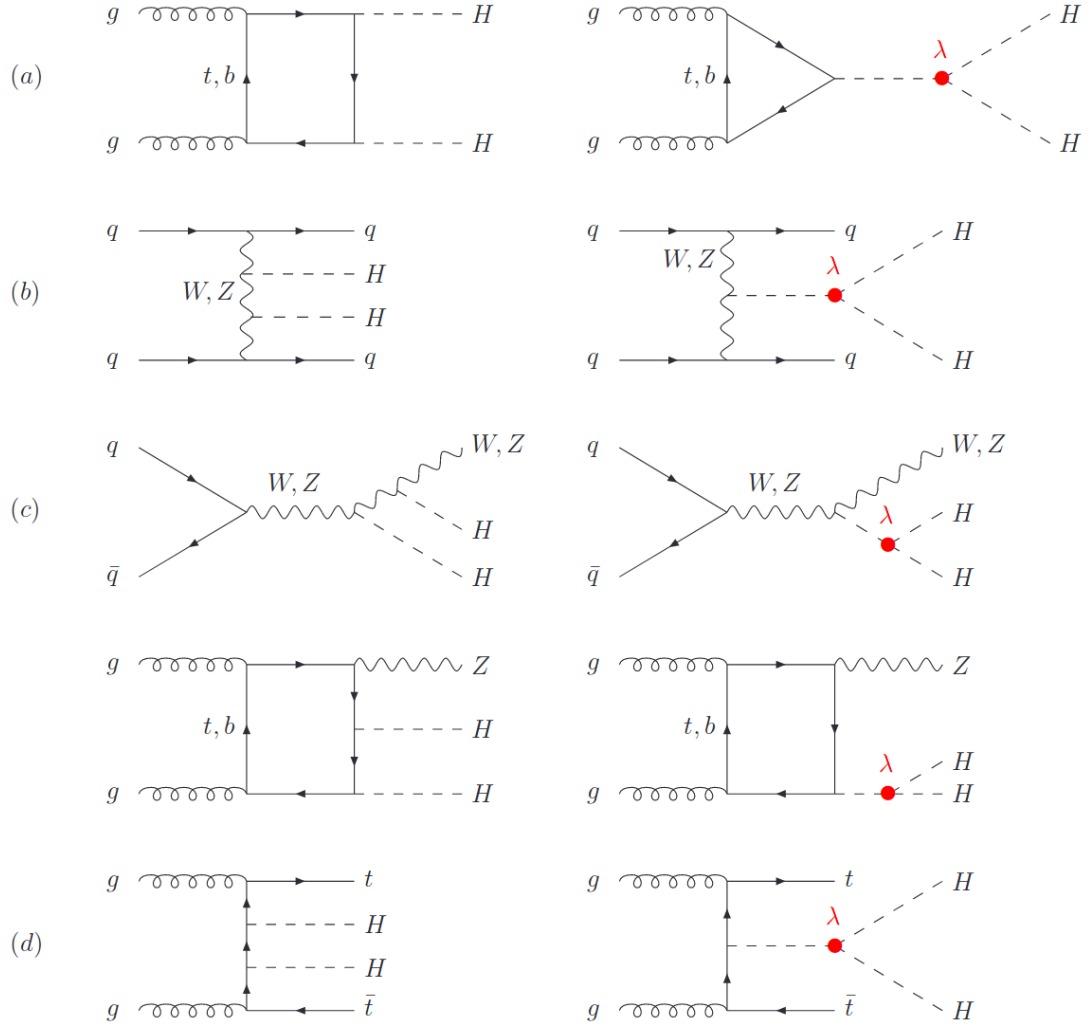


Figure 1.8: Diagrams contributing to Higgs pair production: (a) gluon fusion, (b) vector-boson fusion, (c) double Higgs-strahlung and (d) double Higgs bremsstrahlung off top quarks. The trilinear Higgs coupling contribution is marked in red [29].

Production mode	Cross section [fb] + scale unc.	PDF+ α_s unc.
$ggF\ HH$	$31.05^{+2.2\%}_{-5.0\%}$	$\pm 3.0\%$
$VBF\ HH$	$1.73^{+0.03\%}_{-0.04\%}$	$\pm 2.1\%$
$Z\ HH$	$0.363^{+3.4\%}_{-2.7\%}$	$\pm 1.9\%$
$W^+\ HH$	$0.329^{+0.32\%}_{-0.41\%}$	$\pm 2.2\%$
$W^-\ HH$	$0.173^{+1.2\%}_{-1.3\%}$	$\pm 2.8\%$
$t\bar{t}\ HH$	$0.775^{+1.5\%}_{-4.3\%}$	$\pm 3.2\%$
$tj\ HH$	$0.0289^{+5.5\%}_{-3.6\%}$	$\pm 4.7\%$

Table 1.4: Cross sections (in fb) of the main Higgs boson pair production processes for $m_H = 125$ GeV in proton-proton collisions at $\sqrt{s} = 13$ TeV. The QCD scale factorization and renormalization scale uncertainties, together with the uncertainties on the PDFs and on the α_s computation are also reported [29].

The gluon-gluon fusion is the dominant Higgs boson pair production mechanism, which accounts for more than 90% of the total cross section of the process. In Figure 1.8a are visible two destructively interfering box and triangle diagrams, where only in the triangle diagram appears the trilinear Higgs coupling. The interference between the two diagrams leads to a cross section value three order of magnitude smaller than the one of single Higgs production.

The vector boson fusion process is much less probable than the ggF one, but it is important since it is particularly sensitive to the quartic coupling between the Higgs bosons and vector bosons c_{2V} (also known as c_{VVHH}). The other production processes have a so little cross section that are not currently experimentally accessible.

The decay channels are similar to the single Higgs ones, with the only difference that here there are two decaying Higgs bosons. In Table 1.5 are summarised the main decay channels with their branching ratios.

Decay channel	Branching ratio
$HH \rightarrow b\bar{b}b\bar{b}$	3.37×10^{-1}
$HH \rightarrow b\bar{b}W^+W^-$	2.50×10^{-1}
$HH \rightarrow b\bar{b}\tau^+\tau^-$	7.27×10^{-2}
$HH \rightarrow W^+W^-W^+W^-$	4.63×10^{-2}
$HH \rightarrow b\bar{b}\gamma\gamma$	2.64×10^{-3}
$HH \rightarrow W^+W^-\gamma\gamma$	9.77×10^{-4}

Table 1.5: Branching ratios for the main decay channels of HH processes, considering an Higgs boson with $m_H = 125.09$ GeV.

4 Beyond the Standard Model

The Standard Model of particle physics is a highly predictive theory, that has been tested and validated with several experimental measurements that have confirmed its predictions. The discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments is the last example of the validity of this model. However the Standard Model is known to be an incomplete theory. Some theoretical issues suggest that this theory is not complete and that there should be a more fundamental theory. These are in particular the fact that the model has a not negligible number of free parameters that are only determined by the experimental measurements. No prediction is made on these parameters by the SM, like the vacuum expectation value of the Higgs field and the Higgs boson couplings to fermions (producing the fermion masses and the CKM matrix).

The main argument in support of an incomplete SM is the Hierarchy problem, known also as naturalness of the Higgs mass. This problem arises from the relatively small measured value of the Higgs boson mass. The electroweak symmetry breaking scale $O(10^2 \text{ GeV})$ and the Planck scale $O(10^{19} \text{ GeV})$ are separated by many orders of magnitude. The Higgs mass is set at Planck scale and then modified by all higher order radiative corrections, that determine the final Higgs mass. First order corrections to the Higgs mass can be written as

$$\delta m_H = -\frac{\lambda_f^2}{16\pi^2} \left[2\Lambda^2 + O\left(m_f^2 \ln \frac{\Lambda}{m_f}\right) \right] \quad (1.34)$$

where λ_f is the Yukawa coupling and Λ is the cut-off energy scale at which the model ceases to be valid. The mass correction δm_H is quadratically divergent in Λ , for this

reason it is problematic if the SM is assumed to be valid up to the Plank scale. However, this divergence does not represent an issue in the model when the mass correction is of the same order of the mass itself, leading to a unnatural fine tuned cancellation of terms. In fact, in order to have an Higgs boson with $m_H = 125$ GeV, these corrections must balance the bare mass over sixteen orders of magnitude, requiring a particular fine-tuning of the parameters, that is possible, but extremely unlikely. The hierarchy problem would require new physics at the TeV scale, assuming a cut-off at $\Lambda \sim O(1\text{TeV})$, or an extended theory to solve the divergence problem.

Beside theoretical issues, the incompleteness of the SM is also clear by phenomena observed in nature that are not explained by this model. Some examples are here listed:

- Matter-antimatter asymmetry measured in cosmological and astrophysical observations is not explained by the too small CP violation in the electroweak interactions. The huge imbalance between baryonic and antibaryonic matter, generated in the Big Bang, should be then described by a new physical process, not described in the SM.
- The matter described by the SM is only 5% of the total energy of the universe. Cosmological and astrophysical observations have measured that the remaining part is made by Dark Matter and Dark Energy, accounting for 27% and 68% of the total energy respectively, whose nature is not described by the SM.
- In the SM, neutrinos are massless particles, the observation of neutrino oscillations require, instead, that neutrinos do have mass, although very small. Neutrino mass terms can be added with a little extension of the SM.
- Gravitational interaction is not included in the SM. Thus a unified description of the four fundamental forces is missing, although the SM was never intended to include gravity. However, embedding the SM in a more fundamental framework is theoretically very appealing. The inclusion of gravity in the theory would require its quantisation and an extension of the SM at the gravitational scales $O(10^{19}$ GeV), where it does not work at the current state.

To describe all these uncovered features, many beyond the Standard Model theories have been proposed. In particular, some are related to the presence of additional heavier Higgs bosons. The simplest way to include such a new particle in the SM, is by its extension adding a new real singlet scalar field [30, 31], heavier than the SM Higgs and that can, therefore, decay via Higgs boson pair production processes. The new scalar potential can be expressed like:

$$V(\phi, S) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 + \frac{a_1}{2} \phi^\dagger \phi S + \frac{a_2}{2} \phi^\dagger \phi S^2 + b_1 S + \frac{b_2}{2} S^2 + \frac{b_3}{2} S^3 + \frac{b_4}{4} S^4 \quad (1.35)$$

where the first two terms are the Higgs potential in Equation 1.12, $S = (v_S + s)/\sqrt{2}$ is a gauge singlet scalar, v_S is the vacuum expectation value of S , and s is the new scalar boson. After the electroweak symmetry breaking, the new scalar S and the Higgs mix, resulting in two mass eigenstates with masses $m_{1,2}$ and $m_2 \geq m_1$, where in our case $m_1 = m_H$ is the SM Higgs boson mass and m_2 is the mass of the new scalar particle. If the mass of the new scalar $m_2 > 2m_1$, the process $h_2 \rightarrow h_1 h_1$ of the production of two on-shell Higgs bosons from the decay of the new heavier scalar is allowed. If $2m_1 > m_2 > m_1$, the on-shell production is not allowed but the process can still happen with

one of the two produced Higgs produced off-shell: $h_2 \rightarrow h_1 h_1^*$. This BSM process would then increase the cross section of the Higgs pair production processes, and any deviation from the SM value with an enhanced cross section would be an indication of a possible presence of such a new scalar particle.

Chapter 2

The Large Hadron Collider and the ATLAS Experiment

1 The Large Hadron Collider

The Large Hadron Collider (LHC) [32] is the largest and most powerful particle accelerator in the world. It is located at the CERN (Conseil Européen pour la Recherche Nucléaire) in Geneva, near the border between France and Switzerland, and it is in operation since 2008. It consists of a 27 km long ring, placed at 100 m of depth underground. Inside the ring, there are two high energy beams of particles, travelling in opposite directions with a speed close to the speed of light, in two separate pipes. These tubes are kept in a ultra-high vacuum and they are placed inside the magnet system of the accelerator. The accelerator is composed, in fact, by superconductive magnets providing a magnetic field up to 8.33 T, needed to curve the trajectory of the particles inside the beam pipes to follow the curvature of the ring. The acceleration of the particles is provided by the use of intense electric fields generated in the so called radio-frequency cavities, that are placed along the ring length. For the magnets are used coils of particular wires working in superconductive regime while maintained at very low temperatures. For this scope, it is used a liquid helium cryogenic system, that keeps the magnets at a stable temperature of $\simeq 2$ K. The magnets used in the LHC ring have different dimensions and type depending on their purpose. 1232 magnetic dipoles 15 meters long are used to curve the particle beams, and 392 magnetic quadrupoles, with a length between 5 and 7 meters, are dedicated to the focusing of the beams. The focusing is needed to reduce as much as possible the transverse dimensions of the beams, of the order of some tens of microns, in the interaction points, to enhance the probabilities of having a collision between the particles coming from the two beams. The collisions occurs in four points of the ring, where the two beams are intercepted, that are called interaction points (IP). In the same position, the caverns of the LHC's experiments are placed in order to detect the products of the particle's collisions. These experiments are ATLAS [33], CMS [34], LHCb [35] and ALICE [36].

1.1 The accelerator complex

The particles used in the LHC are protons or ions of heavy nuclei, like lead. The source of the LHC's protons is a bottle of gaseous hydrogen. The electrons of the hydrogen are removed by an electric field, allowing the injection of protons in the accelerator chain:

- Linear Accelerator 2 (Linac 2): linear accelerator made by radio-frequency cavities that brings the protons to an energy of 50 MeV;

- Proton Synchrotron Booster (PSB): synchrotron that pushes the energy of the protons up to 1.4 GeV;
- Proton Synchrotron (PS): 628 meters long synchrotron that accelerates again the protons coming from the PSBoost up to an energy of 25 GeV. The PS can also accelerate α particles, oxygen ions, electrons, positrons and anti-protons.
- Super Proton Synchrotron (SPS): synchrotron of about 7 Km of circumference, that accelerates the proton beams up to 450 GeV and injects them in the two LHC's beam pipes.

When entering in the LHC, the protons are separated in two beams travelling in opposite directions. They are then accelerated to the maximum energy, that currently is 6.5 TeV, but that will be elevated to 7 TeV in the next run of the machine. Along their travel in the accelerators, protons are grouped in bunches to have an higher interaction probability during the collision. In the LHC, when accelerated, the proton bunches are separated by a time interval of 25 ns and can contain up to $\sim 10^{11}$ protons. When these conditions are reached and the protons are accelerated to their maximum energy, they are made to collide. This collision happens between the two opposite-directed proton bunches, giving rise the the so-called Bunch Crossing (BC) in the four experiment's caverns. In the CERN accelerator complex are present also many smaller experiments. A schematic view of it is shown in Figure 2.1.

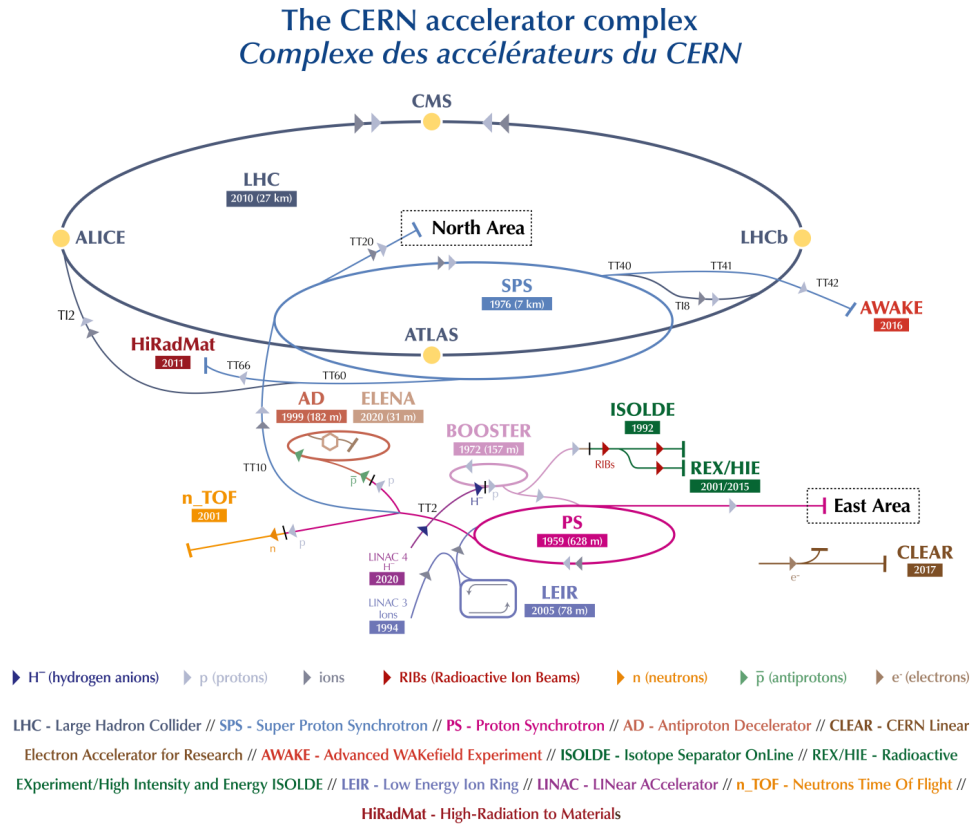


Figure 2.1: The CERN accelerator complex. The LHC is the last ring (dark blue line) in a complex chain of particle accelerators. The position of the four main experiments is shown on the LHC ring [37].

1.2 Luminosity and pile-up

The main work parameters of a particle accelerator are the center of mass energy and the instantaneous luminosity. The instantaneous luminosity $\mathcal{L}$ is a quantity that depends on the beam parameters of the accelerator. It is defined as:

$$\mathcal{L} = n_b f_{rev} \frac{N_1 N_2}{4\pi \Sigma_X \Sigma_Y} \quad (2.1)$$

where n_b is the number of proton bunches, N_1 and N_2 are the number of particles per bunch, and f_{rev} is the revolution frequency of the bunches in the LHC tunnel. The maximum number of bunches is $\simeq 3560$, that is the maximum number of positions available in the LHC ring, having the bunches separated every 25 ns. Actually, some of the bunches are empty for technical reasons, and the number of filled bunches is usually about $n_b \simeq 2500$. The frequency f_{rev} , multiplied by the total number of filled bunches, gives the frequency of the collisions: $f_{coll} = 40$ MHz. Σ_X e Σ_Y are, instead, parameters related to the transverse dimensions of the beams. They are the width of the Gaussian distributions of the particle's position inside the beam, in the two transverse directions. The values of the beam parameters of the LHC machine during the Run 2 are shown in Table 2.1.

Parameter	2015	2016	2017	2018
Maximum number of colliding bunch pairs (n_b)	2232	2208	2544/1909	2544
Bunch spacing (ns)	25	25	25/8b4e	25
Typical bunch population (10^{11} protons)	1.1	1.1	1.1/1.2	1.1
β^* (m)	0.8	0.4	0.3	0.3–0.25
Peak luminosity $\mathcal{L}_{peak}$ ($10^{33} \text{ cm}^{-2} \text{ s}^{-1}$)	5	13	16	19
Peak number of inelastic interactions/crossing ($\langle \mu \rangle$)	~ 16	~ 41	$\sim 45/60$	~ 55
Luminosity-weighted mean inelastic interactions/crossing	13	25	38	36
Total delivered integrated luminosity (fb^{-1})	4.0	38.5	50.2	63.4

Table 2.1: LHC beam parameters during the Run 2 [38].

The number of events produced per second R (rate), for a process with cross section σ , is proportional to the instantaneous luminosity $\mathcal{L}$:

$$R = \sigma \times \mathcal{L} \quad (2.2)$$

Therefore, an higher luminosity allows to accumulate more statistics given a time interval. This quantity is called integrated luminosity L , and it is related to the number of events collected N :

$$N = \sigma \times L = \sigma \times \int \mathcal{L} dt \quad (2.3)$$

The LHC was designed to be able to reach a center of mass energy of 14 TeV and an instantaneous luminosity of about $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. It has started working from 2009 until the end of 2012 (Run 1), producing proton-proton collisions with a center of mass energy of 7 TeV, then grown to 8 TeV, carrying to the discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments. Between 2013 and 2015, the operations have been suspended in order to carry some updates in the detectors. After this shutdown, the collider has started a new data taking run, the Run 2, started in April 2015 with an higher beam energy of 6.5 TeV in order to have a center of mass energy of 13 TeV, and ended in November 2018. The instantaneous luminosity has reached its design value in 2016 and has arrived to the record value of $2.1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ in 2018. The integrated luminosity collected by the ATLAS detector in the different years

of data taking and during the Run 2, describing the total amount of the statistics good for physics analyses, are shown in Figure 2.2a and 2.2b respectively.

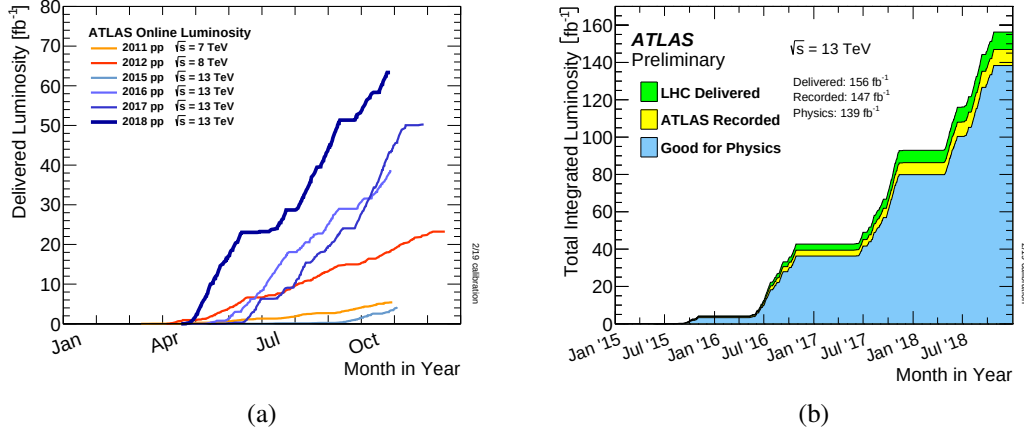


Figure 2.2: Integrated luminosity collected in the different years of data taking by the ATLAS detector (a) and during the Run 2 (b) [39].

In a bunch crossing, given the high density of the beam bunches and the high frequency of collisions, many proton-proton interactions may occur simultaneously. In this case several interaction points are produced, resulting in an higher number of particles emerging from the collision. This particular phenomenon is known as pileup and it is dependant by the luminosity of the beams:

$$\langle \mu \rangle = \frac{\mathcal{L}_{bunch} \times \sigma_{inel}}{f_{rev}} \quad (2.4)$$

where $\mathcal{L}_{bunch}$ is the instantaneous luminosity per bunch, σ_{inel} is the inelastic cross section for pp interaction. The distribution of the mean number of interactions is shown in Figure 2.3 for the data taking years of the Run 2. It corresponds to the mean of a Poisson distribution of the number of interactions per bunch crossing calculated for each bunch. As it is visible from Figure 2.3, in 2017 and 2018 the mean number of interactions per bunch crossing has been higher, reaching values around 40 interactions per collision, due to the increase of luminosity of the LHC machine.

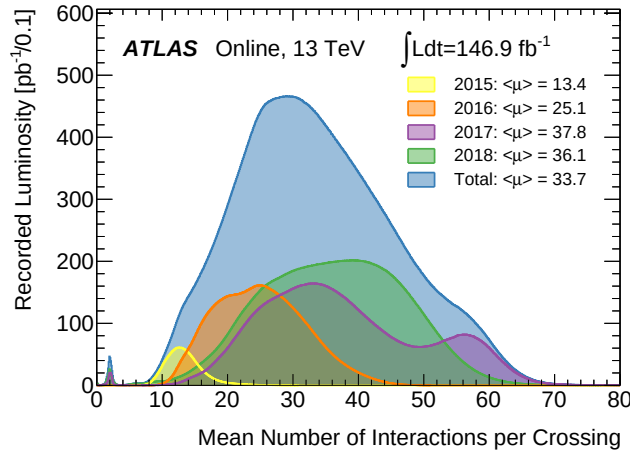


Figure 2.3: Distribution of the mean number of interactions per crossing for the pp collision data at 13 TeV centre-of-mass energy from 2015 to 2018 [39].

2 The ATLAS Experiment at the LHC

The ATLAS experiment is one of the four main experiments of the Large Hadron Collider. It is a general purpose particle detector with a cylindrical symmetry around the beam pipes, designed to study the pp interactions occurring at its center. The dimensions are large: it has a length of 44 m, a diameter of 25 m and a weight of 7000 tons, which make it the largest particle detector ever built. Its layered structure and its dimensions are visible in Figure 2.4.

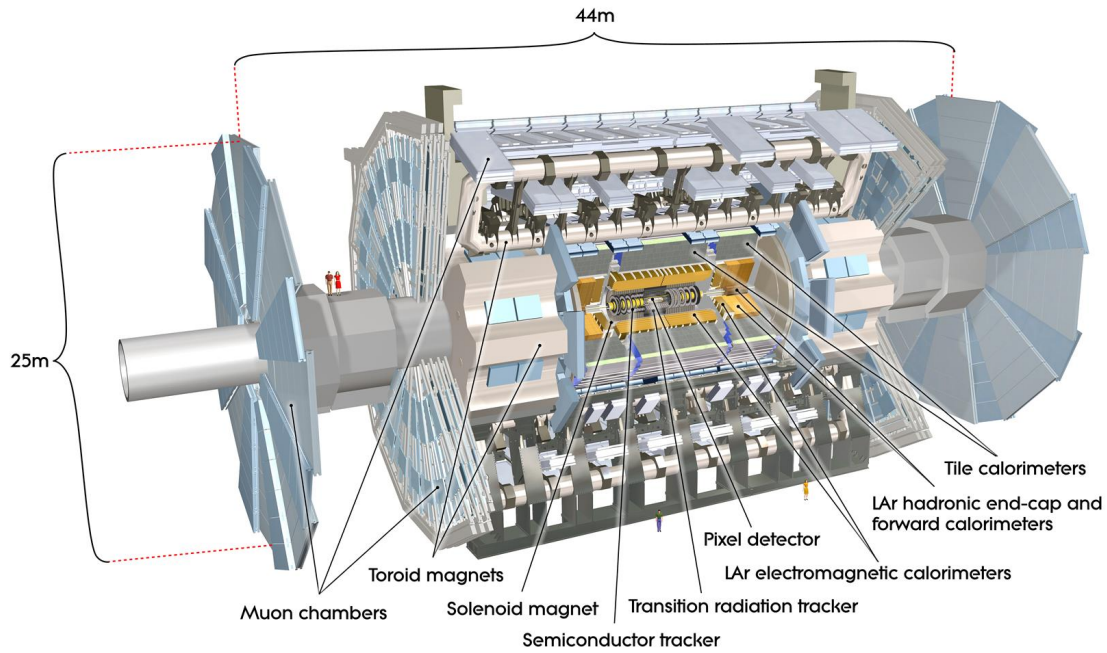


Figure 2.4: Illustration of the ATLAS detector. The dimensions and the several sub-systems are also shown [33].

2.1 The ATLAS detector

The ATLAS detector [33] is structured in three concentric cylindrical sub-detector systems which surround the interaction point. Close to the beam pipe there is the Inner Detector (ID), that allows to reconstruct the tracks and measure the momentum of all the charged particles produced in collisions. It provides also the identification of the interaction vertices. The calorimeter system is placed around the ID and it is composed by an electromagnetic calorimeter (ECAL), dedicated to the identification of electromagnetic showers, and a hadronic calorimeter (HCAL), needed to identify and measure the energy of hadronic jets. Finally, in the outermost part of the detector, there is the Muon Spectrometer (MS), dedicated to the identification and high precision measurement of muons and their momentum. These sub-detectors are divided longitudinally in three regions: the central part, called "barrel", and the two edges of the cylinder, called "end-caps", resulting in a geometric acceptance close to 4π steradians in solid angle.

The particles need to be curved in order to measure their momentum. For this purpose in the ATLAS detector is present a complex magnetic system, composed by a central solenoid that encloses the ID, and three large external toroids, one placed in the barrel and one in each of the two end-caps.

2.1.1 Coordinate system

In Figure 2.5 a picture of the ATLAS coordinate system is shown. It is a xyz right-handed reference system centered in the nominal interaction point of the beams. The z -axis is oriented along the direction of the beams. The x -axis points the center of the LHC ring, and forms, with the vertical y -axis, the transverse plane to the beam direction.

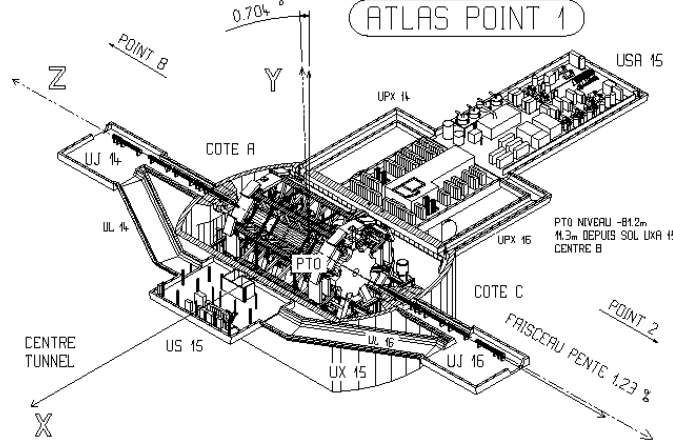


Figure 2.5: The ATLAS coordinate system [33].

This reference system is usually defined by cylindrical coordinates: the azimuth angle ϕ , measured around the beam, and the polar angle θ , measured with respect to the beam axis. The radial distance measured from the origin in the $x - y$ plane is denoted as R , and the longitudinal as z .

In this coordinate system, useful kinematic variables can be defined to be invariant for Lorentz boost along the longitudinal axis, as in hadron colliders the initial z -momentum of the system is unknown. An example is the rapidity, which is defined as:

$$y = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) \quad (2.5)$$

where E is the energy of the particle and p_z is its momentum along the z -axis. In the limit where the particle is travelling close to the speed of light, or equivalently in the approximation that the mass of the particle is negligible, the rapidity can be approximated with the pseudorapidity, defined as:

$$\eta = -\log \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (2.6)$$

In the following, the momentum of a particle in the transverse plane to the beam direction will be referred as to $p_T = p \sin(\theta)$. A commonly used quantity is also the angular distance between objects in the $\eta - \phi$ plane, defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (2.7)$$

2.1.2 Magnet system

The ATLAS magnet system [40] is made of four superconducting magnets. The central solenoid is a 2.4 m superconductive coil that provides a 2 T axial magnetic field along

the z -axis, to the inner detector. The three air-core toroids are composed by eight superconductive rectangular coils each, and generate a toroidal magnetic field, with an intensity between 0.1 and 3 T, for the Muon Spectrometer. In Figure 2.6 a schematic picture of the ATLAS magnetic system is shown.

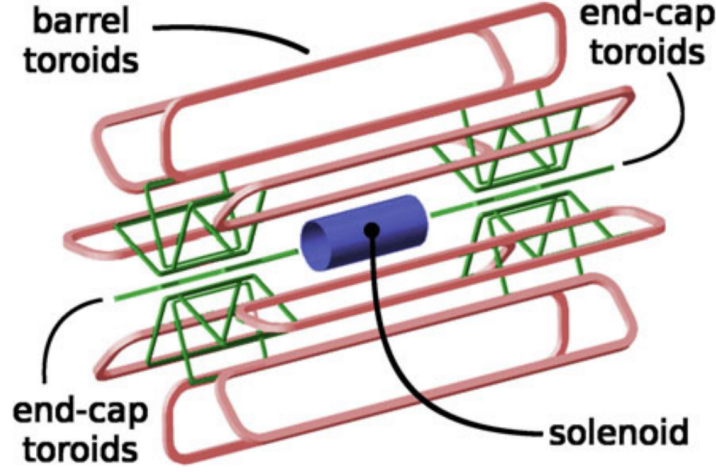


Figure 2.6: A schematic picture of the ATLAS magnetic system layout.

These magnetic fields are needed to measure the momentum of the charged particles that pass through the detectors. The trajectory of a particle with charge q crossing a magnetic field of intensity B with speed v is bent by the Lorentz force $\vec{F}_L = q\vec{v} \times \vec{B}$. The particle follows, then, a helical trajectory with a circumference of radius R in the plane orthogonal to $\vec{B}$. Thus, the momentum component of a charged track that is perpendicular to a uniform magnetic field can be estimated by measuring its bending radius in the magnetic field:

$$p_T[\text{GeV}] \simeq 0.3 B[\text{T}] \times R[\text{m}] \quad (2.8)$$

The axial magnetic field along the z -axis curves the particles inside the solenoid in the $x-y$ transverse plane. The barrel toroid provides 1.5 to 5.5 Tm of bending power in the pseudorapidity range $0 < |\eta| < 1.4$, with a magnetic field directed along the tangential direction of the circumferences centered on the z -axis. The end-cap toroids are two smaller toroids that provide a bending power of approximately 1 to 7.5 Tm in the region $1.6 < |\eta| < 2.7$, with a magnetic field in the $R-\phi$ plane. The bending power is lower in the transition regions where the two magnets overlap, i.e. $1.4 < |\eta| < 1.6$ [33]. Because of their design, all the three toroids curve the particle's trajectories in the $R-z$ plane. With this configuration, the magnetic field is always perpendicular to the particle's momentum, and this allows to measure their total momentum p . Moreover, in the barrel region, this configuration provides a constant resolution of the transverse momentum in η ; this because the field integral along the trajectory grows with η , thus compensating for the same p_T the larger impulse of the forward emitted particles.

2.1.3 Inner Detector

The Inner Detector [41, 42] is the innermost sub-detector of the ATLAS experiment. It is designed for the reconstruction of the tracks of charged particles produced in the pp collisions, and for the identification of the primary vertices of the interactions and of possible secondary vertices generated from the decay of long-lived particles. The ID is

placed inside the central solenoid, subject to a magnetic field of 2 T of intensity, which allows to measure the momentum and the charge of the particles from the curvature of their trajectory, reconstructed through high-resolution position measurements, usually referred as hits. Its cylindrical structure, visible in Figure 2.7a, has a diameter of 2.1 m and a length of 6.2 m, and it is composed by three different particle detectors: the pixel detector and the Insertable B-Layer (IBL), the semiconductor tracker (SCT) and the transition radiation tracker (TRT). The three sub-detectors, visible in Figure 2.7b, are placed in multi-layer concentric cylinders around the beam pipes in the barrel region, while in the forward region they are disposed in disks shaped layers perpendicular to the z -axis. This configuration allows to cover the $|\eta| < 2.5$ region around the interaction point.

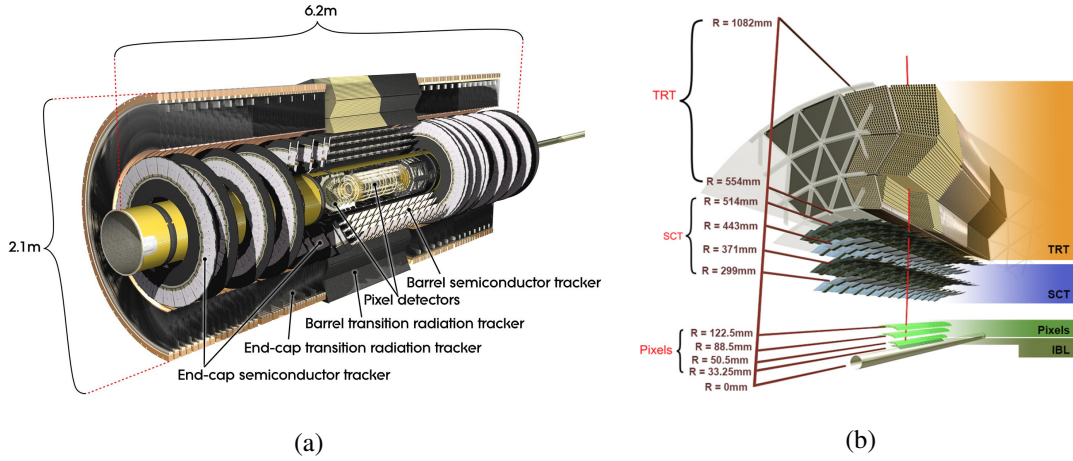


Figure 2.7: Layout of the ATLAS Inner Detector (a) [43]. A zoomed view of the ATLAS Inner Detector, describing its three sub-detectors: the pixel detector and the Insertable B-Layer (IBL), the semiconductor tracker (SCT) and the transition radiation tracker (TRT) [44].

The inner detector provides accurate and efficient tracking for charged particles with $p_T > 0.5$ GeV within $|\eta| < 2.5$, with a transverse momentum resolution of:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\% \quad (2.9)$$

In the following there is a more detailed description of the different technologies used in the ID.

Pixel detector The silicon pixel detector is the sub-detector closest to the beam pipes. It is composed of layers of silicon pixels in order to reach a very high granularity, that is needed for resolving primary and secondary interaction vertices. In the barrel region, it is made of three cylindrical layers positioned at the radial distances of 50.5, 88.5 and 122.5 mm. In the end-caps it is made of three disks perpendicular to the beams at the longitudinal distances of 49.5, 58.0 and 65.0 mm from the nominal IP. An additional pixel layer, the Insertable B-Layer (IBL), was installed in 2014 in the barrel region at a radius of 33 mm from the beam axis. This new layer provides an additional point of measurement very close to the interaction point, that, together with the B-Layer at 50.5 mm, allows a precise identification of the secondary vertices of decaying particles, improving the identification of jets coming from b-quark hadronisation.

The pixel layers are segmented in $R - \phi$ and z with typically four pixel layers crossed by each track. The pixel sensors have a minimal size in $R - \phi \times z$ of $50 \times 400 \mu\text{m}^2$. This results in a total of about 92 million pixels in the system, with an intrinsic resolution of

10 μm in $R - \phi$, and 115 μm in z in the barrel, and of 10 μm in $R - \phi$ and 115 μm in R in the end-caps.

Semiconductor tracker (SCT) The semiconductor tracker (SCT) is a detector made of silicon strips. The structure of this sub-detector is analog to the one of the pixel detector, with concentric cylinders around the beam axis in the barrel, and disks perpendicular to the beam axis in the end-caps, covering the same region of $|\eta| < 2.5$. In the SCT eight layers of stereo strips with an average pitch of 80 μm are used, that provide four space points intercepted by the tracks. In the barrel region, one set of strips is parallel to the beam axis, while the stereo strips are inclined by a small angle of 40 mrad. In the end-cap region, the detectors have a set of strips running radially and a set of stereo strips at an angle of 40 mrad. This configuration allows to measure both the coordinates with an intrinsic precision of 17 μm in $R - \phi$, and 580 μm in z and R .

Transition radiation tracker (TRT) Transition radiation tracker (TRT) is placed in the outermost part of the ID. This detector is composed by straw tubes of 4 mm diameter, filled with a gas mixture of 70% Xe, 27% CO₂, and 3% O₂. In the centre of each tube there is a gold-plated tungsten wire of 31 μm diameter, at ground potential, acting as a anode, while the walls are kept at a voltage of -1.5 kV. In this way each tube acts as a small proportional counter, producing a low amplitude signal on the anode when crossed by an ionising particle. Polypropylene fibers are also interleaved between the drift tubes in order to provide the emission of transition radiation. This radiation is absorbed by the Xenon present in the gas mixture, leading to a high amplitude signal in the TRT electronics, that can be distinguished from low amplitude ionisation signal. Particles emit transition radiation according to the speed they have passing through several layers of material with different refraction indices. As a result, lighter particles, that are high relativistic, have a higher probability of emitting transition photons, with respect to heavier particles, allowing the TRT to identify the electrons.

In the barrel region, the straws are parallel to the beam axis and are 144 cm long, with their wires divided into two halves, approximately at $\eta = 0$. In the end-cap region, the 37 cm long straws are arranged radially in wheels. The total number of TRT readout channels is approximately 351,000. With this configuration, the TRT covers the region $|\eta| < 2.0$ and provides $R - \phi$ information, for which it has an intrinsic precision of 130 μm per straw [33].

2.1.4 Calorimeter system

The ATLAS calorimeter system [45] is shown in Figure 2.8. It is composed by two sub-systems: the electromagnetic calorimeter (EM) and the hadronic calorimeter. The whole system is designed to be hermetic and with a fine segmentation for the reconstruction of photons, electrons and hadronic jets, as well as for the measure of the missing transverse energy (MET) needed to detect the neutrinos and other possible invisible BSM particles.

The calorimeter is placed outside the ID and the central solenoid, and cover the region $|\eta| < 4.9$. The system has a cylindrical shape, with the electromagnetic calorimeter, designed to measure the energy of electron and photons, in the inner part, and the hadronic calorimeter, dedicated to the energy measure of the hadrons, in the outer part.

The calorimeters are designed to provide a good containment of the electromagnetic and hadronic showers, as well as to limit leakage of particles into the muon system. Therefore, the total thickness of the EM calorimeter is > 22 radiation lengths (X_0) in

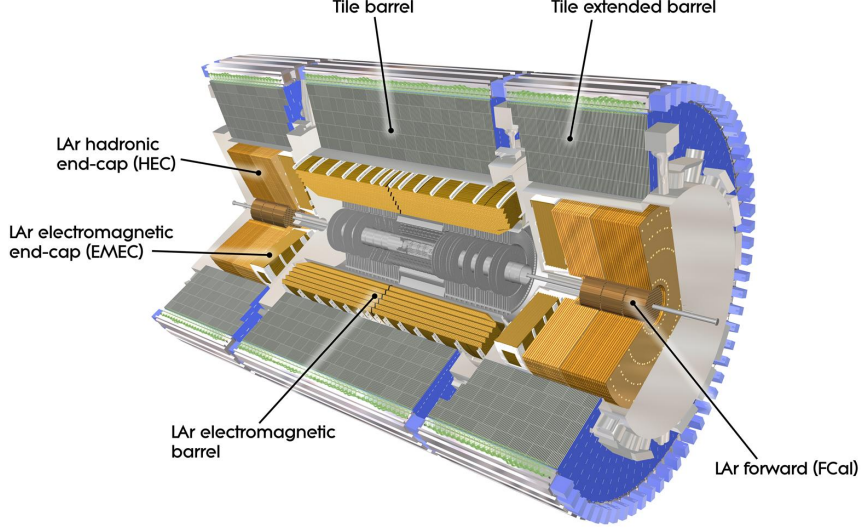


Figure 2.8: The ATLAS calorimeter system, with the description of the electromagnetic and hadronic sub-systems [33].

the barrel and $> 24 X_0$ in the end-caps, while the hadronic calorimeter has a thickness of about 10 interaction lengths (λ_I).

Electromagnetic calorimeter The EM calorimeter is a sampling lead-liquid Argon (LAr) calorimeter [46]. It is made by layers of lead absorber alternating with holes filled with LAr, with an accordion shape that provides complete ϕ symmetry without azimuthal cracks. The calorimeter is divided into the barrel, that covers the region $|\eta| < 1.475$, and into the two end-caps, covering the region $1.375 < |\eta| < 3.2$. The barrel calorimeter is split in two halves with a small gap of 4 mm at $z = 0$, while the two end-cap calorimeters are divided into two coaxial wheels, an external one covering the region $1.375 < |\eta| < 2.5$, and an internal one covering the region $2.5 < |\eta| < 3.2$. Additional material needed for the instrumentation and the cooling of the detector creates a “crack” region at $1.375 < |\eta| < 1.52$, where the energy resolution is significantly degraded. In the radial direction, the EM calorimeter is segmented in three layers: a pre-sampler with very high granularity in η , capable to reconstruct neutral pions decaying to two photons and particles whose shower already started in the inner detector. After the pre-sampler there are longer towers with high granularity for the detection of the bulk of the EM showers, and allowing the measurements of the η and ϕ coordinates. The last layer detects showers generated from hadrons that start their shower inside the EM calorimeter. The energy resolution of the ATLAS EM calorimeter is:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \quad (2.10)$$

Hadronic calorimeter The ATLAS hadronic calorimeter is designed to measure the energy and direction of hadrons produced by the hadronisation of quarks and gluons. It surrounds the EM calorimeter and it is composed of three different typologies of detectors:

1. **Tile Calorimeter:** it is a sampling calorimeter made by a steel absorber and tiles of plastic scintillators used as active material. It is placed next to the EM calorimeter

in the barrel and covers the region $|\eta| < 1$ [47]. Its extensions in the end-caps cover, instead the region $0.8 < |\eta| < 1.7$.

2. LAr Hadronic End-cap Calorimeter (HEC): it is composed of layers of copper absorber alternating with liquid Argon gaps used as active material. Its structure is made of two independent wheels for each end-cap, placed next to the EM calorimeter end-caps, covering the region $1.5 < |\eta| < 3.2$, with a little overlap with the Tile calorimeter.
3. LAr Forward Calorimeter (FCal): it covers the region $3.1 < |\eta| < 4.9$ and it is about $10 X_0$ deep. It is composed of three modules. The first uses copper as absorber and LAr as active material, and it is optimized for electromagnetic measurements. The other two modules are designed for hadronic measurements. They have tungsten as a passive material, chosen for its high density to provide containment and minimise the lateral spread of hadronic showers.

The energy resolution of the ATLAS hadronic calorimeter in the barrel and in the end-caps is:

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \quad (2.11)$$

while in the forward region it is:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \quad (2.12)$$

2.1.5 Muon Spectrometer

The Muon Spectrometer (MS) [48] is the outermost part of the ATLAS detector. It is dedicated to the identification and the measurement of the momentum of the particles that escape the calorimeters.

The MS covers the region in $|\eta| < 2.7$ and provides the trigger of the muons in the region $|\eta| < 2.4$. The structure of the Muon Spectrometer is shown in Figure 2.9. It is divided in the barrel region, defined by $|\eta| < 1.0$, and in the two end-cap regions, in $1.0 < |\eta| < 2.7$. In the region $1.0 < |\eta| < 1.4$, known as transition region, the magnetic deflection is provided by the combination of the fields of the toroids in the barrel and in the end-caps. This configuration of the magnets allows to have the magnetic field always orthogonal to the muon's trajectories, minimizing the effect of the multiple scattering which would degrade the resolution.

The MS is composed by two groups of detectors. One group is designed to measure the position of the muons with high precision in the plane where the muon is deflected by the magnetic field. These precision is needed to allow an accurate measurement of the muon momentum. These detectors are the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC), and are called tracking chambers or precision chambers. The other group of detectors is dedicated to the trigger system. These detectors have a high timing resolution and allow to identify the collision of the detected muon. They are the Resistive Plate Chambers (RPC) and the Thin gap Chambers (TGC), called also trigger chambers.

In Figure 2.9 is visible the disposition of the detectors. In the barrel, the RPC and the MDT chambers are placed before, after and inside the eight toroid coils, disposed in three layers at 5.5 m, 7.5 m and 10 m from the z-axis. In the end-caps, the detectors are placed in three wheels orthogonal to the z-axis, the Small Wheel before the end-cap

toroids, the Big Wheel after the toroids, and the Outer Wheel at ~ 6 m after the Big Wheel. The wheel structure is made by eight ϕ sectors, each one of those is composed by a small sector and a large sector, which have a little overlap in ϕ to minimize the acceptance holes. In the three wheels, the MDT chambers are installed for the precision measurement of the position along the radial coordinate. The TGC chambers are placed in the Large Wheel and, from 2015, in the Small Wheel, and provide, instead, the trigger and the measurement of the ϕ coordinate. In the Small Wheel are present also the CSC for the precision measurement of the tracks emitted in the region $2.0 < |\eta| < 2.7$. In this region of the Small Wheel, the particle rate is much higher than the MDT recommended value, 150 Hz/cm^2 , and therefore the CSC are used, since they are able to work with rates up to 1000 Hz/cm^2 .

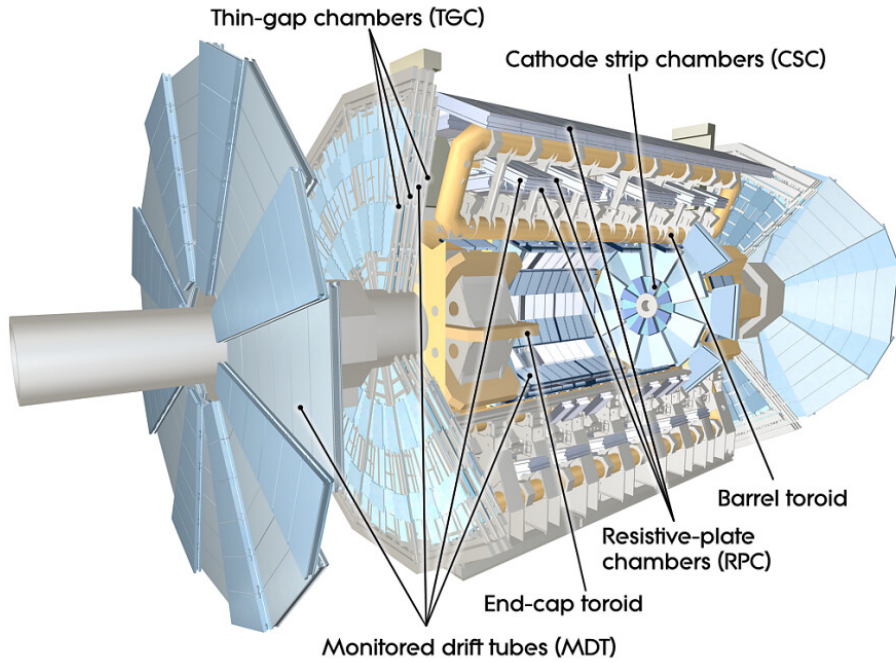


Figure 2.9: Cut-away view of the ATLAS Muon Spectrometer [33].

A brief description of the four types of detectors used in the MS is reported in the following.

Monitored Drift Tubes The MDT are gaseous detectors made of aluminium tubes with 3 cm of diameter and a length ranging from 1 to 6 m. The gas mixture used is 93% - 7% Argon-CO₂ at a pressure of 3 bar. At the center of the tube, there is a $50 \mu\text{m}$ wide tungsten wire, acting as a anode, put at an high-voltage of 3080 V. The tubes, instead, are grounded; due to this value of potential difference between wire and tube, the detector works in the proportional regime. The tubes are disposed to measure the precision coordinates η in the barrel and R in the end-caps.

The MDT signal is mainly given by the ions produced in the ionization occurring inside the tube at the passing of a charged particle. This signal is too slow to identify the BC of collision, not allowing their use in a L1 trigger. From the drift of the electrons to the anode, instead, it is possible to measure very precisely the position of the passage of the particle, that is the main purpose of these detectors. The average position resolution is $80 \mu\text{m}$ per single tube, while it is $50 \mu\text{m}$ for a chamber made by 3+3 or 4+4 tubes.

Cathode Strip Chambers The CSC are multi-wire proportional chambers with a segmented cathode made of strips. The wires are disposed in radial direction, while the strips are orthogonal to the wires on one cathode, and parallel to the wires on the other cathode. This structure allows to measure both the coordinates from the distribution of the charged induced on the strips, with a resolution of $60\ \mu\text{m}$.

Resistive Plate Chambers The RPC is the detector used in the ATLAS barrel trigger. They are gaseous detectors made of two parallel resistive plates with a gap of 2 mm filled with the gas mixture 97%-3%-0.3% $\text{C}_2\text{H}_2\text{F}_4$ - C_4H_{10} - SF_6 at atmospheric pressure. The plates work as anode and cathode with a uniform high-voltage of $\simeq 10\ \text{KV}$. Two planes of orthogonal strips, with average pitch of 30 mm, are placed on each chamber in order to measure both the η and the ϕ coordinates. The detector works in avalanche regime, where the electrons produced in the ionization are accelerated by the electric field producing showers on the anode. The signal is very fast, since it is induced on the strips by the drift of the electrons in the gap and their deposit on the plates. The RPC have a timing resolution of 1.5 ns, and can therefore identify the BC of the event, allowing their use in the trigger system. The spatial resolution is determined by the pitch of the strips and it is of the order of 1 cm.

Thin Gap Chambers In the MS end-caps, due to the high rate of particles, the RPC chambers are substituted by the Thin Gap Chambers. The TGC are multi-wire chambers with resistive cathode. The distance between the anode wire and the cathode (1.4 mm) is smaller than the wire-wire distance (1.8 mm), with the wires having a diameter of $\sim 50\ \mu\text{m}$. The TGC are filled with a gaseous mixture made of 55% of CO_2 and 45% of $\text{n-C}_5\text{H}_{12}$. The electric field configuration and the little distance between the wires allow an excellent timing resolution. The TGC are, in fact, used for the trigger in the end-cap region of the MS. Moreover, they provide the measurement of the azimuth coordinate, that is not measured by the MDT chambers.

2.1.6 Trigger system

The Trigger and Data Acquisition (TDAQ) system [49] is an essential component of the ATLAS experiment. It has the crucial task of deciding in real time whether to record data from a given collision, resulting in a huge impact on the datasets available in physics analyses. This reduction of the recorded data is needed due to the limitation in the data storage and the impressive event rate of the LHC collisions, happening every 25 ns. Fortunately, most of the produced data are not of interest for the ATLAS physics program, as the rate is dominated by low-pT inelastic and diffractive collisions. However, the ATLAS trigger system must provide high efficiency in the selection of interesting physics data. The whole system is then designed to find a balance between these requirements.

The ATLAS Run-2 TDAQ system is built on two levels of online selection, as shown in Figure 2.10: a first hardware-based level (L1), for a coarse reduction of the event rate, and a second software-based level (HLT), where the final decision on the selection of the event is made.

L1 trigger The hardware L1 trigger uses only a small fraction of the information delivered by the detector to take its decision in less than $2.5\ \mu\text{s}$. It exploits data with reduced granularity coming from custom electronics of the calorimeters and the muon detectors, in dedicated regions of interest (RoI). The calorimeters provide information

about clusters of energy deposits and missing transverse energy calculations, while the MS provides information from trigger chambers about transverse momentum and track position. The L1 trigger decision can reduce in this way the event rate of 40 MHz up to 100 kHz.

HLT trigger The high level trigger refines the decisions taken in the L1 trigger, integrating the RoI data with the full detector information. It is a software level that runs complex trigger algorithms to decide whether to select the events, with a two steps procedure. In the first step, called Level 2, dedicated fast trigger algorithms are used to provide early rejection, in which partial ID information, only inside identified RoI at L1, is incorporated in the trigger. Each L1 muon candidate is integrated with MDT data preforming a track fit extrapolated to the ID. Raw calorimetric informations are reconstructed by fast algorithms into cluster and cell objects. The second step, called Event Filter, consists in more precise and more CPU-intensive algorithms, similar to those used for offline reconstruction, to reconstruct the objects in jet, electron, and photon candidates, and perform the final selection. The HLT has a processing time of 200 ms and reduces the event rate up to 1 kHz.

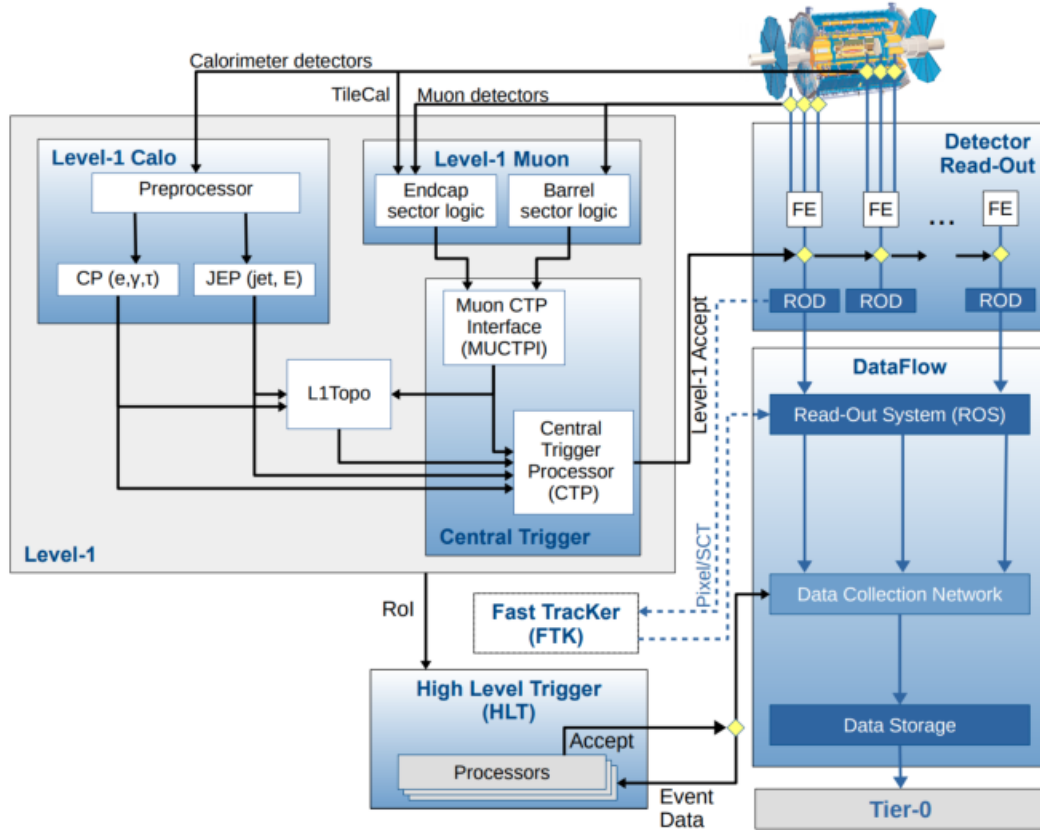


Figure 2.10: Trigger and data acquisition scheme of the Run 2 of the ATLAS detector [49].

2.2 Physics objects definition and reconstruction

Standard object definitions are usually recommended by the ATLAS collaboration to be used in physics analyses, and the analysis described in this work follow these recommendations. In this section are, therefore, described the definitions and the reconstruction techniques of the different physics objects in the ATLAS detector. The reconstruction is a complex process designed to identify the particles produced in the event

starting from track and cluster information, built using all the sub-detectors signals. The charged tracks emerging from the collisions are reconstructed together with the primary and secondary vertices of interaction, as described in Section 2.2.1. The reconstruction and identification of electrons, muons and taus are given in Section 2.2.2, 2.2.3 and 2.2.4, respectively. The hadronic jets reconstruction is presented in Section 2.2.5 together with the description of the b-tagging identification algorithm. Finally, the building method of the missing transverse energy is described in Section 2.2.6, while the overlap removal criteria between the different reconstructed objects is discussed in Section 2.2.7.

2.2.1 Tracks and vertices

As described in Section 2.1.3, charged particles leave several hits in the different sub-detectors of the ID during their passage, that are used to reconstruct the tracks. Track reconstruction in ATLAS [50] is performed in two stages: a loose track candidate is searched from combination of cluster of hits, then, a stringent ambiguity-solver algorithm compares and rates the individual tracks by assigning a relative track score to each track, selecting the best candidates.

The hits are first grouped in clusters for each pixel and SCT detector layer. From these clusters, three-dimensional space-points are created, representing the point where the charged particle has passed. Three space-points define a track seed. The seed, integrated with additional space-points from the remaining layers of the pixel and SCT detectors which are compatible with the preliminary trajectory, is used from a combinatorial Kalman filter [51] to build a track candidate. Some track candidates may share some of their space-points. This situation is solved by the ambiguity solver algorithm. Here, track candidates are processed individually in descending order of a track score, that is higher for tracks that more likely correctly represent the trajectory of a charged primary particle, selecting candidates with the largest score. This score is assigned depending on the track quality, looking at the momentum, cluster multiplicity, holes and χ^2 of the track fit. Track candidates are rejected if they fail to meet basic quality criteria, and, finally, the tracks are extended into the TRT. By the use of the full information of the three detectors, the tracks are fitted once again with a high-resolution fit to extract the final track parameters.

The final reconstructed tracks are described by five parameters:

- the impact parameter d_0 , defined as the distance of closest approach in the transverse plane of the track to the primary vertex;
- the z coordinate of the point where the track is closest to the interaction region z_0 ;
- η and ϕ of the outgoing particle;
- the track momentum or curvature q/p_T .

Interaction vertices can be reconstructed after track reconstruction [52]. The reconstruction of primary vertices is done in two steps. Reconstructed tracks are used to find a vertex candidate from their crossing point by the primary vertex finding algorithm. Then, the vertex position is reconstructed an adaptive vertex fitting algorithm [53], which takes as input the seed position and the tracks around it. Tracks incompatible with the vertex are used to seed a new vertex and the procedure is repeated until all tracks are associated of no additional vertex can be found. Among all the reconstructed vertices in a event, the vertex with the highest $\sum_{tracks} (p_T^{track})^2$ is selected as the primary vertex.

2.2.2 Electrons

The characteristic signature of electrons consists in a track in the ID together with a narrow shower in the EM calorimeter. An electron is, therefore, defined as an object consisting of a cluster built from energy deposits in the calorimeter and a matched track (or tracks) [54]. Tracks are reconstructed according to the procedure described in Section 2.2.1, while energy deposit clusters are formed from topologically connected EM and hadronic calorimeter cells using a cluster based algorithm. The algorithm first removes the cells not passing some noise thresholds, and then iteratively merges the neighbouring cells in the so called topo-clusters. Electron reconstruction only uses the energy from cells in the EM calorimeter, called EM energy, except in the transition region of $1.37 < |\eta| < 1.63$, where the energy measured in the hadronic calorimeter is considered. The EM fraction (fEM) is defined as the ratio of the EM energy to the total cluster energy. Clusters with less than 400 MeV are discarded and a preselection requirement of $\text{fEM} > 0.5$ is applied, as it rejects $\sim 60\%$ of pile-up clusters without affecting the selection efficiency of true electron topo-clusters. The ID tracks are then re-fitted accounting for additional energy loss to improve the estimation of the track parameters, and then matched to the selected clusters, called EM topo-clusters. The algorithm also builds conversion vertices using the tracks with the higher probability to be electron tracks as determined by the TRT. Two-track $\gamma \rightarrow e^+e^-$ conversion vertices are reconstructed from two opposite-charge tracks forming a vertex consistent with that of a massless particle, while single-track vertices are essentially tracks without hits in the innermost sensitive layers [54]. The reconstructed conversion vertices are then matched to the EM topo-clusters.

The final electron object is based on dynamic, variable-size clusters, called superclusters. EM topo-clusters are used as seed cluster candidates; the neighbouring EM topo-clusters are identified as satellite cluster candidates, which may emerge from bremsstrahlung radiation or topo-cluster splitting, and added to the seed candidates to the final superclusters. Finally the electron objects to be used for analyses are built. The four-momentum of the electrons is computed using information from both the final calibrated energy supercluster and the best track matched to the original seed cluster. The energy is given by the final calibrated supercluster, while the ϕ and η directions are taken from the corresponding track parameters with respect to the beam-line.

The reconstructed electron candidates contain a high contamination of electrons from photon conversions, non-isolated electrons from in-jets decays and jets faking electrons. To identify prompt electrons, a multivariate likelihood-based discriminant is used. The discriminating variables used by the likelihood identification are the longitudinal and transverse shower profiles, the track quality, the track and cluster positions to match in η and ϕ and the presence of high-threshold TRT hits. From cuts applied to the discriminator output, three levels of identification are provided in order of increasing electron purity and decreasing electron efficiency, labelled as Loose, Medium and Tight. The efficiencies for the three working points are shown in Figure 2.11.

An isolation criterion is also defined for the reconstructed electrons to further suppress the mis-identification, quantifying the energy of the particles produced around the electron candidate. The isolation variable used for reconstructed electrons is a track-based isolation, $p_T^{\text{varcone30}}$, defined as the sum of p_T of all tracks, satisfying quality requirements, within a cone of $\Delta R = \min(0.3, 10 \text{ GeV}/p_T)$ around the candidate electron track. A calorimeter-based isolation variable is also used, E_T^{cone20} . It is defined as the sum of the transverse energy of topological clusters in a cone of size $\Delta R = 0.2$ around the electron, after subtracting the contribution from the energy deposit of the electron itself and correcting for pile-up effects. Five isolation working points, listed in

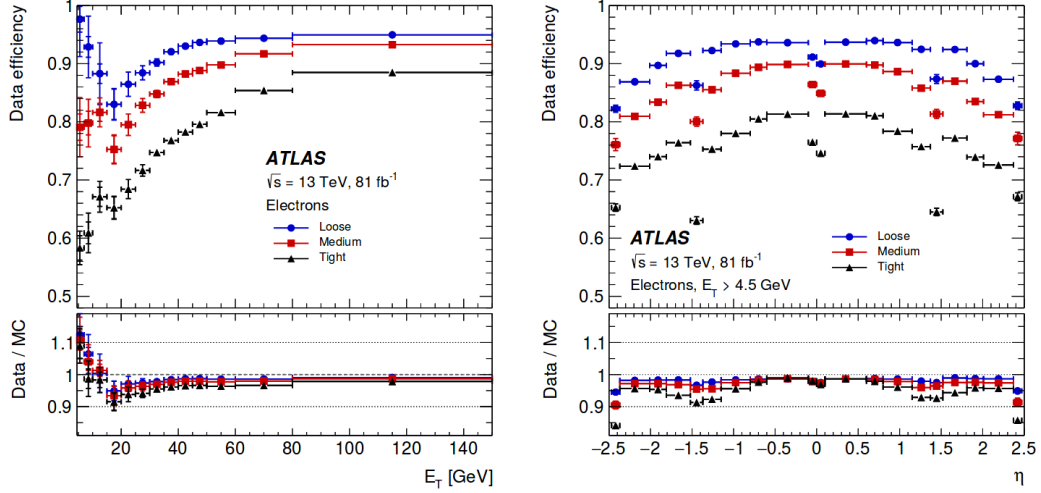


Figure 2.11: The electron identification efficiency in $Z \rightarrow ee$ events in data as a function of E_T (left) and as a function of η (right) for the Loose, Medium and Tight operating points [54].

Table 2.2 are defined using these two variables.

Isolation WP	Calorimeter isolation	Track isolation
HighPtCaloOnly	$E_T^{\text{cone20}} < \max(0.015 \cdot p_T, 3.5 \text{ GeV})$	-
TightTrackOnly_VarRad	-	$p_{T,\text{TTVA}}^{\text{varcone30}} / p_T < 0.06$
TightTrackOnly_FixedRad	-	$p_{T,\text{TTVA}}^{\text{varcone30}} / p_T < 0.06$ (if < 50 GeV) $p_{T,\text{TTVA}}^{\text{cone20}} / p_T < 0.06$ (if ≥ 50 GeV)
Tight_VarRad	$E_T^{\text{cone20}} / p_T < 0.06$	$p_{T,\text{TTVA}}^{\text{varcone30}} / p_T < 0.06$
Loose_VarRad	$E_T^{\text{cone20}} / p_T < 0.2$	$p_{T,\text{TTVA}}^{\text{varcone30}} / p_T < 0.15$

Table 2.2: Definition of the electron isolation working points. The pedix "TTVA" stands for track-to-vertex association and indicates that the track was used in the vertex fit, or satisfies $|\Delta z_0| \sin \theta < 3 \text{ mm}$.

2.2.3 Muons

Muons give a characteristic clean track in the ID and in the MS, while deposit a small amount of energy in the calorimeter system.

The muon reconstruction is first done independently in the ID and in the MS. In the ID, muons are reconstructed like any other charged particle as described in Section 2.2.1. In the MS, the reconstruction starts with the "segment" finding, i.e. a short straight-line track, in the two precision sub-detectors, MDT and CSC, using a Hough Transform [55] to search for hits aligned on a trajectory in the bending plane of the detector. The second coordinate is provided to MDT segments by RPC or TGC hits. Muon track candidates are then built by fitting together hits from segments in different layers. A track candidate is accepted if the χ^2 of the fit satisfies the selection criteria. Hits providing large contributions to the χ^2 are removed and the track fit is repeated. A hit recovery procedure is also performed looking for additional hits consistent with the candidate trajectory, and the fit is repeated if additional hits are found.

At this stage, the segments from the ID and the MS are combined, using also calorimeter information. Depending on the type of combination, four muon types are defined:

- Extrapolated muons (ME): muons reconstructed exclusively from hits in the MS.

- Combined muons: muons reconstructed with a global refit of the hits from both the ID and MS subdetectors, used to reconstruct independently tracks in the ID and MS. During the global fit procedure, MS hits may be added to or removed from the track to improve the fit quality. The standard procedure follows an outside-in pattern recognition, in which the muons are first reconstructed in the MS and then extrapolated inward and matched to an ID track. An inside-out combined reconstruction, in which ID tracks are extrapolated outward and matched to MS tracks, is used as a complementary approach [56]. This alternative approach has advantages for muons with low p_T , which lose a significant part of their energy in the calorimeter and cannot be reconstructed in the MS because of small number of hits or big deviation of the track due to energy loss and multiple scattering. This approach is also advantageous where the muon spectrometer coverage is incomplete because of layout issues or malfunctioning chambers [57]. The combined muons are the muons with the best quality, but are available only in the region $|\eta| < 2.5$.
- Segment-tagged muons: ID tracks are extrapolated outwards to the precision plane of the muon segments. Then, a matching is performed using position and angle in the precision rz -plane and the position in the xy -plane. Finally, additional cuts on the quality of the segments are applied to reduce the contribution of fake muons.
- Calorimeter-tagged muons: the algorithm search for a match between ID tracks and energy deposits compatible with muons in the calorimeter. If the match is found, a cut is applied to define the ID track as a muon.

After reconstruction, high-quality muon candidates are selected by a set of requirements on the number of hits in the different ID and MS sub-detectors stations, the track fit properties, and the match between the individual measurements in the two detector systems [58]. Five muon identification working points are then defined for use in physics analyses:

- Loose: maximizes the reconstruction efficiency while providing good quality muon tracks. All muon types are used.
- Medium: minimizes the systematic uncertainties associated with muon reconstruction and calibration. Only combined and extrapolated tracks are used, with some selection criteria on the number of hits in the tracks [58].
- Tight: maximises the purity of muons at the expenses of the efficiency. Only CB muons with hits in at least two stations of the MS and satisfying the Medium selection criteria are considered.
- High- p_T : provides the best momentum measurement for the p_T range of hundreds of GeV up to several TeV.
- Low- p_T : maximizes the efficiency for muons down to $p_T = 3$ GeV while maintaining a reasonable fake rate.

The efficiencies for Medium and Tight muons are shown in Figure 2.12.

Isolation requirements on the muons are necessary to select muons originating from the decay of heavy particles, such as W , Z , or Higgs bosons, while rejecting muons coming from semileptonic decays, which are embedded in jets. The muon isolation variables, $p_T^{\text{varcone30}}$ and $E_T^{\text{topocone20}}$, are defined in a similar way to the electron isolation

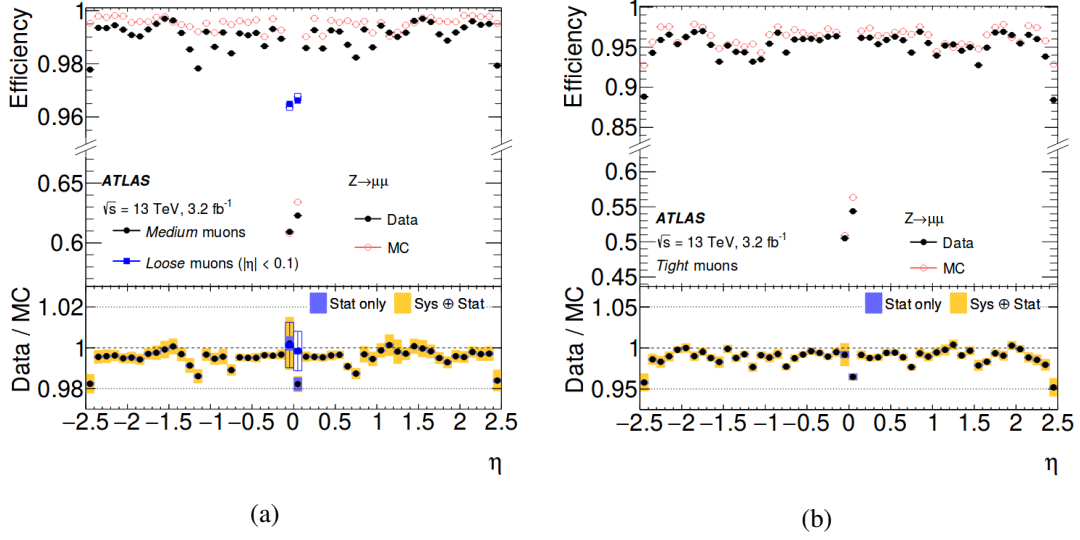


Figure 2.12: Muon reconstruction efficiency as a function of η measured in $Z \rightarrow \mu\mu$ events for Medium (a) and Tight (b) muon working points. The plot (a) also shows the efficiency of the Loose WP in the region $|\eta| < 0.1$ where the Loose and Medium selections differ significantly [56].

variables. An additional particle-flow based variable is defined as the transverse energy of neutral particle-flow objects in a cone of size $\Delta R = 0.2$ around the muon, labelled as E_T^{neflow20} . Five isolation working points, listed in Table 2.3 are defined using these two variables.

Isolation WP	Definition	Track p_T requirement
PflowLoose*	$(p_T^{\text{varcone30}} + 0.4 \cdot E_T^{\text{neflow20}}) < 0.16 \cdot p_T^\mu$	$p_T > 500 \text{ MeV}$
PflowTight*	$(p_T^{\text{varcone30}} + 0.4 \cdot E_T^{\text{neflow20}}) < 0.045 \cdot p_T^\mu$	$p_T > 500 \text{ MeV}$
Loose*	$p_T^{\text{varcone30}} < 0.15 \cdot p_T^\mu, E_T^{\text{topocone20}} < 0.3 \cdot p_T^\mu$	$p_T > 1 \text{ GeV}$
Tight*	$p_T^{\text{varcone30}} < 0.04 \cdot p_T^\mu, E_T^{\text{topocone20}} < 0.15 \cdot p_T^\mu$	$p_T > 1 \text{ GeV}$
HighPtTrackOnly	$p_T^{\text{cone20}} < 1.25 \text{ GeV}$	$p_T > 1 \text{ GeV}$
TightTrackOnly*	$p_T^{\text{varcone30}} < 0.06 \cdot p_T^\mu$	$p_T > 1 \text{ GeV}$

Table 2.3: Definitions of the muon isolation WPs. The criteria used are listed in the second column, while the requirement on the minimal track p_T is shown in the third column. WPs marked with * exist in two variants: one with the cone ΔR parameter shrinking following $\min(10 \text{ GeV}/p_T^\mu, 0.3)$ for any p_T^μ , the other remaining constant at $\Delta R = 0.2$ for $p_T^\mu > 50 \text{ GeV}$ [58].

2.2.4 Taus

The reconstruction of hadronic tau candidates ($\tau_{\text{had}}^{\text{vis}}$) is described in detail in Ref. [59]. The candidates are seeded by jets formed using the anti- k_t clustering algorithm [60], with a distance parameter of 0.4. A set of boosted decision trees is used to classify all tracks within $R = 0.4$ of the $\tau_{\text{had}}^{\text{vis}}$ axis into core and isolation tracks, depending on their p_T , the number of hits in the tracking detectors and their transverse and longitudinal impact parameters with respect to the tau vertex.

Dedicated algorithms are used to identify the visible decay products of hadronic tau decays, and, due to the distinct signatures of 1- and 3-prong hadronic tau decays, the tau identification is split into dedicated algorithms for 1- and 3-track $\tau_{\text{had}}^{\text{vis}}$. In particular

a recurrent neural network (RNN) classifier is used [61]. Depending on cuts on RNN score, four working points with increasing background rejection (Very loose, Loose, Medium and Tight) are defined to be used by physics analyses. The corresponding signal selection efficiencies and rejection powers are given in Table 2.4.

Working point	Signal efficiency		Background rejection RNN	
	1-prong	3-prong	1-prong	3-prong
Tight	60%	45%	70	700
Medium	75%	60%	35	240
Loose	85%	75%	21	90
Very loose	95%	95%	9.9	16

Table 2.4: List of tau identification working points with fixed true τ_{had}^{vis} selection efficiencies and the corresponding background rejection factors for misidentified τ_{had}^{vis} in dijet events for the RNN classifier, both for 1-prong and 3-prong selections [61].

2.2.5 Hadronic Jets

At hadron colliders, quarks and gluons are also produced in the collisions. Due to the QCD colour confinement, quarks and gluons, which are colored, hadronise and produce collimated showers of particles, called hadronic jets.

The jet reconstruction has the goal to combine these produced particles into a physics object that gives information about the initial parton. The jet reconstruction in ATLAS can be performed in two ways: using only calorimeter information, or via a more recent method that combines calorimeter and ID informations. The first method produces the so called electromagnetic topological jets (EMTopoJets). Calorimeter cells are grouped by a three-dimensional topological clustering algorithm in the so called topological clusters (topo-clusters) of energy deposits. The topo-clusters are used as inputs to a jet finding algorithm that reconstructs the EMTopoJets using the anti- k_t algorithm [60] with a distance parameter $R = 0.4$. The anti- k_t clustering algorithm sequentially combines topo-clusters into larger objects based on the momentum-weighted distance between two clusters, given by:

$$d_{i,j} = \min \left(k_{ti}^{2p}, k_{tj}^{2p} \right) \frac{\Delta_{ij}^2}{R^2} \quad (2.13)$$

and the momentum-weighted distance between a cluster and the beam, given by:

$$d_{i,B} = k_{ti}^{2p} \quad (2.14)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and y_i and ϕ_i are respectively the rapidity and the azimuth of particle i , while k_{ti} is the transverse momentum. R is distance radius parameter and p is a parameter of the anti- k_t algorithm that is set to -1. The algorithm identifies the minimum distance between d_{ij} and d_{iB} , starting from the entity i with the highest momentum as seed. If $d_{ij} < d_{iB}$, i and j are combined into a single pseudo-jet. Otherwise, i is considered as a final state and is removed from the list of entities. The distances are then recalculated between the objects in the new updated list, and the procedure is repeated until no entities are left. For construction, this algorithm tends to cluster soft particles with hard objects before clustering soft particles together. The algorithm is therefore sensitive only to hard particles proximity at disadvantage of soft radiation, resulting in conical-shape jets, as visible in Figure 2.13.

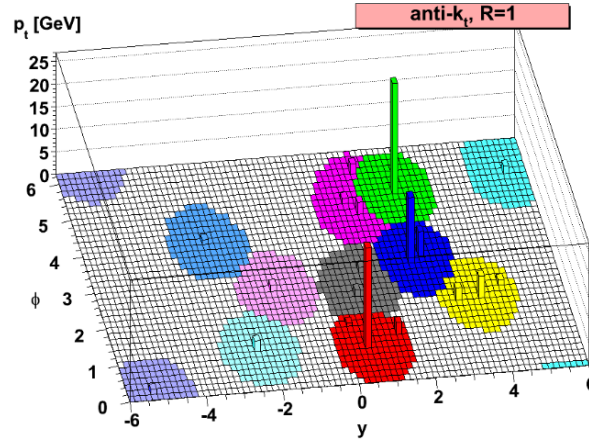


Figure 2.13: Illustration of the active catchment areas of the resulting hard jets clustered with the anti- k_t clustering algorithm [60].

The second reconstruction method used in ATLAS is based on particle-flow techniques [62]. The ID tracking system provides more precise measurements of charged particle energies and positions, while at high energies, the calorimeter's energy resolution is superior to the tracker's momentum resolution. This technique combines the ID tracking information for charged particles with the calorimeter information for both charged and neutral particles for an optimal event reconstruction. Both the hadronic jets and the soft activity, i.e. additional hadronic activity below the threshold used in jet reconstruction, reconstruction is performed with this combination. Outside the geometrical acceptance of the ID, only the calorimeter information is available and therefore only topo-clusters are used as inputs to the particle flow jet reconstruction. In the region $|\eta| < 2.5$ the energy of the charged particle is double-counted. To avoid this, the charged track's signal in the calorimeter is identified and subtracted on a cell-based basis, ideally leaving only a calorimeter measurement of the electrically neutral particles. This procedure is done only for well reconstructed ID tracks, satisfying stringent requirements on the number of hits in the reconstructed track. The main challenge of this technique, for which it is optimised, is to accurately subtract all track's energy, without removing any energy deposited by any other particle.

Jet objects are then calibrated in the energy scale using truth simulated jets. The calibration, known as Jet Energy Scale (JES), is needed to correct for non-linearities, energy loss in inactive areas, leakage, mis-reconstruction and different energy scales in the electromagnetic and hadronic calorimeters. A correction due to pile-up is also applied, since the energy deposition in jets is affected by the presence of multiple pp collisions in the same bunch crossing as well as residual signals from other bunch crossings.

Moreover, the Jet Vertex Tagger (JVT) tool is used in ATLAS for additional pile-up suppression [63]. This technique uses a multivariate combination of pile-up sensitive track information to discriminate between pile-up jets and jets originating from the primary vertex.

b-tagging The identification of jets produced by b -quark hadronisation, known as b -jets, is of high importance in order to discriminate these particular jets from the larger background of jets containing other quark flavours. Many physics processes of interest contain in their signature a b -jet, like the decay of the Higgs boson into bottom quark pairs, or the decay of a top quark into a bottom quark and a W that is present in many

SM and BSM processes.

b-hadrons have lifetimes of the order of 1.6 ps, which allows the hadron to have a small but significant mean flight length in the detector before the decay. This length is of the order of 2-3 mm from the primary vertex and it is followed by the b-hadron decay vertex, usually referred to as secondary vertex. The secondary vertex and the tracks arising from it can be detected in the first layers of the inner detector. The distance between the secondary and the primary vertex, the mass of the particles associated to the vertex and the impact parameter d_0 of the tracks are useful quantities used for the identification of b-jets. A schematic view of decay of a b-hadron into a jet is shown in Figure 2.14.

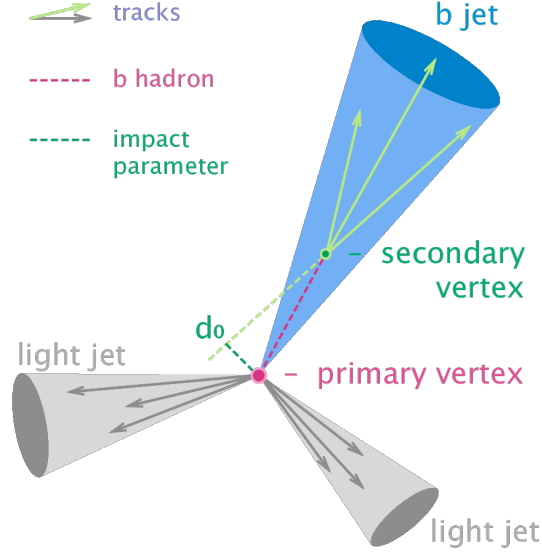


Figure 2.14: Schematic view of a b-hadron decay inside a jet resulting in a secondary vertex with three charged particle tracks. The track impact parameter d_0 and two additional light-jets are also shown.

In ATLAS several algorithms are used to perform the identification of the b-jets and the rejection of c-jets and light-jets, known as flavour tagging algorithms [64]. There are two types of flavor tagging algorithms: the low level algorithms and the high level algorithms. The low level algorithms use information coming from low level reconstructed objects, such as tracks and ID hits clusters. The high level taggers use the low level tagger outputs, such as reconstructed secondary vertices or jet flavor probabilities, to decide if a given jet is a b, c or light-flavor jet.

The high level algorithms are two, MV2 and DL1, and differ mainly on their architecture. MV2 is based on a boosted decision tree, while DL1 is based on a deep neural network and provides an output with three different probabilities (p_b , p_c and p_u) that are combined to define a final discriminant:

$$d_{DL1} = \log \frac{p_b}{f_c \times p_c + (1 - f_c) \times p_u} \quad (2.15)$$

where f_c controls the importance of c-jets discrimination in the tagger and is optimized to improve the performance for both light and c-jets rejection. DL1r is a variant of DL1, that contains an additional low level algorithm as input. This provides, as visible in Figure 2.15, a better rejection of c-jets and light-jets compared to the other two taggers.

Several working points based on the efficiency in b-jet identification are provided for each b-tagging algorithm. Usually these working point are defined to provide 60%,

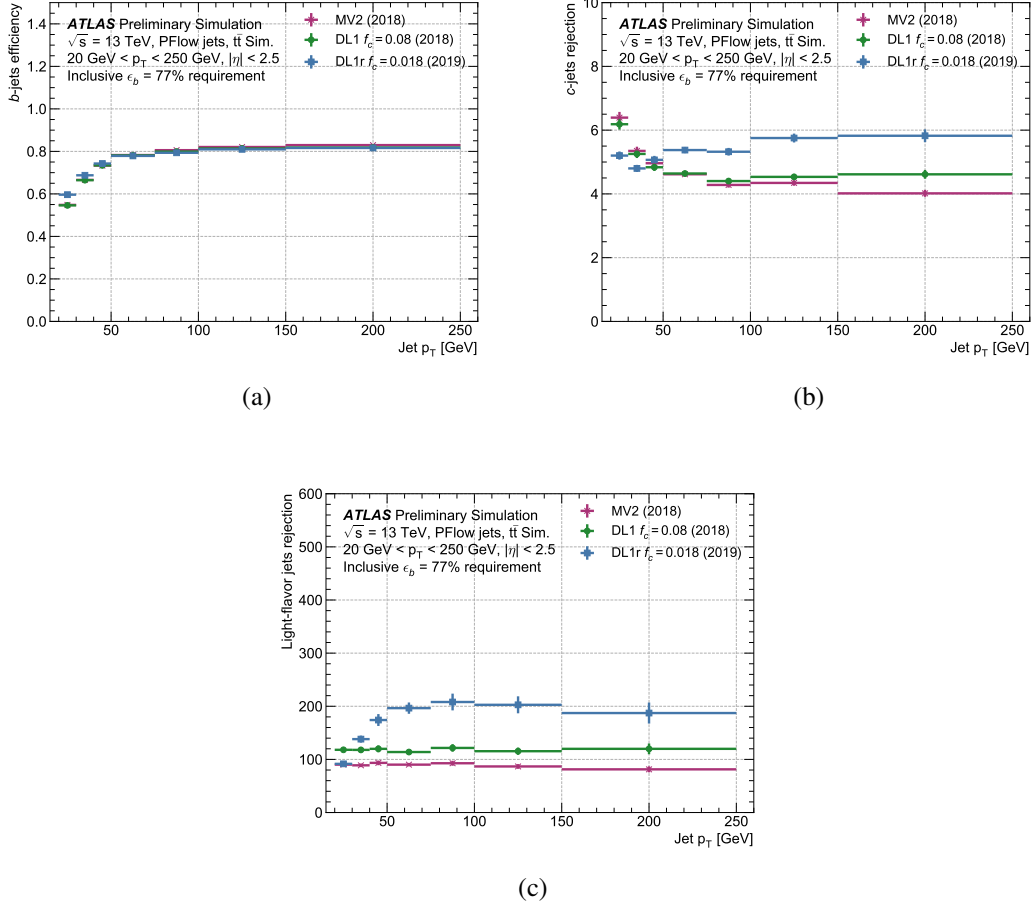


Figure 2.15: (a) Efficiency of identifying b -jets as a function of jet p_T for an inclusive $\epsilon_b = 77\%$ efficiency requirement, for MV2, DL1 and DL1r algorithms. Rate of (b) c -jets and (c) light-flavor jets rejection ($1/\epsilon_{c,u}$) as a function of jet p_T for an inclusive $\epsilon_b = 77\%$ efficiency requirement, for MV2, DL1 and DL1r algorithms.

70%, 77% and 85% b -tagging efficiency. Jets passing the selected b -tagging algorithm working point are commonly called b -tagged jets, while if not passing the cut are called light-jets.

2.2.6 Missing Transverse Energy

Since neutrinos, or other invisible particles, are not directly detectable, they appear only as an imbalance in transverse momentum, known as missing transverse momentum (MET, or E_T^{miss}). The initial state of LHC collisions is well defined only in the transverse plane. Since the proton beams are aligned with the longitudinal axis, the transverse momentum of the partons inside the proton is negligible. For this reason the initial state transverse momentum can be approximated with zero. For the longitudinal momentum this is not true since the momentum fraction x_1 and x_2 carried by the two interacting partons is unknown. The momentum conservation of all the produced particles, then, can be imposed only in the transverse plane. The E_T^{miss} is defined as:

$$-E_T^{\text{miss}} = \sum p_T^e + \sum p_T^\mu + \sum p_T^\gamma + \sum p_T^\tau + \sum p_T^{\text{jet}} + \sum p_T^{\text{soft}} \quad (2.16)$$

E_T^{miss} is computed by using all the objects passing a baseline, usually loose, selection, depending on the specific analysis. Fully calibrated electrons, muons, photons, hadron-

ically decaying τ -leptons, and jets reconstructed from calorimeter energy deposits and charged-particle tracks are used in the MET reconstruction [65]. These are combined with the so called MET soft-term, that is the soft hadronic activity measured by reconstructed charged-particle tracks matched to the primary vertex not associated with the already counted reconstructed objects. Possible double counting of contributions from reconstructed charged-particle tracks from the ID, energy deposits in the calorimeter, and reconstructed muons from the MS is avoided by applying a signal ambiguity resolution procedure which rejects already used signals when combining the various E_T^{miss} contributions.

2.2.7 Overlap Removal

In order to uniquely identify the physics objects, overlapping objects are removed according to the so called overlap removal procedure. This procedure is a standard procedure performed in ATLAS analyses and it is here described:

1. Electron-Muon overlap :
If a muon is calo-tagged and shares an ID track with an electron, the muon is removed. While, if the muon is not calo-tagged, the electron is removed.
2. Jet-Electron overlap : Jet Removal
If a surviving electron and a jet overlap within a $\Delta R(\text{calo-jet}, \text{electron}) < 0.2$, the jet is removed.
3. Jet-Electron overlap : Electron Removal
If an electron and a jet overlap within a $\Delta R(\text{electron}, \text{calo-jet}) < 0.4$, the electron is removed.
4. Jet-Muon overlap : Jet Removal
If a muon and a jet overlap with $\Delta R(\text{calo-jet}, \text{muon}) < 0.2$, the jet is removed, unless it passes any of the following criteria:
 - the number of tracks in the jet are more than two;
 - the jet is b-tagged.
5. Jet-Muon overlap : Muon Removal
If a muon and a jet overlap with $\Delta R(\text{muon}, \text{calo-jet}) < 0.4$, the muon is removed.

Chapter 3

Higgs boson at the Large Hadron Collider

In this Chapter the state of the art of single Higgs and Higgs pair measurements is presented. These measurements are among the most important in the physics program of the LHC experiments both during Run 1 and Run 2 periods. An overview of the Higgs boson discovery in the main decay channels is given in Section 1, together with a summary of the latest measurements of mass, width and cross section of the SM Higgs boson. The state of the art of the Higgs pair searches, with a brief description of the analysis strategies performed in the different decay channels, is reported in Section 2.

1 Discovery and measurements

The discovery of the Higgs boson has been the main result of the LHC Run 1. On 4th July 2012, the discovery of a particle compatible with the SM Higgs boson has been announced by the ATLAS and CMS experiments. The new particle was observed using LHC pp collisions data at a centre-of-mass energy of $\sqrt{s} = 7$ TeV in 2011 and $\sqrt{s} = 8$ TeV in 2012. The ATLAS experiment observation was given by the combination of the decay channels $H \rightarrow ZZ^* \rightarrow 4l$, $H \rightarrow \gamma\gamma$ and $H \rightarrow WW^* \rightarrow l\nu l\nu$ with a significance of 5.9σ [23]. The CMS experiment observation was obtained with the combination of the decay channels $H \rightarrow ZZ^* \rightarrow 4l$, $H \rightarrow \gamma\gamma$ and $H \rightarrow WW^* \rightarrow l\nu l\nu$, $H \rightarrow b\bar{b}$ and $H \rightarrow \tau^+\tau^-$ with a significance of 5.0σ [24]. These measurements were obtained evaluating the cross sections of the Higgs production process via gluon-gluon fusion (ggF), performing cuts on kinematic variables to select the desired final state for each channel. The invariant mass distributions of the main discovery decay channels for ATLAS and CMS are shown in Figure 3.1 and 3.2 respectively. The mass of the observed particle was measured using $H \rightarrow ZZ^* \rightarrow 4l$ and $H \rightarrow \gamma\gamma$ channels, which provide a high mass resolution of 1-2% and a high sensitivity. The combined ATLAS+CMS Run 1 measured mass is currently the reference value of the Higgs boson mass, and it is:

$$m_H = 125.09 \pm 0.24 \text{ GeV} \quad (3.1)$$

where the total, statistical plus systematic, uncertainty is indicated.

The discovery was followed by a detailed exploration of properties of the Higgs boson during both Run 1 and Run 2 of the LHC. A summary of several measurements of the mass of the Higgs boson performed by ATLAS and CMS in different periods of data taking is reported in Figure 3.3.

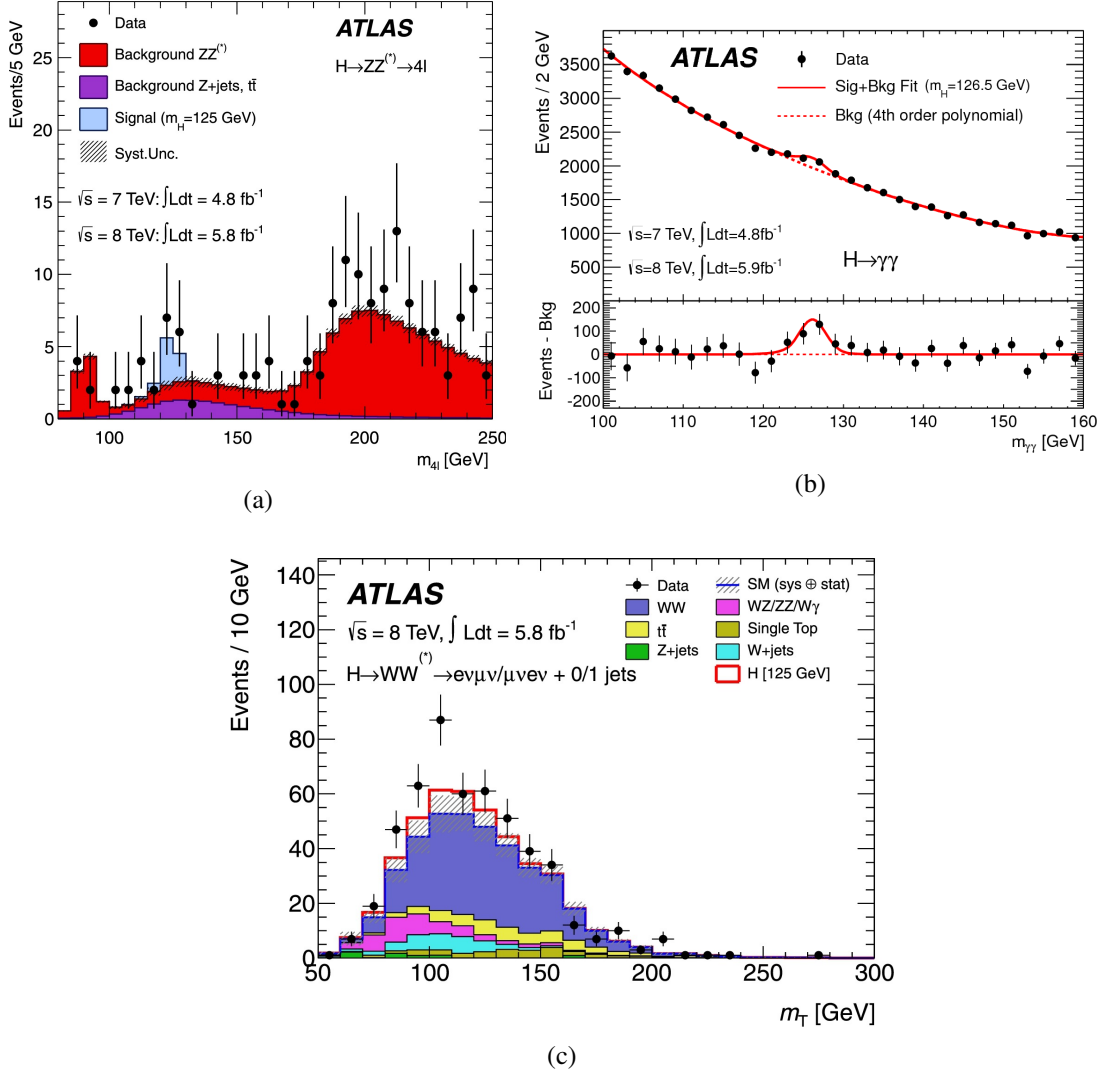


Figure 3.1: The distributions of the four-lepton invariant mass m_{4l} (a), of the invariant mass of diphoton candidates $m_{\gamma\gamma}$ (b), and of the transverse mass m_T of the $e\mu/\mu e$ channels combined (c), after all the selections for the combined 7 TeV and 8 TeV data sample [23]. The two channels of the $H \rightarrow WW$ analysis differs for the p_T ordering of the leptons: in the $e\mu$ channel the leading- p_T lepton is the electron, while the opposite is for the μe channel.

In the SM, the Higgs boson width is very precisely predicted once the Higgs boson mass is known. For a mass of 125.1 GeV, the Higgs boson has a very narrow width of 4.2 MeV. It is dominated by the fermionic decays partial width at approximately 75%, while the vector boson modes are suppressed and contribute 25% only [3]. Direct on-shell measurements of the Higgs-boson width are limited by detector resolution that brings a mass resolution of 1-2 GeV, much larger than the expected SM Higgs boson width. Also indirect width measurements are possible, exploiting off-shell production of the Higgs boson, for which vector bosons and top-quark decay products become on-shell. A combination of on- and off-shell measurements allows to obtain the best limits on the Higgs boson width, and it has been performed both by ATLAS and CMS. ATLAS, using $ZZ \rightarrow 4l$, and $ZZ \rightarrow 2l2\nu$ final states and data corresponding to an integrated luminosity of 36.1 fb^{-1} , has put a limit on the Higgs boson width [66] of:

$$\Gamma_H < 14.4 \text{ MeV at 95\% CL} \quad (3.2)$$

The CMS limit at 95% CL for the Higgs-boson width from on-shell and off-shell

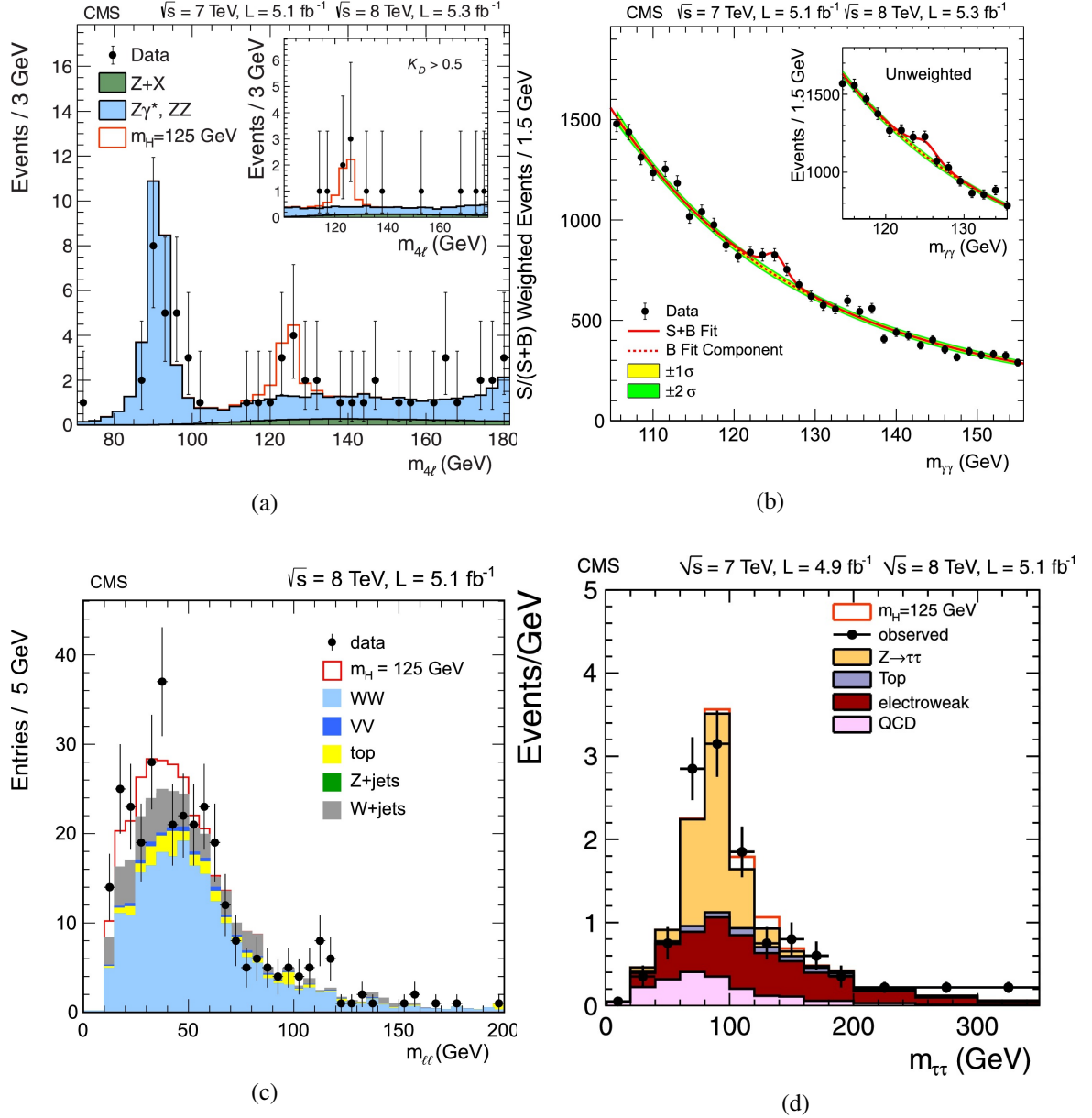


Figure 3.2: The distributions of the four-lepton invariant mass $m_{4\ell}$ (a), of the invariant mass of diphoton candidates $m_{\gamma\gamma}$ (b), of the dilepton invariant mass $m_{\ell\ell}$ (c), and of the ditau invariant mass (d), after all the selections. The luminosity used in each channel is given in each plot [24].

Higgs boson production in the four-lepton final state using an integrated luminosity of 80.2 fb^{-1} , under the assumption of SM-like couplings, is [67]:

$$0.08 < \Gamma_H < 9.16 \text{ MeV} \quad (3.3)$$

New analyses, extending the list of the studied decay channels, have been performed exploiting the higher luminosity provided by the Run 2 data of the LHC collisions. Moreover, the higher available statistics has opened the possibility to perform searches of Higgs boson candidates generated via production modes different from the gluon-gluon fusion, providing new observations of such a particle produced via vector boson fusion (VBF), associated production with a vector boson (VH) and top-quark pairs (ttH). An overview of the latest ATLAS and CMS measurements of cross section times branching ratios for the different association between production modes and decay channels is shown in Figure 3.4a and 3.4b respectively. Moreover, new analyses exploiting the

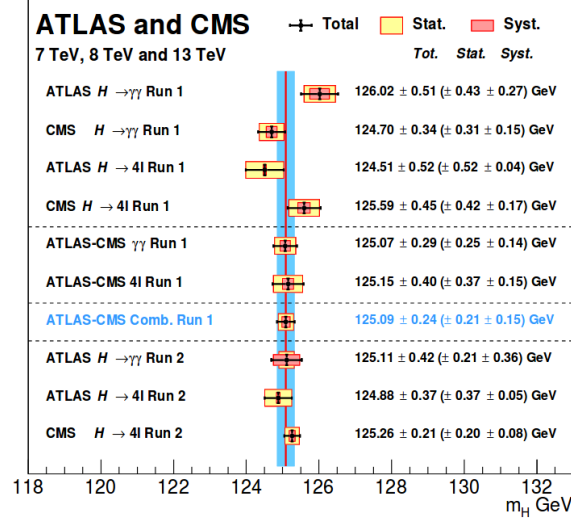


Figure 3.3: Summary of CMS and ATLAS Higgs boson mass measurements in the $\gamma\gamma$ and ZZ channels in Run 1 and Run 2 [3].

full Run 2 data are currently ongoing both in ATLAS and CMS and will provide more precise measurements of all the properties of the Higgs boson.

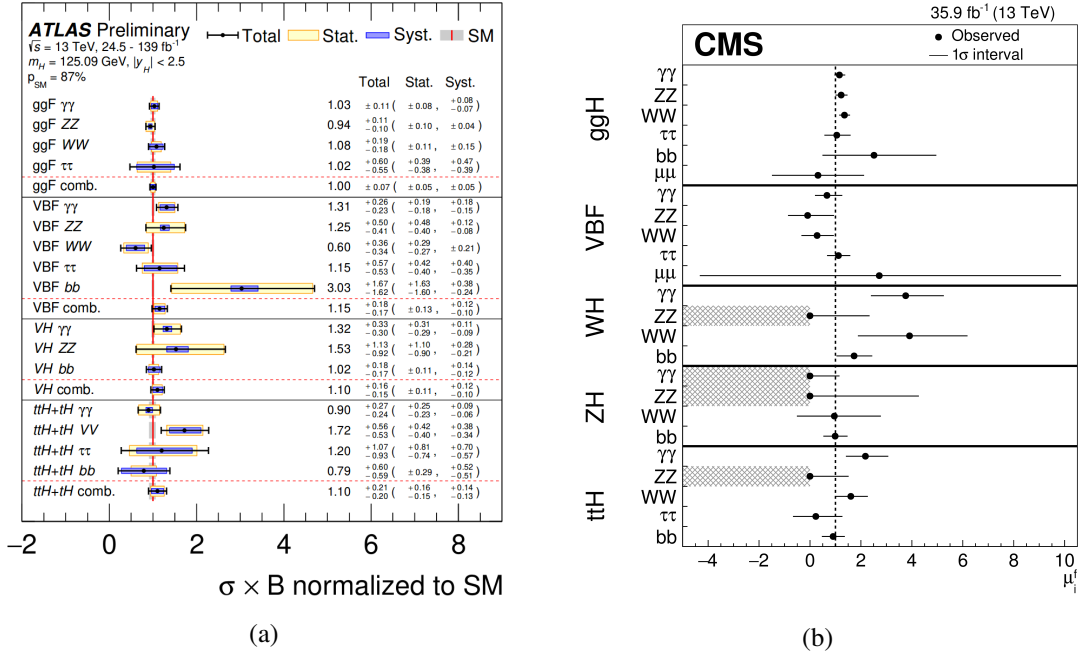


Figure 3.4: (a) Cross sections times BR for ggF, VBF, VH and $t\bar{t}H + tH$ production in each relevant decay mode, normalized to their SM predictions. The values are obtained from a simultaneous fit to all channels. Combined results for each production mode are also shown. The total, systematic and statistical uncertainties are shown for each measurement. The gray bands show the theory uncertainties in the predictions [68].

(b) Summary plot of the fit to the production-decay signal strength¹ products. The points indicate the best fit values while the horizontal bars indicate the 1 σ CL intervals. The hatched areas indicate signal strengths that are restricted to non negative values [69].

¹The signal strength is defined as the ratio between the cross section of the searched theory σ_{th} and the actual observed cross-section σ : $\mu = \frac{\sigma}{\sigma_{th}}$.

2 Higgs-pair production: state of the art

The searches for Higgs boson pair production have started in Run 1 with several ATLAS and CMS analyses, exploiting the data collected at a center-of-mass energy of pp collisions of $\sqrt{s} = 8$ TeV [70–74]. After these first studies, limited by the very low statistics for such a rare process, many new ATLAS and CMS analyses have been performed with partial Run 2 data. These analyses have taken advantage from both the increased statistics and from the higher center-of-mass energy of the proton-proton collisions, which during the Run 2 was $\sqrt{s} = 13$ TeV. The higher center-of-mass energy provides, in fact, a larger cross section for Higgs boson pair production process, $\sigma(pp \rightarrow HH)$, going from 9.441 fb at $\sqrt{s} = 8$ TeV to 31.05 fb at $\sqrt{s} = 13$ TeV for the gluon-gluon fusion production mode [75]. The Run 2 analyses and their results are briefly described in Section 2.1 and 2.2 for ATLAS and CMS respectively. Analyses focusing on $b\bar{b}VV$ decay channel are discussed in Chapter 4 as they can be directly compared to the analysis that is the object of this work.

2.1 ATLAS results

ATLAS HH early Run 2 searches, using the data collected at $\sqrt{s} = 13$ TeV in 2015 and 2016, corresponding to an integrated luminosity up to 36.1 fb^{-1} , have been performed in six decay channels: $b\bar{b}b\bar{b}$ [76], $b\bar{b}\tau^+\tau^-$ [77], $b\bar{b}\gamma\gamma$ [78], $b\bar{b}WW^*$ [79], $\gamma\gamma WW^*$ [80], $WW^{(*)}WW^{(*)}$ [81]. The analyses have focused only on ggF production mode, the one with the higher cross section, due to the limited statistics. Furthermore, a combination of these six channels has been performed [82], obtaining a better limit on the HH production cross section. A brief description of the analyses and of their results is reported here, except for $b\bar{b}WW^*$ analysis that is described in Chapter 4.

$b\bar{b}b\bar{b}$ analysis The $b\bar{b}b\bar{b}$ analysis [76] looks for final states with at least four b-tagged jets reconstructed using the anti- κ_t algorithm with $R = 0.4$. Two Higgs boson candidates, each composed of two b-tagged jets with invariant masses close to m_H , are selected and combined to form the Higgs boson pair system. The invariant mass of the Higgs boson candidate pair system, $m_{HH} = m_{4j}$, is used as the final discriminant between signal and the backgrounds, which are mainly QCD multijets and $t\bar{t}$. The SM HH production was expected to result in an excess in the tail of the m_{4j} spectrum, on which the fit to estimate the HH production cross section is performed. The dataset is split according to the years 2015 and 2016, and then statistically combined taking into account the different trigger algorithms used in 2015 and 2016. In part of the 2016 data period, inefficiencies in the online vertex reconstruction affected b-jet triggers that were used in the analysis for the event selection, reducing the 2016 integrated luminosity to 24.3 fb^{-1} and the total available integrated luminosity to 27.5 fb^{-1} . The shape of the m_{4j} distribution is shown in Figure 3.5 for 2015 and 2016 data. The observed 95% CL upper limit is $\sigma(pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}) < 147 \text{ fb}$. This value, compared with the SM prediction for gluon-gluon fusion produced HH of $\sigma(pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}) = 11.3^{+0.9}_{-1.0} \text{ fb}$, results in an observed upper limit on the cross section of 12.9 times the SM.

$b\bar{b}\tau\tau$ analysis The $b\bar{b}\tau\tau$ analysis [77] selects events looking for two particular final states, where one or two of the τ leptons decay hadronically. These events are, therefore, characterized by one electron or muon and one τ_{had}^{vis} , as defined in Section 2.2.4, of opposite charge, or two τ_{had}^{vis} of opposite charge, plus two b-tagged jets and E_T^{miss} . The selection criteria are slightly different for $\tau_{lep}\tau_{had}$ and $\tau_{had}\tau_{had}$ channels, optimizing the selection of such final states with the use of single-lepton and lepton plus τ_{had} triggers

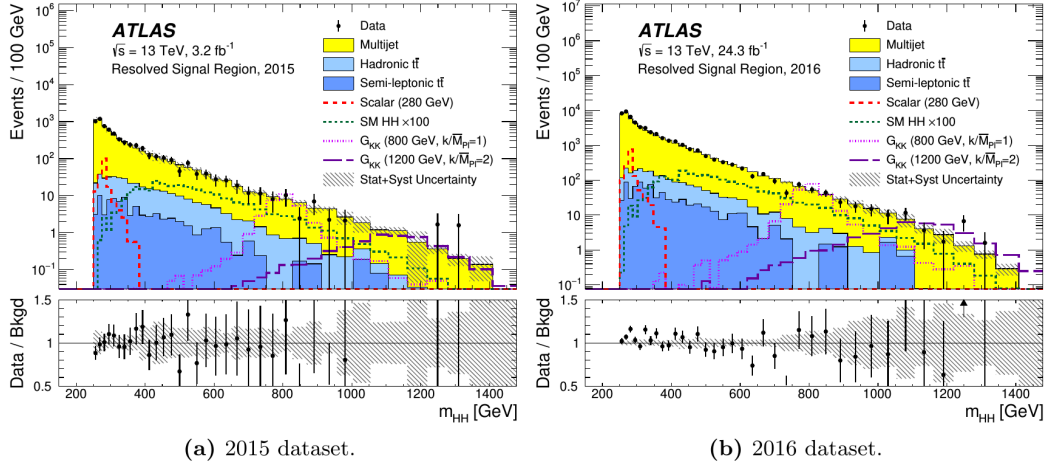


Figure 3.5: Distributions of m_{4j} in the signal region of the $b\bar{b}b\bar{b}$ analysis for (a) 2015 data and (b) 2016 data, compared to the predicted backgrounds. The hatched bands represent the combined statistical and systematic uncertainties in the total background estimates. The expected signal distributions of SM non-resonant HH production ($\times 100$) is also shown [76].

or single- τ_{had} or di- τ_{had} triggers, respectively. For signal region events the invariant mass of the di- τ system, $m_{\tau\tau}$, is required to be greater than 60 GeV with the addition of the presence of two b-tagged jets. Boosted Decision Trees are used in the analysis to improve the separation of signal from the main background sources, taking in input a set of kinematics variables. The BDTs are trained against $t\bar{t}$, $Z \rightarrow \tau\tau$, and multijet events for the $\tau_{had}\tau_{had}$ channel, and only against the main $t\bar{t}$ background for $\tau_{lep}\tau_{had}$ channel. Their output distributions are used then in the final fit to extract the signal cross section. As no significant excess over the expected background was observed, upper limits were set on non-resonant Higgs boson pair production at 95% CL using the CLs method [83], summarised in Table 3.1. The observed limit was 30.9 fb, 12.7 times the SM prediction.

		Observed	-1σ	Expected	$+1\sigma$
$\tau_{lep}\tau_{had}$	$\sigma(HH \rightarrow b\bar{b}\tau\tau)$ [fb]	57	49.9	69	96
	σ/σ_{SM}	23.5	20.5	28.4	39.5
	$\sigma(HH \rightarrow b\bar{b}\tau\tau)$ [fb]	40.0	30.6	42.4	59
$\tau_{had}\tau_{had}$	σ/σ_{SM}	16.4	12.5	17.4	24.2
	$\sigma(HH \rightarrow b\bar{b}\tau\tau)$ [fb]	30.9	26.0	36.1	50
Combination	σ/σ_{SM}	12.7	10.7	14.8	20.6

Table 3.1: Observed and expected upper limits on the production cross section times the $HH \rightarrow b\bar{b}\tau\tau$ branching ratio for non-resonant HH at 95% CL, and their ratios to the SM prediction. The $\pm 1\sigma$ variations about the expected limit are also shown [77].

$b\bar{b}\gamma\gamma$ analysis The $b\bar{b}\gamma\gamma$ analysis [78] looks for final states with at least two photons and at least two jets, one or two of which are tagged as b-jets. The diphoton invariant mass is initially required to fall within a broad mass window of $105 \text{ GeV} < m_{\gamma\gamma} < 160 \text{ GeV}$. In order to remain orthogonal to the ATLAS search for $HH \rightarrow b\bar{b}b\bar{b}$, any event with more than two b-jets is rejected. The selected events are divided into two signal categories, characterized by exactly one or two b-tagged jets. In the 1-tag category, the second jet is chosen using a boosted decision tree that considers every pair of jets in the training events to find the correct jet pair looking at jet kinematic distributions. Further

requirements are then made on the p_T of the jets and on the mass of the dijet system. The highest- p_T and the next-highest- p_T jets are required to have $p_T > 100$ GeV and $p_T > 30$ GeV, respectively, with $90 \text{ GeV} < m_{jj} < 140$ GeV, keeping 5.8% and 3.9% of the expected signal events for the 2-tag and 1-tag category respectively. The signal is extracted using a fit to the diphoton invariant mass, $m_{\gamma\gamma}$, distribution of the selected events. The signal consists of a narrow peak around m_H superimposed on a smoothly falling background. The observed data are in good agreement with the data-driven background expectation across all categories, with the number of observed events in data compatible with the number of expected background events within the calculated uncertainties. The 95% CL upper limit for the non-resonant Higgs boson pair cross section is therefore obtained, and this upper limit, together with $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands is shown in Figure 3.6. The observed (expected) value is 0.73 ($0.93^{+0.37}_{-0.27}$) pb, corresponding to 22 (28^{+12}_{-8}) times the SM prediction.

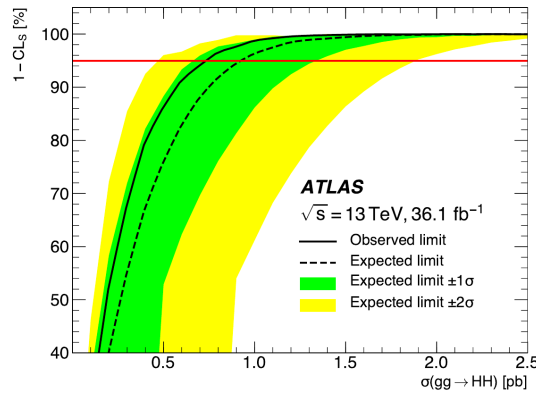


Figure 3.6: The expected and observed 95% CL limits on the non-resonant production cross section $\sigma(gg \rightarrow HH)$ for the $b\bar{b}\gamma\gamma$ analysis. The red line indicates the 95% confidence level. The intersection of this line with the observed, expected, and $\pm 1\sigma$ and $\pm 2\sigma$ bands is the location of the limits. The red line indicates the theoretical uncertainty of this prediction [78].

$\gamma\gamma WW^*$ analysis The $\gamma\gamma WW^*$ analysis [80] looks for final states with two isolated photons, at least two jets, no b-tagged jets, and at least one charged e/μ lepton. The jets and the leptons are required with p_T larger than 25 and 10 GeV respectively. The two photon candidates with the leading (sub-leading) E_T are required to satisfy $E_T/m_{\gamma\gamma} > 0.35$ (0.25), and the transverse momentum of the diphoton system is required to be $p_T^{\gamma\gamma} > 100$ GeV, keeping more than 70% of signal events. The signal region is defined by adding a requirement on the invariant diphoton mass of $105 \text{ GeV} < m_{\gamma\gamma} < 160$ GeV around m_H , in addition to the selection criteria above, selecting 8.5% of the expected signal events. The signal would be directly visible with one narrow peak in the $m_{\gamma\gamma}$ distribution corresponding to one Higgs boson, and with the combination of two jets and at least one charged lepton consistent with the $H \rightarrow WW^*$ decay for the other Higgs boson. A fit to $m_{\gamma\gamma}$ distribution is performed in the signal region to extract the signal yield. No excess was found and an observed (expected) limit was set on the $HH \rightarrow \gamma\gamma WW^*$ cross section of 7.5 (5.3) fb, corresponding to 230 (160) times the expected SM cross section, as shown in Table 3.2.

$WW^{(*)}WW^{(*)}$ analysis The $WW^{(*)}WW^{(*)}$ analysis [81] targets events with three possible final states, depending on the number of leptons: $WW^{(*)}WW^{(*)} \rightarrow \ell\nu + \ell\nu + 4q$, $WW^{(*)}WW^{(*)} \rightarrow \ell\nu + \ell\nu + \ell\nu + 2q$ and $WW^{(*)}WW^{(*)} \rightarrow \ell\nu + \ell\nu + \ell\nu + \ell\nu$, where ℓ

	$+2\sigma$	$+1\sigma$	Median	-1σ	-2σ	Observed
Upper limits on $\sigma(HH)$ (pb)	12	8.0	5.4	3.9	2.9	7.7
Upper limits on $\sigma(HH) \times B(\gamma\gamma WW^*)$ (fb)	12	7.8	5.3	3.8	2.8	7.5
Ratios of limits over the SM $\sigma(HH)$	360	240	160	120	87	230

Table 3.2: The 95% CL upper limits for the non-resonant production and the ratios of the limits to the SM cross section value for the $\gamma\gamma WW^*$ analysis. The $\pm 1\sigma$ and $\pm 2\sigma$ intervals around the median limit are also reported [80].

is either an electron or a muon. Events are selected requiring to pass single-lepton or dilepton triggers, and are classified in channels, defined according to the number of reconstructed leptons and according to the charge and flavour of the leptons. Events in the di-lepton channel are required to have exactly two leptons with the same electric charge, and are splitted in three categories: ee , $\mu\mu$ and $e\mu$. The three lepton channel events are required to have exactly three leptons with a summed electric charge $i \in q_i = \pm 1$, while in the four lepton channel they are analysed in four event categories, based on the multiplicity of same flavour and opposite charge lepton pairs and the mass of the 4-lepton system. In order to suppress top quark and $Z \rightarrow ee$ backgrounds, events are removed if they contain b-tagged jets or a same-flavour lepton pair with an invariant mass, m_{ll} , near the Z boson mass $|m - m_Z| < 10$ GeV. The event yields observed in the data is in agreement with the SM background only expectation, therefore the analysis has set upper limits on the HH signal cross section. The combined fit of the nine event categories, visible in Figure 3.7, yields an observed (expected) limit on the cross section for non-resonant HH production of 160 (120) times the SM prediction.

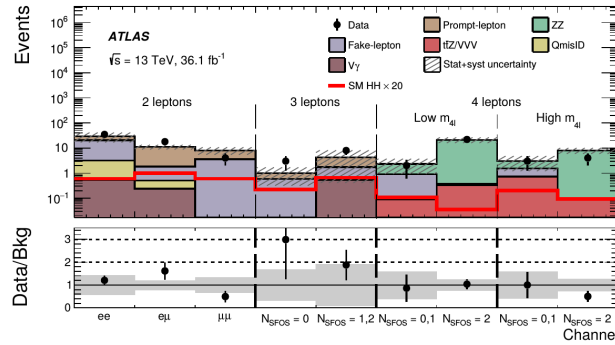


Figure 3.7: Expected and observed yields in each channel after all selection criteria for the non-resonant HH production search of the $WW^{(*)}WW^{(*)}$ analysis. The label NSFOS indicates the number of same-flavour, opposite-sign lepton pairs in the channel. Low and high $m_{4\ell}$ indicates $m_{4\ell} < 180$ GeV and $m_{4\ell} > 180$ GeV, respectively. The shaded band in the ratio plot shows the systematic uncertainty in the background estimate. The signal is scaled by a factor of 20 [81].

Combination of early Run 2 analyses The combination of searches for non-resonant Higgs-boson pair production [82] includes all the six decay channels described above. The analyses provide events in several signal regions, which are either orthogonal by construction or have negligible overlap. The statistical combination is performed with a simultaneous fit to the data in all the signal regions, for the extraction of the signal cross section, using the CLs approach. The detector systematic uncertainties and the uncertainties on the signal acceptance are correlated across all final states, while theoretical and modelling systematic uncertainties are not correlated due to the negligible overlap between the analyses. Since no statistically significant excess of events above

the Standard Model predictions is found, an observed (expected) upper limit is set on the gluon-gluon fusion $pp \rightarrow HH$ cross section of 6.9 (10) times the Standard Model prediction at 95% confidence level. The upper limits at 95% CL on the cross section of the ggF SM HH production normalised to its SM expectation for the six decay channels and their combination are shown in Figure 3.8. Here is also reported the 36.1 fb^{-1} $b\bar{b}WW^*$ result, that is worse than the other channels since it was not optimised for the non-resonant HH production search, and whose optimisation is the target of this work. In addition, upper limits on the $pp \rightarrow HH$ cross section are also computed as a function of the Higgs boson self-coupling modifier κ_λ , by combining the three most sensitive channels, $b\bar{b}b\bar{b}$, $b\bar{b}\tau\tau$ and $b\bar{b}\gamma\gamma$. The result is an observed (expected) exclusion limit on κ_λ values outside the range $-5.0 < \kappa_\lambda < 12.0$ ($-5.8 < \kappa_\lambda < 12.0$) at 95% confidence level.

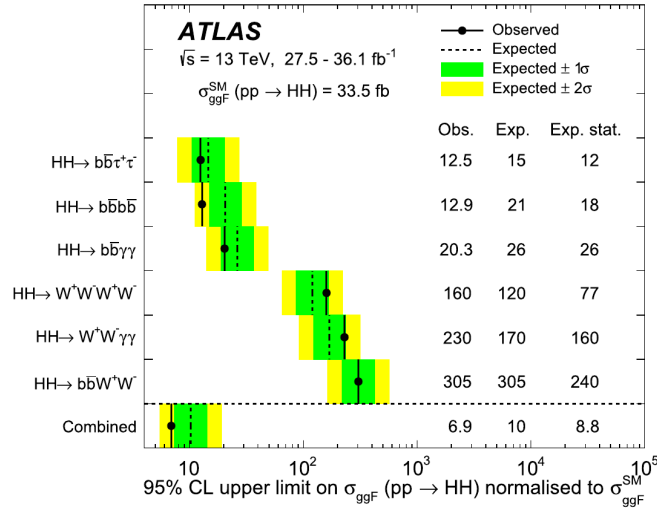


Figure 3.8: Upper limits at 95% CL on the cross section of the ggF SM HH production normalised to its SM expectation $\sigma_{\text{ggF}}^{\text{SM}}(pp \rightarrow HH)$ from the $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, $WW^{(*)}WW^{(*)}$, $\gamma\gamma WW^*$ and $b\bar{b}WW^*$ searches, and their statistical combination. The column “Obs.” lists the observed limits, “Exp.” the expected limits with all statistical and systematic uncertainties, and “Exp. stat.” the expected limits obtained including only statistical uncertainties in the fit [82].

After these early Run 2 analyses, efforts have been renewed to exploit the full Run 2 data collected by the ATLAS detector, corresponding to 139 fb^{-1} , about four times the previous available luminosity. Both ggF and VBF production modes are currently under study in several decay channels. The VBF production mode, in particular, involves three different types of couplings: the Higgs boson self-coupling (λ_{HHH}), the Higgs-boson-vector-boson coupling (λ_{VVH}) and the di-vector-boson-di-Higgs-boson (λ_{VVHH}) coupling. This last coupling is available only via VBF production processes, and thus its modifier κ_{2V} can be constrained by VBF Higgs pair production searches, while ggF analyses can provide better constraints on λ_{HHH} and λ_{VVH} . Currently, the only public result is provided by the VBF $HH \rightarrow b\bar{b}b\bar{b}$ analysis [84].

VBF $HH \rightarrow b\bar{b}b\bar{b}$ analysis This analysis [84] uses 127 fb^{-1} for the search of Higgs boson pairs produced through diagrams such as those presented in Figure 1.8b, and using the dominant $H \rightarrow b\bar{b}$ decay mode. Differently from the ggF, the VBF process is characterised by the presence of two additional jets with a large rapidity gap resulting from quarks from which a vector boson is radiated. In [84], events are selected with exactly 4 b-tagged jets, with the addition of two forward ($|\eta| > 2.0$) light-jets, one for each side

of the detector, for the VBF topology. Additional kinematic cuts are applied in order to suppress multijet and $t\bar{t}$ backgrounds and select the HH signal. The mass of the four selected b-jets, m_{4b} , after all the event selection cuts, is used as the final discriminant for the cross section extraction, and it is visible in Figure 3.9. As no significant excess over the background prediction is observed, exclusion limits are computed using the CLs method, providing an observed (expected) limit of 1460 (950) fb on the HH production cross section via VBF, corresponding to 840 (550) times the SM value. Limits are also calculated as a function of κ_{2V} , excluding values of $\kappa_{2V} < -0.76$ or $\kappa_{2V} > 2.90$, while the expected exclusion region is $\kappa_{2V} < -0.91$ or $\kappa_{2V} > 3.11$, putting a first limit on this important Higgs boson coupling.

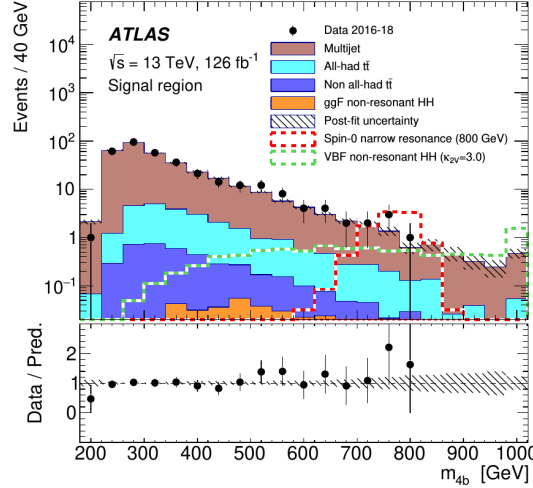


Figure 3.9: Distribution of m_{4b} invariant mass of the selected HH candidates of the VBF $HH \rightarrow b\bar{b}b\bar{b}$ analysis. The non-resonant signal at $\kappa_{2V} = 3$ is also shown overlaid on the expected background [84].

2.2 CMS results

Also the CMS collaboration has performed early Run 2 di-Higgs searches, using the data collected at $\sqrt{s} = 13$ TeV in 2015 and 2016, corresponding to an integrated luminosity up to 35.9 fb^{-1} . Four decay channels have been studied focusing on ggF production mode: $b\bar{b}b\bar{b}$ [85], $b\bar{b}\tau\tau$ [86] and $b\bar{b}\gamma\gamma$ [87], $b\bar{b}\ell\nu\ell\nu$ [88]. As done by ATLAS, a combination of these four channels has been performed [89], in order to obtain a better limit on the ggF HH production cross section. The analyses are here briefly described, except for $b\bar{b}\ell\nu\ell\nu$ analysis that is described in Chapter 4.

$b\bar{b}b\bar{b}$ analysis The $b\bar{b}b\bar{b}$ analysis [85] is focused on the search for HH production with both Higgs bosons decaying into bottom quark pairs, resulting in four resolved hadronic jets. Events are selected if four or more jets are b-tagged, contributing to a significant reduction of the QCD multijet background. Two Higgs boson candidates are then reconstructed by pairing the b-jets and a boosted decision tree (BDT) classifier is trained for a better discrimination of the signal over the background. The BDT exploits a set of kinematic observable and properties related to the jets, the Higgs candidates and the HH system. The output distribution of the BDT is then compared with the expected background searching for a signal contribution. As no excess over the predicted SM background is found, an upper limit on the $HH \rightarrow b\bar{b}b\bar{b}$ cross section is extracted using

the CLs method, with an observed (expected) value of 847 (419) fb. The obtained upper limits are reported in Table 3.3.

Category	Observed	Expected	-2 s.d.	-1 s.d.	+1 s.d.	+2 s.d.
SM $HH \rightarrow b\bar{b}b\bar{b}$	847	419	221	297	601	834

Table 3.3: The observed and expected upper limits on $\sigma(pp \rightarrow HH \rightarrow b\bar{b}b\bar{b})$ in the SM at 95% CL in units of fb [85].

$b\bar{b}\tau\tau$ analysis The $b\bar{b}\tau\tau$ analysis [86] considers three final states of the $\tau\tau$ lepton pair, covering about 88% of the possible $\tau\tau$ -system decays: one of the two τ leptons is required to decay into hadrons and a neutrino (τ_h), while the other can decay either to the same final state, or into an electron (τ_e) or a muon (τ_μ) and neutrinos. The two selected τ leptons are required to have opposite electric charge, and, if additional isolated leptons are present, the event is discarded to reduce the $Z/\gamma^* \rightarrow \ell\ell$ background contribution. Two additional $R=0.4$ anti- κ_t jets are also required, splitting the events in two categories depending on the number of b-tagged jets (one or two b-tags). The selected events are classified using a BDT that exploits a set of kinematic observables to discriminate signal from $t\bar{t}$ background. Then, the transverse mass m_{T2} , defined as the largest mass of the parent particle (in this case the top quark) that is compatible with the kinematic constraints of the event, is used in the final signal extraction. The m_{T2} variable has an upper-bound at m_t for the irreducible background process $t\bar{t} \rightarrow b\bar{b}WW \rightarrow b\bar{b}\tau\nu_\tau\tau\nu_\tau$, while it can assume larger values for the HH signal where the τ and the b-jet do not originate from the same parent particle. A binned maximum likelihood fit is performed simultaneously in the signal regions defined in this search for the three final states considered. As no significant excess of signal over the background prediction is observed, exclusion 95% CL upper limits are set on the cross section for Higgs boson pair production using the CLs method. The observed (expected) 95% CL upper limit on the $HH \rightarrow b\bar{b}\tau\tau$ cross section amounts to 75.4 (61.0) fb, corresponding to about 30 (25) times the SM prediction.

$b\bar{b}\gamma\gamma$ analysis The $b\bar{b}\gamma\gamma$ analysis [87] selects events with at least two identified photon candidates and two jets. Photons are required to pass the following criteria: $100 < m_{\gamma\gamma} < 180$ GeV, $p_T^{\gamma_1}/m_{\gamma\gamma} > 1/3$ and $p_T^{\gamma_2}/m_{\gamma\gamma} > 1/4$, where the photon pair with the highest p_T is chosen as the $H \rightarrow \gamma\gamma$ candidate. The $H \rightarrow b\bar{b}$ candidate is built using the two jets with the highest b-tagging score and satisfying $70 < m_{jj} < 190$ GeV. The two system candidates are then combined to form an HH candidate using the definition $\tilde{M}_X = m_{\gamma\gamma jj} - (m_{jj} - m_H) - (m_{\gamma\gamma} - m_H)$, that is corrected to mitigate the dijet and diphoton energy resolutions assuming the event as a signal event. A high- and a low-mass regions are defined grouping the events with $\tilde{M}_X$ larger or smaller than 350 GeV, respectively. Then a BDT is trained for each region, using jet b-tagging variables, three helicity angles ($HH, jj, \gamma\gamma$) and HH transverse balance variables. The BDT output classifies events as more signal- or continuum-background-like, separating events into high, medium and low purity categories, where the low purity category events are then discarded. Different from the ATLAS analysis, the signal extraction is here performed simultaneously in the $m_{\gamma\gamma}$ and m_{jj} distributions with a 2D profile Likelihood fit. No evidence for HH production is observed in the data. Upper limits on the production cross section of $HH \rightarrow b\bar{b}\gamma\gamma$ are computed using the CLs method. The observed (expected) 95% CL upper limit is 2.0 (1.6) fb, and 0.79 (0.63) pb for the total ggF HH production cross section. This last result corresponds to 24 (19) times the SM prediction.

Combination of early Run 2 analyses As done by ATLAS, a statistical combination of the results obtained in each channel has been performed to increase the sensitivity of the Higgs pair production searches and derive better upper limit on its cross section [89]. The four $b\bar{b}b\bar{b}$, $b\bar{b}\tau\tau$, $b\bar{b}\gamma\gamma$ and $b\bar{b}\ell\nu\ell\nu$ channels have been considered in the combination. This combination has allowed the CMS experiment to set an observed (expected) upper limit on the HH production at 22 (13) times the SM prediction, using data collected in 2015 and 2016 at 13 TeV. The results for each channel and for the combination is summarised in Figure 3.10.

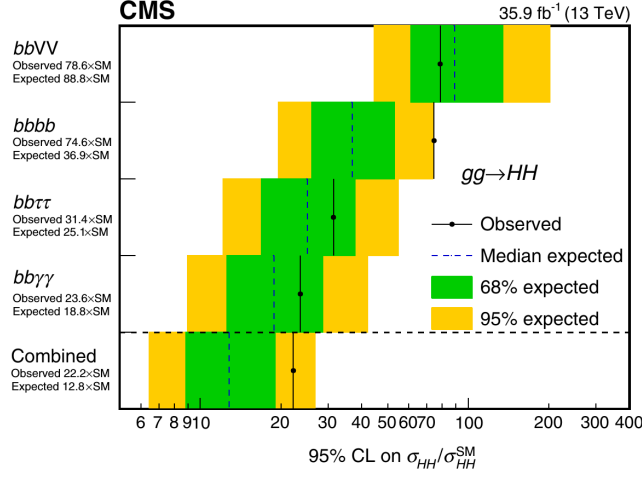


Figure 3.10: The 95% CL upper limits on the signal strength $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$ for the $b\bar{b}\ell\nu\ell\nu$, $b\bar{b}b\bar{b}$, $b\bar{b}\tau\tau$ and $b\bar{b}\gamma\gamma$ analyses, and their statistical combination. The inner (green) band and the outer (yellow) band indicate the regions containing 68 and 95%, respectively, of the limits on μ expected under the background-only hypothesis [89].

Finally, upper limits on the $pp \rightarrow HH$ cross section are also computed as a function of the Higgs boson self-coupling modifier κ_λ , and κ_λ is observed (expected) to be constrained to the range $-11.8 < \kappa_\lambda < 18.8$ ($-7.1 < \kappa_\lambda < 13.6$) at 95% confidence level.

Chapter 4

Analysis of the Higgs boson pair production in the $b\bar{b}WW^*$ decay channel

This Chapter describes the analysis of the SM process of Higgs boson pair production in the $b\bar{b}WW^*$ decay channel. This analysis has already been performed, with partial Run 2 data corresponding to 2015 and 2016 data taking periods, by ATLAS and CMS, looking at the single lepton and di-lepton final states respectively. These first analyses are briefly described in Section 1 together with the full Run 2 result of the $bbll$ analysis, recently published by the ATLAS collaboration, to give an overview of the searches in this particular decay channel.

The object of this work is the search for Higgs boson pair production in the single-lepton $b\bar{b}WW^*$ channel. The choice of the single-lepton final state is discussed in Section 2, describing the objects in this particular final state and the main background sources. A brief overview of the analysis strategy is also reported in Section 2. Data and simulated signal and background samples used in this work are described in Section 3, while the object and event selection is reported in detail in Section 4. This work makes use of Machine Learning techniques to exploit their discriminating power to maximise the sensitivity of the analysis. A general introduction to these techniques and the description of the approach used in this work is presented in Section 5. The estimation of the background contribution in the analysis and the systematic uncertainties considered in the analysis are described in Section 6 and 7 respectively. Finally, the statistical model used to obtain the final results of this search is described in Section 8. The summary of the results and further possible improvements of the analysis are discussed in Section 9.

1 Previous measurements

The $b\bar{b}VV$ decay channel under study in this thesis work has been already analysed both by ATLAS and CMS with data collected at a center-of-mass energy of pp collisions of $\sqrt{s} = 13$ TeV. CMS has focused only on the di-lepton final state, accounting for both $H \rightarrow WW^*$ and $H \rightarrow ZZ$ decays, exploiting partial Run 2 data corresponding to a luminosity of 35.9 fb^{-1} [88]. On the other hand, ATLAS has first performed a search in the single-lepton $b\bar{b}WW^*$ channel with a luminosity of 36.1 fb^{-1} [79], and then it has exploited the data collected corresponding to the full Run 2 luminosity of 139 fb^{-1} focusing on the di-lepton final state [90], as done previously by CMS.

These analyses are briefly described in the following in order to expose the state

of the art of the searches for di-Higgs production processes decaying into the channel under study in this work. Analysis strategies of these previous works have been studied with the aim of understanding the characteristics and the choices made to optimise the sensitivity of the analyses on such a rare process as the Higgs pair production process is. Interesting selections and procedures have been adopted and optimised as it will be described in the following of the Chapter, trying to improve as much as possible the sensitivity of the analysis subject of this work.

CMS $bbll$ analysis The analysis [88] considers the final states with two leptons coming from the decays of W or Z bosons, produced in the decay of one Higgs boson, having the second Higgs boson decaying in b-quark pairs. The object selection consists of two opposite sign leptons and two b-tagged jets. Data are collected with a set of dilepton triggers, which have asymmetric p_T threshold for leading and subleading leptons. Then, events with two oppositely charged leptons (e^+e^- , $\mu^+\mu^-$, $e^\pm\mu^\mp$) are selected, requiring lepton p_T to be above the corresponding trigger threshold: 25 GeV and 15 GeV are required for leading and subleading leptons of ee and μe events, 20 GeV and 10 GeV for $\mu\mu$ events, and 25 GeV and 10 GeV for $e\mu$ events. Electrons in the pseudorapidity range $|\eta| < 2.5$ and muons in the range $|\eta| < 2.4$ are considered, asking for isolation of each of the two leptons. Jets are selected with $p_T > 20$ GeV and are tagged as b-jets if they pass the 70% efficiency working point of a combined multivariate b-tagging algorithm. Two jets with the highest combined multivariate algorithm outputs are selected among all possible dijet combinations. A requirement of $12 < m_{\ell\ell} < m_Z - 15$ GeV is applied to remove the large background at the Z boson peak as well as the high- $m_{\ell\ell}$ tail of the DY and $t\bar{t}$ processes. Signal-to-background separation is improved with the use of a deep neural network (DNN) discriminator trained on signal and on the irreducible $t\bar{t}$ background. Kinematic variables are provided as input to the DNN: $m_{\ell\ell}$, $\Delta R_{\ell\ell}$, ΔR_{jj} , $\Delta\phi(\ell\ell, jj)$, $p_T^{\ell\ell}$, p_T^{jj} , $\min\Delta R_{\ell j}$, and the transverse mass of the system made by the two leptons and E_T^{miss} . DNN discriminator is evaluated in three m_{jj} regions, building two background enriched control regions for $m_{jj} < 75$ GeV and $m_{jj} \geq 140$ GeV, and one signal region for $m_{jj} \in [75, 140)$ GeV. In each region and in each decay channel (e^+e^- , $\mu^+\mu^-$, $e^\pm\mu^\mp$), DNN output is used as final discriminant in a binned maximum likelihood fit. No significant excess above background predictions was seen, and an upper limit was set at 95% CL on the product of the Higgs boson pair production cross section and branching fraction for $HH \rightarrow bbVV \rightarrow bbl\nu\ell\nu$ using the asymptotic CLs, combining the e^+e^- , $\mu^+\mu^-$ and $e^\pm\mu^\mp$ channels. The observed upper limit on the SM $HH \rightarrow bbVV \rightarrow bbl\nu\ell\nu$ cross section is found to be 72 fb, in agreement with an expected upper limit of 81^{+42}_{-25} fb, corresponding to ~ 80 times the SM prediction.

ATLAS single-lepton $bbWW^*$ analysis This analysis focused on the search of Higgs pair production processes looking at the single-lepton $bbWW^*$ decay channel [79]. The analysis performed in this thesis work is the new updated and optimised iteration of the 36.1 fb^{-1} analysis described in ref [79], which have been taken for reference in the development of the work. The analysis strategy used followed a cut-based approach, selecting events passing a set of cuts on kinematic distributions built from the topology of signal signature. Events are first selected with single-lepton triggers, and are required to have exactly one isolated trigger matched lepton (e/μ) with $p_T > 27$ GeV. Jets with $p_T > 20$ GeV are selected, and they are tagged as b-jets if passing a b-tagging algorithm with 85% efficiency. Events are required to have exactly two b-tagged jets, which form the Higgs boson candidate, and two or more light-jets, considering only the three leading ones and building the W boson candidate from the jet pair with the

lowest ΔR_{jj} . The event kinematics of the $H \rightarrow WW^* \rightarrow lvqq$ topology can be fully reconstructed by computing the longitudinal momentum of the neutrino (p_z^v), which is the only unknown component among all the four-momenta of the final state objects. This component is obtained from a quadratic equation exploiting the $m_H = 125$ GeV constraint on WW^* system, allowing to reconstruct and exploit kinematic variables of WW^* and HH systems. Kinematic selections, based on differences in the kinematics of $t\bar{t}$ and signal events, are used to suppress $t\bar{t}$ background. The Higgs mass constraint is used to define cuts on invariant masses of WW^* and bb systems: $m_{WW^*} < 130$ GeV and $m_{bb} \in [105, 135]$ GeV. Additional cuts are made on the transverse momentum of WW^* and bb systems, $p_T^{WW^*} > 250$ GeV and $p_T^{bb} > 300$ GeV, and on the MET of the event, $E_T^{miss} > 25$ GeV, to further reduce the large $t\bar{t}$ background. In this analysis the presence of a signal is indicated by an excess of events over the SM prediction for the background yield in the signal regions. A simultaneous maximum-likelihood fit is performed for the number of events in the signal region and three control regions, defined to take the normalisation of $t\bar{t}$ and multijet background contribution. No significant excess over the expectation is observed and the results are used to evaluate an upper limit at the 95% confidence level on the production cross section times the branching fraction for the signal hypotheses under consideration. Exclusion limits are then calculated with the CLs method, obtaining an observed (expected) upper limit of $\sigma(HH \rightarrow b\bar{b}WW^*) < 2.5$ ($2.5_{-0.7}^{+1.0}$) pb at 95% CL, corresponding to ~ 300 times the SM prediction.

ATLAS $bb\ell\ell$ analysis The analysis is focused on the $bb\ell v\ell v$ final state, where ℓ refers to a either an electron or a muon, using full Run 2 data collected with the ATLAS detector corresponding to a total integrated luminosity of 139 fb^{-1} [90]. Machine-learning techniques based on neural network architectures are used to construct an event-level classifier trained to distinguish between the HH signal and SM backgrounds, $t\bar{t} + Wt$, $Z \rightarrow \ell\ell$ and $Z \rightarrow \tau\tau$. Candidate events are selected with exactly two oppositely charged leptons, isolated electrons or muons with $p_T > 10$ GeV, and at least two signal b-tagged jets with $p_T > 20$ GeV. A multi-class deep neural network is trained using 35 kinematic variables, which are provided as inputs to the classifier, ranging from momentum components of the visible final-state objects to observables using event-wide information, mainly angular variables that are useful to discriminate HH signal from SM background as the final state objects from each of the Higgs bosons in the signal tend to be near to each other. The classifier produces four outputs with their sum constrained to one, and each constrained to values between 0 and 1. These outputs indicate the probability of how the event likely belongs to a particular class. The main discriminant in the analysis, d_{HH} , is then constructed by combining the four output probabilities. A cut of $d_{HH} > 5.5$ is used together with requirements of $m_{\ell\ell} \in (20, 60)$ GeV and $m_{bb} \in (110, 140)$ GeV in order to further suppress background contributions. Two signal regions are defined by splitting the selected events into same-flavour (ee or $\mu\mu$), or different-flavour ($e\mu$) channels. A profile-likelihood fit is performed on d_{HH} distribution on the two signal regions and on two control regions defined to normalise top and Z backgrounds. No significant excess of events over the expected SM background was observed and upper limits was set on the SM Higgs boson pair production using the CLs method. The observed (expected) limit at 95% CL is 1.2 (0.9) pb, corresponding to 40 (29) times the SM prediction.

The limit obtained in ATLAS $36.1 \text{ fb}^{-1} bbWW^*$ analysis was not competitive with respect to the others analyses focusing on different channels previously described in Chapter 3. This has been the one of the key motivations for the development of a new

analysis, starting from truth-level studies with the aim of providing an optimisation of the selection and of the sensitivity of the search, which has been the target of the work described in this thesis. The use of DNN classifiers was found to be very effective in the $b\bar{b}l\bar{l}$ analyses described above. For this reason, the strategy of using machine learning techniques in addition to a cut based selection of the events was adopted in the new $b\bar{b}WW^*$ analysis. This strategy allowed to improve the sensitivity of the analysis with respect to the previous result [79] as it will be described in the rest of the Chapter.

2 Single lepton final state overview

The $b\bar{b}WW^*$ decay channel can appear in three different final states, all hadronic, all leptonic and semi-leptonic, depending on the decay process of the two W bosons. Although it has the largest branching fraction, the all hadronic final state is experimentally difficult to observe due to the huge QCD induced multijet background. This background can be mitigated by requiring the presence of at least a lepton, electron or muon, in the final state. The semi-leptonic final state, on which this work is focused, is, therefore, a compromise between signal efficiency and background reduction, that allows to maintain a good branching fraction of the $HH \rightarrow b\bar{b}WW^*$ events of about 30%. Tau leptons, due to their possible decay in hadrons, are usually not considered as leptons in the analyses, which prefer to rely only on electrons and muons. In this final state the dominant background is the $t\bar{t}$ background which is a so called irreducible background as its final state is composed by the same type and number of objects of the signal final state. In addition to the main $t\bar{t}$, smaller background contributions come from W/Z +jets, multijet and single top-quark production processes.

A search for Higgs boson pair production with one Higgs boson decaying via $H \rightarrow b\bar{b}$, and the other via $H \rightarrow WW^* \rightarrow lvqq$, where l is either an electron or a muon, is reported in this work. The semi-leptonic final state of interest of this analysis is composed by two b-jets, two light-jets, one lepton and missing energy coming from the emission of a neutrino in the W boson decay. For the kinematics of this non-resonant Standard Model process, the objects in the final states present almost no overlap. This topology is often referred to as resolved topology. Figure 4.1 shows a Feynman diagram of the non-resonant HH production via gluon-gluon fusion with the subsequent decays of the Higgs bosons to a $b\bar{b}$ and a WW^* pairs searched for in this analysis.

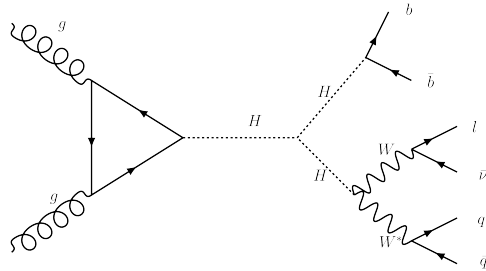


Figure 4.1: Feynman diagram of non-resonant HH production via gluon-gluon fusion, decaying in the $b\bar{b}WW^*$ single lepton final state.

The challenge of this search comes from the huge contribution of the $t\bar{t}$ irreducible background that requires a powerful background reduction. A possible approach is to exploit the capabilities of machine learning (ML) techniques in obtaining the best discrimination between signal and backgrounds. The application of ML algorithms is used in this work after a preliminary event selection, providing powerful discriminating

variables that categorise the events as signal- and background-like. These distributions are used to categorise the events in signal-enriched and background-enriched regions, and to evaluate the sensitivity of the analysis, as it is described in the following of this Chapter.

3 Data and simulated samples

In this section the data and the Monte Carlo (MC) simulated samples used in this analysis are described. Monte Carlo samples are commonly used by physics analyses to simulate the signal and background processes of interest. The analysis is constructed and the cuts are optimised having these simulated samples as a reference. The data are used only in a second stage of the analysis, for the estimation of data driven background and the normalisation and validation of the background simulated contributions, and of course for the final evaluation of the results of the analysis.

3.1 Data samples

The analysis presented in this work uses the pp collisions data collected during the full Run 2 period at the center-of-mass energy of 13 TeV and recorded with the ATLAS detector at the Large Hadron Collider. The total integrated luminosity is $139.0 \pm 2.4 \text{ fb}^{-1}$, with a relative uncertainty of 1.7% [38]. This value corresponds to the integrated luminosity collected in the runs with stable beams between 2015 and 2018 where all the sub-detectors conditions were optimal for the data acquisition. This collection of runs, known as Good Run List, provide the fraction of the collected data that have the required quality for their use in the physics analyses.

3.2 Monte Carlo samples

Monte Carlo generators provide complete simulation of high-energy collisions, like the one occurring at LHC. These generators have several components, dedicated to the description of different physics processes that occur in particle interactions. The simulation starts with the Matrix Element (ME) calculation of the hard scatter. Parton Distribution Functions (PDFs) are used to describe the interacting partons, while the probability distribution of produced partons are given by lowest order perturbation theory. The initial and final state radiation are simulated computing the probabilities of photon and gluon emission from the incoming and outgoing partons. At this stage, the emitted gluons can generate new radiation via their splitting, leading to an extended shower that is described by the parton shower (PS). The splitting probability is evaluated step-by-step, determining the lowering of the momentum transfer scale, up to a point where perturbation theory breaks down. The most commonly used generators are Pythia [91–93], Sherpa [94] and Herwig [95, 96] that provide leading order ME hard processes, and differentiate for the use different ordering of the produced particles (q^2 , p_T or angular), providing better description of particular physics processes. These generators contain also the simulation of the hadronisation process, that is based on non-perturbative QCD calculations, providing the production of the final state hadrons and the decays of the unstable particles. There are also generators that provide Next-to-leading order (NLO) ME, such as Powheg [97] and MadGraph5_aMcAtNLO [98], that are usually used to describe the hard scatter and matched with the parton shower generators.

All MC simulated samples for background processes are passed through the full ATLAS detector simulation [99] based on GEANT4 [100] while the signal samples

are passed through ATLAS fast simulation framework, Atfast II (AF2) [101]. Simulated events are processed with the same reconstruction algorithms used for data and described in Chapter 2. The pileup effect is modelled in the simulated samples with the overlay of additional pp collisions generated with Pythia8.186 [92]. The number of overlaid collisions is such that the distribution of the number of interactions per pp bunch crossing in the simulation roughly matches the one observed in data.

As the pileup is depending on the data taking periods, for each simulated process, three Monte-Carlo campaigns are produced (mc16a, mc16d, and mc16e) with different assumptions on the distribution of pileup. mc16a events are simulated to roughly match the 2015-2016 data taking conditions, mc16d to match the 2017, and mc16e to match the 2018. However, the actual data taking conditions of the ATLAS detector differ between those assumed in the MC samples. For this reason, a pileup reweighting procedure is performed to account for these differences between data and simulations using an event weight to recover the discrepancies on an event by event basis.

3.2.1 Signal samples

The signal samples used in this analysis simulate the ggF non-resonant Higgs boson pair production processes decaying in the $b\bar{b}WW^*$ channel. The final state simulated is the one with one W boson decaying hadronically and the other W boson decaying in a lepton and a neutrino. Events are first generated at Next-to-Leading-Order with MadGraph5_aMCatNLO [98] using the FTApprox method. The CT10 parton distribution function [102] is used in the event generation, which is interfaced to Herwig7 [96] using the H7-UE-MMHT tune for underlying events and the H7-MMHT2014LO tune for parton shower and hadronisation. Events are generated with an effective Lagrangian in the infinite top-quark mass approximation, and reweighting the generated events with form factors that take into account the finite mass of the top quark. In addition, scale factors depending on the HH invariant mass m_{HH} at the generator level were applied to match the MC m_{HH} distribution with NLO and NNLO calculations that compute exact finite m_t contributions [75, 103–107]. The final state is semi-leptonic, where one of the W bosons is required to decay into a charged lepton and a neutrino, while the other into two quarks. The cross section of the process is calculated using the reference values visible in Table 1.4, where the production cross sections are calculated at NNLO at $\sqrt{s} = 13$ TeV for $m_H = 125$ GeV. The ggF sample cross section is then calculated applying the H and W boson branching ratios for the simulated decay channels and are reported in Table 4.1 together with the details of the signal sample generators. These cross section assume all the possible leptonic decay of the W boson as in the sample both electron, muon and tau W -decays are simulated. However, final state containing a τ is not considered as signal in this analysis. Therefore, in order to evaluate the expected signal yields, the cross section in Table 4.1 must be scaled by a factor 2/3, as the leptonic W decay have the same BR for all the lepton flavors. With these scaled cross sections, the expected number of signal events is ~ 316 for ggF production mode.

Process	Generator	Parton shower	PDFs	$\sigma \times \text{BR}$	Yields
ggF $HH \rightarrow b\bar{b}WW^* \rightarrow b\bar{b}l\nu q\bar{q}$	MADGRAPH5_AMCATNLO	HERWIG7	CT10	3.41 fb	474

Table 4.1: Details of the di-Higgs signal sample used in the analysis. Here l stands for electron, muon or tau, as the sample is generated with the inclusive leptonic decay of the W boson. Monte Carlo generators for hard scatter process and parton shower, as well as PDF set used are reported. Cross section times branching ratio and the expected event yield with an integrated luminosity of 139 fb^{-1} is also shown.

3.2.2 Background samples

With the exception of the QCD multijet background, that is described in Section 6, MC simulated events are used to estimate SM backgrounds and the signal acceptances. Table 4.2 summarises the MC samples used for background estimation with details of the generators, PDFs and underlying event tune used. These processes are considered for their possibility of mimic the signal, either with the same final state, as the $t\bar{t}$ does, or with a slightly different final state that, with object mis-identification or non-detection, can pass the event selection requirement.

Process	Generator	Parton Shower	PDF	Tune	$\sigma \times \text{BR}$ [fb]	Yields
$t\bar{t}$	POWHEG-BOX	PYTHIA8	NNL	A14	8.32×10^5	1.16×10^8
Single-top	POWHEG-BOX	PYTHIA8	NNL	A14	1.45×10^5	4.98×10^6
$t\bar{t}W$	aMCATNLO	PYTHIA8	NL	A14	600.9	8.35×10^4
$t\bar{t}Z$ ($Z \rightarrow \nu\nu$)	aMCATNLO	PYTHIA8	NL	A14	172.1	2.39×10^4
W +jets ($W \rightarrow l\nu$)	SHERPA 2.2.1		NNL		2.01×10^7	2.79×10^9
Z +jets ($Z \rightarrow ll$)	SHERPA 2.2.1		NNL		1.89×10^6	2.63×10^8
Diboson (VV)	SHERPA 2.2.1 / 2.2.2		NNL		4.73×10^4	6.57×10^6
ggF H	POWHEG	PYTHIA8	CT10	AZNLO	4.86×10^4	6.75×10^6
VBF H	POWHEG	PYTHIA8	NNL	AZNLO	3.78×10^3	5.25×10^5
WH ($H \rightarrow b\bar{b}$)	POWHEG MINLO	PYTHIA8	NNL	AZNLO	89.9	1.25×10^4
ZH	POWHEG MINLO	PYTHIA8	NNL	AZNLO	759.7	1.06×10^5
$t\bar{t}H$	POWHEG-BOX	PYTHIA8	NL	A14	525.1	7.30×10^4

Table 4.2: Monte Carlo nominal samples used for background estimation. Monte Carlo generators for hard scatter process and parton shower, as well as PDF set and tune used are reported for each sample, where NNL and NL PDFs stand for NNPDF3.0 NNLO and NNPDF3.0 NLO, respectively. Cross section times branching ratio and the expected event yield with an integrated luminosity of 139 fb^{-1} are shown for each background.

Top processes The $t\bar{t}$ and single top-quark samples are generated using Powheg with the Powheg-Box [97, 108] framework and the NNPDF3.0 NLO PDF set [109]. Pythia8 is used for showering and hadronisation. The A14 set of tuned underlying event parameters [110] was used. For the settings of the properties of the bottomed and charmed hadron decays, EvtGen [111] is used. The mass of the top quark is set to $m_t = 172.5 \text{ GeV}$, while the H_{damp} parameter, used to regulate the high- p_T radiation in Powheg, is set to 1.5 times the mass of the top quark for good data/MC agreement [112]. The $t\bar{t}$ cross section for pp collisions at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ is $\sigma_{t\bar{t}} = 831.76_{-29.20}^{+19.77}(\text{scale}) \pm 35.06(\text{PDF} + \alpha_s) \text{ pb}$ for a top quark mass of 172.5 GeV [113].

Top-antitop pairs can also be produced in association with a vector boson, where the W boson is emitted in initial state radiation, while the Z boson can be emitted both in initial and in final state radiation. These processes are usually called $t\bar{t}W$ and $t\bar{t}Z$, or in general $t\bar{t}V$, and can contribute as a smaller top background in the final state of interest of this analysis. They are simulated with aMCATNLO using the NNPDF3.0 PDF set at next-to-leading order, interfaced with Pythia8 parton shower and using the A14 tune for the underlying event.

Each process of single top-quark (t -channel, s -channel and Wt -channel) is generated separately. The cross section of single-top t -channel and s -channel processes are calculated at NLO in QCD using the Hathor v2.1 program [114, 115]. For the Wt -channel the NNLO approximate predictions are calculated with the prescriptions in Ref. [116, 117].

W/Z+jets processes The processes of $W \rightarrow \ell\nu$ and $Z \rightarrow \ell\ell$ associated with jets (V+jets) are simulated with Sherpa2.2.1 using the NNPDF3.0 NNLO PDF set and showered by the Sherpa built-in implementation which has matrix elements for up to 2 additional jets at NLO and up to 4 additional jets at LO. The cross section to normalise the simulations are calculated at NNLO accuracy in QCD and include EW corrections at NLO accuracy [118].

Diboson processes The diboson (VV , or in detail WW , WZ , ZZ) leptonic processes and the loop-induced diboson inclusive processes are generated with Sherpa2.2.2 event generator with the NNPDF3.0 NNLO PDF set used in conjunction with the built-in Sherpa's parton shower tuning. The diboson semi-leptonic processes are instead generated with Sherpa2.2.1. The cross section for the normalisation of the simulated samples are calculated at NLO [119].

Single Higgs processes The single Higgs processes can contribute as a background source due to the possible similar final state to the signal one, and the small cross section of Higgs pair production process. All the Higgs production modes are considered. The ggF H sample, with the inclusive decay of the Higgs boson, is generated with Powheg using the CT10 PDF set [120], interfaced to Pythia8 for parton shower. The ttH process is simulated in three splitted samples depending on the top decay, all-hadronic, semi-leptonic and di-leptonic, with the inclusive Higgs decay. The samples are generated using Powheg with the Powheg-Box framework interfaced to Pythia8, using the A14 tune. The VBF sample is generated with Powheg plus Pythia8, as the WH and ZH samples. The VH samples are generated with up to two additional jets, using the MiNLO method [121] with NLO calculations. VBF and ZH samples consider inclusive H and Z boson decays, while the WH sample consider only $H \rightarrow b\bar{b}$ and $W \rightarrow l\nu$ decays, as the resulting final state is very similar to the signal one. ggF , VBF and VH samples use the AZNLO tune [122], which is a Pythia8 tune used to shower Powheg processes involving electroweak boson production (W, Z, H, VV) that provides a very good description of ISR in the low and medium p_T region.

4 Object and event selection

In this Section the selection applied to the reconstructed object and the criteria used in the event selection of this analysis work is described. The selection requirements are chosen in such a way as to keep the higher possible efficiency on signal events, given their small expected number compared to the overwhelming $t\bar{t}$ background.

4.1 Object selection

As the final state of interest of the analysis is composed by two b-jets, two light-jets, one lepton and a neutrino, the physics object considered in the selection are jets, leptons and E_T^{miss} . B-tagging algorithms are applied to reconstructed hadronic jets in order to recognize the $H \rightarrow b\bar{b}$ decay. Identification and isolation requirements are applied to select electrons and muons, while a veto is defined to reject the events with reconstructed hadronic tau leptons. No requirements are, instead, made on the E_T^{miss} of the events.

As already mentioned, QCD multijet background is estimated using a data driven method. As it will be described in Section 6, this method requires an object selection different from the one used for the definition of signal-like events. Therefore, for

electrons and muons in particular, two set of selection requirements are defined in the following, that will be referred as to "Loose" and "Signal" selection. All these requirements are summarised in the following of this Section for each of the physics object considered.

Electrons Two sets of electron selections are defined. They are denoted as LooseElectron and SignalElectron. The selections are defined as follows:

- **LooseElectron:** The electron p_T is required to be greater than 5 GeV. The electron cluster should be in the range of $|\eta| < 2.47$. Medium likelihood identification is applied in this criteria, the LHMedium ID working point described in Chapter 2. Impact parameter significance, $|d_0^{\text{sig}}| = d_0/\sigma_{d_0}$, less than 5 standard deviations, and $|\Delta z_0^{\text{IBL}} \sin \theta| < 0.5$ mm are also required, where IBL refers to the ATLAS Insertable B -Layer.
- **SignalElectron:** The electron candidate is required to pass the LooseElectron selection with its p_T required to be greater than 10 GeV. The electron cluster should be in the range of $|\eta| < 2.47$ but excluded from the crack region ($1.37 < |\eta| < 1.52$). Medium likelihood identification is applied in SignalElectron criteria with the impact parameter significance required to be less than 5. In addition, the electron is required to be isolated by passing the Loose_VarRad isolation working point, already defined in Chapter 2.

The choice of the ID and isolation working point has been driven by truth studies, performed in this work, comparing the reconstructed electrons with the true leptons present in the Monte Carlo signal events. The two requirements were: to have high efficiency in selecting a reconstructed electron in events with a true electron, and to have a low efficiency in selecting an electron in events with a true muon, giving priority to the first requirement in order to keep the largest possible efficiency on the already small number of signal events. This approach mitigates the contribution of fake leptons in the selection and shows the good performance of the working points used. The two requirements are shown for possible combination of electron ID and isolation WPs in Figure 4.2a and 4.2b, respectively. As it is visible, LHMedium ID (m ID) and Loose_VarRad isolation (l iso), show an efficiency of 84.6% and a good robustness against "fake" electrons, selecting 4.4% of these events. The choice of using LHMedium ID and not LHLoose ID comes from the possible contribution of fake electrons from hadronic jets, that is much more likely to happen with the more relaxed requirements of the LHLoose ID working point. Conversely, LHTight ID, while providing the best identification performance, exhibits the worst efficiency, affecting the selection of signal events. The combination of LHMedium ID and TightTrackOnly_VarRad isolation WP (tt iso) also shows good performances, with 80.7% efficiency and 1.4% fakes rejection; however, this choice was not possible due to unavailable calibration of this WP for the timescale of this thesis work.

A summary of electron selection requirements is shown in Table 4.3.

Electron Selection	p_T	$ \eta $	ID	$ d_0^{\text{sig}} $	$ \Delta z_0^{\text{IBL}} \sin \theta $	Isolation
LooseElectron	> 5 GeV	< 2.47	LHMedium	< 5	< 0.5 mm	-
SignalElectron	> 10 GeV	< 2.47 and $\notin [1.37, 1.52]$	LHMedium	< 5	< 0.5 mm	Loose_VarRad

Table 4.3: Electron selection requirements.

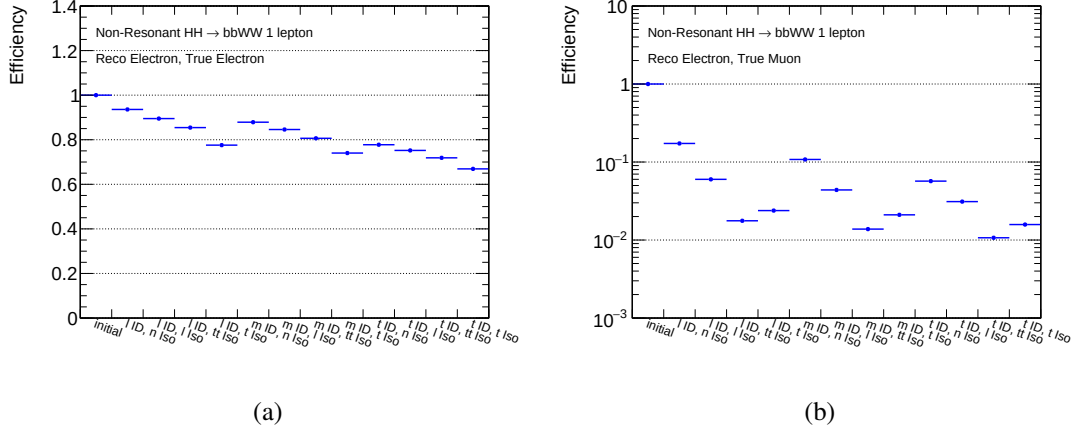


Figure 4.2: Selection efficiencies of reconstructed electrons for events with true electron (a) or with true muon (2) are shown for several combinations of electron ID and isolation WPs: LHLoose (l ID), LHMedium (m ID), LHTight (t ID) IDs, and no isolation (n iso), Loose_VarRad (l iso), TightTrackOnly_VarRad (tt iso), Tight_VarRad (t iso).

Muons Two sets of muon selections are defined. They are denoted as LooseMuon and SignalMuon. The selections are defined as the following:

- **LooseMuon:** The muon p_T is required to be greater than 5 GeV, while the reconstructed track should be in the range of $|\eta| < 2.7$. Medium identification working point, described in Chapter 2, is required. Impact parameter significance, $|d_0^{\text{sig}}| = d_0/\sigma_{d_0}$, less than 3 standard deviations, and $|\Delta z_0^{\text{IBL}} \sin \theta| < 0.5$ mm are also required.
- **SignalMuon:** The muon candidate is required to pass the LooseMuon selection with its p_T required to be greater than 10 GeV. The same ID working point, η range, impact parameter significance and $|\Delta z_0^{\text{IBL}} \sin \theta|$ of the Loose selection are valid. In addition, the muon is required to be isolated by passing the TightTrackOnly_VarRad isolation working point.

The same studies as the electron case have been performed also for muons, determining the choice of the ID and isolation working point. Same considerations and requirements were done for this choice. Efficiencies for events with true muons and with true electrons are shown for the several combinations of muon ID and isolation WPs in Figure 4.3a and 4.3b, respectively. As it is visible, Medium ID (m ID) and TightTrackOnly_VarRad isolation (tt iso), show an efficiency of 87.9% and a good robustness against "fake" muons, selecting only 4.3% of these events, where TightTrackOnly_VarRad isolation was preferred to PFlowLoose isolation for the higher fakes rejection, being at the same level of efficiency.

A summary of muon selections is shown in Table 4.4.

Muon Selection	p_T	$ \eta $	ID	$ d_0^{\text{sig}} $	$ \Delta z_0^{\text{IBL}} \sin \theta $	Isolation
LooseMuon	> 5 GeV	< 2.7	Medium	< 3	< 0.5 mm	-
SignalMuon	> 10 GeV	< 2.7	Medium	< 3	< 0.5 mm	TightTrackOnly_VarRad

Table 4.4: Muon selection requirements.

Taus In order to be orthogonal to the $HH \rightarrow bb\tau\tau$ analysis, a veto is applied to the events with at least one $\tau_{\text{had}}^{\text{vis}}$ candidate reconstructed. This orthogonality is needed

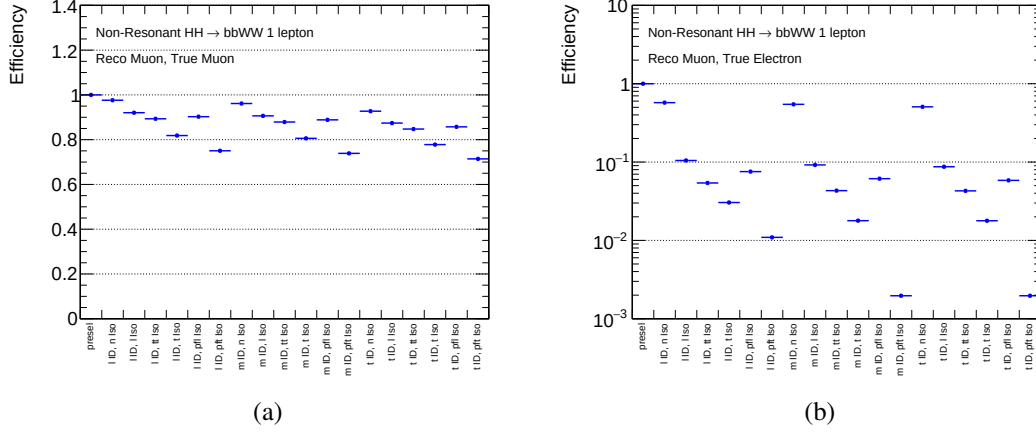


Figure 4.3: Selection efficiencies of reconstructed muons for events with true muon (a) or with true electron (2) are shown for several combinations of muon ID and isolation WPs: Loose (l ID), Medium (m ID), Tight (t ID) IDs, and no isolation (n iso), FCLoose (l iso), TightTrackOnly (tt iso), FCTight (t iso), PFlowLoose (pfl iso), PFlowTight (pft iso).

to allow a possible combination of the results in the different analysis channels. The same requirements for the selection in $HH \rightarrow bb\tau\tau$ analysis are used: $p_T > 20$ GeV, $|\eta| < 2.5$, one or three tracks, unit charge, and passing the RNNLoose working point (see Section 2.2.4). This results in an efficiency loss smaller than 0.5% for signal events, as will be shown in the following, thus not affecting the analysis sensitivity.

Jets Jets are reconstructed using the anti κ_t algorithm with a distance parameter $R = 0.4$ and using the Particle-Flow (PFlow) algorithm [62]. Jet cleaning is applied to remove events with jets built from noisy calorimeter cells or non-collision backgrounds. To avoid selecting jets originating from pile-up interactions a "jet vertex tagger" (JVT) criterion [63] is applied for jets with $20 < p_T < 60$ GeV and $|\eta| < 2.4$ requiring a JVT > 0.5 cut. In order to be selected, jets which pass the jet cleaning and JVT criteria, are further required to have $p_T > 20$ GeV and $|\eta| < 2.5$.

Jets produced by b -quark hadronisation (b -jets) are identified using the DL1r (see Section 2.2.5) b -tagging discriminant using the 77% efficiency working point. Selected jets are labeled "b-tagged jets" if they pass the DL1r 77% fixed-cut working point cut and labeled as "light-jets" if they fail the cut. Studies have been carried out in this work to compare the performance of the b -tagging working points to make the final choice. This comparison is visible in Figure 4.4, where the number of b-tagged jets for different b-tagging algorithms and WPs is shown. The DL1r 77% efficiency WP has been chosen since it shows the highest fraction of events with two b-tagged jets, which are used for the event selection, and provides also a good fraction of events with 1 b-tagged jet, which are used for the fakes estimation. These requirements will be shown in detail in the following Sections 4.2.4 and 6.2.

PFlow reconstructed jets (PFlowJets) have been preferred to standard EMTopoJets (see Section 2.2.5) for the higher number of reconstructed jets, especially in the lower p_T range interesting for the final state of this analysis. Comparisons between EMTopoJets and PFlowJets have been performed in this work and are shown in Figure 4.5 for signal events, where the b-tagging algorithm is MV2c10 for EMTopo and DL1r for PFlow jets. PFlow jets show an higher multiplicity in the range between 20 and 30 GeV, and a higher number of events with two b-tagged jets. Looking at the m_{bb} distribution in Figure 4.5d, it is visible that these additional b-tagged jets contribute to a better resolution of the invariant mass peak, useful to reconstruct the $H \rightarrow bb$ decay.

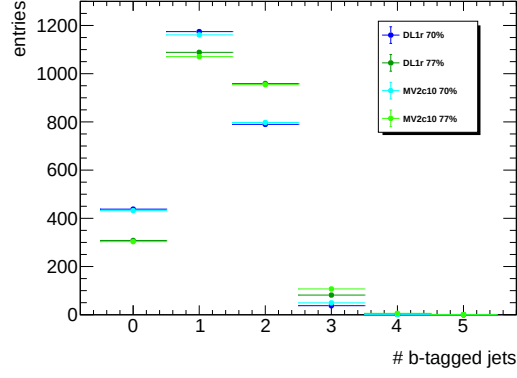


Figure 4.4: Number of b-tagged jets for different b-tagging algorithms and WPs.

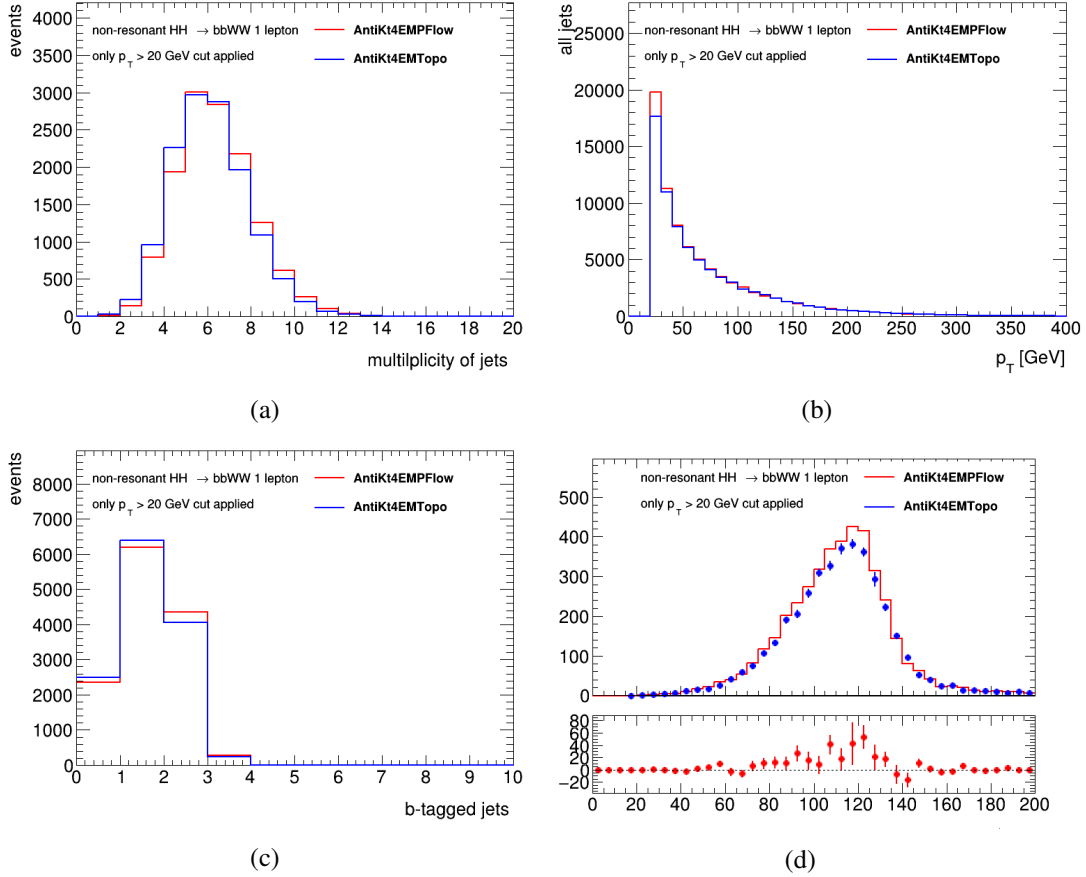
Figure 4.5: Multiplicity of reconstructed jets (a), p_T (b), number of b-tagged jets (c), and invariant mass of the two b-tagged jets (d) for events selected with jets reconstructed using EMTopo and tagged using MV2c10 and with jets reconstructed using Particle-Flow and tagged using DL1r algorithm at 77% efficiency WP. In (d) the ratio of the two distributions is also shown.

Table 4.5 summarizes the jet selection requirements chosen after all the checks done in this section.

The difference in the efficiency of b -tagging between data and simulation is taken into account by applying scale factors provided by the ATLAS Flavour Tagging Combined Performance group.

MET As described in Chapter 2, neutrinos are not directly detectable in ATLAS. The one present in the final state of this analysis, therefore, appears only as an imbalance in

Requirements	Signal Jets
JetCollection	AntiKt4EMPFLOW_BTagging201903
p_T	20 GeV
$ \eta $	< 2.5
Quality	not "bad" jet
Pile-up jet removal	JVT Tight working point
b -tagging	DL1r 77% fixed-cut WP

Table 4.5: Selection for jets with radius parameter $R = 0.4$.

transverse momentum. In this analysis, MET is computed by using electrons that pass the LooseElectron selection, muons passing the LooseMuon selection and selected jets. Photons and hadronically decaying taus are included in the E_T^{miss} calculation as jets since they are not used explicitly in the event reconstruction.

4.2 Event selection

This section details the stages of the event reconstruction and the progression towards the final selection which defines the analysis regions, which are defined in Section 6.1. The event selection follows the composition of the analysed final state, and in particular the presence of a lepton and two b -jets. Among all the events collected by the ATLAS detector, only some of them have the characteristics that are present in the signal events. A first reduction of this amount of events is done by taking into account only the events that pass some specific trigger. The triggers used in the different periods of the Run 2 are summarised in Section 4.2.1. Events passing the triggers are then required to pass a set of selection requirements that are defined to select possible signal candidates, and also events needed for the estimation of the multijet background in dedicated control regions. These requirements are discussed in Section 4.2.3 and 4.2.4 together with the construction of systems of objects that are used in the event reconstruction. The identification of the pair of light jets coming from the W boson decay is an example. This is one of the challenges of this final state as the one W boson is off-shell, and thus there is no W mass constraint that helps the reconstruction. In Section 4.2.5, additional kinematic variables, that have been studied and adapted to the $bbWW$ single lepton final state in order to provide a better discrimination between signal and $t\bar{t}$ events, are described.

4.2.1 Trigger requirements

Events are selected using unprescaled single lepton triggers and MET triggers. These high level triggers (HLT) select events with at least one muon, one electron or with an E_T^{miss} passing a particular threshold among all the events collected by the ATLAS detector. The choice of the use of this type of triggers is driven by the need of selecting events with low p_T leptons, since they arise from the decay of a W boson, that is produced off-shell half of the time in HH events. The triggers with lowest threshold are chosen among all available triggers. The list of triggers used in this analysis is reported in Table 4.6, divided per dataset year. The naming scheme of the trigger algorithms is made of components separated by underscores, and starts with HLT. Some of them are useful to better understand the list in Table 4.6:

- Trigger type: e (electron), mu (muon), xe (MET). Refers to the type of object on which the trigger is focused. If a number is present in front of the trigger type, it

stands for the number of objects that are triggered: in case of a di-muon trigger it would start with 2mu.

- **Threshold:** is the number after the trigger type, indicating the p_T threshold in GeV of the algorithm.
- **Lepton ID:** online lepton ID working point are evaluated for each trigger candidate. These working points follow the naming scheme of the offline ID working points described in Chapter 2, like `lhmedium` or `lhloose`. Muon triggers do not present this component as the default medium ID is used in all the muon HLT triggers. Electron triggers sometimes present the `nod0` tag, which refers to an alignment-robust likelihood tune that does not use d_0 information.
- **Lepton isolation:** online lepton isolation is also evaluated with three working points in particular: `iloose`, `ivarloose` or `ivarmedium`. The first one stands for loose isolation cutting on $p_T^{cone20}/p_T < 0.12$. `ivarloose` defines a variable cone size loose isolation, with a cut on $p_T^{varcone30}/p_T < 0.16$. While `ivarmedium` defines a tighter variable cone size isolation, with a cut on $p_T^{varcone30}/p_T < 0.07$.
- **Level 1 seed:** HLT triggers are seeded by a level 1 (L1) trigger algorithm that has selected a trigger candidate in a defined region of interest. L1 selection is based only on the p_T of the object, while at HLT level, the seed object selection is refined with additional requirements. Each type of trigger uses a default L1 seed. If a seed different from the default one is used, the L1 seed is placed at the end of the algorithm name.
- **MET trigger specific tags:** give the description of the MET reconstruction performed, depending on the object used, and on calibration and corrections applied:
 - `tc` = MET from clusters, either with local cluster weighting calibration (`tc_lcw`) or calibrated at EM scale (`tc_em`);
 - `mht` = MET from jets with soft corrections;
 - `wEFmu` = muon correction at Event Filter level (EF);
 - `pufit` = MET with pile-up fit using clusters as inputs;
 - `pueta` = MET with pile-up subtraction using clusters as inputs.

The triggers listed in Table 4.6 were used in particular data taking periods of each year. They were not all active simultaneously. Events are selected with a logical OR between the triggers that were applied in the same data taking period, checking first if an event is selected by single lepton triggers, and then, only in the negative case, checking if it passes MET triggers. This hierarchy is used since MET triggers have larger uncertainties on their scale factors, since they are not centrally provided by ATLAS due to the MET trigger's phase space dependence, that is different for each analysis. The fraction of signal events selected by the three types of triggers used for each year of data taking, is summarised in Table 4.7 and 4.8 for events with final states containing a true electron or a true muon, respectively.

Single lepton triggers collect about 30-40% of signal events, but are limited by the quite high p_T thresholds that are not lower than 26 GeV, with the only exception of 2015. In order to recover events, otherwise lost, MET triggers have been added whether respect to the 36 fb^{-1} analysis [79]. These algorithms can provide an additional fraction of selected signal events ranging from 7 to 13% among the different years, providing higher signal statistics in the analysis signal region. The fraction of signal events selected by each group of triggers changes among the four years due to different algorithm thresholds and due to improvements in object reconstruction performances during Run 2.

Dataset	Trigger items
2015	HLT_e24_lhmedium_L1EM20VH
	HLT_e60_lhmedium
	HLT_e120_lhloose
	HLT_mu20_iloose_L1MU15
	HLT_mu50
	HLT_xe70_mht
2016	HLT_e26_lhtight_nod0_ivarloose
	HLT_e60_lhmedium_nod0
	HLT_e140_lhloose_nod0
	HLT_mu26_ivarmedium
	HLT_mu50
	HLT_xe80_tc_lcw_L1XE50
	HLT_xe90_tc_lcw_wEFMu_L1XE50
	HLT_xe90_mht_L1XE50
	HLT_xe90_mht_wEFMu_L1XE50
	HLT_xe100_L1XE50
	HLT_xe100_tc_em_L1XE50
	HLT_xe110_pueta_L1XE50
	HLT_xe110_mht_L1XE50
	HLT_xe120_pueta
	HLT_xe120_pufit
	HLT_xe120_tc_lcw_L1XE50
2017	HLT_e26_lhtight_nod0_ivarloose
	HLT_e60_lhmedium_nod0
	HLT_e140_lhloose_nod0
	HLT_mu26_ivarmedium
	HLT_mu50
	HLT_xe110_pufit_L1XE50
	HLT_xe110_pufit_L1XE55
2018	HLT_e26_lhtight_nod0_ivarloose
	HLT_e60_lhmedium_nod0
	HLT_e140_lhloose_nod0
	HLT_mu26_ivarmedium
	HLT_mu50
	HLT_xe110_pufit_xe65_L1XE50
	HLT_xe110_pufit_xe70_L1XE50
	HLT_xe120_pufit_L1XE50

Table 4.6: Summary of unprescaled trigger items used for Run 2 data. Default level 1 seeds are L1MU15, L1EM20 and L1XE50 in 2015, and L1MU20, L1EM22 and L1XE50 in 2016-2018.

Trigger type	2015	2016	2017	2018
Single electron	34.6%	33.5%	37.7%	42.2%
Single muon	2.4%	2.4%	2.5%	3.0%
MET	7.0%	9.4%	9.5%	8.8%
Total	44.0%	45.3%	49.7%	54.0%

Table 4.7: Fraction of selected signal events by single electron, single muon and MET triggers, for each year of data taking. Only events with final state containing a true electron are considered here. The total fraction of selected events, given by the sum of trigger's contributions, is also shown. Events selected here by single muon triggers are rejected in the event selection when requiring the trigger match.

Trigger type	2015	2016	2017	2018
Single electron	0.0%	0.0%	0.0%	0.0%
Single muon	31.4%	31.5%	33.8%	37.1%
MET	9.1%	11.8%	12.9%	12.4%
Total	40.5%	43.3%	46.8%	49.7%

Table 4.8: Fraction of selected signal events by single electron, single muon and MET triggers, for each year of data taking. Only events with final state containing a true muon are considered here. The total fraction of selected events, given by the sum of trigger's contributions, is also shown.

4.2.2 ATLAS derivations

The size of MC simulated samples and real collected data is often very large. A common ATLAS procedure, named "derivation", is done to reduce the size of the datasets before their use in physics analyses. This procedure consists in four operations:

- Skimming: removes whole events
- Thinning: removes whole objects from within an event, but keeps the rest of the event
- Slimming: removes information from within objects, but keeps the rest of the object
- Augmentation: adds information not found in the input data

There are several types of possible derivations, that differs for the selection requirements that are applied. The choice of the derivation therefore depends on the final state of interest of the analysis and on the object and event information that are needed for their use in the analysis. The main requirements are: type of objects included, selection threshold of these objects, trigger information included and event selection cuts applied. Usually all this cuts are loose cuts, that have the aim of discarding all the data not needed in standard physics analyses. For this analysis a check of the signal events kept by a set of potential derivations has been performed. The selected derivation provides electrons, muons and jets with the desired p_T thresholds, as well as single-lepton and MET trigger information. A summary of the object requirements for the chosen derivation is reported in Table 4.9.

Derivation skimming is then performed requiring one of the possible combinations:

- Single lepton trigger firing, at least 1 muon_{HL} or electron_{HL}, at least 4 jet₁₅.

Object definition	p_T [GeV]	$ \eta $	ID
muonSL	$\in [2.5, 25]$	< 2.7	Loose
muonHL	≥ 25	< 2.7	Loose
electronSL	$\in [2.5, 25]$	< 2.6	LHLoose
electronHL	≥ 25	< 2.6	LHLoose
jet15	> 15	< 2.8	
jet25	> 25	< 2.8	
jet200	> 200	< 2.8	

Table 4.9: Selection requirements applied at derivation level.

- Single lepton or MET trigger firing, at least 1 muonHL or electronHL, at least 2 jet25.
- MET trigger firing, at least 1 muonSL or electronSL, at least 2 jet25.
- MET trigger firing, at least 1 lepton among the four possible definitions, at least 1 jet200.

The efficiency of this selection requirement is reported in Table 4.10 as the first selection step, since it is performed on the initial set of events in each MC and data sample. As it is visible from Table 4.10, derivation skimming have a large impact on signal sample, discarding 48% of the events. The main reason of this effect is that $W \rightarrow \tau \nu$ decays are present in the signal sample and are rejected by the derivation skimming. This corresponds to $\sim 35\%$ of initial events discarded. The remaining fraction mostly depends on the requirement of having a single-lepton or a MET trigger firing the event. However, these requirements are standard and are also applied between pre-selection and event selection, as can be visible comparing the cuts defined above with the ones in Section 4.2.3 and 4.2.4. For this reason, event rejected at this stage would be rejected in the following steps of the analysis selection, not affecting the sensitivity of the analysis.

4.2.3 Event pre-selection

A pre-selection is applied in order to keep all the events useful for all the analysis regions, both signal region and signal depleted control regions used for background estimation. The following selection requirements are applied at the pre-selection level to the recorded events:

- In order to assure good data quality, events with bad detector conditions, namely where large part of the detectors were missing from data acquisition due to problems during a run, or when the performance of the detectors were affected by large noise, have been rejected from the data analysis. Only events contained in the Good Run List are used, as described in Section 3.1. Moreover, incomplete events or events with bad detector information are rejected.
- The presence of a primary vertex with at least two tracks is required. Among all primary vertices, the one with the highest $\sum p_{T,trk}^2$, where $p_{T,trk}$ is the transverse momentum of tracks associated with the vertex, is retained as the primary interaction vertex.
- The presence of at least one LooseElectron or LooseMuon, as defined in Section 4.1, is required.

- The presence of at least 4 jets, of which at least 1 is b -tagged, as defined in Section 4.1, is required.
- The application of the tau veto is performed, rejecting events with at least one tau lepton satisfying the requirements defined in Section 4.1.

A cut-flow of the event pre-selection is reported in Table 4.10, showing the fraction of events selected after each requirement. The expected event yields after all selection cuts, at an integrated luminosity of 139 fb^{-1} , are also shown.

Sample	Skim	≥ 1 loose lep	≥ 4 jets	≥ 1 b-jet	Tau veto	Event yields
ggF Signal	0.5176	0.4054	0.2915	0.2612	0.2551	120.9
$t\bar{t}$	0.4730	0.3878	0.2439	0.2196	0.2076	24011443.5
Single-top	0.3909	0.3327	0.0760	0.0640	0.0612	1234414.9
$t\bar{t}W$	0.5456	0.4257	0.3489	0.3080	0.2832	23656.1
$t\bar{t}Z$	0.6231	0.3394	0.2365	0.2102	0.1993	4767.0
W +jets	0.2998	0.2458	0.0631	0.0152	0.0148	41396778.9
Z +jets	0.5335	0.4898	0.1441	0.0466	0.0430	11305610.1
Diboson	0.5285	0.5035	0.0358	0.0068	0.0057	37581.3
ggF H	0.0467	0.0341	0.0038	0.0008	0.0008	5129.9
VBF H	0.0864	0.0526	0.0062	0.0015	0.0014	729.5
WH	0.4727	0.4044	0.0794	0.0709	0.0685	856.4
ZH	0.1619	0.1077	0.0287	0.0138	0.0124	1311.1
$t\bar{t}H$	0.5878	0.4448	0.4150	0.3919	0.3591	26210.4

Table 4.10: Fraction of MC events selected by each step of the pre-selection, normalised to the total number of events in the input sample. Skim stands for the derivation level’s event selection. The expected event yields with an integrated luminosity of 139 fb^{-1} is also reported for each sample in the last column.

After the pre-selection, the surviving events are split in two regions, tight and loose, where the first corresponds to signal-like events, while the second collects all the events that are needed for the multijet background data-driven estimation, described in Section 6.

4.2.4 Signal topology selection

After the pre-selection, a tighter event selection is done to keep only the events satisfying the requirements of the signature of interest for this analysis, that, as already described, is composed by one charged lepton (e/μ), two b -tagged jets, two light-jets and MET. The selection requires:

- Exactly one SignalElectron or SignalMuon in the event.
- Exactly two b -tagged jets, using the DL1r b -tagging algorithm at 77% working point.
- At least two light-jets, already required in the pre-selection.
- If the event is selected by a single lepton trigger, the SignalLepton is trigger matched to the corresponding HLT object which fires the trigger, and an additional p_T requirement, of 1 GeV for electrons or 5% for muons, above the HLT trigger threshold is made on the matched lepton in order to use the trigger scale factors.

- If the firing trigger is a single electron trigger that requires a LHTight electron, SignalElectrons must pass the LHTight ID WP.
- If the event is triggered by MET triggers, a cut on the MET of 5% above the trigger threshold of the firing HLT trigger is applied.

A cut-flow of the event selection is reported in Table 4.11, showing the fraction of events selected after each step of the selection requirements. The expected event yields after all selection cuts, at an integrated luminosity of 139 fb^{-1} , are also shown.

Sample	Trigger	2 b-jets	1 SignalLepton	Yields
ggF signal	0.6407	0.3456	0.3348	40.4
$t\bar{t}$	0.6485	0.2471	0.2409	5762309.5
Single-top	0.3874	0.1401	0.1374	346574.3
$t\bar{t}W/Z$	0.7632	0.3131	0.2937	7849.8
W +jets	0.3190	0.0105	0.0083	298484.3
Z +jets	0.3180	0.0045	0.0039	40318.7
Diboson	0.4592	0.0473	0.0463	12697.3
$ggF H$	0.2757	0.0291	0.0271	138.6
$VBF H$	0.1810	0.0463	0.0440	32.1
WH	0.4880	0.2805	0.2662	901.1
ZH	0.4334	0.0851	0.0806	163.8
$t\bar{t}H$	0.7709	0.1777	0.1696	4411.8

Table 4.11: Fraction of MC events selected by each step of the event selection, normalised to the total number of events surviving the pre-selection cuts. The trigger cut requires at least a firing trigger among the ones in Table 4.6 and the trigger matching requirements. The expected event yields with an integrated luminosity of 139 fb^{-1} is also reported for each sample.

After all the selection cuts, only $\sim 8.5\%$ of signal events, evaluated from the total number of expected $HH \rightarrow bbWW \rightarrow bbl\nu qq$ events, are kept, corresponding to about 40 expected signal events in selected data. As visible from Table 4.11, the applied cuts have the effect of decreasing the number of selected background events. The cut requiring exactly one SignalLepton allows to reduce all the background processes with hadronic or multi-leptonic final states. The requirement of having exactly two b-tagged jets provides the larger effect in reducing the electroweak processes, such as W/Z +jets and dibosons, while the top background is less affected. These cuts affect also the signal, but with a smaller reduction of the events compared to the majority of the background sources. $t\bar{t}$ processes are still the larger background source, as expected, because of the final state identical to that of the HH signal. The huge presence of this background is the main reason of using machine learning techniques to perform a further selection step. In this way, exploiting the information carried by several kinematic distribution obtained by the combination of the final state selected objects, it is possible to recognise with a good accuracy the parent process of the event. Therefore, a classification of the events in signal and background-like categories provides a powerful discrimination and a reduction of this background. Besides $t\bar{t}$ processes, the main background sources are W +jets and single-top production processes, that, having a large initial cross-section, are still present in large quantity among the selected events.

Multijet background is not present in Table 4.11 since its estimation is done looking at a different region obtained with different selection criteria. This other selection is referred to as "loose" selection in the following to distinguish it from the event selection described in this Section that is also referred to as "tight" selection. Events not satisfying the requirements listed above are asked to pass the requirements for being included

in the "loose" region for the multijet background data-driven estimation, as described in Section 6. If also those requirements are not satisfied, the events are discarded. A first rough estimation of the multijet contribution may be done by subtracting all the expected events coming from MC simulated backgrounds to the number of selected data events. However, MC predicted events of major backgrounds, like $t\bar{t}$ and W +jets needs to be normalised to data in dedicated control regions, as will be described in Section 6.1, before doing such a comparison. For this reason and in order to keep the data blinded in regions where we expect the presence of signal events, the cut-flows of the selection of the collected data are not shown here. A general comment can still be done at this stage. Although one lepton and two b-jets are required, this background is still present due to the selection of fake leptons. This fake events arise mainly from a wrong identification of energy deposit clusters, identified as electrons, and from the tagging of soft light-jets as b-jets, where the b-tagging algorithms have a smaller background rejection, as visible from Figure 2.15. For this reason, a robust estimation of this background contribution is needed and it is described in Section 6.2.

4.2.5 Higgsness and Topness

The Higgsness and the Topness are variables based on the kinematics of HH and $t\bar{t}$ processes, applied for the first time in the $HH \rightarrow bbWW^* \rightarrow bbl\nu l\nu$ channel by CMS [123]. These variables have been modified in this work, taking into account the different kinematics of the process, to be used in the $HH \rightarrow bbWW^* \rightarrow bbqq l\nu$ channel [124]:

$$\mathbf{T} = \min(\chi_{12}^2, \chi_{21}^2) \quad \text{where: } \chi_{ij}^2 = \min_{p_z^\nu} \left[\frac{(m_{b_l\nu}^2 - m_t^2)^2}{\sigma_t^4} + \frac{(m_{l\nu}^2 - m_W^2)^2}{\sigma_W^4} + \frac{(m_{b_jqq}^2 - m_t^2)^2}{\sigma_t^4} + \frac{(m_{qq}^2 - m_W^2)^2}{\sigma_W^4} \right] \quad (4.1)$$

$$\mathbf{H} = \min_{p_z^\nu} \left[\frac{(m_{l\nu qq}^2 - m_H^2)^2}{\sigma_H^4} + \min \left(\frac{(m_{l\nu}^2 - m_W^2)^2}{\sigma_W^4} + \frac{(m_{qq}^2 - m_{W^*_{peak}}^2)^2}{\sigma_{W^*}^4}, \frac{(m_{qq}^2 - m_W^2)^2}{\sigma_W^4} + \frac{(m_{l\nu}^2 - m_{W^*_{peak}}^2)^2}{\sigma_{W^*}^4} \right) \right] \quad (4.2)$$

where the mass values are set to $m_t = 173$ GeV, $m_W = 80.379$ GeV, $m_H = 125.18$ GeV, and $m_{W^*_{peak}}$, that is the invariant mass peak of the off-shell W , is set to 40 GeV. In truth level studies, the sigma values are set to: $\sigma_H = 2$ GeV, $\sigma_t = 5$ GeV, $\sigma_W = 5$ GeV, $\sigma_{W^*} = 5$ GeV. These values are the same used in [123] and are associated with experimental uncertainties and intrinsic particle widths.

Both equations are functions of the longitudinal momentum of the neutrino p_z^ν , on which these functions are minimised. The initial value of p_z^ν is given from the constraint on H and W mass, respectively for Higgsness and Topness, since they are functions evaluated in the hypothesis of HH and $t\bar{t}$ events:

$$m_H^2 = E_{l\nu qq}^2 - p_{l\nu qq}^2 \quad (4.3)$$

$$m_W^2 = E_{l\nu}^2 - p_{l\nu}^2 \quad (4.4)$$

In reconstructed events, MET components are used as neutrino's in Equation 4.3 and 4.4, leaving the longitudinal component of the neutrino momentum as unknown variable of the two equations. Neutrino p_z can be reconstructed using the relations:

$$E^\nu = \sqrt{p_T^2 + p_z^2} \quad (4.5)$$

$$p_x^\nu = p_T \cos(\phi) \quad (4.6)$$

$$p_y^\nu = p_T \sin(\phi) \quad (4.7)$$

where ϕ is the azimuthal angle of the E_T^{miss} , E^ν the neutrino energy, p_x and p_y the two transverse spatial components of the neutrino momentum. Being Equations 4.3 and 4.4 quadratic expressions in p_z , they can have two real, one real or two complex solutions. In the last case only the real part of the complex solution is taken into account, therefore a single value of p_z is obtained. In the first case the solution with the neutrino direction closest to the charged lepton is selected.

Higgsness and Topness distributions in a two-dimensional plane ($\log(H)$, $\log(T)$) are shown in Figure 4.6a and 4.6b for truth-level simulated signal and $t\bar{t}$ background samples respectively. Red lines are drawn to give a visible reference for a possible separation between signal and background: with these cuts the surviving fraction of the signal is $\sim 69\%$ and the background surviving fraction is $\sim 1.2\%$.

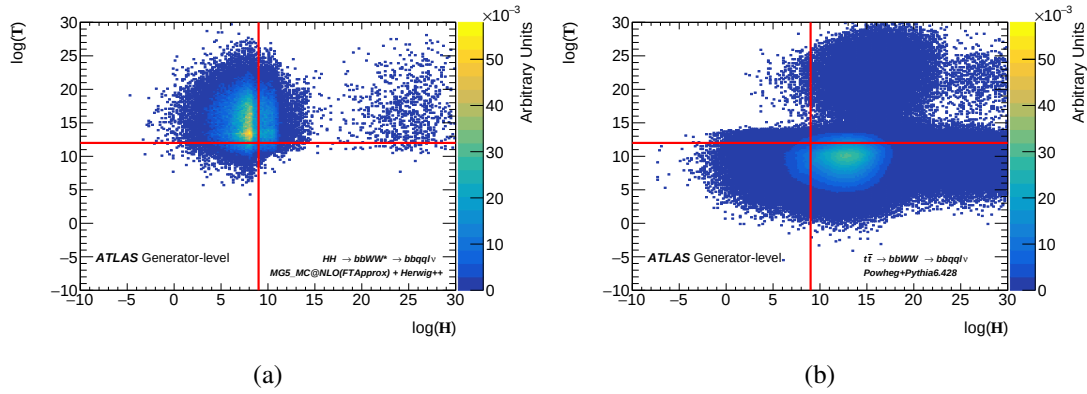


Figure 4.6: Distribution of Higgsness and Topness in a two-dimensional plane ($\log(H)$, $\log(T)$) for simulated signal $HH \rightarrow bbWW^* \rightarrow bbqqlv$ (a) and background $t\bar{t} \rightarrow bbWW \rightarrow bbqqlv$ (b) events without selection requirements at generator level. The signal sample was generated with MG5_MC@NLO(FTApprox) + Herwig++, while the background sample was generated with Powheg + Pythia6.428. The distributions are normalised to unit area. Red lines are drawn to give a visible reference for a possible separation between signal and background [124].

At reconstruction level, detector smearing and resolutions affect the invariant mass resolution of the systems considered in Equation 4.2 and 4.1. These effects must be considered in the choice of the σ values, and, therefore, they are taken from fits to the peaks of invariant mass distributions for W , W^* , t and H , providing the values: $\sigma_H = 18$ GeV, $\sigma_t = 30$ GeV, $\sigma_W = 15$ GeV, $\sigma_{W^*} = 18$ GeV. The distributions used with the fit superimposed are shown in Figure 4.7. Signal sample has been used to perform the fit of H invariant mass distribution on m_{bb} , while $t\bar{t}$ sample has been used for fitting m_{bjj} and m_{jj} in order to extract σ_t and σ_W respectively.

The use of these fitted sigma values has been tested, whose result is visible in Figure 4.8, where it is shown that the variables defined with the fitted sigma values provide a better discrimination power at reconstruction level, mitigating the large effect of the detector resolution.

However, the poorer invariant mass resolution affects the discrimination power of the distributions when used alone, smearing the peaks visible in Figure 4.6, as it will be seen in the following Section. This issue is partially recovered by the use of these variables within machine learning algorithms, where they provide still a useful information, carrying correlations with other object kinematic observables.

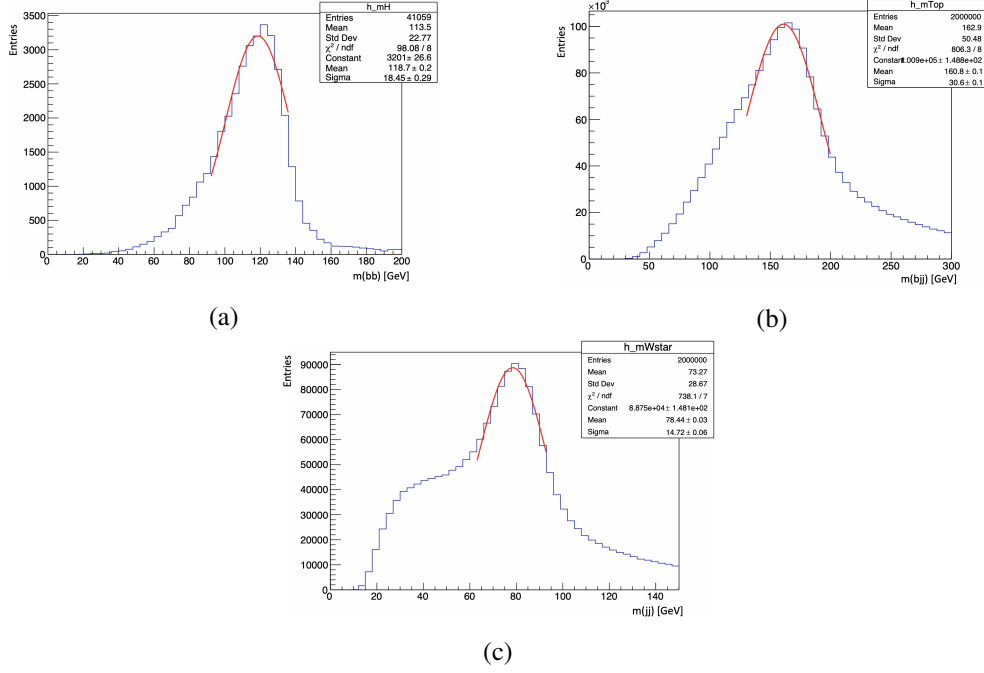


Figure 4.7: Invariant mass distributions of m_{bb} (a) of signal events, and m_{bjj} (b) and m_{jj} (c) of $t\bar{t}$ events. Gaussian fits on the peaks of the distributions are done to extract σ_H , σ_t and σ_W respectively.

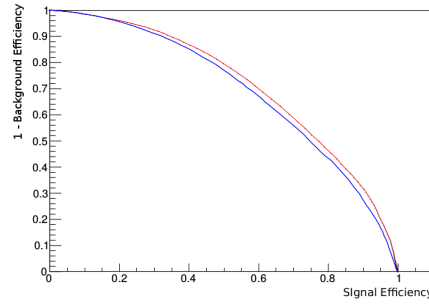


Figure 4.8: Signal efficiency as a function of background rejection for the ratio of $\log(H)$ over $\log(T)$ distribution on reconstruction level signal events. Higgsness and Topness variables are here defined with standard sigma values (blue line) and fitted sigma values (red line).

5 Machine Learning based analysis

Artificial Neural Networks (ANN or NN) [125–127] are used in the analysis to improve the separation between signal and dominant backgrounds, and the overall sensitivity of the analysis. This Machine Learning (ML) technique has been developed, optimised and implemented in the analysis framework for this thesis work. Multivariate and ML algorithms are now commonly used in high-energy physics searches, where the process of interest has a very small cross section compared to the one of the background processes miming its signature. In these conditions, the use of standard cut-based analyses is not optimal, since the simple rectangular cuts that are applied on a set of discriminating variables may not provide a good separation between signal and background. In these processes, common kinematic variables of the final state objects or of particle systems may show a very little or no shape difference for signal and background distributions, resulting in a poor rejection of background. The use of ML techniques can provide better separation between signal and background compared to cut-based analyses. These techniques can make use of information regarding linear and non-linear

correlations between the discriminating input variables. Moreover, while in a cut-based analysis it is only possible to classify events in separate categories, ML classification algorithms can also provide a classification output distribution that can be itself used as final discriminating variable, that provides a larger separation power. This output discriminant can be used in the final fit to increase the sensitivity of the analysis.

The choice made in this thesis work of using a combination of NN output distributions as the final discriminant variable in the fit for this analysis resulted in a significantly improved sensitivity compared to the one obtained with other analysis strategies that were tested during the analysis development, such as cut-based approach or the use of Boosted Decision Trees (see Section 5.1.1 and Appendix A).

5.1 Introduction to Machine Learning techniques

Machine Learning is a particular sub-group of Artificial Intelligence which is devoted to the development of algorithms that learn from the experience. There are different categories of ML depending on the way the learning process is done: supervised, unsupervised and reinforced learning. In supervised learning, the learning process is done by providing examples in input to the algorithm, together with the real class of which each data belongs to, called label. Unsupervised learning leaves the algorithm to find the relations between the inputs, and to create clusters of similar data with topological or probabilistic methods, without having any information of the class to which the data belongs. In Reinforcement learning the learning process is based on Markov chains, which is a stochastic model describing a sequence of possible events where the probability of each event depends only on the state of the previous event and not on the ones of the previous events. Here, Markov chains are applied to software agent actions in an unknown environment: each action determines a positive or negative reward by the environment, and the model tries to maximise the cumulative rewards during the exploration process. Among these learning paradigms, the main one used in physics analyses is the supervised learning. Since it is the one used in this thesis work, we will focus only on this type of ML.

In a ML model, for each input object, a set of observables that characterise it, called features, are provided. The algorithm, then, exploits the information contained in the features and the correlation between them, in order to recognise the different classes of input examples, producing a set of rules for the classification. The correct or incorrect assignment of an object to a class is rewarded with a weight that is assigned to the extracted rule. This weight is then used in a new step of the learning process, called training, to correct the rules and recover the correct classification of the misclassified events. The training is therefore an iterative procedure, that allows a gradual improvement of the capabilities of the classification algorithm. At the end of the training, a set of final rules are produced, which can be applied to an independent set of data to obtain their classification.

The training process tries to reduce the classification error to the minimum value, but this does not guarantee a good performance of the model during its application on new data. Input training samples have always a fixed dimension and the information extracted from them can be biased, meaning that the used features are particular of the training sample but are not useful in general to separate signal and background. For this reason a good classifier must avoid the so called overtraining, guaranteeing good generalisation capabilities, being robust against statistic and systematic noise of the input datasets.

ML techniques can be applied to a plenty of possible problems, like classification of objects, face recognition, image analysis, search engines and many others, where the

manual development of a standard algorithm is challenging and not sufficient for the problem's requests. Different types of ML algorithm, called models, have been developed. The difference among them is the approach used to extract the information from the input data and its processing in the training. In high-energy physics the applications of ML are many, like the possibility of discriminating signal over background for event selection, particle identification or flavour tagging, and the fast simulation of detector response to different types of particles. In physics data analyses the most used models are based on Boosted Decision Trees (BDT) and Artificial Neural Networks, for which a brief description is given in Section 5.1.1 and 5.1.2.

5.1.1 Boosted Decision Trees

A decision tree [128] is a binary-tree structured classifier, as schematically illustrated in Figure 4.9. A set of variables x_i of a training sample, in which the event category (signal or background) is known, are taken as inputs. Binary decisions are then taken on one single variable at a time, and events are placed in two possible groups, called nodes, depending on the outcome of the decision, that is usually a simple cut. For each node, a new decision on a different variable is taken, and the separation of the events continues until a stop criterion is fulfilled. This criterion can be the maximal allowed number of subsequent decisions that can be taken, which corresponds to the depth of the tree, or the training can stop when in a node are present less than a configurable percentage of training events. The nodes at the bottom end of the tree are called leaves, and are labeled S for signal and B for background depending on the majority of events that end up in the respective nodes. The phase space is split in this way into many regions that are classified as signal or background, that are then grouped together forming two final groups of signal-like and background-like events.

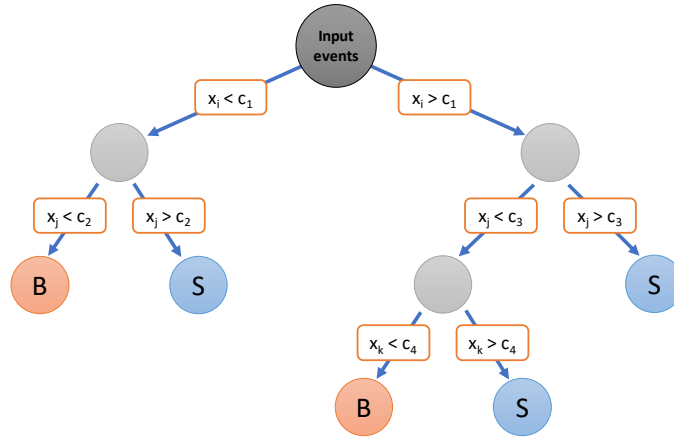


Figure 4.9: Schematic view of a decision tree.

Decisions during the training are taken according to various separation criteria: the standard criteria is the Gini-Index (G), defined as:

$$G = p \cdot (1 - p) \quad \text{with} \quad p = \frac{S}{S + B} \quad (4.8)$$

where p is the purity of the sample after the cut. As said, for each node a splitting criterion is applied, based on a cut on a single variable. The variable and the value of the cut for a particular node are selected during the training in order to optimise the increase in the separation index between the parent node and the sum of the indices of the two daughter nodes, weighted by their relative fraction of events. This optimisation is

done by scanning over the variable range with a configurable granularity, selecting the cut value that gives the best separation between signal and background. The training procedure can be extended from one single tree to a group of several trees, each with different weights and different selection chains, forming a forest, all derived from the same training sample and processed in sequence. This technique is called boosting and is based on the re-weighting of the events for each tree, giving them a different importance in each tree. Boosting increases the stability of the response of the classifier with respect to statistical fluctuations and it is able to considerably improve the separation performance compared to a single decision tree. Several boosting algorithms can be used, the most common one is the so called "Adaptive" boosting, for which at the end of each training step, events that were misclassified are given an higher event weight, by multiplying it by a common boost weight α . These set of weights are then used as starting weights in the training of the following tree. The boost weight is derived from the misclassification rate, err , of the previous tree:

$$\alpha = \frac{1 - err}{err} \quad (4.9)$$

The weights of the entire event sample are then renormalised such that the sum of weights remains constant. The result of an individual classifier $h(x)$, with x being the set of input variables, is encoded for signal and background as $h(x) = +1$ and -1 respectively. Then, the boosted event classification $y_{Boost}(x)$ is given by [128]:

$$y_{Boost}(x) = \frac{1}{N_{collection}} \sum_i^{N_{collection}} \ln(\alpha_i) \cdot h_i(x) \quad (4.10)$$

where the sum is done over all classifiers in the collection. Small and large values for $y_{Boost}(x)$ indicate a background-like and signal-like event respectively. This BDT classification output function shows then a different shape of the distribution for signal-like and background-like events, and can be used as discriminating variable or can be fitted to extract the sensitivity of the analysis. The BDT score distributions for the training sample and a testing, independent, sample are then compared to evaluate the possible overtraining. If the algorithm is overtrained the shape of the distributions for training and test samples are different, and the separation performance of the BDT results to be much less powerful on the testing sample.

A BDT was trained using the TMVA tool [128] during preparatory work of this analysis for a comparison of the reachable sensitivity with respect to the cut-based approach used in the 36 fb^{-1} analysis. Signal and $t\bar{t}$ background truth level kinematic variables were used in the training. The cuts used in the previous cut-based analysis were optimised, keeping the same set of kinematic variables. In the BDT a new set of 15 kinematic variables were used, including the new Higgsness and Topness variables. Cuts simulating common requirements on reconstructed objects were also applied on truth level particles. After the optimisation of both the models, the BDT was found to be a factor ~ 4 more sensitive than the cut-based approach. This was one of the main reason for proceeding with the development of a multivariate analysis strategy for the search of this process. More details about this study are discussed in Appendix A.

5.1.2 Artificial Neural Networks

Artificial Neural Networks [125–127] are a mathematical model based on the simulation of biological neural network working processes. It consists in a interconnected group of artificial neurons, which are mathematical functions that receive one or more inputs, each carrying a weight, and combine them through a weighted sum to produce

an output. The sum is then passed through a non-linear function, known as activation function. Activation functions can have several shapes, both linear or non-linear, like sigmoid and tanh, or also step functions like the rectified linear unit (ReLU), which is defined by:

$$\text{ReLU}(x) := \max(0, x) \quad (4.11)$$

These functions usually are monotonically increasing, continuous and differentiable, even this last requirement can be relaxed as in the case of ReLU function which has an undefined derivative in zero.

The ANN analyses all the inputs in parallel using layers of several artificial neurons. These layers are organised in the following way:

- An input layer contains a number of nodes equal to the number of input features and containing their information.
- A number $N \geq 1$ of "hidden" layers are placed after the input layer. These layers can be made by a different number of nodes from the one of the input layer, and among the hidden layers themselves, where these nodes are artificial neurons with a non-linear activation function.
- An output layer, with a number of nodes depending on the number of the output classes of the model, closes the chain of layers. Also here the nodes are artificial neurons, providing the output response of the model.

Connections between nodes belonging to two neighbour layers are the only ones allowed. Schematic views of an artificial neuron and of the structure of an artificial Neural Network are visible in Figure 4.10a and 4.10b. This type of ANN, that is also the more common one, is named Feed-Forward Multilayer Perceptron (MLP).

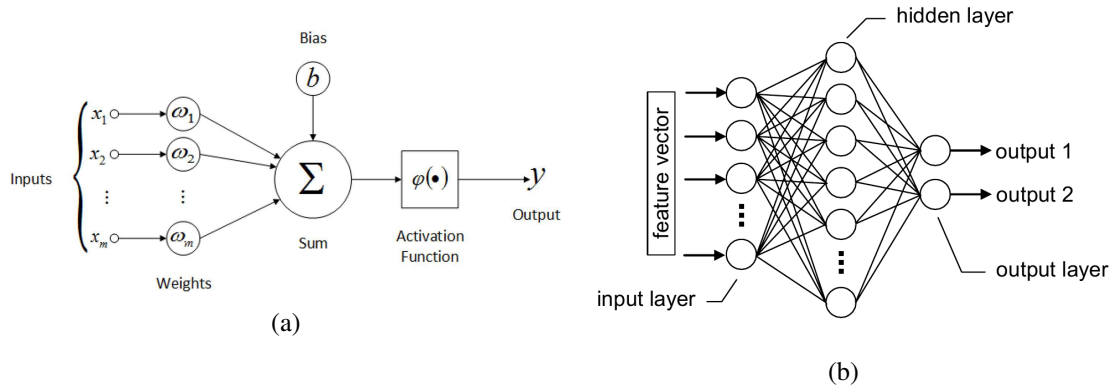


Figure 4.10: Schematic views of the structures of an artificial neuron (a) and of an Artificial Neural Network (b).

ANN behaves like adaptive system: it modifies its structure based on the information flowing through the network during the learning phase.

During NN training, N events $x_i (i = 1, \dots, N)$ are provided to the network. For each training event, the network response $y_{NN}(i)$ is evaluated. This response can be expressed, for simplicity for a network with one hidden layer, one neuron in the output layer and a linear activation function in the output layer, by:

$$y_{NN}(i) = \sum_{k=1}^{n_h} A \left(\sum_{j=1}^{n_{var}} x_j^i \omega_{jk}^{(1)} \right) \cdot \omega_{k1}^{(2)} \quad (4.12)$$

where n_{var} and n_h are the number of nodes of the input and hidden layer respectively, A is the activation function of the hidden layer, $\omega_{jk}^{(1)}$ is the weight between the j -th input layer neuron and the k -th hidden layer neuron, and $\omega_{k1}^{(2)}$ is the weight between the k -th hidden layer neuron and the output layer neuron.

The network response is compared to the expected output $Y_i \in [0, 1]$, where 0 is for background and 1 is for signal events respectively. A measure of the distance Δ between $y_{NN}(i)$ and Y_i is then defined as:

$$\Delta = \sum_{i=1}^N \frac{1}{2} (y_{NN}(i) - Y_i)^2 \quad (4.13)$$

The distance function used in an ANN is also known as loss function. The set of optimised weights ω is defined as the one that minimise the distance Δ . This minimisation is usually obtained with the so called gradient descent method, where the weights are updated on a small distance basis in the $\nabla_{\omega} \Delta$ direction:

$$\omega^{(\rho+1)} = \omega^{(\rho)} - \eta \nabla_{\omega} \Delta = \omega^{(\rho)} - \eta \frac{\partial \Delta}{\partial \omega} \quad (4.14)$$

where η is the step size, or learning rate of the minimisation, and ρ is the index that tracks the weight update iterations. MLP utilises a supervised learning called "back-propagation" for training, which works in two phases: a forward phase, where the network weights are fixed and the input vector is propagated through the layers of the network (function signal), and a backward phase, in which the Δ obtained at the end of the network is propagated layer by layer through the network, but in the opposite direction of the previous phase (error signal). Every neuron, in hidden and output layers, receives and evaluates both function and error signals. In this way the information on the error made by the network classification is provided to the nodes of the network for the update of their weights to minimize the loss. This update can be done in different ways. The most common one is the batch training, where the weights are updated only after that all the training events have been processed. Every iteration of this type is called epoch of the training, at the start of which the input events are shuffled in order to learn differently and to optimise the performances. A variation of this method is the stochastic training, in which the weights are updated after the processing of a subset of training events, chosen randomly. This procedure allows a much faster training process and it is used especially for big multilayer networks.

Many parameters of the network model can be configured in order to optimise the performances, like: number of input features, number of hidden layers, nodes per hidden layer, layer activation function, initial weights value, and learning rate. The number of output nodes depends on the number of classes in the classification problem. The initial weights value must be set different from zero and should break the symmetry of the network. For this scope, different probability density functions are used to initialise the weights of the nodes.

As already discussed, the overtraining is a critical problem in ML. It can easily arise during the training of multilayer neural networks with a big number of nodes per layer. One solution is provided by the so called dropout. This method consists in setting to zero a fraction, called dropout rate, of node weights, chosen randomly at the beginning of every training epoch. This results in lighter networks with a smaller number of free parameters, and so less exposed to possible overfitting of the training sample, and in a big improvement the performances of deep networks. The learning is, in fact, preserved by the random choice of the nodes that are switched off, being different at every training epoch, that creates different networks at every training step, maintaining

all the nodes information available during the training process. A schematic description of the dropout process is shown in Figure 4.11.

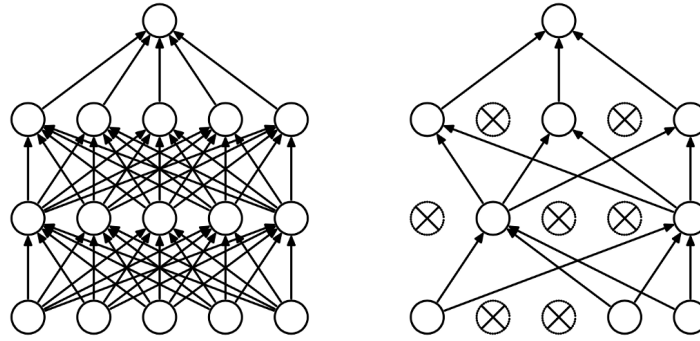


Figure 4.11: Comparison of a standard neural network before (left) and after (right) the application of the dropout.

Deep Neural Networks When the network structure is made by more than two hidden layers, it is usually referred to as deep neural network (DNN). Deep network architectures result more efficient in representing some classes of functions. Problems that can be represented with a polynomial number of nodes organised in k layers, may require an exponential number of nodes with $k - 1$ layers [129]. In these cases, a deep architecture allows to use a smaller number of neurons, and therefore a smaller number of weights and free parameters to be updated at every iteration. DNN are used in many applications that present complex problems, like image processing and big data analysis.

Another advantage of deep architectures is that every layer with a non-linear activation function corresponds to a transformation of the information coming from the previous layer. This results in an increased level of abstraction at every layer, that improves the capacity of selecting and amplifying the relevant aspects of the input for the purposes of classification.

DNNs suffer of overtraining by definition. In this architectures the use of regularisation methods, like dropout, are of crucial importance.

In DNN it is usually used the ReLU activation function, defined in Equation 4.11, in the several hidden layers. The computational time of such a function is much smaller with respect to other activation functions: there is only the need of comparison, addition and multiplication, and no more complex operations are needed. This provides several advantages since the number of neurons is often large, allowing faster and effective training of deep neural architectures on large and complex datasets, compared to sigmoid or similar activation functions.

5.2 Deep Neural Network model for the $bbWW^*$ analysis

As previously discussed, the analysis described in this thesis work makes use of MVA techniques to better discriminate the huge background of its final state. The aim is to classify not only signal events, but also two of the main background sources, $t\bar{t}$ and $W + jets$, using a multi output classifier. The choice of these backgrounds is driven by the large contribution that they show after the event selection, as it can be seen in Table 4.11, and by the different final state topology that these two processes show among them and with respect to the signal. Final state object's properties, in fact, differ between these three processes, as well as the decay processes that occur producing their

final states. All these informations can be exploited with the use of ML techniques, as it will be discussed in the following, to correctly classify data and apply a further selection, to increase the signal over background ratio and to increase the sensitivity of the analysis. The use of neural networks is one of the best choices for this purpose, for the advantages described in Section 5.1. The analysis strategy makes use of a deep learning neural network based approach, trained with three class labels for the three processes whose classification is required following the previous discussion:

- Non-resonant HH production
- $t\bar{t}$
- $W + jets$

Single-top background, although having a contribution comparable with $W+jets$ after the event selection, was not considered in the training as the classification of $t\bar{t}$ background was expected to be valid also for single-top events.

The neural network model is built using the Keras library [130] and TensorFlow [131] as a backend. The network inputs are first standardized with the use of Scikit-Learn tools [132] to transform the variables to look like standard normally distributed data, e.g. Gaussian with 0 mean and unit variance, and make easier and more effective the work of the estimator. For example, a feature that has a variance that is orders of magnitude larger than the other variables might dominate in the training and bias the estimator, which could be unable to learn from other features as expected.

The input standardized variables are then passed through five fully-connected layers, also referred to as dense layers. The first four layers have 512 nodes, while the last one has 256 nodes. Between each pair of subsequent dense layers there is a dropout layer, and an additional dropout layer is present also between the last hidden layer and the output layer. The architecture of the model is completed with a final layer with the 3 output nodes and softmax activation function, which is defined by:

$$f_{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad \text{for } i \in [1, K] \quad (4.15)$$

Softmax function transforms a K -dimensional input vector z in a K -dimensional vector $f_{softmax}(z)$ containing values bounded between 0 and 1, whose sum is normalised to unity.

Each of the dense layers in the network have their weights randomly initialised by sampling from a truncated normal distribution centered on zero with standard deviation of $\sqrt{2/(N_{inputs} + N_{outputs})}$, where N_{inputs} is the number of input features and $N_{outputs}$ is the number of output nodes of the layer. Output layer weights are instead initialised by sampling from a uniform distribution. The activation functions used for the dense layers are rectified linear units, ReLU.

Softmax activation function of the output layer allows one to interpret the outputs as each representing the probability for the output's associated class. Since the sum over all outputs of a layer with softmax activation function is equal to unity, the outputs of the softmax layer can be seen as a probability distribution. For this reason it is commonly used for multi-class neural network classifiers, and in the following we will refer to the outputs of our network model as p_c , where c has three values for each of the output classes: $c \in [HH, t\bar{t}, W + jets]$.

Dropout layer randomly sets inputs to zero with a frequency of rate (dropout rate) at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by $1/(1 - rate)$ such that the sum over all inputs is unchanged [130]. At

each forward pass through the network, the dropout layers in practice disable random portions of the network, presenting in this way a modified thinner network.

For training, the loss metric used in the model is the categorical cross-entropy with the use of the "Adam" optimization algorithm [133]. Categorical cross-entropy is a function composed by a softmax activation function combined with a cross-entropy function CE . The cross-entropy function is defined by:

$$CE = - \sum_{c=1}^C t_c \log(f(z_c)) \quad (4.16)$$

Where t_c and z_c are the class label and the NN score for each class c in C . $f(z_c)$ refers to the application of an activation function the NN scores, that happens before the cross-entropy loss computation. Categorical cross-entropy definition is the same of Equation 4.16, where $f(z_c)$ is the result of the application of the softmax activation function to the NN scores. Using this loss, NN is trained to output a probability over the C event categories. In multi-class classification, for one event only the relative class a keeps its term in the loss, so there is only one element of the target vector t which is not zero, that is $t_c = t_a = 1$, and all the other elements of the sum are discarded. The resulting equation is therefore for one class a :

$$CE = -\log\left(\frac{e^{z_a}}{\sum_{j=1}^K e^{z_j}}\right) \quad (4.17)$$

The "Adam" optimization algorithm [133] is an extension to stochastic gradient descent, defined in Equation 4.14, that is often used in deep network models. Differently from stochastic gradient descent, in Adam algorithm the learning rate changes during training. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients [133]. Adam adapts parameter learning rates based on the average first moment, the mean, and the average of the second moments, the variance, of the gradients. In this way this algorithm combines different optimization algorithms present in the literature, providing an optimiser that can handle sparse gradients on noisy problems.

A summary of the architecture configuration and the parameters used in the DNN model construction is summarised in Table 4.12.

Configuration/parameter	Setting
Number of input features	36
Number of hidden layers (nodes per layer)	5 (512 + 512 + 512 + 512 + 256)
Number of output classes	3
Dropout layers (dropout rate)	5 (0.5 + 0.5 + 0.5 + 0.5 + 0.5)
Activation function	ReLU (hidden layers) Softmax (output layer)
Layer weight initialiser	Truncated Normal distribution (hidden layers) Uniform distribution (output layer)
Loss function	Categorical cross-entropy
Optimiser (learning rate)	Adam (0.0005)

Table 4.12: Summary of the parameters used in the analysis DNN model.

The parameters summarised in Table 4.12 have been chosen after studying several possible configurations, by varying some of the main architecture parameters like number

of nodes per layer, number of layers, and learning rate of the optimisation algorithm. Such a study is reported in Table 4.13, where the final configuration has been selected by looking at the set of parameters with the larger test accuracy, defined as the ratio of the number of correct predictions over the total number of predictions made by the trained DNN model on an independent set of events, as will be described in Section 5.2.2.

Number of hidden layers (nodes per layer)	Learning rate	Test accuracy [%]
3 (512 + 512 + 512)	0.0005	77.98
4 (512 + 512 + 512 + 512)	0.0005	78.03
5 (512 + 512 + 512 + 512 + 256) *	0.0005	78.33
5 (128 + 128 + 128 + 128 + 128)	0.0005	77.71
5 (256 + 256 + 256 + 256 + 256)	0.0005	77.75
5 (512 + 512 + 512 + 512 + 256)	0.001	77.86
5 (512 + 512 + 512 + 512 + 256)	0.0001	77.60 overtraining

Table 4.13: Scan of model architecture parameter configurations and relative test accuracy. The parameters not present in the table are set in the same way that is reported in Table 4.12. The final selected configuration is marked with *. The configuration with learning rate 0.0001 was showing overtraining and it has been discarded.

5.2.1 Input features

A set of 36 features for the three input samples has been provided as input to the network model. These kinematic distributions are related to single physics objects or to systems of physics objects of particular importance for the $bbWW$ final state. Leading and sub-leading object have been defined for b-jets and light-jets, and are related to their ranking in decreasing p_T . A description of the input features is here given:

- Leading b-jet ($bjet1$): transverse momentum, η , and ϕ components distributions are used and shown in Figure 4.12. For signal and $t\bar{t}$ events p_T^{bjet1} is larger than for W -jets as in this last case the jets comes from the emission in the initial state of quarks and gluons that are usually much softer than jets arising from Higgs or top-quark decays. Also the η distribution shows a larger presence of these jets in the forward directions, while jets coming from hard objects mostly populate the central detector region.
- Sub-leading b-jet ($bjet2$): transverse momentum, η , and ϕ components distributions are used and shown in Figure 4.13. The same conclusions of the leading b-jet are valid here, where it is visible an enhanced peak in p_T^{bjet2} for W -jets due to the fact that having a second b-tagged jet in such events is less likely and mostly given by emission of very soft quarks and gluons.
- bb system: invariant mass m_{bb} of the system made by the two selected b-tagged jets, and difference in angular distribution between the two objects, ΔR_{bb} and $\Delta\phi_{bb}$ are provided as inputs and are shown in Figure 4.14. ΔR is defined in equation 2.7, while $\Delta\phi$ is simply defined as:

$$\Delta\phi_{A,B} = |\phi_A - \phi_B| \quad (4.18)$$

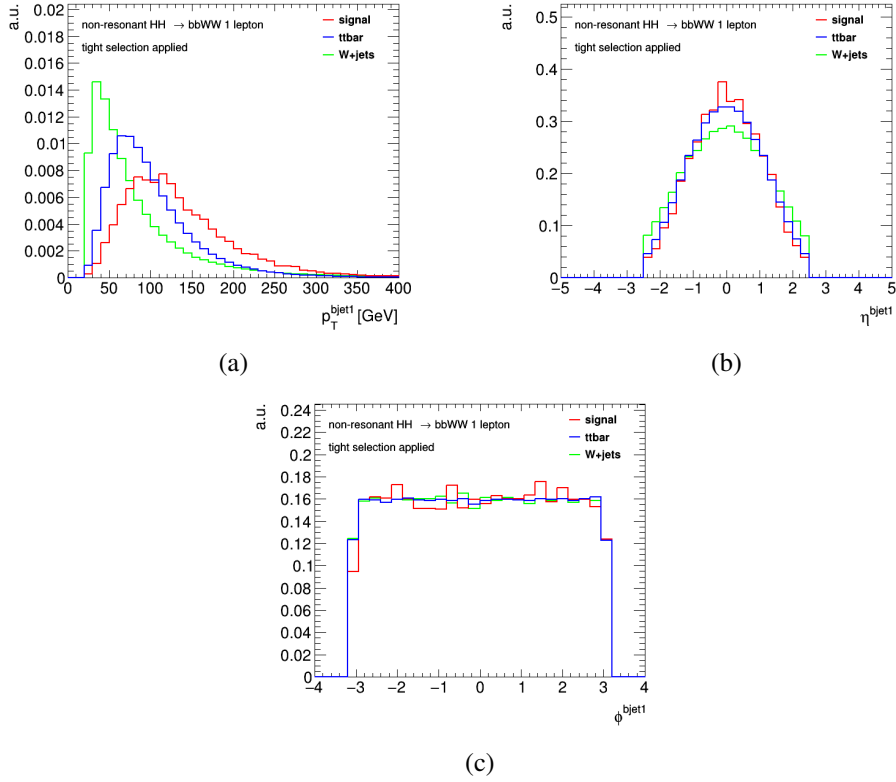


Figure 4.12: p_T (a), η (b) and ϕ (c) distributions of the leading p_T b-tagged jet, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

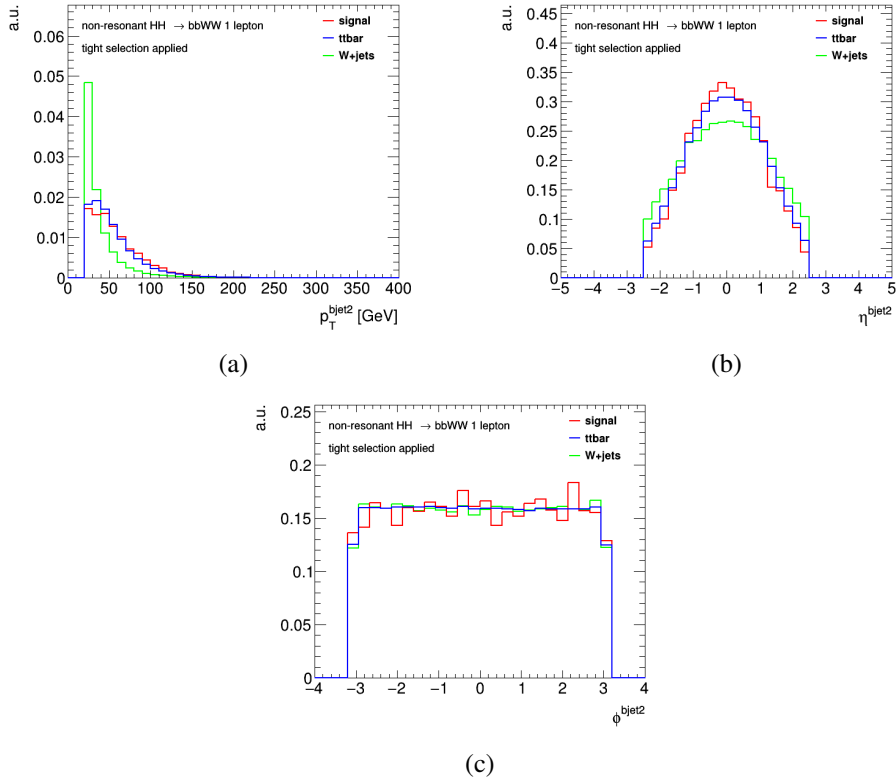


Figure 4.13: p_T (a), η (b) and ϕ (c) distributions of the sub-leading p_T b-tagged jet, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

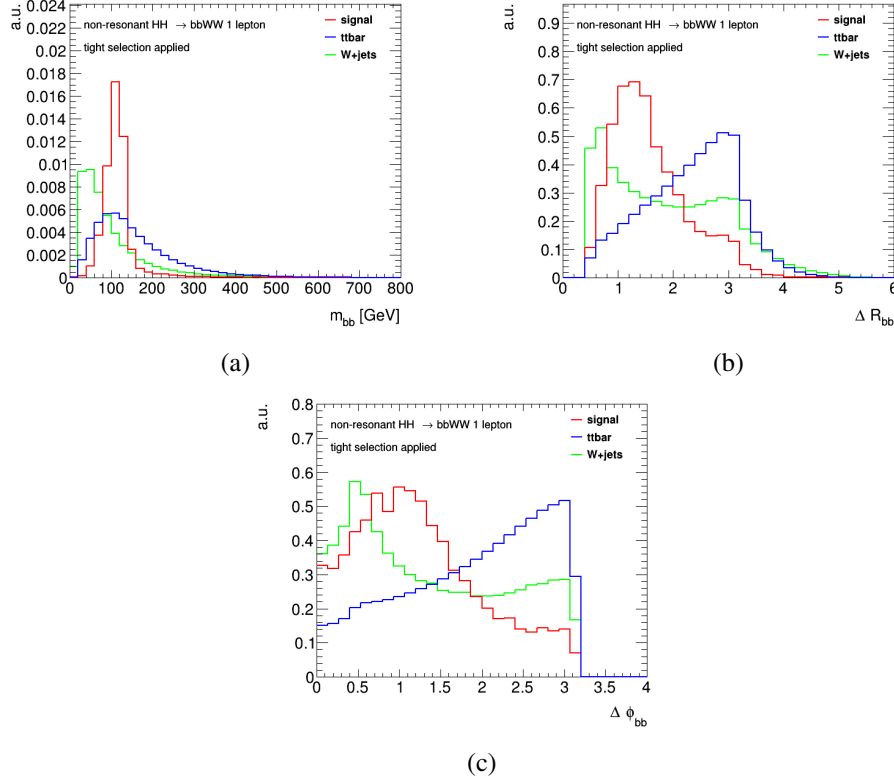


Figure 4.14: Invariant mass (a), ΔR_{bb} (b) and $\Delta\phi_{bb}$ (c) distributions for the two b-tagged jets system, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection. Distributions are normalised to have unit area.

This system reconstructs the Higgs boson decaying in a b-jet pair. As it is visible from Figure 4.14 these distributions provide a big discriminating power between the three processes. For signal events we expect to have m_{bb} peaked at 125 GeV, while for $t\bar{t}$ the distribution is smeared due to the fact that the two b-jets comes from two different parent objects. In W +jets events, the invariant mass is much smaller since it comes from the combination of two soft objects as described above. Angular distributions are very useful. In $t\bar{t}$ decays we expect to have two b-jets in opposite directions, giving the peaks at π in Figures 4.14b and 4.14c, while in signal events we expect a small opening angle between the two, as they are coming from the decay of the same Higgs boson. Also from W +jets events the opening angle is small as the b-jets can arise from the splitting of an emitted gluon.

- Leading light-jet ($jet1$): transverse momentum, η , and ϕ components. Distributions are shown in Figure 4.15. Signal jets are softer than signal b-jets as they come from the decay of one W , which can be produced off-shell half of the times. $t\bar{t}$ light-jets, instead, comes from the hadronic decay of W bosons, which are always produced on-shell in the top-quark decay. W +jets distributions are similar to the ones of the b-jets as expected, as the emitted quarks can be of every flavor.
- Sub-leading light-jet ($jet2$): p_T , η , and ϕ components. Distributions are shown in Figure 4.16 and follow the same conclusions given for the leading light-jet.
- Third-leading light-jet ($jet3$): transverse momentum, η , and ϕ components. Distributions are shown in Figure 4.17. Events may not have a third light-jet as the event selection does not requires it. In this case a null four-vector is set for this

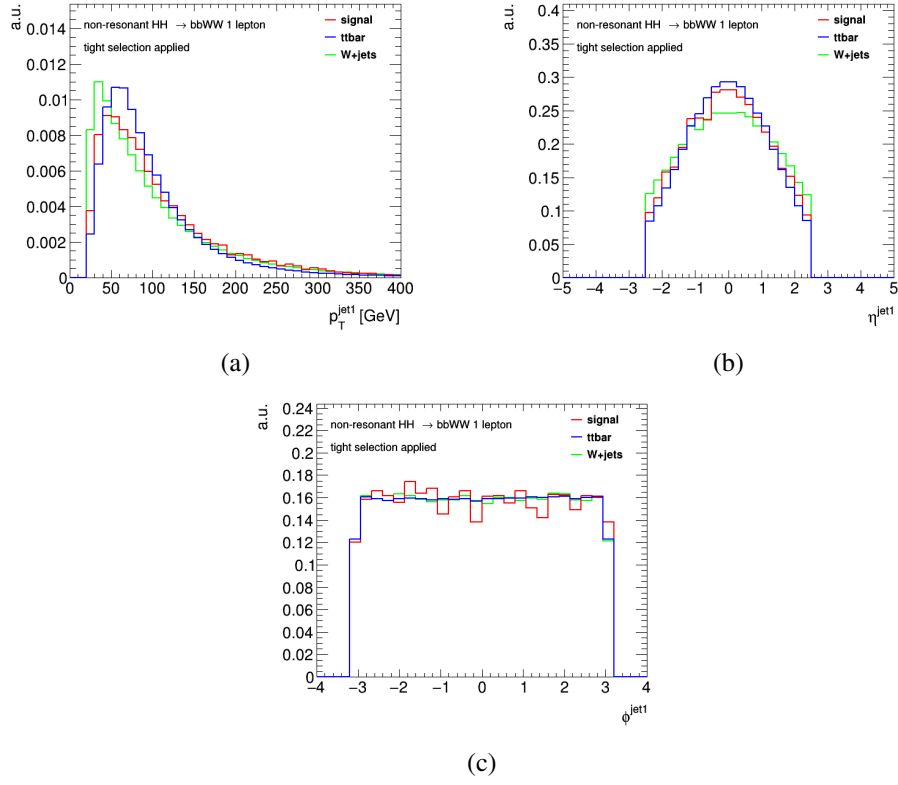


Figure 4.15: p_T (a), η (b) and ϕ (c) distributions of the leading p_T light-jet, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

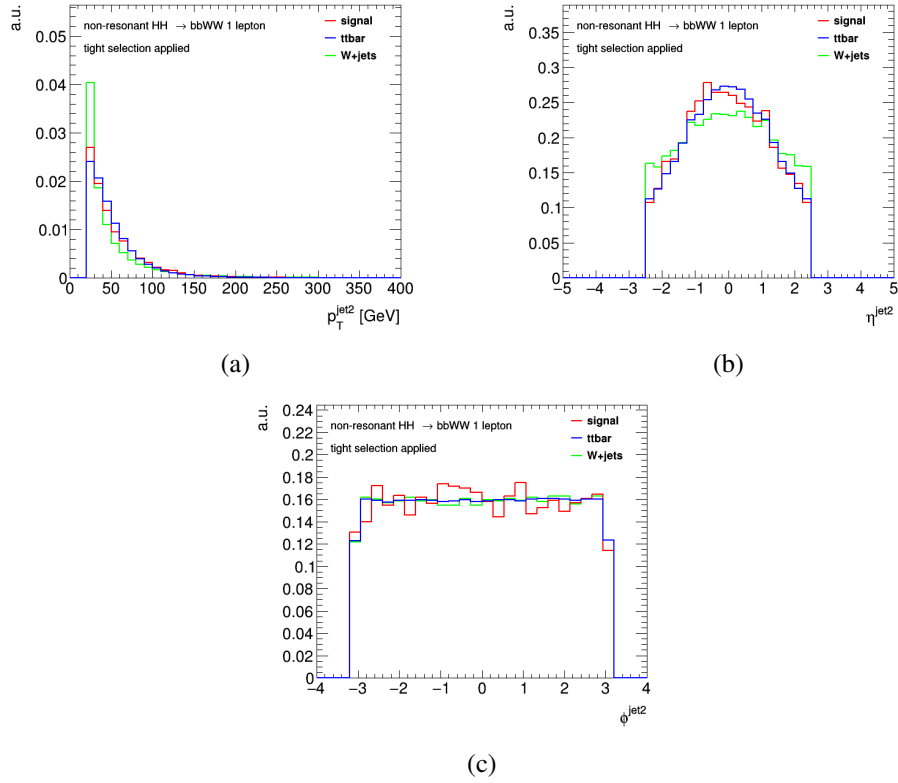


Figure 4.16: p_T (a), η (b) and ϕ (c) distributions of the sub-leading p_T light-jet, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

object, that provides the peaks at zero visible in Figure 4.17. This non-physical information does not affect the training of the classifier as it is symmetric and can be used to recognise events with a smaller number of jets. Sub-leading and third-leading light-jet features are provided to the classifier in order to allow an optimised combination of the objects to discriminate the three processes from the different decay kinematics.

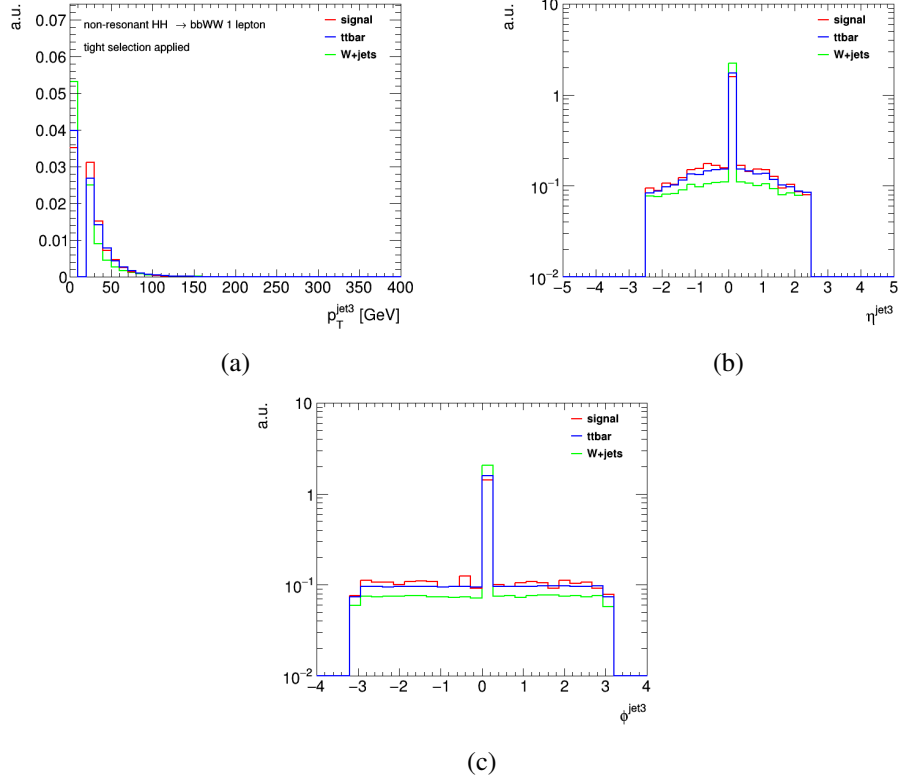


Figure 4.17: p_T (a), η (b) and ϕ (c) distributions of the third-leading p_T light-jet, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area. Events with only two light-jets provide the peaks at zero.

- Number of all jets in the event, b-tagged and light-jets. Its distribution is shown in Figure 4.18, showing a smaller number of jets for W +jets events, as expected as the probability of emitting more than the four jets already required in the event selection is small. This information is correlated with the peak in the third-leading light-jet, as those events corresponds to the ones that have only four selected jets in the final state, and therefore without a third light-jet.
- Lepton: transverse momentum, η , and ϕ components, with the addition of the information of the flavor of the selected lepton. This last feature is provided as a two bin only distribution: one corresponding to the electron and the other to the muon, depending on the selected SignalLepton. Distributions are shown in Figure 4.19.
- Missing E_T (MET): magnitude of the missing transverse energy E_T^{miss} , and its ϕ direction. Distributions are shown in Figure 4.20.
- Lepton+MET system: transverse mass m_T , and difference in ϕ direction ($\Delta\phi$) between the selected SignalLepton and the MET. The transverse mass is defined

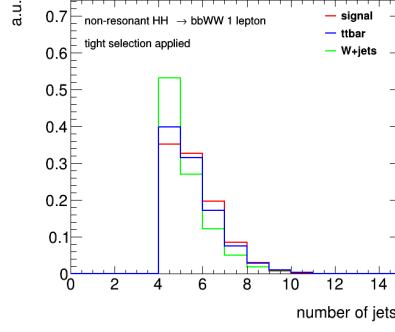


Figure 4.18: Distribution of the total number of jets in the event, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

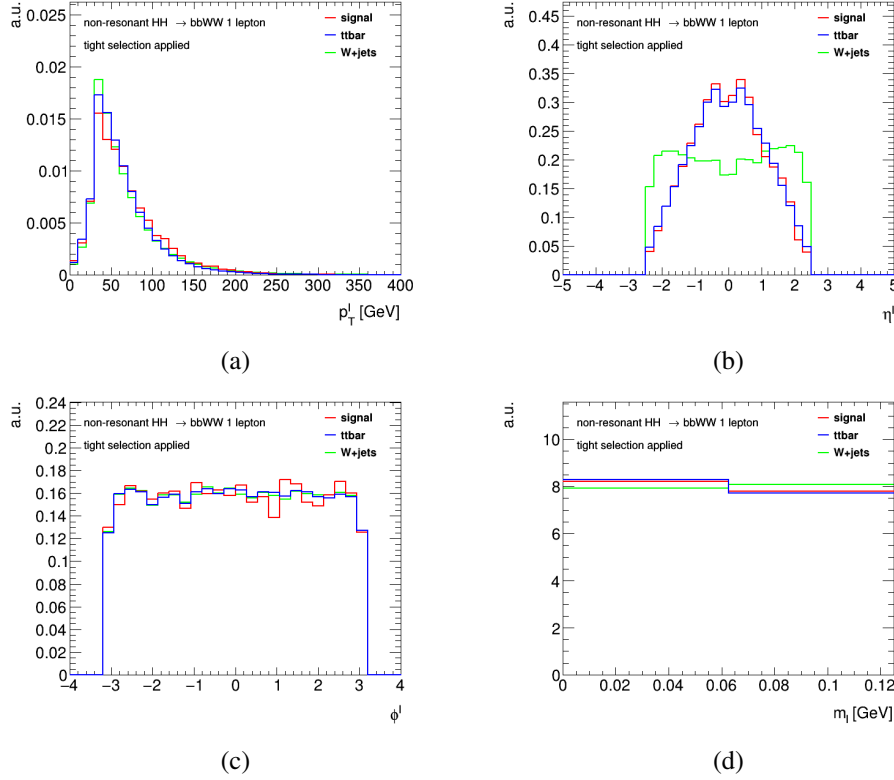


Figure 4.19: p_T (a), η (b), ϕ (c) and mass (d) distributions of the selected SignalLepton, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area. Mass distribution (d) is given in two bins to provide the information of the lepton flavor (e/μ).

as:

$$m_T = \sqrt{\left(\sum_i E_T^i\right)^2 - (p_x^i)^2 - (p_y^i)^2} \quad (4.19)$$

where the sum is made on the objects contained in the system. This system reconstructs the leptonic decay of one W boson. Distributions are shown in Figure 4.21, where $m_T^{l\nu}$ have smaller values due to the decay of both on-shell and off-shell W bosons.

- WW system: transverse mass of the system m_T^{WW} , calculated by using Equation 4.19 with the hadronically decaying W (W_{had}) and the lepton, where W_{had} is built combining the two light-jets closest in ΔR among the three leading ones

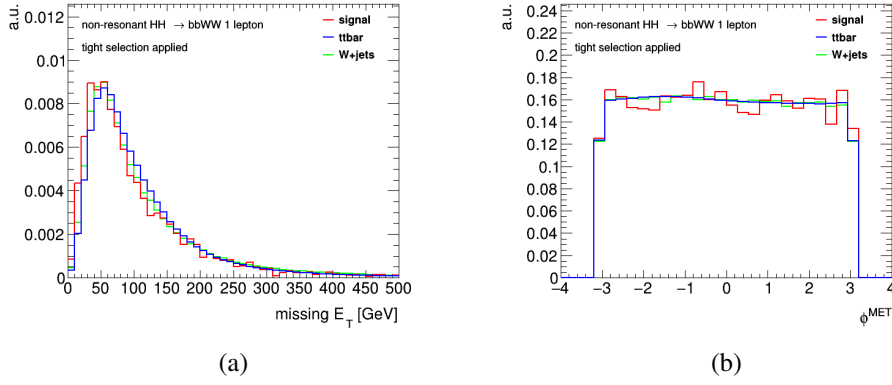


Figure 4.20: Magnitude (a) and ϕ direction (b) distributions of the missing E_T of the event, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

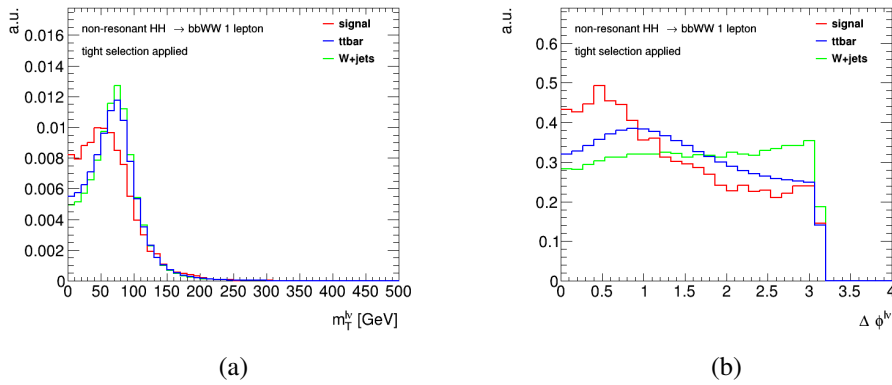


Figure 4.21: Distributions of the lepton plus MET system transverse mass (a), and $\Delta\phi$ between lepton and MET direction (b), for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

described above. Its distribution is shown in Figure 4.22, which for the signal process shows a mass peak compatible with the $H \rightarrow WW^*$ decay, while for $t\bar{t}$ background the peak is shifted to higher values.

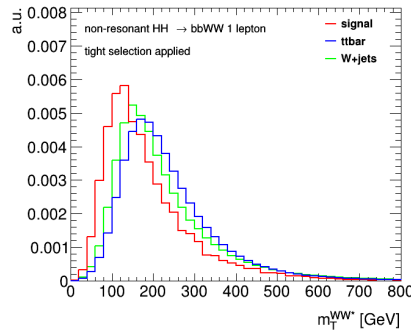


Figure 4.22: Transverse mass distribution of the WW system, for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area.

- Higgsness and Topness: logarithmic distributions of the variables defined in Section 4.2.5, built with sigma values obtained by fit to the invariant mass peaks at reconstruction level. Distributions are shown in Figure 4.23 together with the correlation plots between Higgsness and Topness distributions for signal, $t\bar{t}$ and W +jets samples.

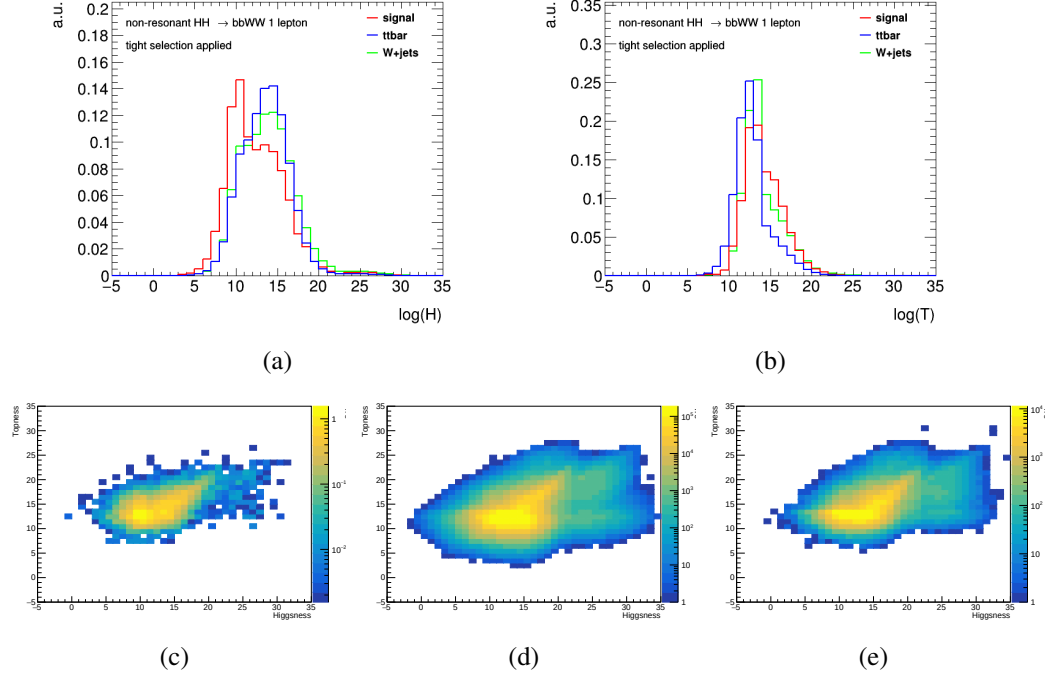


Figure 4.23: Higgsness (a) and Topness (b) distributions for signal, $t\bar{t}$ and W +jets events passing the event (tight) selection, normalised to have unit area. Correlation plots between Higgsness and Topness for signal (c), $t\bar{t}$ (d) and W +jets (e) events are also shown.

- Angular relations between different objects. Minimum and maximum ΔR between lepton and the two b-tagged jets, $\Delta R_{max}^{lep,bjet}$ and $\Delta R_{min}^{lep,bjet}$ are used to highlight the difference in top-quark and Higgs decays: top-quark decays into a b-jet and a W , which, if it decays leptonically, produces a lepton close to the b-jet coming from the same top-quark and far from the other one; on the other hand the lepton is expected to be far from the b-jets for signal events as the first comes from the decay of one Higgs, while the two b-jets are produced in the decay of the second Higgs boson.

Minimum and maximum ΔR between lepton and the three leading light-jets, $\Delta R_{max}^{lep,jet}$ and $\Delta R_{min}^{lep,jet}$, for which it is expected to have the lepton close to one of these jets as they come from the same decaying Higgs boson in HH events, while to have it far from all light-jets for $t\bar{t}$ and W +jets events.

Finally, minimum and maximum $\Delta\phi$ between the MET and the two b-tagged jets, $\Delta\phi_{min}^{MET,bjet}$ and $\Delta\phi_{max}^{MET,bjet}$, are also used as show a good shape difference between signal and background events. In $t\bar{t}$ events, MET and b-jet are close as they come from the decay of the same top-quark, while in signal events b-jets and neutrino are produced by the two Higgs bosons, and have opposite direction. Distributions are shown in Figure 4.24.

Some of the distributions clearly show a good separation power between signal and background events. They are mainly high-level variables, built by the combination of physical object information and whose definition is driven by the aim of reconstructing the decay chain of the process studied in the analysis. bb , WW and lepton plus MET systems follow this purpose. Their invariant masses distributions try to discriminate events by looking at the invariant mass peak of the decaying parent particle. Angular distributions are also high-level features, as they are defined in order to highlight the right or wrong association between the objects belonging to a particular decay process. As an example, the ΔR distribution between the two b-tagged jets is expected to peak at

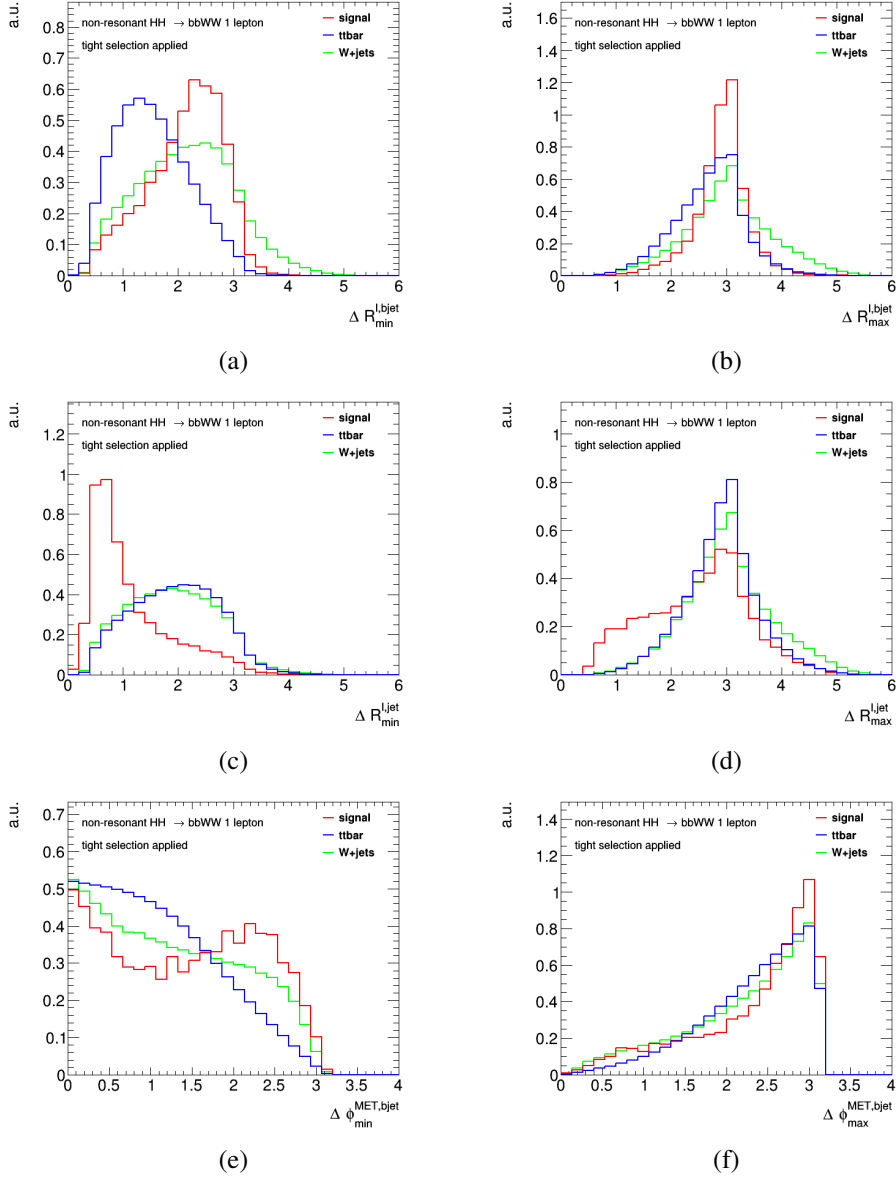


Figure 4.24: Minimum (a) and maximum (b) ΔR distributions between lepton and the two b-tagged jets; minimum (c) and maximum (d) ΔR distributions between lepton and the three leading light-jets; minimum (e) and maximum (f) $\Delta\phi$ between the MET and the two b-tagged jets. Distributions are given for HH , $t\bar{t}$ and W +jets events passing the event (tight) selection, and are normalised to have unit area.

low values for HH events, while it is expected to peak at π for $t\bar{t}$ events, as in the first case the two b-jets are coming from the decay of one Higgs boson, while in the other case they are coming from the decays of two different top quarks emitted in opposite directions. The angular differences between the lepton and the light- and b-tagged jets, as well as the ones between the MET and the b-tagged jets, have the same purpose of providing discriminating power to the classifier, extracted from the different kinematics of the decay chains between signal and background processes.

Other distributions do not show a clear separation between signal and background events. These are usually low level features, like η and ϕ direction of some physics object, that are nevertheless important in the training of the classifier. The information they provide is used to find a correlation with the other input features in order to recognise signal- and background-like events, as can be visible from Figure 4.25, which

shows the correlation between the ϕ components of the leading jet, sub-leading jet and the lepton. These three distributions, taken alone, do not show a separation power between signal and background. However, when used together they provide an important separation based on the different kinematics of HH and $t\bar{t}$ physics processes. The lepton is in fact produced by the same Higgs boson and its ϕ angle is correlated with the one of the two jets arising from the same Higgs boson decay, while for $t\bar{t}$ and W +jets events the lepton and the light-jets are produced by different decays, giving a different type of correlation. This difference can be then exploited by the DNN, increasing its classification performances.

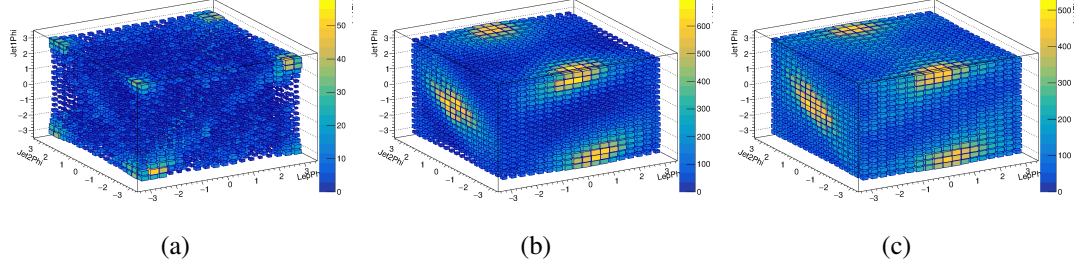


Figure 4.25: Correlation 3D plots between ϕ_{jet1} , ϕ_{jet2} and ϕ_{lep} for signal (a), $t\bar{t}$ (b) and W +jets (c) events. A color palette is used to show the number of entries in each 3D bin.

Another example is the correlation between the third-leading light-jet and the number of jets. In this particular case, for events with only four jets, and therefore without a third light-jet, the $jet3$ four-vector is set as a null vector, giving the peaks visible in Figure 4.17. This "non-physical" information is recognised by the classifier through the correlation between this variable and the feature of Figure 4.18. In Figure 4.26, the linear correlation between the p_T of the third-leading light-jet and the number of jets is visible.

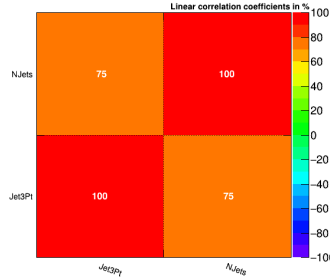


Figure 4.26: Linear correlation coefficients between the features describing the p_T of the third-leading light-jet and the number of jets.

These correlations are deeply exploited in neural network models during the update of the node's weights happening in the training epochs. The small discriminating power can be also increased by the network applying non-linear transformations of the input information. The importance of low level features was also checked by removing some of them from the input feature list, and finding a lower classification power of the DNN model. In Table 4.14 a breakdown study of the input features is reported. From the set of 36 features described above, some with poor visible discriminating power are removed. The relative test accuracy is evaluated and is reported in the table, showing that the configuration with 36 input features gives the larger test accuracy. In particular the use of the third leading jet information is useful, in combination with other features with small separation power, to improve the performances of the model. Moreover, Higgsness and Topness variables as well as m_T^{WW} use only two among the three leading

light-jets in their definition, thus having the information of the third leading light jet in the DNN training avoids this bias in the event reconstruction. All the 36 input features, then, are important for training the model and for obtaining good classification performances, which are discussed in the following Section.

Number of input features	Features removed	Test accuracy [%]
36	-	78.33
33	jet3 (p_T, η, ϕ)	77.12
32	jet3 (p_T, η, ϕ), m_{lep}	78.06
30	jet3 (p_T, η, ϕ), m_{lep} , ϕ_{bjet1} , ϕ_{bjet2}	78.10
28	jet3 (p_T, η, ϕ), m_{lep} , ϕ_{bjet1} , ϕ_{bjet2} , ϕ_{jet1} , ϕ_{jet2}	78.09
26	jet3 (p_T, η, ϕ), m_{lep} , ϕ_{bjet1} , ϕ_{bjet2} , ϕ_{jet1} , ϕ_{jet2} , ϕ_{lep} , ϕ_{MET}	77.87

Table 4.14: Test accuracy of DNN model with several sets of input features. The number of input features and the ones removed from the initial set of 36 variables is reported for each test.

5.2.2 DNN model performances

Training performances are evaluated by monitoring the evolution of loss, which is the summation of the errors made for each input example, and classification accuracy along the epochs. For this purpose, input samples are split in three orthogonal groups, training, validation and test samples, to evaluate the model on independent sets of data. In this analysis the relative fractions of events for each of the three groups have been chosen as follows: input sample is first split to create the test sample, selecting randomly 20% of the events; then, training and validation samples are created splitting the remaining events with a 80:20 ratio. The number of events in training and validation sample is determined by the signal input sample statistics, as it is the smaller one. In this way, the network model receives in input an equal amount of training, validation and test events for each signal and background sample. Spare background events are used for the final evaluation of the network after training phase. MC event weights are also provided to the network after their normalisation to 1, in order to keep the training independent from the cross section of the different processes. Classification accuracy is defined as the ratio of the number of correct predictions over the total number of predictions made. Training is stopped after an arbitrary number of epochs, or it can be stopped when there is no more improvement in the classification performance. This improvement is evaluated comparing the loss of training events to the one of validation events. After every epoch, the updated value of the loss is compared to the one of the previous epoch. When the validation loss reaches its lowest value, the training continues for a number of epochs, waiting for a new lower value of the validation loss to decide whether to stop or not. This procedure is followed to avoid the problem of the overtraining, since the training loss and accuracy continue their improvement epoch by epoch, increasing their gap from the validation values. The evolution of training and validation loss and accuracy are shown in Figure 4.27a and 4.27b respectively, where it is visible the agreement of the validation parameters with the training ones during the training epochs, and in particular at the epoch 23 in which our model reached the minimum validation loss value. The training was stopped after 10 more epochs without improvements in the validation loss.

At the end of the training phase, the weights extracted in the epoch with the lowest value of validation loss are kept for testing the model. Test phase consists in applying the network model built with the selected weights to the independent set of test events. At this stage, two new values of loss and classification accuracy are obtained, named

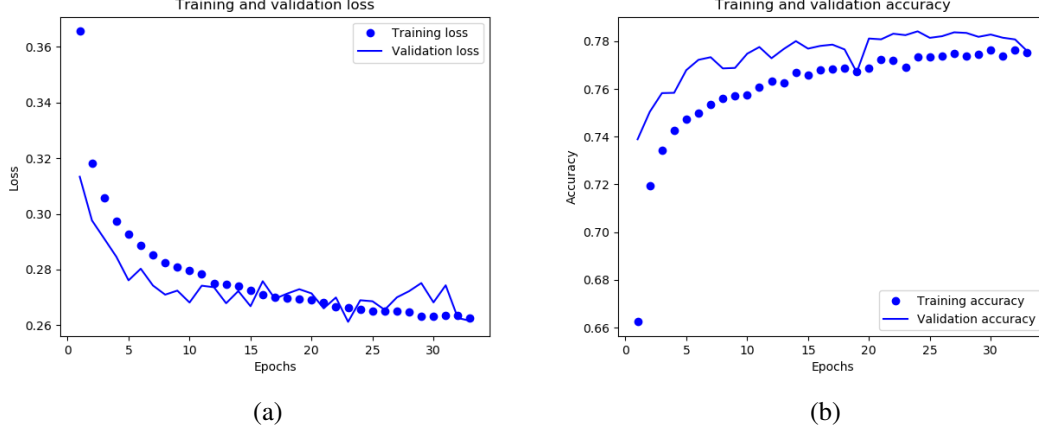


Figure 4.27: Training history of loss (a) and accuracy (b). The set of weights used in the final model is the one obtained in the 23th training epoch.

test loss and test accuracy, which define the performances of the neural network model. A summary of loss and accuracy values of the different phases, training, validation and test, of the analysis network model construction is reported in Table 4.15. Test loss has a value smaller than training and validation ones, confirming the goodness of training epoch chosen.

	Loss	Accuracy [%]
Training	0.2664	76.91
Validation	0.2613	78.25
Test	0.2575	78.33

Table 4.15: Summary of loss and accuracy values of the different phases of the analysis network model construction. Training and validation loss and accuracy at the epoch with the lowest validation loss. Test loss and accuracy obtained with the evaluation of the trained model on test sample.

A better visualisation of the classification goodness can be obtained with the use of a confusion matrix, shown in Table 4.16, with respect to classification accuracy. In such a matrix, each column represents the instances in a predicted class, while each row represents the instances in an actual class. The events in the intersection of a predicted and actual class are referred as true positive, or true negative if the true class is a background class, and represents the correctly classified event. Instances in the intersection of two different classes, are instead wrongly classified events and are referred as to false positive or false negative: false positive if a background events is predicted as signal, while false negative if a true signal event is predicted as a background-like one. In multi-class problems the construction of a confusion matrix needs some adjustment. In particular, the one in Table 4.16 is built assigning the events in the class with the larger predicted probability.

For our DNN model the true positive rate is $\sim 87\%$, while a fraction of $\sim 9\%$ of $t\bar{t}$ and $W + jets$ events are predicted as signal-like ones, but these values are only based on the most probable predicted class for each event.

The distributions obtained for the three class probabilities, visible in Figure 4.28, allow to better select regions of interest with an higher concentration of signal events and a smaller contribution of background processes.

As it is visible, the trained network model performs a correct classification of each type

	Predicted HH	Predicted $t\bar{t}$	Predicted $W + jets$
True HH	86.95	5.68	7.37
True $t\bar{t}$	9.02	70.78	20.20
True $W + jets$	9.51	13.29	77.20

Table 4.16: Confusion matrix obtained in the evaluation of the trained DNN model on test sample. For each row, the fraction, expressed in percentage, of events predicted in that class are split according to the true category of which they belong to.

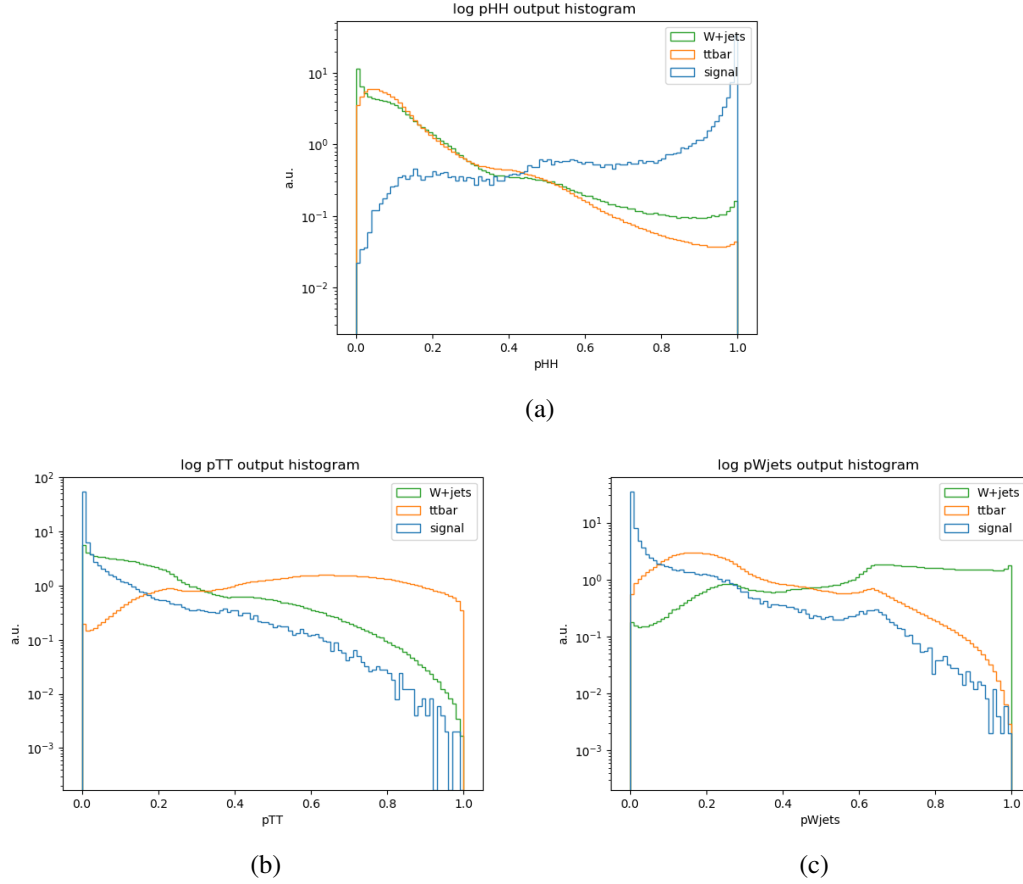


Figure 4.28: DNN model output probabilities p_{HH} (a), $p_{t\bar{t}}$ (b) and p_{W+jets} (c), produced in the evaluation of the DNN trained model for each test (and spare) event of the three input samples. Entries are normalised to have unit area for each sample distribution.

of event most of the times. Each class probability distributions shows in fact value mostly close to 1 for events of the relative class, and smaller values, shifted towards zero, for events of a different class.

Output probabilities can be combined to obtain a better discriminating variable, named d_{HH} , which is visible in Figure 4.29 and has the following definition:

$$d_{HH} = \log \left(\frac{p_{HH}}{p_{t\bar{t}} + p_{W+jets}} \right) = \log \left(\frac{p_{HH}}{1 - p_{HH}} \right) \quad (4.20)$$

This variable shows a greater discriminating power, with signal events mostly present in its positive range, while background events shows peaks at $d_{HH} \simeq -3$ and an exponential descent for larger values. The d_{HH} variable has been chosen for the definition of the analysis regions, and used for the final fit of the analysis. These two topics are discussed in Section 6.1 and 8.2 respectively.

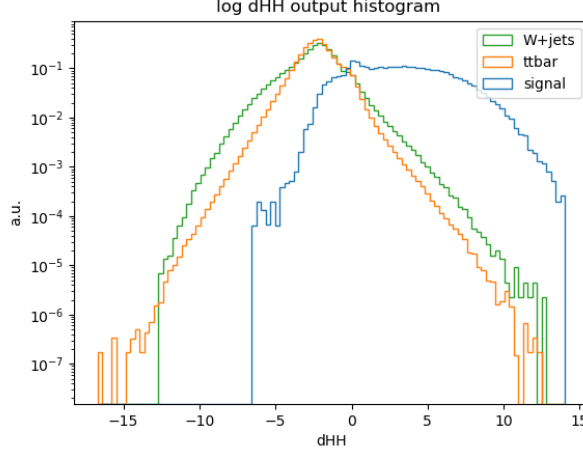


Figure 4.29: Shape of d_{HH} discriminating variable, obtained with the combination of DNN model output probabilities, on test (and spare) events. Entries are normalised to have unit area for each sample distribution.

6 Estimation of backgrounds

The background composition in the phase space defined by the event selection is made mainly by $t\bar{t}$, single-top and W +jets events, as visible in Table 4.11. In addition there are smaller contributions by Z +jets, diboson and single Higgs production processes. All background contributions with prompt lepton, i.e. a lepton originating from the main collision or as a direct product of the decay of the particles produced in the main collision, are estimated using MC simulation. The normalisation of $t\bar{t}$ and W +jets backgrounds, being the ones with the larger contribution after the event selection as already seen in Section 4.2.4, is then corrected by comparing data and predicted MC events in dedicated control regions, defined in Section 6.1. The normalisation procedure is described in Section 6.3, where these control regions are taken as a single bin histograms in the fit.

Processes with misidentified leptons, which consist of fakes and non-prompt leptons, are less accurately described in the MC and therefore a data-driven background estimation is needed. Fake electrons arise from two background sources. They are either mis-reconstructed jets, mostly originating from light quark or gluon fragmentation, or real electrons arising from secondary decays of light and heavy flavour mesons within jets. Fake muons might arise from in-flight decays of mesons inside jets, with the semi-leptonic decay of B meson as one of the largest contributions. Gluon fragmentation, hadronisation and meson decays are described in the simulation within the parton showering algorithm, which has a limited scope of validity. Furthermore, for an accurate description of the multijet background, which represents a large portion of the fake leptons, higher order terms in the perturbation series are necessary, which also makes it difficult to generate a reliable parton-level description. For these reasons a data-driven approach is adopted to describe these processes, and it is reported in Section 6.2.

6.1 Definition of the analysis regions

Events satisfying the requirements listed in Section 4.2 are split in several analysis regions. These regions are built dividing the phase space in areas with different event type population, by using powerful discriminating variables. A signal region (SR) is built defining a set of cuts that select the majority of the expected signal events, reducing

as much as possible the background contribution. This region is used then in the final fit for the extraction of the parameter of interest of the analysis. Besides SR, signal depleted control regions (CR) are defined inverting some cuts of the SR to focus on particular background sources, with the aim of checking the consistency of the SM background MC predictions with the observed data and of providing a normalisation of their contribution. These regions show a minimal presence of signal events to avoid the contribution of eventual selected signal events in the normalisation of background sources. In this work three control regions are defined to normalise the three most important background sources, here summarised and described:

- TopCR: control region with an high purity of $t\bar{t}$ events, used to normalise $t\bar{t}$ background, and defined by:

$$d_{HH} < -3 \quad \text{and} \quad p_{t\bar{t}} > 0.95 \quad (4.21)$$

The cut value $d_{HH} < -3$ was chosen in order to keep less than 1% of the signal in this and in the other two CRs, and to have a good statistics of $t\bar{t}$ and W +jets events. As it is visible from Figure 4.29 and 4.30a, in fact, signal events populate the d_{HH} range above -3, and the peaks of $t\bar{t}$ and W +jets are placed in the range $-4 < d_{HH} < -3$. The additional cut value on $p_{t\bar{t}}$ has been chosen looking at Figure 4.28b and 4.30b, in order to exploit the large purity of $t\bar{t}$ background events visible in the right-end edge of their range.

- WjetsCR: control region a big fraction of W + jets events, used to normalise W +jets background, and defined by:

$$d_{HH} < -3 \quad \text{and} \quad p_{W+jets} > 0.85 \quad (4.22)$$

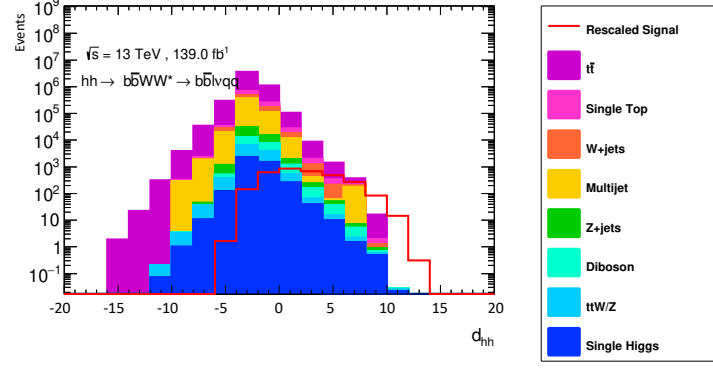
The choice of the cut value on p_{W+jets} follows the same reasoning done for $p_{t\bar{t}}$ cut, and have been chosen looking at the large W +jets purity in the right-end edge of the range of the distributions shown in Figure 4.28c and 4.30c.

- dHHCR: control region orthogonal to TopCR and WjetsCR, built for the normalisation of multijet events, and defined by:

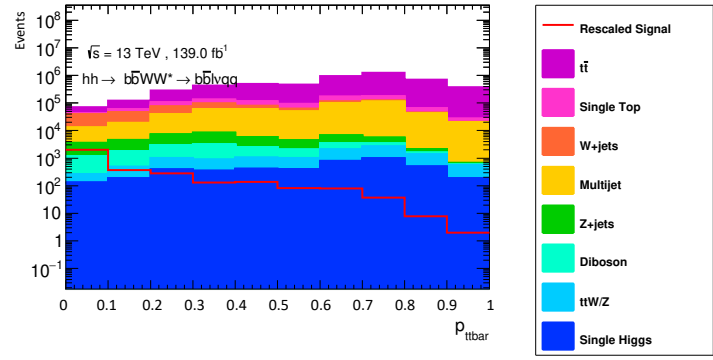
$$d_{HH} < -3 \quad \text{and} \quad p_{t\bar{t}} \leq 0.95 \quad \text{and} \quad p_{W+jets} \leq 0.85 \quad (4.23)$$

By construction, as $p_{t\bar{t}}$ and p_{W+jets} are defined as the probability of an event to be classified as $t\bar{t}$ - and W +jets-like, these cuts are selecting events which are tagged as $t\bar{t}$ and W +jets events at 95% and 85% probability, respectively, resulting in 98.6% of $t\bar{t}$ events in TopCR and 44.9% of W +jets events in WjetsCR, as visible in Table 4.17 which reports the background compositions in each analysis region.

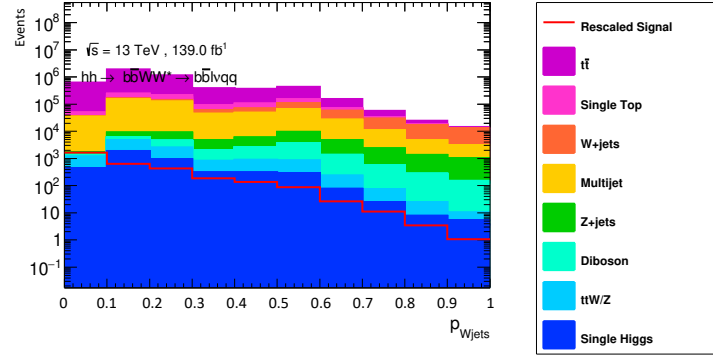
In these control regions, p_{W+jets} distribution is used to correct the observed differences between MC and data. A fit is performed leaving as a free parameter the normalisation factor of $t\bar{t}$ W -jets and multijet backgrounds, in order to adjust the MC prediction and to obtain a match with the fitted data, as will be described in detail in Section 6.3. In order to validate the MC background estimate obtained with the application of the normalisation factors, an orthogonal region, called validation region (VR) is defined. This region is required to have a negligible amount of signal events as it is not used in the final fit, but needs to be close to the signal region in order to confirm that the background estimation is valid also in the signal region phase space. The signal region, instead, is obtained by inverting the d_{HH} cut, selecting the range above $d_{HH} \geq -3$ where there are 99.4% of the selected signal events. An additional cut on the Topness ($\log(T)$) variable is used to define the validation region, removing only 1% of the signal from the signal region. The definition of these two regions is here summarised:



(a)



(b)



(c)

Figure 4.30: Stacked distributions of the DNN output variables, d_{HH} (a), $p_{t\bar{t}}$ (b) and p_{Wjets} (c). Signal events are scaled times 100 and superimposed to background events.

- dHHSR: signal region containing 98.4% of the expected signal events passing the event selection, and defined by:

$$d_{HH} \geq -3 \quad \text{and} \quad \log(T) > 10 \quad (4.24)$$

- dHHTopnessVR: validation region built for the validation of background modelling, containing less than 1% of signal events, and defined by:

$$d_{HH} \geq -3 \quad \text{and} \quad \log(T) \leq 10 \quad (4.25)$$

The expected event yields for each sample considered in the analysis is reported in Table 4.17 for the five analysis regions defined above.

Sample	TopCR	WjetsCR	dHHCR	dHHTopnessVR	dHHSR
ggF signal	0.002	0.05	0.2	0.4	39.7
$t\bar{t}$	168864.3	42444.7	1175221.3	106787.9	4268991.3
Single-top	1809.2	6684.6	47125.6	3543.8	287411.1
$t\bar{t}W/Z$	164.7	50.6	1159.9	98.5	6376.1
W +jets	86.1	55533.0	40589.3	1531.4	200744.5
Z +jets	23.6	5141.1	5072.6	279.7	29801.7
Diboson	6.1	689.3	1053.2	80.8	10867.9
$ggF H$	0.0	11.9	20.6	0.5	105.6
VBF H	0.0	2.3	3.4	0.2	26.2
WH	0.7	6.3	36.1	6.7	851.3
ZH	0.0	3.0	7.5	0.9	152.4
$t\bar{t}H$	60.7	17.3	492.8	52.3	3788.7
Multijet	179.4	13201.8	17320.2	1032.5	104601.8

Table 4.17: Expected event yields for each sample in analysis CRs, VR and SR. Multijet estimated yields are obtained with the procedure described in Section 6.2.

6.2 Multijet background estimate

The *fake-factor method* has been used for the estimation of multijet and non-prompt background contributions in this analysis. This method is a data-driven extrapolation technique for electron or muon fakes which are taken from control regions rich in this kind of background and with high statistical power. This method is also known as ABCD method [134], as it uses three orthogonal control regions (the B, C, and D regions) to estimate the background contribution in the A region, where the separation between these regions is granted by the application of orthogonal cuts. The A region is defined by the requirements of the standard event selection listed in Section 4.2, requiring exactly 2 b-tagged jets, two or more light-jets, exactly one SignalLepton and a single-lepton or a MET trigger that fires the event. Therefore, the A region contains all the five regions defined in Section 6.1, for which the multijet background is required to be estimated. In this analysis a control region rich in multijet events has been built by loosening some of the requirements: events triggered by single-lepton triggers, with only 1 b-tagged jets and exactly two leptons, with at least the LooseLepton definition, are selected and fill this control region called FakesCR. The number of b-tagged jets has been relaxed to increase the statistics of multijet background events in the control regions, while the use of dilepton final states follows a particular requirement that is described in the following of the Section. A comparison of the different requirements for the standard event selection of the A region, referred to as "tight" selection in the following, and the one that defines the FakesCR are summarised in Table 4.18

	Trigger	N_ℓ	$N(\text{jet})$	$N(\text{b-jet})$
Tight selection	Single lepton or MET	1 SignalLepton	≥ 4	2
FakesCR	Single lepton	2 at least LooseLeptons	≥ 4	1

Table 4.18: Comparison of tight selection and FakesCR selection requirements.

The four ABCD regions are then obtained with an additional requirement on these two

selections. Fake leptons are normally less isolated than prompt leptons, therefore the isolation requirement has been used to split FakesCR into B and D regions. In the same way, C region is obtained by changing the requirement on the lepton definition used in the tight selection. The two lepton definitions, LooseLepton and SignalLepton introduced in Section 4.1 and visible in Tables 4.3 and 4.4, are used here for this purpose: events with LooseLeptons, which do not pass the isolation criteria, are therefore possible candidates of being fake events and fall in regions C and D, while events with SignalLeptons fall in regions A and B. A schematic view of the definition of the four ABCD regions is given in Figure 4.31.

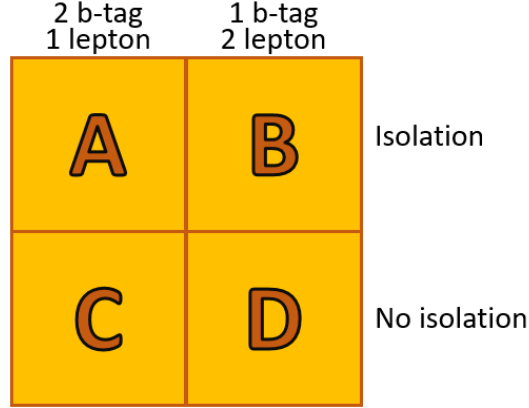


Figure 4.31: Schematic view of the ABCD regions definition.

Now, as the events in FakesCR contain two leptons, it is possible for the same event to have one lepton falling in region B and one in region D; this because the requirement is not an event property but rather a lepton property, which means that it is possible to fill lepton distributions in these regions but not jet distributions, for example. However, this is not a problem for the purpose of the fake-factor method used in this analysis as only lepton distributions are used. A weight, called the *fake-factor* (FF), is calculated for each lepton from dilepton regions B and D. FF is related to the misidentification probability for a fake lepton to satisfy the selection isolation requirement of a SignalLepton. The misidentification probability, sometimes referred to as *fake rate*, is defined as:

$$f = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}} = \frac{N_B}{N_B + N_D} \quad (4.26)$$

where N_{pass} and N_{fail} are the amounts of leptons satisfying and failing the isolation requirement, and therefore corresponding to leptons in regions B and D respectively. The FF is then defined as:

$$FF(p_T, \eta) = \frac{f}{1 - f} = \frac{N_{\text{pass}}}{N_{\text{fail}}} = \frac{N_B(p_T, \eta)}{N_D(p_T, \eta)} \quad (4.27)$$

FFs are measured in this analysis as a function of lepton's p_T and η to better describe the expected distribution of the multijet background. The yield of the expected multijet background in the A region is then given by the extrapolation from region C events, weighted by the FF , which is applied to each event according to the kinematic properties of the lepton:

$$N_A(p_T, \eta) = N_C(p_T, \eta) \times \frac{N_B(p_T, \eta)}{N_D(p_T, \eta)} = N_C(p_T, \eta) \times FF(p_T, \eta) \quad (4.28)$$

The lepton yields in each region N_i of Equations 4.26, 4.27 and 4.28, with $i \in [B, C, D]$, are calculated subtracting, bin by bin in lepton's p_T and η distributions, all MC backgrounds from observed data with the following equation:

$$N_i = N_i^{data} - N_i^{MC,bkg} \quad (4.29)$$

The calculation of Equation 4.28 is done in parallel for each of the analysis regions defined in Section 6.1. The region C, in fact, is split in the corresponding analysis regions containing exactly one LooseLepton, instead of a SignalLepton as in the case of region A. In this way, the multijet background estimated distributions are provided for each of the analysis regions.

Fake factor method often suffers from large systematic uncertainties due to trigger bias, if the triggers used for the measurement of the FFs do not match the triggers defining the signal region. For example, both low p_T 2016-2018 single electron trigger HLT_e26_lhtight_nod0_ivarloose and low p_T 2017-2018 single muon trigger HLT_mu26_ivarmedium contain an isolation requirement. Since FF is the ratio of the events with an isolated lepton by events for which the lepton fails isolation requirement, its value is significantly different for the region defined by one of these two triggers and for an isolation independent trigger. Moreover, due to isolation requirement in some of the single lepton triggers used in the analysis, visible in Table 4.6, different FFs have to be used for single lepton and MET triggered events. One possibility would be to define two regions triggered by single lepton and MET triggers respectively in order to derive FFs, but finding a fake enriched region triggered by MET is not an easy task, and the small statistics in such a region would result in large fluctuations of the estimate. Therefore, *tag and probe* method has been used inside the fake-factor method described above to overcome this particular issue.

Tag and probe method Tag and probe method is a special case of the fake-factor method, that uses a control sample with exactly two reconstructed leptons, and therefore exploits the events in B and D regions defined above, that have been constructed following this particular requirement of this approach. Each event of the control sample contains one *tag* and one *probe* lepton. Tag lepton is the trigger matched lepton, while probe lepton is trigger independent and can, therefore, be used to obtain trigger unbiased FFs. If both leptons fire the trigger, higher p_T lepton is denoted as tag. In FakesCR, events containing two leptons with same-flavour and opposite-sign have been vetoed in order to reduce the Drell-Yan contribution, and MET trigger has not been used to select the events, since the definition of the tag lepton would not be possible due to a missing trigger match of the lepton. Moreover, truth-reco object matching has been performed for MC events independently on each tag and probe lepton of regions B and D in order to discard fake leptons from the FF calculation. Two sets of FFs have been derived in order to obtain the estimation of multijet background in the phase space of this analysis: tag and probe FFs. Tag FFs are obtained from the tag lepton kinematic distributions and are used to estimate fakes in the single lepton region, while probe lepton FFs are used to estimate fakes in the MET region. Both sets of FFs have been calculated using Equation 4.27. Tag and probe lepton p_T distributions are shown in Figure 4.32 and Figure 4.33.

Differences between data and MC events can be seen as the contribution of the multijet background, whose estimation is the target of this method. As it is visible, this contribution is larger for low p_T values, where we expect to have the majority of fake leptons. Equation 4.29 can be described from these figures: new histograms are built calculating the difference between data and MC for each distribution in Figure 4.32 and

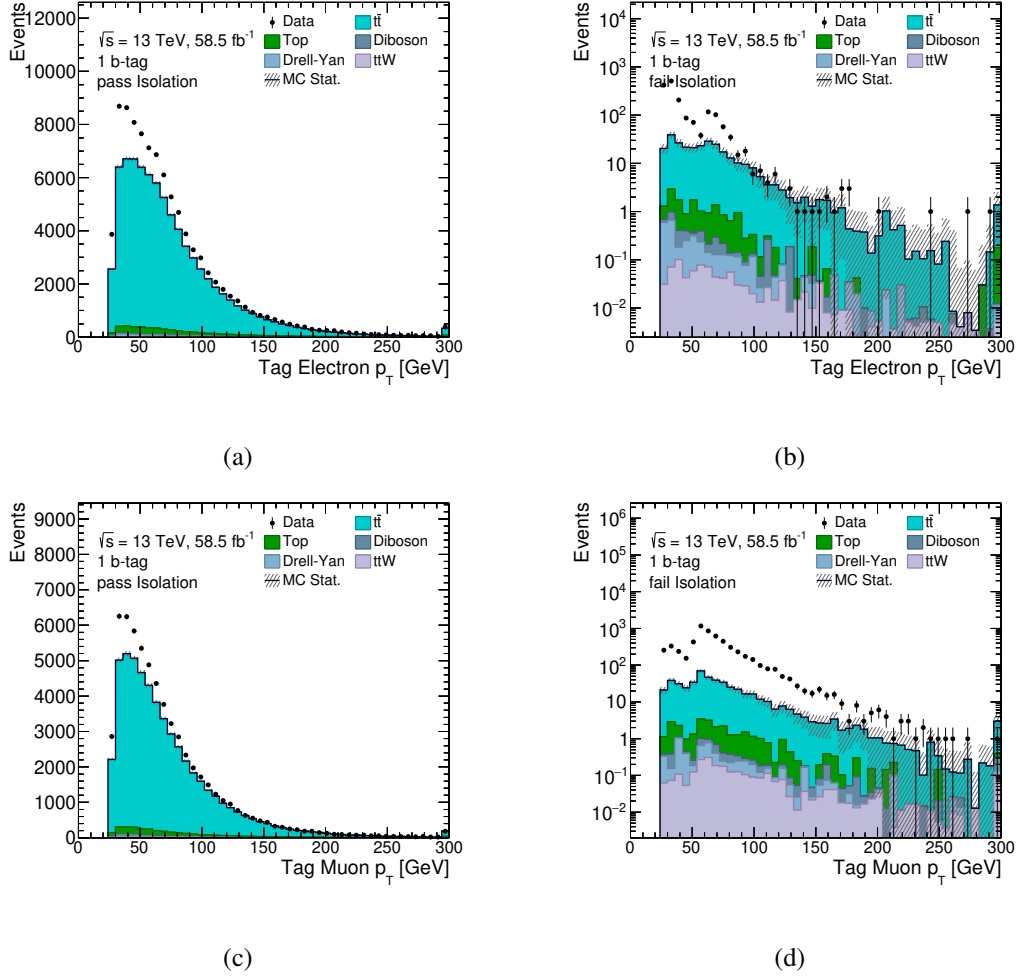


Figure 4.32: Distributions of the tag lepton's p_T from FakesCR. Region B (a) and region D (b) electron, and region B (c) and region D (d) muon are shown respectively. The cut at 25 GeV comes from the single lepton trigger threshold, see Section 4.2.1.

Figure 4.33. Then, FFs are evaluated dividing bin by bin these new histograms, applying Equation 4.27, and therefore dividing the region B distribution over the region D one.

Fake factors FFs are measured in the FakesCR, where the two leptons are classified as tag and probe, as described in the previous paragraph. As already described, FFs are evaluated from lepton kinematic distributions, in particular from p_T and η distributions of tag and probe leptons. Therefore, FFs are binned in p_T and η . The η region has been split into two bins, one for the barrel region and one for the end-cap. The p_T bins have been chosen in a way, that the statistical error is kept fairly constant along the whole range. An additional requirement is asked for the p_T binning: since single electron triggers have a different ID requirement when they pass the threshold of 60 GeV, as between HLT_e26_lhtight_nod0_ivarloose and HLT_e60_lhmedium_nod0 which have LHTight and LHMedium ID requirement, it is asked not to have bins crossing the 61 GeV value in order to keep the population of the two triggered events mostly separated, as visible in Figure 4.34. This threshold value is used since the trigger matched electrons are required to have 1 GeV above the trigger threshold, and therefore the HLT_e60_lhmedium_nod0 trigger can select only events having electrons with $p_T > 61$ GeV.

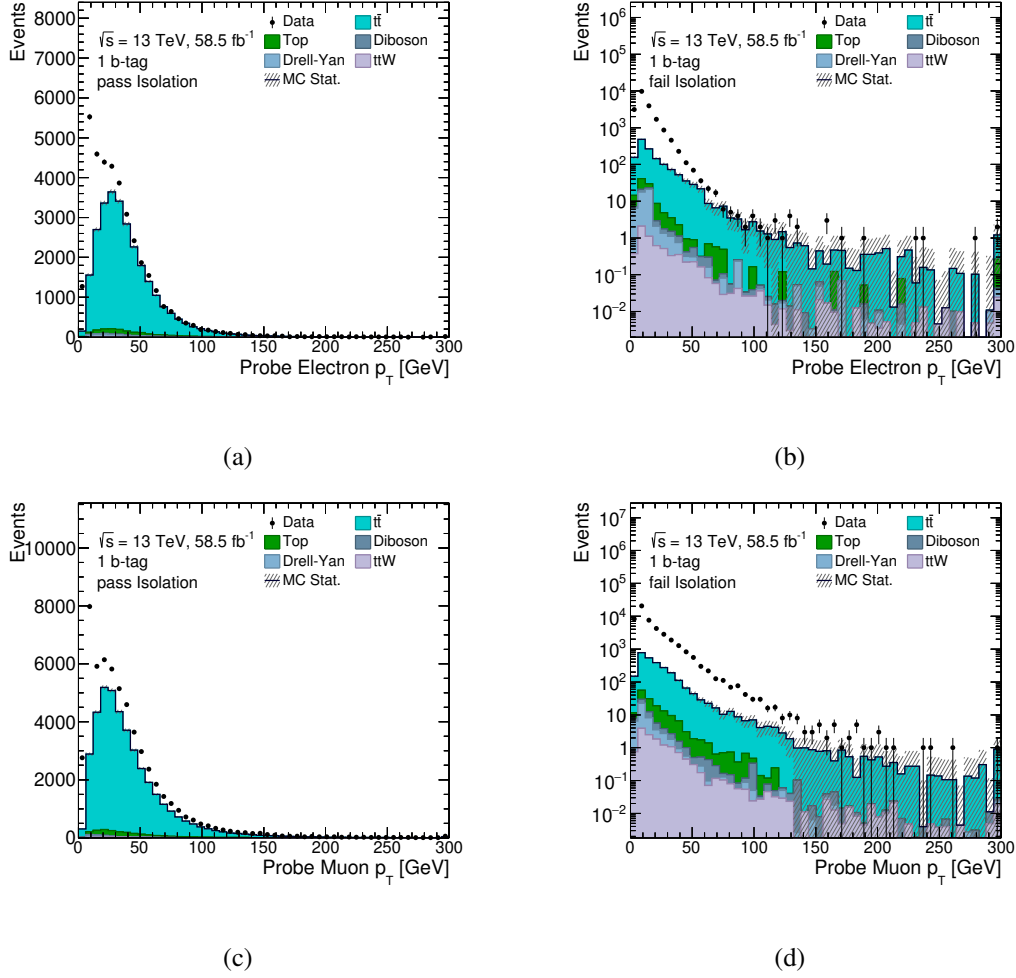


Figure 4.33: Distributions of the probe lepton's p_T from FakesCR. Region B (a) and region D (b) electron, and region B (c) and region D (d) muon are shown respectively.

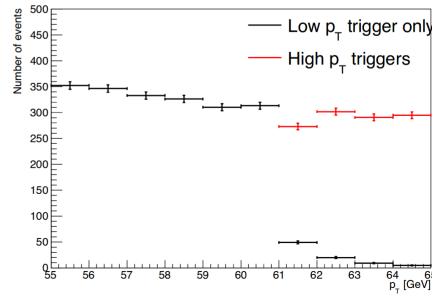


Figure 4.34: Transverse momentum of electron triggered $t\bar{t}$ MC events in dHHCR, showing the different population of triggered events below and above 61 GeV.

The obtained tag and probe FFs distributions are fully reported in Appendix B, and an example is given in Figure 4.35 for reference, showing the FFs distributions for tag and probe electron in the barrel region of the detector. As it is visible, tag electron FF distribution shows a step at 61 GeV due to transition region from low- p_T to high- p_T single-electron triggers, while probe lepton FFs do not show this feature being trigger independent. The larger values of tag FFs are due to the triggers as they follow the grow of the efficiency curve of the trigger algorithm. Muon FFs have instead smaller values as the fakes contribution is smaller for muon final states.

As already described, the multijet background estimate in the A region is obtained by

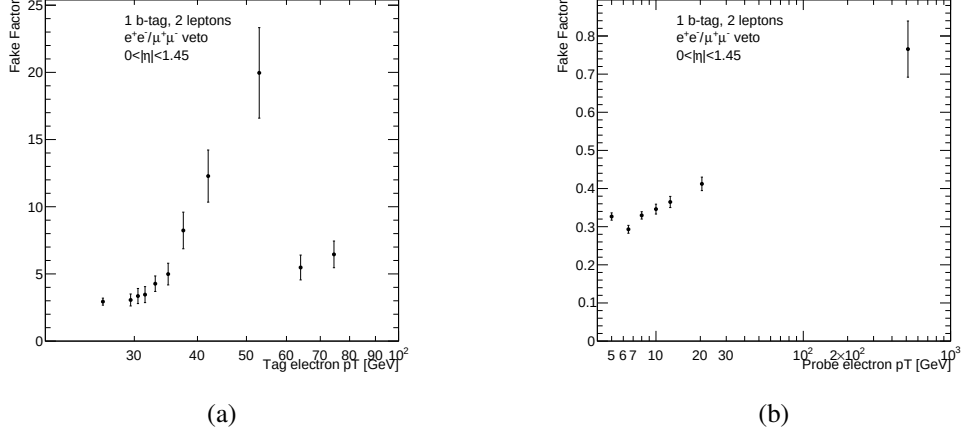


Figure 4.35: Fake factor p_T distributions for tag (a) and probe (b) electron in the barrel region of the detector.

scaling C region events following equation 4.28. Example distributions of ΔR^{bb} and $p_{W jets}$ in the C region. for events selected with dHHCR requirements, are reported in Figure 4.36.

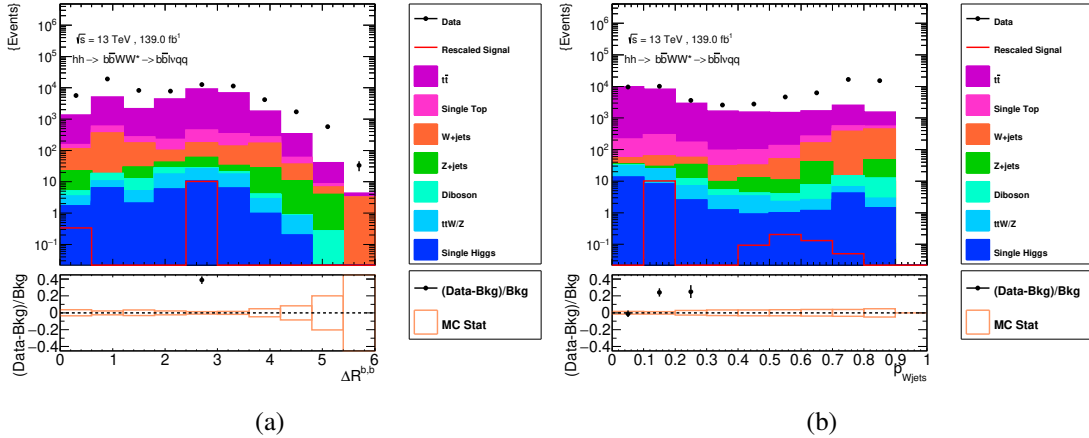


Figure 4.36: Comparison between C region data and MC events using ΔR^{bb} (a) and $p_{W jets}$ (b) distributions. The selection of dHHCR is applied on top of C region requirements. Multijet background is not present in the distributions as it extracted from the difference between data and MC events as described in the text.

The presence of multijet background events in this region is visible from the difference between data and MC distributions. The same distribution of this background is expected in the dHHCR A region. A visualisation of the result of this method can be seen by comparing Figure 4.36b with the same $p_{W jets}$ distribution shown in the following Section 6.3 in Figure 4.37 for dHHCR events, selected with the standard event selection and therefore relative to the A region.

6.3 Background normalisation

As described in Section 6.2, multijet background expected distributions are obtained using Equation 4.28 by subtracting MC background events from data and scaling this event yield in the C region by the fake factors evaluated from B and D regions using the tag and probe method. The normalisation of this estimate is then corrected by comparing data and predicted MC events in the dHHCR control region defined in Section 6.1.

In addition to the multijet background estimate, the MC prediction of the main backgrounds of the analysis, $t\bar{t}$ and W +jets, are also normalised to data in the dedicated TopCR and WjetsCR control regions. The normalisation factors are extracted performing a simultaneous fit of these three control regions to account for the correlations of the background contribution among the different regions. TopCR and WjetsCR are used with only one bin in order to reduce the impact of the control regions on the systematic nuisance parameter pulls and overconstraints (see Section 8.2). In dHHCR, the shape of p_{Wjets} distribution for $t\bar{t}$, W +jets and multijet backgrounds are different, as visible from Figure 4.30c, this feature can be exploited by a binned fit, using the model described in Section 8.2. The range of p_{Wjets} distribution has been optimised to provide only the relevant part of the distribution where background's shape difference is more pronounced, and removing the lower range of the distribution $0 < p_{Wjets} < 0.5$. Moreover, this range has been removed because there is a large presence of $t\bar{t}$ background, that may affect the purpose of the fit, which would give a major importance to correct data/MC in these bins with large statistics, giving less weight to the shape information in the relevant range $0.5 < p_{Wjets} < 0.85$.

An iterative procedure has been followed in the multijet estimation, which it is linked to the background normalisation described in this Section. MC events are subtracted from data, as described by Equation 4.29, therefore, MC predictions have an impact in the FFs calculation. The iterative procedure consists in doing a first estimation of the FFs, that is used to obtain the expected multijet background distributions in the region A, considering the background contribution as it comes from MC predictions. Then, the estimated multijet distribution is used together with data and predicted MC background distributions in a binned Likelihood fit that exploits only the information of the control regions to extract normalisation factors for $t\bar{t}$, W +jets and multijet backgrounds, as described above. After this first iteration, $t\bar{t}$ and W +jets normalisation factors are used to scale MC predictions in a new round of the tag and probe method, in order to extract new updated values of the fake factors, and, therefore, to obtain a corrected prediction of the multijet background contribution in the analysis regions. Systematic uncertainties are not considered in this CR-only fit. This choice was made to derive normalisation factors taking into account only the statistical error of the background distributions. Systematic uncertainties, instead, will be derived from the global fit described in Section 8.2, which includes the dHHSR together with all the analysis control regions, performing a simultaneous fit of the different distributions.

After the iteration procedure, a new CR-only fit is performed to extract the final values of the normalisation factors of the three backgrounds, which are reported in Table 4.19.

Background process	Normalisation factor
$t\bar{t}$	0.946 ± 0.002
W +jets	1.17 ± 0.08
Multijet	1.30 ± 0.12

Table 4.19: Normalisation factors for $t\bar{t}$, W +jets and multijet background processes extracted from the CR-only fit on p_{Wjets} distribution.

Post-fit distribution for p_{Wjets} variable in dHHCR is shown in Figure 4.37. The agreement between data and MC predictions is good in the full range of the distribution. A small discrepancy of $\sim 5\%$ is present in the first bin of the distribution, not used in the fit, where there is almost only $t\bar{t}$ background. This means that the fit is probably overestimating $t\bar{t}$ background to better describe the data in the rest of the distribution.

This is actually what the fit is doing, not having the possibility at this stage of using systematic uncertainties to adjust the result.

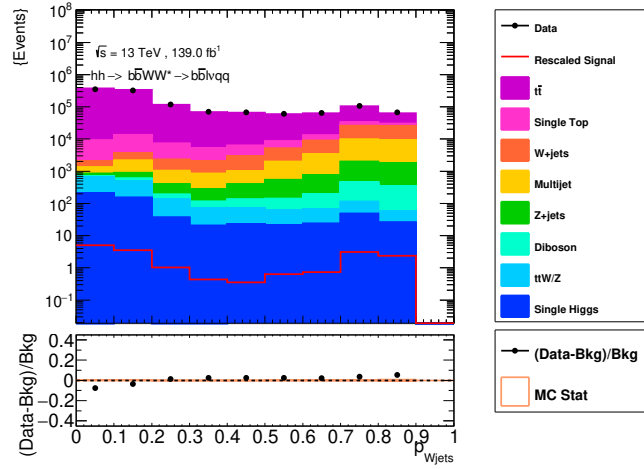


Figure 4.37: Post-fit distribution of p_{Wjets} distribution in dHHCR. All background components are scaled to their normalisations as determined in the fit.

The goodness of this normalisation procedure can be assessed by looking at validation distributions. They are obtained by applying the normalisation factors in Table 4.19 to the MC predicted events of the corresponding process. Here are given some of these validation distributions, taken from different control and validation regions, for some interesting variable among the ones used in the DNN training. In particular ΔR^{bb} is one of the most interesting, having a defined localisation of the multijet background for its larger values. The agreement between data and scaled MC background predictions is visible in the distributions shown in Figure 4.38, validating in this way the CRs-only fit procedure and its results. The agreement is good in all the range of each distribution, with some small discrepancy in bins with lower statistics. It needs to be remarked that at this stage no systematic uncertainty is considered, and therefore some small discrepancy may be recovered at a later stage of the analysis when accounting for detector and modelling systematic uncertainties.

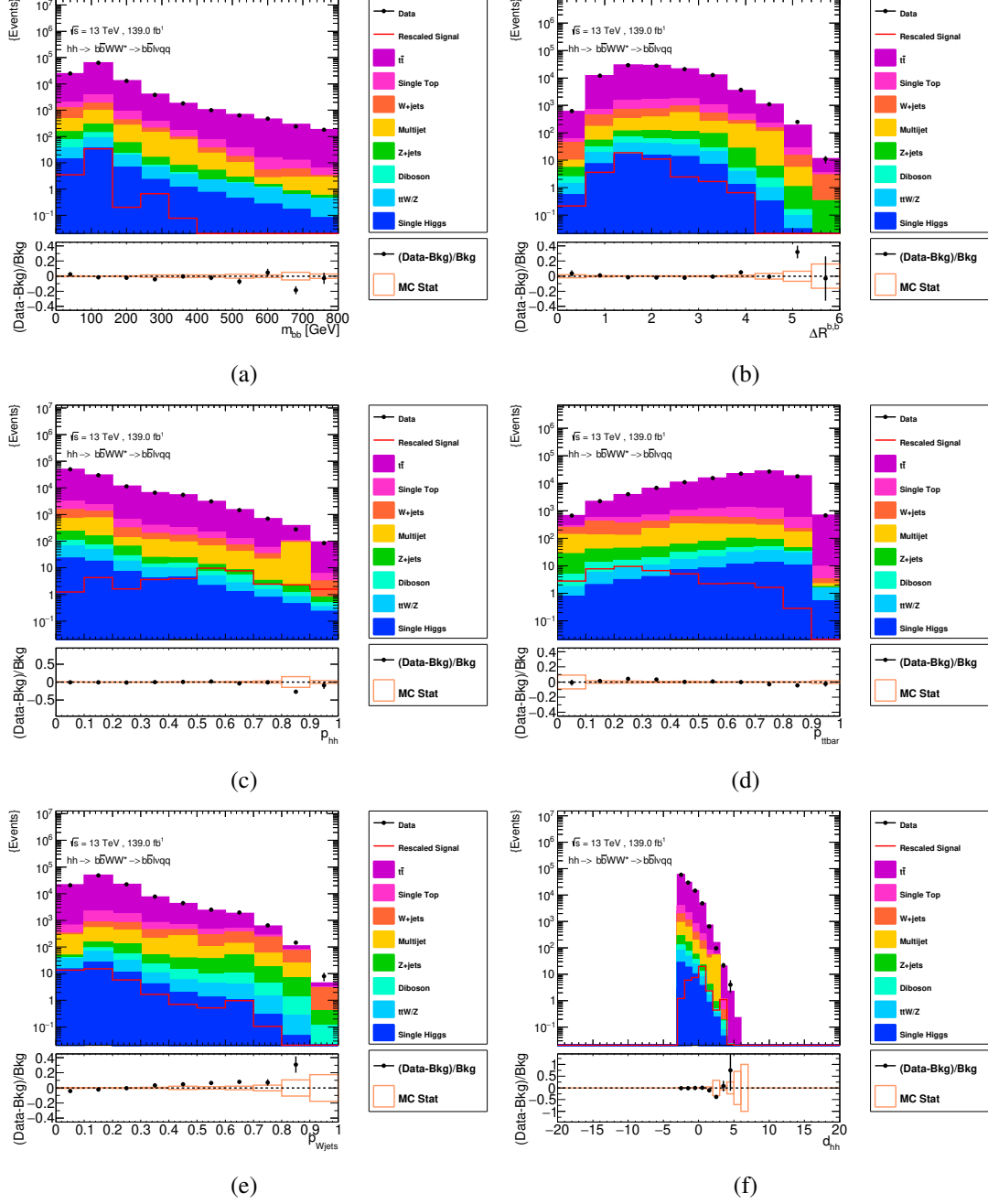


Figure 4.38: Post-fit distribution of m_{bb} (a), ΔR^{bb} (b), p_{HH} (c), $p_{t\bar{t}}$ (d), p_{Wjets} (e) and d_{HH} (f) distribution in dHHTopnessVR. All background components are scaled to their normalisations as determined in the CRs-only fit. The SM $HH \rightarrow bbWW \rightarrow bblvqq$ signal distribution is also shown with a normalisation scaled to 100 times the SM expectation.

7 Systematic uncertainties

All the uncertainties that are introduced by the analysis techniques and assumptions are referred to as systematic uncertainties. Depending on the source of such uncertainties, three main groups are usually defined: experimental uncertainties, theoretical uncertainties and modelling uncertainties. The first ones are related to the detector response and to the object reconstruction and identification, while the last ones are related to the MC simulated background and signal processes modelling, and by consequence also to data-driven background estimates. Theoretical uncertainties, instead, are related to the assumptions made on the cross-section and branching ratio of the processes, used in the normalisation of the MC sample yields. Finally, modelling uncertainties are related to the settings, and their possible variations, used in the production of the simulated physics process sample. The effect of these uncertainties is evaluated either with two-sided variations or with one-side variations of the parameter to which they are related. Two-sided systematics are evaluated by considering the up- and down- variations of the parameter under study by its uncertainty. For modelling uncertainties the systematic is then assumed to be symmetric and the larger difference between the two variations and the nominal is taken as the systematic uncertainty, while for experimental systematics up and down variations are used normally. One-sided systematics, instead, are evaluated by doing a single variation and using the full difference to the nominal value as the uncertainty, which is then symmetrised around the nominal value. All these uncertainties must be propagated through the analysis to evaluate their contribution to the error on the parameter of interest of the final fit.

Experimental uncertainties are discussed in Section 7.1, while theoretical and modelling uncertainties are discussed in Section 7.2. All these systematic uncertainties are included in the statistical analysis fit, which is described in Section 8.2.

7.1 Experimental uncertainties

Experimental uncertainties are common for all the ATLAS analyses, since they depend on characteristics and calibrations of the detector, on data taking conditions and on the object reconstruction and identification techniques. ATLAS Collaboration provides a set of recommendations for their treatment in physics analyses, which have been followed in this work.

The final state selected in this analysis contains muons, electrons, MET and jets. Reconstructed objects are corrected to take into account the differences between MC simulations and data in identification, isolation and trigger selection. This is done by applying scale factors derived from the comparison of these distributions between MC and special data samples very clean for the purpose of determining the systematic under study; for example $Z \rightarrow e^+e^-$ and $\mu^+\mu^-$ events are used to determine electron and muon scale factors. All the selection made on these objects are, therefore, affected by the uncertainties of the applied corrections. Moreover, corrections related to energy scales and resolutions are applied by smearing or re-scaling the objects energies.

A list of all the experimental uncertainties included in the analysis is summarised in Table 4.20, and reported in details in Appendix D. These uncertainties can contribute in different ways: applying an overall normalisation on the distribution, or both applying a normalisation and modifying the shape of the MC predicted distribution. All the systematic uncertainties listed in Table 4.20 are treated as two-sided systematics, with the exception of MET related systematics, and enter as nuisance parameters in the fit that is described in Section 8.

Systematic uncertainty	Type	NPs
Luminosity	N	1
Pile-up	N	1
Electron scale and resolution	NS	3
Electron ID, reconstruction, isolation and trigger efficiency	N	4
Muon scale, ID and reconstruction	NS	5
Muon reconstruction, isolation, track-to-vertex association efficiency	N	8
Tau energy scale	NS	5
Tau ID and reconstruction efficiency	N	14
Tau electron-veto efficiency	N	3
Jet Energy Scale (JES)	NS	33
Jet Energy Resolution (JER)	NS	14
Jet Vertex tagging efficiency	N	2
Jet Flavor tagging efficiency	N	13
Track reconstruction	NS	5
Track efficiency	N	2
MET energy scale	NS	2
MET resolution	NS	2

Table 4.20: List of experimental systematic uncertainties accounted in the analysis. Type is labelled "N" when the systematic is taken as a normalisation, or "NS" when it is taken both as normalisation and shape. The number of nuisance parameters that enter in the fit described in Section 8 is also shown in the table for each systematic group.

A brief description of the sources of experimental systematics is reported in the following.

Luminosity and pile-up systematics The size and shape of the interaction region provide the value of the instantaneous luminosity and its uncertainty. Uncertainties in the measurements, then, affect the calculated value of the total integrated luminosity. The luminosity scale is calibrated using dedicated Van Der Meer scans in each year, which consist in recording the relative interaction rates as a function of x - y beam-separation. The use of complementary measurements from several luminosity-sensitive detectors, provide then the extrapolation to the physics regime. The uncertainty on the integrated luminosity recorded by the ATLAS experiment for the full Run 2 dataset, used in this analysis, is 1.7%. The extrapolation of the calibration from the low-luminosity Van Der Meer scans to the high-luminosity physics conditions brings the largest contribution to the uncertainty [38]. This uncertainty is applied to all the samples whose normalisations are derived from MC simulation. Furthermore, the pile-up reweighting procedure brings an additional uncertainty, as large as the rescaling, that is considered as an overall normalisation correction to MC simulated samples.

Trigger systematics Trigger efficiency scale factors are calculated as a function of the p_T of the triggering object in order to correct the difference in trigger efficiency between simulation and data. Uncertainties on the trigger scale factors (SFs) are propagated as systematic source that may affect the application of the trigger selection. For MET triggers, there are no ATLAS centrally provided SFs, since their value depends on the phase-space of each analysis. For this reason, scale factors have been evaluated for this analysis, comparing the E_T^{miss} distribution of MET triggered data events to the ones

of MC simulated events. Trigger active in the same periods have been specifically combined to evaluate the SFs on the application of the logic OR of these algorithms, as it is done in the event selection procedure. A summary of these SFs is reported in Appendix C.

Electron systematics Uncertainties related to electrons arise from uncertainties in reconstruction, identification and isolation efficiencies, coming from the discrepancies between MC events and observed data distributions. A correction is applied with scale factors derived from $Z \rightarrow ee$ decays [54], and their uncertainties are treated as an overall normalisation variation of the algorithm efficiency. Moreover, uncertainties related to electron and photon energy calibration and momentum scale are included, which, instead, contribute also to a smearing of the p_T and $|\eta|$ distributions of the selected electrons.

Muon systematics Muon systematic uncertainties arise from several sources. Reconstruction in the Inner Detector and Muon Spectrometer, identification and isolation efficiencies contribute to these uncertainties as well as the momentum scale and resolution of the sagitta, which is used in the measurement of the muon p_T . Scale factors are applied to ID and isolation efficiencies to correct the MC predictions to the observed data using $J/\Psi \rightarrow \mu\mu$ and $Z \rightarrow \mu\mu$ decays [58]. Their uncertainties are propagated through the analysis selection and they are treated as a normalisation variation of the ID or isolation algorithms. Momentum scale and sagitta resolution uncertainties appear, instead, as a smearing of the transverse momentum resolution, determining a variation of the shape of the p_T spectrum of the selected muons.

Tau systematics Even if τ leptons are not selected in the analysis, a veto is applied when hadronic- τ s are reconstructed and identified with the requirement described in Section 4.1. Variations of the number of vetoed τ s result in an higher or lower number of selected events, and, therefore, systematic uncertainties related to hadronic- τ s are taken into account in this analysis, although their contribution might be small. Hadronic- τ reconstruction and identification efficiency is corrected in MC using scale factors derived as a function of the $\tau_{had}^{vis} p_T$ to account for differences between simulation and data. Uncertainties on the efficiency scale factors are included as normalisation systematic uncertainties. The τ_{had}^{vis} energy scale is calibrated after reconstruction, and the residual difference between simulation and data and the uncertainties on the energy scale corrections are included as systematic uncertainties. Dedicated uncertainties on the τ -electron overlap removal are also included. All these uncertainties have been used in the $bb\tau\tau$ analysis [77] and in its replication targeting full Run 2 data, and therefore they have been taken into account here, although their contribution is expected to be small.

Jet systematics As described in Chapter 2, jet energies need to be calibrated after reconstruction. Uncertainties arise from the energy scale corrections (JES), which depends on p_T and $|\eta|$ of the jet, and jet energy resolution (JER), which considers the differences in energy resolution between simulations and data [62]. These uncertainties depends also on the jet flavour (BJES, Flavour Response and Composition), i.e. on the flavor of the quark or the gluon that initiated the jets, which is of particular interest in this analysis as the requirements on b-tagged and not b-tagged jets are important cuts of the event selection of this work. All these jet systematics have been evaluated in this

analysis accounting for different nuisance parameters for each source, with the addition of the uncertainty on the JVT cut, which depends on the calibration of the tagger efficiency.

b-tagging systematics Scale factors are applied in MC events to account for flavour-tagging algorithm efficiency differences between simulation and data [64]. They are derived separately for b , c , and light-flavour jets as a function of the jet p_T and η . The different sources of uncertainties of these correction factors are decomposed into uncorrelated components, resulting in four uncertainties for b -jets, five uncertainties for c -jets, and four uncertainties for light-flavour jets. B-tagging efficiency uncertainties are one of the main systematic uncertainty sources that affect this analysis, as the requirement on the number of b-tagged jets in the event is one of the stringent cut applied in the event selection. A variation in the efficiency of the b-tagging algorithm results in different number of selected events and in a possible rejection of signal events, decreasing the sensitivity of the analysis.

MET systematics Systematic uncertainties related to all the objects described above, which are used to calculate the E_T^{miss} , are propagated to its calculation. Additional systematic uncertainties are taken into account for the scale and the resolution of the E_T^{miss} soft term, in order to account for differences between simulations and data using $Z \rightarrow \mu\mu$ final states without jets [65]. Scale uncertainties correspond to scaling the soft term magnitude up or down in the direction longitudinal to the vector sum of the hard objects (p_T^{Hard}). Resolution uncertainties correspond to a smearing of the soft term magnitude in the direction parallel or perpendicular to the direction of p_T^{Hard} . These four components are treated as one-sided variations, following the prescription of ATLAS collaboration.

7.2 Theoretical and modelling uncertainties

Theoretical and modelling uncertainties are related to our knowledge of Standard Model and eventual BSM signal processes. This knowledge is propagated in the MC simulation of such processes, and the deeper it is the bigger is the consistency of the MC predictions to the observed data. The sources of these uncertainties depend on the settings used in the theoretical cross section calculation and in the MC simulations.

Theoretical uncertainties on the cross section calculation only affect the normalisation of the MC samples, and therefore are analysis-independent. They arise from the different assumptions and choices made on PDF sets, factorisation and renormalisation scales, α_s as well as on the mass of the top-quark for some particular process, like the HH production cross section which is determined mainly by top-quark loop gluon-gluon fusion process.

A summary of the theoretical systematic uncertainties considered in this thesis work is given in Table 4.21. Uncertainties are related to the cross sections used for MC sample normalisation, and in particular, as described above, to the parameters used in their theoretical calculations.

Modelling uncertainties, instead, takes into account the use of different matrix element generator, parton shower model and tuning parameters in the MC simulation of SM processes, as well as variations of the PDF set used, variations on the factorisation and renormalisation scale parameters and ISR and FSR tuning in the simulation. The use of different generators results both in a different normalisation and in modified shapes of kinematic distribution of the simulated samples. Therefore, the selection

Process	Systematic	+1 σ	-1 σ
ggF signal	XS_SCALE	2.2 %	5.0 %
	XS_PDF	2.1 %	2.1 %
	XS_ α_s	2.1 %	2.1 %
	XS_ m_t	2.6 %	2.6 %
$t\bar{t}$	XS_PDF	4.2 %	4.2 %
	XS_SCALE	2.4 %	3.5 %
	XS_ m_t	2.8 %	2.7 %
Single-top s-channel	XS_PDF	2.6 %	2.6 %
	XS_SCALE	2.8 %	2.3 %
	XS_ m_t	2.2 %	2.1 %
Single-top t-channel	XS_PDF	2.8 %	2.8 %
	XS_SCALE	3.1 %	2.1 %
	XS_ m_t	0.8 %	0.8 %
Single-top Wt	XS_PDF	2.5 %	2.5 %
	XS_SCALE	4.7 %	4.7 %
$t\bar{t}W$	XS_PDF	2.0 %	2.0 %
	XS_SCALE	12.9 %	11.5 %
	XS_ m_t	2.7 %	2.7 %
$t\bar{t}Z$	XS_PDF	2.8 %	2.8 %
	XS_SCALE	9.6 %	11.3 %
	XS_ m_t	9.8 %	9.8 %
W +jets	XS_PDF	2.6 %	2.7 %
	XS_SCALE	1.3 %	1.3 %
	XS_ α_s	1.4 %	1.4 %
Z +jets	XS_PDF	2.6 %	2.6 %
	XS_SCALE	1.6 %	1.6 %
	XS_ α_s	1.6 %	1.6 %
Diboson	XS_TOTAL	7.5 %	5.8 %
ggH	XS_SCALE	3.9 %	3.9 %
	XS_PDF	1.9 %	1.9 %
	XS_ α_s	2.6 %	2.6 %
VBF H	XS_SCALE	0.4 %	0.3 %
	XS_PDF	2.1 %	2.1 %
	XS_ α_s	0.5 %	0.5 %
WH	XS_SCALE	0.5 %	0.7 %
	XS_PDF	1.7 %	1.7 %
	XS_ α_s	0.9 %	0.9 %
ZH	XS_SCALE	3.8 %	3.1 %
	XS_PDF	1.3 %	1.3 %
	XS_ α_s	0.9 %	0.9 %
$t\bar{t}H$	XS_SCALE	5.8 %	9.2 %
	XS_PDF	3.0 %	3.0 %
	XS_ α_s	2.0 %	2.0 %
All Higgs samples	BR $H \rightarrow b\bar{b}$	1.2 %	1.3 %
(including signal sample)	BR $H \rightarrow WW$	1.6 %	1.5 %

Table 4.21: Summary of theoretical systematics applied in the analysis fit, divided per sample. XS stands for cross-section uncertainties, where the source of the uncertainty is specified: scale (factorisation and renormalisation scales), PDF, α_s , m_t . One σ up and down uncertainties are reported for each systematic in percentage.

applied in the analysis may have a different acceptance on each process, and, for this reason, this effect is analysis-dependant. For larger background processes, like $t\bar{t}$ and W +jets in this analysis, acceptance systematics may have an important impact on the total uncertainty, while for sub-leading backgrounds the impact is much smaller and

usually is not evaluated in the analyses. Acceptance uncertainties are derived by comparing the predictions obtained with the nominal settings of the MC simulation used in the analysis with alternative samples generated with different settings, as described above. The contribution of this systematic source is usually divided into a normalisation acceptance uncertainty, which affects the number of predicted events in the analysis signal region, and a shape uncertainty, which affects the shape of the final discriminant distribution.

7.2.1 $t\bar{t}$ modelling

Modelling uncertainties related to $t\bar{t}$ background are estimated by considering different sources: the initial-state (ISR) and final-state (FSR) gluon emission modelling, the dependence on the choice of the PDF set and the dependence on the renormalisation and factorisation scales used in the simulation of the process. Initial-state and final-state radiation systematic uncertainties are computed by varying the generator parameters from their nominal values to increase or decrease the amount of radiation, acting in particular on α_s . PDF uncertainties are computed using the eigenvectors of the PDF4LHC15 combined PDF set [135], which are made by the envelope of the variations of different PDF sets, CT14, MMHT14 and NNPDF3.0. Scale uncertainties, due to missing higher-order corrections, are computed by independently scaling the renormalisation (μ_R) and factorisation (μ_F) scales in the Powheg+Pythia8 $t\bar{t}$ sample, by a factor of 0.5 and 2, using six combinations of varied weights: ($\mu_F = 1$, $\mu_R = 0.5$), ($\mu_F = 1$, $\mu_R = 2$), ($\mu_F = 0.5$, $\mu_R = 1$), ($\mu_F = 2$, $\mu_R = 1$), ($\mu_F = 0.5$, $\mu_R = 0.5$) and ($\mu_F = 2$, $\mu_R = 2$), where the nominal have both the μ factors equal to one and the two extreme combinations, ($\mu_F = 0.5$, $\mu_R = 2$) and ($\mu_F = 2$, $\mu_R = 0.5$), are not considered.

The estimation of the effect of these modelling systematic sources is done by changing the weights of MC events, which results in modifications of the normalisation and shape of the background distribution with respect to the nominal configuration of the simulation. The analysis selection is then performed normally on these varied samples, evaluating at the end of the selection process, the difference between the varied distributions and the nominal one. PDF systematic contribution is evaluated separately from all the other sources considered, which are referred to as scale systematics. For scale systematics, the variation with the larger difference with the nominal distribution is taken and it is used as a conservative estimation of the systematic variation. This evaluation is done in dHHSR and then it is applied in all the analysis regions. As it is visible from Figure 4.39, the variation relative to the FSR scale, $FSR \mu_R fac = 0.5$, is the one which shows the larger discrepancy from the nominal distribution. This variation is therefore used as the $t\bar{t}$ modelling scale variation in the statistical analysis.

PDF uncertainties are instead evaluated following the prescriptions given in Ref. [135], by using the following master formula to evaluate the PDF uncertainty:

$$\delta^{PDF} \sigma = \sqrt{\sum_{k=1}^N (\sigma^{(k)} - \sigma^{(0)})^2} \quad (4.30)$$

where the sum is made on N PDF eigenvectors, which are 30 for PDF4LHC15, taking the square of the difference between the varied cross-section $\sigma^{(k)}$ to the nominal value $\sigma^{(0)}$. This evaluation is done in each analysis region, obtaining the relative varied distribution for $t\bar{t}$ sample.

For both PDF and scale uncertainties, shape and normalisation components are evaluated separately. The shape information is removed from the analysis control regions in order to obtain the normalisation factors without pulling the systematics, as will

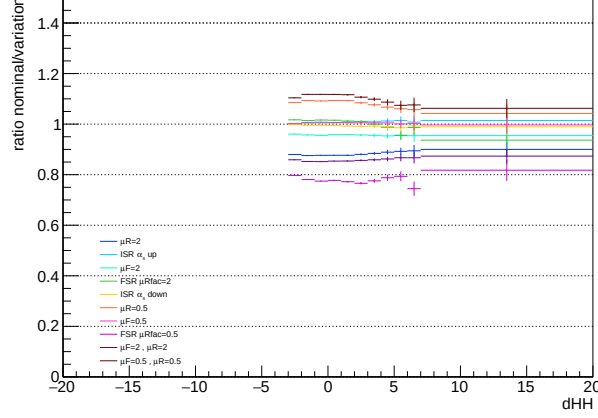


Figure 4.39: Ratio plot between nominal and varied $t\bar{t}$ samples of d_{HH} distribution in dHHSR. Varied samples are built, as described in the text, by using different set of event weights, which are reported in the legend. Bins from $d_{HH} = 7$ are merged to avoid statistical fluctuations in the evaluation of the larger variation.

be described in Section 8.2. Normalisation component is instead considered for each analysis region separately, resulting in the so called acceptance uncertainty.

7.2.2 W +jets modelling

Modelling uncertainties related to W +jets background are estimated by considering scale and PDF variations. To evaluate scale variations, 7-point scale variations stored as internal weights in the MC samples are used. The analysis is run for each of these points applying a reweight to the nominal W +jets simulated sample. Then, the envelope of the differences of the variations around the nominal is taken to obtain the systematic scale uncertainty from the maximum shift of the envelope with respect to the nominal. This envelope is given by the variation with the larger difference with the nominal distribution, as done for $t\bar{t}$ modelling in Section 7.2.1. As it is visible from Figure 4.40, the variation of both factorisation and normalisation scales to 0.5, $\mu F = 0.5$ and $\mu R = 0.5$, is the one which shows the larger discrepancy from the nominal distribution. This variation is therefore used as the W +jets modelling scale variation in the statistical analysis.

The normalisation component of the W +jets scale uncertainty is large in all the analysis regions and overestimates this systematic, as shown in Table 4.22, which summarises the normalisation components of both PDF and scale modelling systematic uncertainties for $t\bar{t}$ and W +jets in all the analysis regions used in the fit described in Section 8.2. It has been observed that this large value was in conflict with the W +jets normalisation factor, creating issues in the fit. As W +jets normalisation is expected to be mainly taken from WjetsCR, the histograms with varied W +jets scale have been rescaled to remove the normalisation component in WjetsCR, and for consistency this rescaling of 35.2% have been performed in all the regions, removing the issue in the fit.

PDF uncertainties are evaluated using the internal weights of the samples, which provide $N = 100$ NNPDF replicas, computing the standard deviation of the PDF eigenset, according to the NNPDF recipe [136], to obtain the combined PDF uncertainty. This evaluation is done in each analysis region, obtaining the relative varied distribution for W +jets sample.

For both PDF and scale uncertainties, shape and normalisation components are evaluated separately. As done for $t\bar{t}$ modelling systematics, the shape information is removed

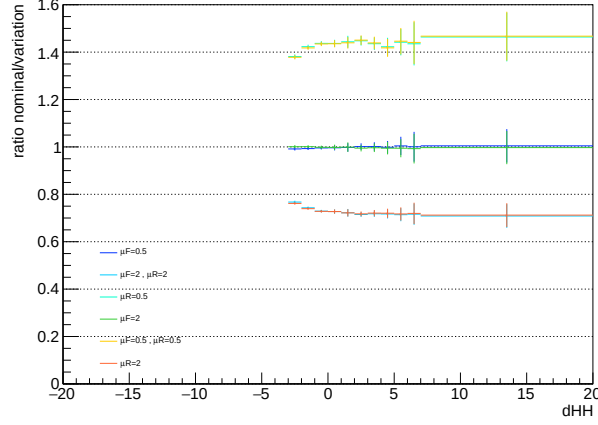


Figure 4.40: Ratio plot between nominal and varied W +jets samples of d_{HH} distribution in dHHSR. Varied samples are built, as described in the text, by using different set of event weights, which are reported in the legend. Bins from $d_{HH} = 7$ are merged to avoid statistical fluctuations in the evaluation of the larger variation.

Systematic	dHHSR	dHHCR	TopCR	WjetsCR	StopCR
MODEL_ttbar_PDF	$\pm 15.1\%$	$\pm 11.5\%$	$\pm 13.8\%$	$\pm 10.8\%$	$\pm 12.6\%$
MODEL_ttbar_SCALE	$\mp 22.8\%$	$\mp 18.4\%$	$\mp 20.4\%$	$\mp 18.4\%$	$\mp 21.3\%$
MODEL_W_PDF	$\pm 1.16\%$	$\pm 1.14\%$	$\pm 1.07\%$	$\pm 1.02\%$	$\pm 0.91\%$
MODEL_W_SCALE	$\pm 44.5\%$	$\pm 33.3\%$	$\pm 44.4\%$	$\pm 35.2\%$	$\pm 40.7\%$

Table 4.22: Normalisation components of PDF and scale modelling systematic uncertainties for $t\bar{t}$ and W +jets samples in all the analysis regions used in the fit described in Section 8.2.

from the analysis control regions, as will be described in Section 8.2, and the normalisation component is considered for each analysis region separately.

8 Statistical analysis

This analysis work is a search for a process predicted by the Standard Model of particle physics, but not yet observed in data. The search for a new process in data analysis is performed by the use of statistical hypothesis test. In order to discover a new signal process, the so called null hypothesis H_0 is defined, which describes all the known processes which constitute the background in the analysis. Then, the level of agreement of the observed data with this null hypothesis is tested, and if the observed data are not compatible with the null hypothesis a new observation can be claimed. If this is not the case, an exclusion limit can be set on some parameter of the new process, like the cross section. In this case, the incompatibility of the background plus signal hypothesis H_1 with the data is tested. The presence of a new signal can be then excluded if there is no compatibility between the observed data and the background plus signal hypothesis.

The results of the hypothesis test can be quantified by computing a p-value, that gives the probability, if the hypothesis H_0 is true, of finding data compatible or greater with respect to the predictions of H_0 . The p-value p_0 is defined by:

$$p_0 = \int_{t^{obs}}^{\infty} f(t|H_0) dt \quad (4.31)$$

where t^{obs} is the observed value in data of the test statistics t and $f(t|H_0)$ is the probabil-

ity distribution function (PDF) of the test statistics t for the given hypothesis H_0 ($t|H_0$). The hypothesis is discarded if the p-value is below a certain threshold, and therefore if the incompatibility $1 - p_0$ is higher than a certain value.

P-value is usually converted into an equivalent significance Z , defined such that a Gaussian distributed variable with Z standard deviations above its mean has an upper-tail probability equal to p :

$$Z = \Phi^{-1}(1 - p) \quad (4.32)$$

where Φ^{-1} is the inverse of the cumulative of the Gaussian distribution. The thresholds used to claim a discovery or an exclusion limit are arbitrary. In high-energy physics there is the convention of using a threshold of $Z = 5$, also referred as to 5σ , for discovery claims, which corresponds to a probability $p = 2.87 \times 10^{-7}$ that the background-only hypothesis is correct. For exclusion purposes, a threshold $p = 0.05$ is used on the p-value evaluated on an alternative hypothesis H_1 , which is known also as the 95% confidence level (CL) and corresponds to $Z = 1.96$. Exclusion limits are usually upper limits on the cross section of the new signal process predicted by H_1 , meaning that such a process, if it exists, is produced with a cross section below this limit quantity at the 95% confidence level.

Hypothesis tests can be performed on simulated events to check the expected sensitivity of the analysis during its development, keeping the analysis blinded, hiding the observed data in the signal region, to avoid biases in the selection. Only at the end of the development of the analysis the signal region is unblinded giving an observed sensitivity evaluated with the hypothesis test on the real observed data.

8.1 Statistical model

The model used in the statistical analysis of this work is based on a binned profile Likelihood ratio used as a test statistic [137] to calculate p-values. The Likelihood function for a counting experiment, as the one performed in this thesis work, is defined as the product of the Poisson probabilities for all the bins of the histogram of the variable on which we want to perform the fit:

$$L(\underline{n}|\mu, \theta) = \prod_{i=1}^N \frac{(\mu s_i + b_i)^{n_i} e^{-(\mu s_i + b_i)}}{n_i!} \prod_{j=1}^K \frac{u_j^{m_j} e^{-u_j}}{m_j!} \quad (4.33)$$

where n_i is the observed number of events in the bin i , while s_i and b_i are the expected number of signal and SM expected background events respectively. The number of expected signal events is multiplied by a quantity μ called signal strength, having the following properties: $\mu = 1$ corresponds to the theory expectation, while $\mu = 0$ corresponds to no effect of the tested theory; any other value corresponds to a different rate for the theory. Therefore, if σ_{th} is the cross section of the searched theory and σ the actual observed cross-section, the signal strength is define by:

$$\mu = \frac{\sigma}{\sigma_{th}} \quad (4.34)$$

In addition to the parameter of interest μ , the signal and background models also contain nuisance parameters θ (NPs), given by the systematic uncertainties, on which $s_i(\theta)$ and $b_i(\theta)$ depends. In Equation 4.33 they enter also as K additional bins related to the control regions of the analysis, with contents m_j and expected values $u_j(\mu, \theta)$ which depend both on nuisance parameters and on the signal-strength. NPs describe systematic uncertainties that can affect the normalisation of the samples, the shape of the

distributions used in the fit or both normalisation and shape. Their values are not taken as known a priori but rather must be fitted (profiled) from the data. The addition of control regions in the fit, in particular, can be used to constrain the nuisance parameters in such a way to reduce their uncertainty, and accordingly to reduce their impact on the final result. However, this constraint should be avoided as the systematic model is not designed on the analysis and can modify the effect of the systematic nuisance parameters in the analysis signal region.

Starting from Equation 4.33, the profile Likelihood ratio can be then defined as:

$$q_\mu = -2 \ln \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})} \quad (4.35)$$

where, for simplicity, we now consider L a function of the parameters, and where new symbols have been introduced: $\hat{\mu}$ and $\hat{\theta}$ are the best values of the parameters obtained by maximizing L , while $\hat{\hat{\theta}}$ are the values of the nuisance parameters obtained by maximizing L at a fixed value of μ . The profile likelihood ratio is a function of μ and shows the behavior of the likelihood for different possible values of the parameter of interest. Higher values of q_μ correspond to increasing incompatibility between the data and the tested signal strength. The test statistics q_μ can be therefore used to perform the hypothesis test, computing the p-value:

$$p_\mu = \int_{q_\mu^{obs}}^{\infty} f(q_\mu | \mu) dq_\mu \quad (4.36)$$

At this point, discovery test or exclusion test can be performed.

Discovery As already mentioned, in order to falsify a null hypothesis H_0 we need to test the background-only hypothesis. This can be done by using the test statistics of Equation 4.35, in the case when $\mu = 0$:

$$q_0 = -2 \ln \frac{L(0, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})} \quad (4.37)$$

and defining its observed p-value:

$$p_0 = \int_{q_0^{obs}}^{\infty} f(q_0 | 0) dq_0 \quad (4.38)$$

If p_0 is below the defined limit of $p = 2.87 \times 10^{-7}$, the background-only hypothesis is falsified as the observed data are incompatible with the H_0 prediction. This corresponds to a 5σ significance and therefore it is possible to claim a discovery.

Exclusion The profile likelihood ratio can be used also for the exclusion of a given theory. In this case Equation 4.35 has $\mu = 1$, and it is defined as:

$$q_1 = -2 \ln \frac{L(1, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})} \quad (4.39)$$

Following this equation, lower is q_1 , more data are compatible with the new theory under consideration, and less data are compatible with the pure background expectations. The probability density function (PDF) of q_1 can be evaluated starting from MC samples. In particular, two PDFs can be generated by using MC samples with $\mu = 1$ for the signal+background hypothesis, or with $\mu = 0$ for the background-only hypothesis,

which are called $f(q_1|1)$ and $f(q_1|0)$ respectively. The separation between $f(q_1|1)$ and $f(q_1|0)$ defines the capability to discriminate the two hypotheses. The expected sensitivity of the analysis, i.e. the expected exclusion limit of the signal hypothesis, can then be assessed by exploiting these probability density functions, using simulated samples. This is an important step in the construction of the analysis itself, as it provides indications on the the goodness of the selections applied in the analysis. The expected sensitivity may indeed change by varying the event selection or the cuts applied on the physics objects, and therefore constitutes an important parameter in the design of the analysis.

The median CL with which the signal, with $\mu = 1$, is excluded in case of a background-only experiment can be calculated as:

$$CL_{s+b}^{exp} = \int_{\tilde{q}_1}^{\infty} f(q_1|1) dq_1 \quad (4.40)$$

where $\tilde{q}_1$ is defined as the median of $f(q_1|0)$, and constitutes a sort of average outcome for a background-only experiment. The smaller is the CL_{s+b}^{exp} , the higher is the capability of the experiment to exclude the signal. The observed confidence level is instead evaluated from the collected data, by calculating the relative q_1^{obs} , and it is defined as:

$$CL_{s+b}^{obs} = \int_{q_1^{obs}}^{\infty} f(q_1|1) dq_1 \quad (4.41)$$

Following the convention, for values of CL_{s+b}^{obs} smaller then 5%, the signal is excluded at 95% CL. The CL_{s+b} , however, shows some problems when an under-fluctuation of the background occurs in the experiment. In this case, it may verifies that the CL_{s+b}^{exp} is high, while the CL_{s+b}^{obs} is small in such a way that the signal hypothesis is rejected.

Due to this behaviour, a different, corrected, method has been proposed and it is now widely employed in particle physics analyses for the exclusion of new physics signals. This method is the so-called modified frequentist approach, also known as CL_s method [83], and exploits both the information given by the test of the signal+background hypothesis and the test of the background-only hypothesis. CL_s is defined as:

$$CL_s = \frac{CL_{s+b}}{CL_b} \quad (4.42)$$

where the term CL_b is defined as:

$$CL_b^{obs} = \int_{q_1^{obs}}^{\infty} f(q_1|0) dq_1 \quad (4.43)$$

Given this definition, CL_s is not affected by background under-fluctuation, as in this case CL_{s+b}^{obs} is small, but CL_b^{obs} is small too. Therefore, we are excluding the signal+background hypothesis and the background-only hypothesis either, resulting in a more conservative method, that makes more difficult the exclusion of a possible signal. If the CL_s is lower than 5% the observed data can exclude the signal with $\mu = 1$ at 95% CL. In physics searches, like this analysis work, exclusion limit are set on the signal strength μ . By scanning the values of the signal strength in the test statistics q_μ , it is possible to determine the value μ^* for which the CL_s is equal to 5%. This value μ^* is the upper limit on the signal strength, meaning that a signal with $\mu \geq \mu^*$ can be excluded at 95% CL, and thus, a signal with a cross-section larger than μ^* times the theory prediction can be excluded at 95% CL.

8.2 Fit model for the $bbWW^*$ analysis

The statistical model built for the analysis done in this work is a binned profile Likelihood ratio fit on the DNN output distributions. Both dHHSR and the three control regions, TopCR, WjetsCR and dHHCR, are taken into account in the fit, which is performed on d_{HH} distribution in dHHSR and on p_{Wjets} distribution in dHHCR, while using the total yields in TopCR and WjetsCR. The choice of fitting p_{Wjets} in dHHCR has already been discussed in Section 6.3, and it is driven by the different shape of the distribution for the main background processes that can be exploited, with the one-bin TopCR and WjetsCR, to obtain the normalisation factors of $t\bar{t}$, W +jets and multijet backgrounds. The binning of d_{HH} in dHHSR has been chosen to have a balance between the number of bins, the expected signal over background ratio in each bin and to keep under control the MC background relative uncertainty. In order to do this, a constant bin width of $d_{HH} = 1$ has been set. Then, the last bins from $d_{HH} = 13$ up to the end of the d_{HH} range have been merged to have a maximum MC relative error of 40%, and to increase the sensitivity of the last bins, as visible from Table 4.23 which shows the expected signal and background yields in each dHHSR bin as well as the corresponding s/b ratio. The values in Table 4.23 are also plotted in Figure 4.41.

d_{HH} range	Exp. signal	Exp. $t\bar{t}$	Exp. W +jets	Exp. single-top	Other Bkg	s/b
[-3,-2]	1.41071	1808622.00	124779.60	125927.38	91091.02	0.000001
[-2,-1]	4.19250	1238184.00	93130.82	96750.82	77344.36	0.000003
[-1,0]	5.39337	541434.60	31938.40	41713.04	28070.96	0.000008
[0,1]	7.90754	189251.10	15659.92	16268.18	12224.80	0.000034
[1,2]	5.93910	40920.21	5243.46	3942.24	2924.59	0.000112
[2,3]	4.36937	14166.18	2260.86	1612.68	1140.68	0.000228
[3,4]	3.66784	5626.36	1094.52	698.07	667.50	0.000454
[4,5]	2.64120	2338.99	521.65	302.65	387.78	0.000744
[5,6]	1.82129	951.96	221.41	126.00	115.84	0.001287
[6,7]	0.99242	371.83	105.42	45.54	106.21	0.001579
[7,8]	0.69476	145.46	42.29	16.86	14.18	0.003175
[8,9]	0.36001	53.51	13.85	5.02	6.28	0.004577
[9,10]	0.18766	15.97	4.86	1.80	1.18	0.007883
[10,11]	0.08760	6.98	1.14	0.24	1.63	0.008771
[11,12]	0.02732	1.74	1.30	0	0.39	0.007962
[12,13]	0.00901	0.47	0.37	0	0.26	0.008210
[13,14]	0.00373	0	0.01	0.13	0.01	0.025556
[14,15]	0.00058	0	0	0	0	-
[15,20]	0	0	0	0	0	-

Table 4.23: Expected signal and background yields in each dHHSR bin, expressed in d_{HH} values. Background yield is shown separately for $t\bar{t}$, W +jets and single-top, while the remaining expected background yield is shown in "Other Bkg" column. MC predictions of $t\bar{t}$, W +jets and multijet backgrounds have been used with their normalisation applied. The corresponding s/b ratio is also shown for each d_{HH} range.

The good performances of the DNN are visible in Table 4.23 and Figure 4.41, where it can be seen that DNN classification allows to reduce background, and in particular the huge $t\bar{t}$ background, from millions of events up to few hundreds in the sensible range of d_{HH} distribution. This result is a confirmation of the optimisation of the network model and of its high classification power that can push the sensitivity of this analysis.

MC background events are expected only up to $d_{HH} = 14$, while a very small fraction of signal events is expected at $14 < d_{HH} < 15$. By merging together these two last

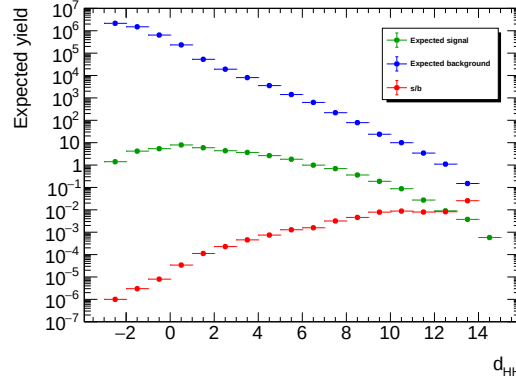


Figure 4.41: Expected signal (green) and background (blue) yields in dHHSR as a function of d_{HH} . The corresponding s/b ratio (red) is also shown for each d_{HH} value. Background yield has been calculated by summing MC predictions, with normalised $t\bar{t}$, W +jets and multijet contributions.

bins, with the remaining range up to $d_{HH} = 20$ that is expected to be empty, the signal over background ratio increases from $s/b = 0.025556$ to $s/b = 0.029513$, providing the bin with the larger sensitivity and determining this binning choice. In addition to the merging of the last bins, the first four bins have been removed from the fitted distribution, removing the range $d_{HH} \in [-3, 1]$. The choice of this range reduction has been made since the number of expected background events is larger than 2×10^5 , having therefore a very small statistical error, with a relative value smaller than 0.2%, that may overconstrain the nuisance parameters. Moreover, the expected signal yield in these four bins is large, about 19 events, but, as visible from Table 4.23, the s/b ratio is very small, less than 0.2% of its maximum value, and therefore this cut is not expected to decrease the sensitivity of the analysis.

All sources of systematic uncertainties described in Section 7 are considered as nuisance parameters (NP) in the profile likelihood. A conservative approach has been used in the treatment of the shape systematic variations. Their effect has been turned off from dHHCR to prevent constraints of the systematics from $p_{W \text{ jets}}$ shape information. In this way, shape systematic variations are fitted only from the signal region, where we want to quantify their effect on the signal strength, being the sensible phase-space region of the analysis. Moreover, this approach provides a better determination of background normalisation factors, which are extracted by the control regions without being affected by systematic shape variations that could modify their values and, as a consequence, affect the result in the signal region. In order to do this, the integral of up and down variations of each sample is calculated and used to scale the nominal sample distribution, leaving only the normalisation component of the systematic uncertainty.

Moreover, to reduce the complexity of the fit, nuisance parameters that have a negligible effect, quantified in a variation with respect to the nominal smaller than 1%, are removed.

The configuration of the fit procedure exposed in this Section has been defined after several preliminary fits done with the use of so called "fake-data" in the signal region. These data are built by summing MC background predictions with normalisation applied for $t\bar{t}$, W +jets and multijet contributions. Fake-data are built in the background-only hypothesis, thus not accounting for the expected signal in the signal region. In this way, we are able to perform a global fit on all the control and signal regions to adjust the settings of the fit, while we keep the blinding of the sensible region in the same time. During these preliminary fits, we have also tried to add a normalisation factor for

the single-top background, having a contribution in the signal region comparable to the one of W +jets and multijet background. Its normalisation was not investigated before since its contribution was assumed to scale similarly to the dominant $t\bar{t}$ background, with the same normalisation factor. However, with this additional normalisation factor, the fit resulted to be more stable, as single-top background events were found to be expected in the bins with the larger s/b ratio of Table 4.23. At this point, the same procedure done for the other normalisation factors has been applied for the single-top. As its normalisation is determined by the shape information of the dHHSR, it has been decided to split this region in two parts: one made by the most sensible bins in the range $d_{HH} \in [6, 20]$, and one made by the remaining bins, having s/b ratio smaller than 5% of s/b maximum value, i.e. the one obtained by merging the last bins. This resulted in having an 8-bins signal region, and a new control region, with a large fraction of the total expected signal, but with small sensitivity, covering the range $d_{HH} \in [1, 6]$. Single-top normalisation factor is expected to be determined by the shape information in this 5-bins control region, which have a good statistics of background events. For this reason the same procedure done in dHHCR of turning off the shape systematic variation have been applied in this new region, which is called in the following "StopCR". Finally, shape component of modelling uncertainties, described in Section 7.2.1 and 7.2.2, has not been considered in the dHHSR as it was observed to be subject to large statistical fluctuations of the MC samples, whose uncertainty is already accounted in the fit. The only modelling systematic used with the shape information in the dHHSR is the W +jets PDF uncertainty, because of to the small value of its variation.

At this point we have prepared all the necessary to perform the global fit on data selected by our analysis, and to unblind the result in the signal region.

8.3 Background-only fit

The analysis done in this thesis work is a search for a SM process not yet observed in data, therefore, a test of the background-only hypothesis is needed to potentially claim a discovery. This test has been done by checking the compatibility of the observed data with the SM background-only hypothesis performing a background-only profile likelihood fit. The fit model as well as the analysis region, distributions and binning used in the fit are the ones described in Section 8.2. Floating parameters of the fit were the four normalisation factors, of $t\bar{t}$, W +jets, multijet and single-top, as well as the nuisance parameters coming from the systematic sources described in Section 7. All these parameters have been evaluated by the fit. The resulting normalisation factors are summarised in Table 4.24.

Background process	Normalisation factor
$t\bar{t}$	0.99 ± 0.02
W +jets	1.65 ± 0.16
Multijet	0.98 ± 0.24
Single-top	1.57 ± 0.19

Table 4.24: Normalisation factors for $t\bar{t}$, W +jets, multijet and single-top background processes extracted from the global background-only fit.

As it is visible, the values and the errors of the normalisation factors are different from the ones obtained in Section 6.3 as an effect of the correlation with the NPs included in the fit. Moreover, single top normalisation factor differs from the $t\bar{t}$ one, confirming that a separate normalisation was needed.

Fit results have been investigated in order to validate the goodness of the fit. Post-fit data over MC prediction agreement has been checked by looking at the overall predicted yields in control and signal regions, which are reported in Table 4.25. As visible from Table 4.25, observed data in dHHSR are compatible to background-only expectation, and show a small statistical over-fluctuation with respect to MC predictions. The agreement between data and MC is good in all the control regions used in the fit, as it can be seen from the overall yields and in post-fit distributions.

Process	TopCR	WjetsCR	dHHCR	StopCR	dHHSR
$t\bar{t}$	144930 ± 530	41160 ± 870	202600 ± 2200	62850 ± 920	582 ± 23
W +jets	134 ± 12	81700 ± 3600	50500 ± 2800	9900 ± 480	190 ± 15
Single-top	2620 ± 320	10000 ± 1100	28100 ± 2900	9800 ± 1100	106 ± 13
Multijet	0	10800 ± 360	14500 ± 3500	1930 ± 560	47.6 ± 0.61
Z +jets	19.7 ± 1.7	4640 ± 380	3370 ± 350	579 ± 47	10.27 ± 0.79
Diboson	6.40 ± 0.84	662 ± 62	750 ± 73	561 ± 49	13.2 ± 1.1
$t\bar{t}W/Z$	160 ± 28	52.1 ± 9.0	187 ± 32	190 ± 33	2.98 ± 0.53
Single Higgs	61.6 ± 5.5	42.8 ± 2.6	120.2 ± 7.9	336 ± 21	11.08 ± 0.61
Total Background	147930 ± 390	149110 ± 390	300150 ± 550	86140 ± 280	915 ± 24
Expected signal	0	0.05 ± 0.01	0.058 ± 0.022	17.5 ± 1.1	2.33 ± 0.14
Data	147899	149079	300224	86097	943

Table 4.25: Post-fit expected number of background events, determined from a background-only fit, compared to the observed number of data events in the analysis regions used in the fit. The expected number of signal events is also reported. Systematic and MC statistics uncertainties are included.

Post-fit distributions of the four control regions used in the fit, reported in Figure 4.42, have been used to check the agreement between data and MC prediction in all the range of the fitted distributions. In particular, in dHHCR and StopCR the data/MC agreement is good in each bin of the distributions used in the fit, as it is visible in Figure 4.42c and 4.42d, validating the goodness of the fit.

Post-fit distributions were also checked for some DNN input variables of particular interest in both control validation and signal regions, as well as DNN output distributions in the validation region. In Figure 4.43 ΔR^{bb} distribution is shown for dHHCR, dHHTopnessVR and dHHSR, used in their original definition of Section 6.1 without applying the additional cuts used in the fit. This distribution, with backgrounds weighted by the normalisation factors of Table 4.24, shows a good data/MC agreement in all the three analysis regions. DNN output distributions and m_{bb} variable are also shown in Figure 4.44 for dHHTopnessVR, as an additional cross check of the goodness of the final fit. As visible from these distributions the data/MC agreement is good in the different analysis regions, and in particular in the dHHTopnessVR, which, being the closest region to the signal region, provides a validation of the background modelling and of the fit also for the dHHSR.

Other checks have been performed to validate the quality of the statistical model and of the background modelling. The NP "pulls" have been checked, which are defined as the difference between the NP estimation from the likelihood maximisation $\hat{\theta}_i$, and the initial value of the NP θ_{i0} , normalised to the value of the uncertainty:

$$Pull(NP_i) = \frac{\hat{\theta}_i - \theta_{i0}}{\Delta\theta_i} \quad (4.44)$$

Pulls have been evaluated performing two different fits: one fit to an Asimov dataset, that is a dataset obtained by setting all observable quantities equal to their expected values, built with $\mu = 0$ and one fit to the observed data with μ fixed to zero in the fit.

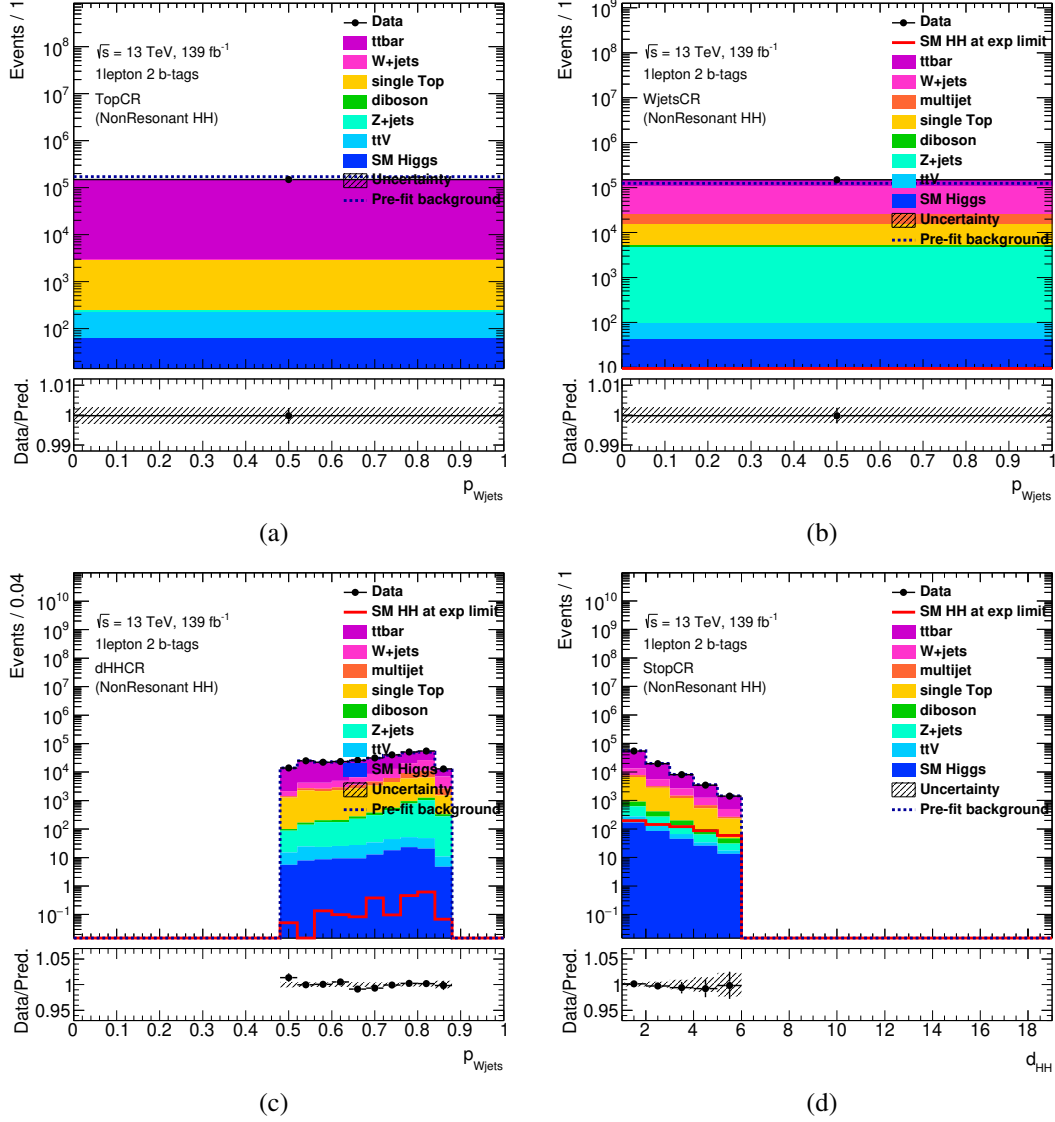


Figure 4.42: Post-fit distribution of p_{Wjets} distribution in TopCR (a), WjetsCR (b) and dHHCR (c), and of d_{HH} distribution in StopCR (d). All background components are scaled to their normalisations as determined in the fit. The pre-fit normalisation of the sum of all background components predicted by the simulated events is given by the dashed blue line. The SM $HH \rightarrow bbWW \rightarrow bblvqq$ signal distribution is also shown with a normalisation scaled to the expected exclusion limit (see Section 9). The shaded bands represent the total post-fit uncertainty with NPs profiled to the data.

The use of $\mu = 0$ is determined by the background only hypothesis which it is tested in the fit. The NP pulls from the fit are shown in Figure 4.45, where NPs names follow ATLAS conventions, which are summarised for reference in Appendix D. Asimov dataset pulls are expected to be centered at zero, with an uncertainty of ± 1 . It can be seen that there are present some overconstraints on theoretical NPs related to $t\bar{t}$ background, visible by the pull's error smaller than 1 in the Asimov dataset pulls. These overconstraints are due to the large statistics of the $t\bar{t}$ background in several regions of the fit, which with its very small statistical uncertainty results in these systematic's constraints.

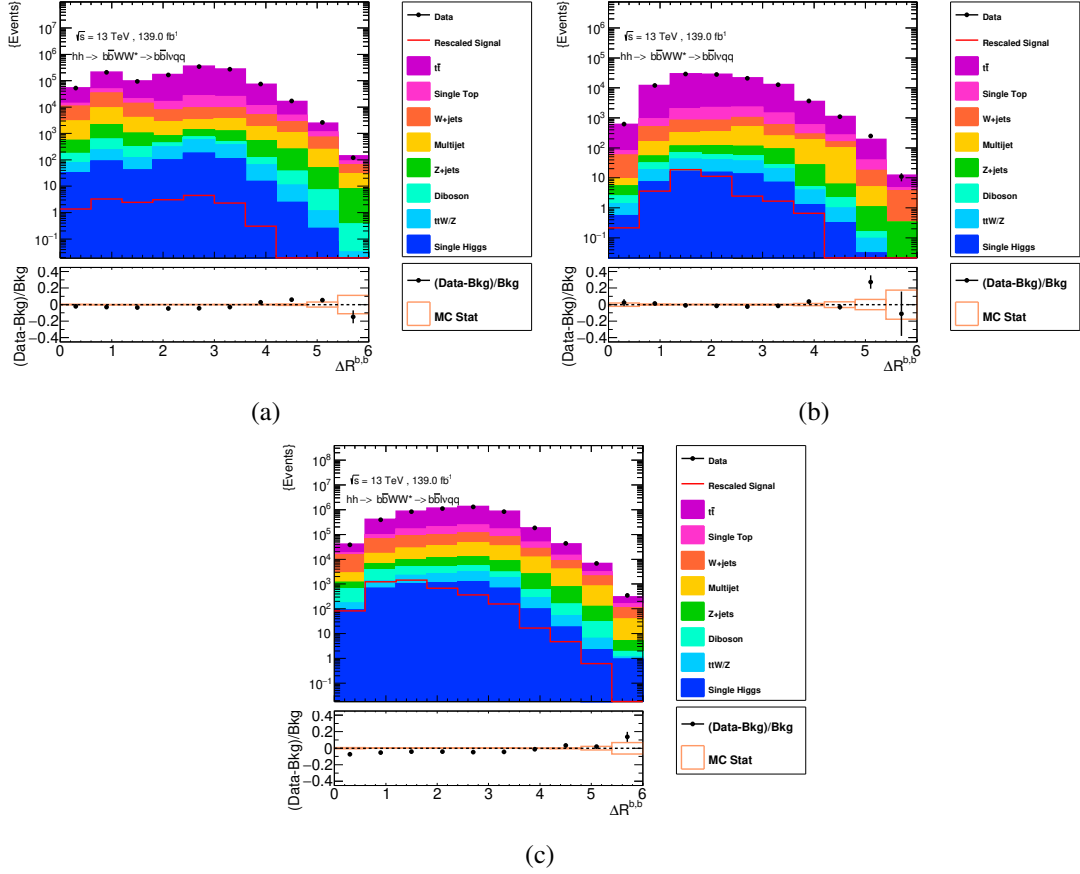


Figure 4.43: Post-fit distribution of ΔR^{bb} distribution in dHHCRCR (a), dHHTopnessVR (b) and dHHSR+StopCR (c). All background components are scaled to their normalisations as determined in the fit. The SM $HH \rightarrow bbWW^* \rightarrow bblvqq$ signal distribution is also shown with a normalisation scaled to 100 times the SM expectation. Systematic uncertainty is not shown in these validation plots.

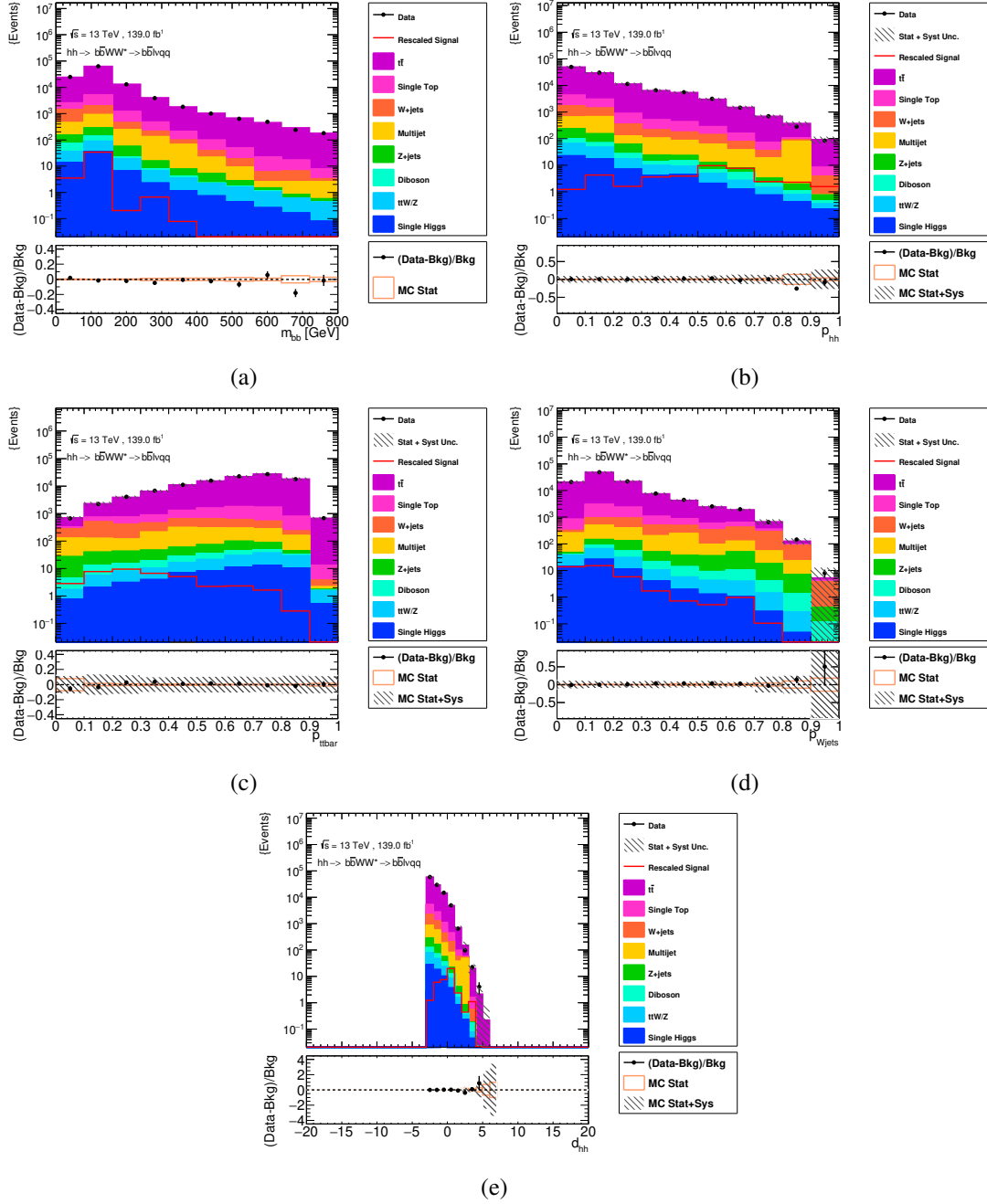


Figure 4.44: Post-fit distribution of m_{bb} (a), p_{HH} (b), $p_{t\bar{t}}$ (c), p_{Wjets} (d) and d_{HH} (e) distribution in dHHTopnessVR. All background components are scaled to their normalisations as determined in the fit. The SM $HH \rightarrow bbWW \rightarrow bblvqq$ signal distribution is also shown with a normalisation scaled to 100 times the SM expectation. Systematic uncertainty is shown in all the distributions apart from m_{bb} validation plot (a).

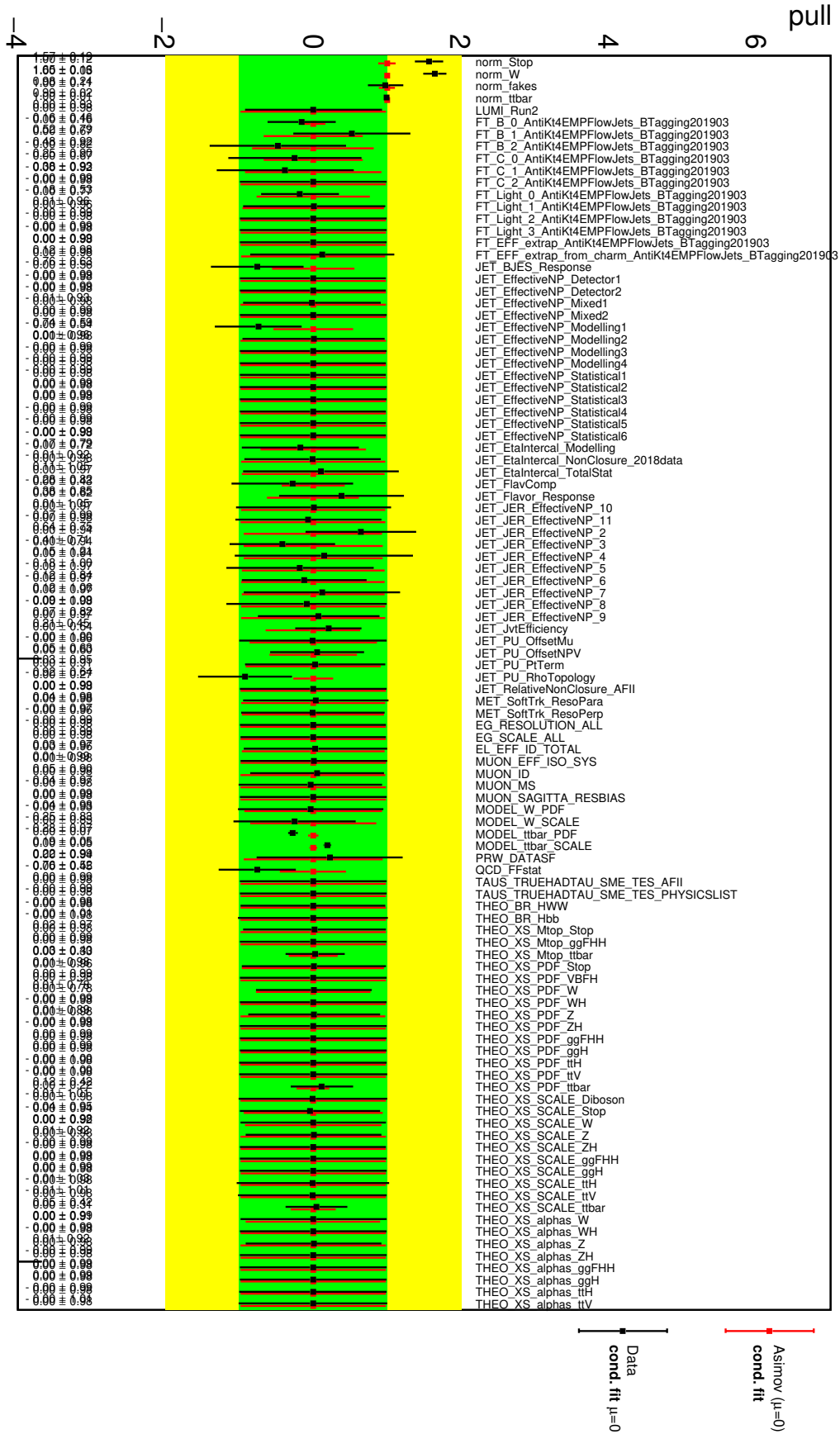


Figure 4.45: Nuisance parameters pulls from the fit to an Asimov dataset with $\mu = 0$ (red) and to the observed data (black). The associated error bars show the fitted uncertainties of the NPs, relative to their nominal uncertainties.

Other overconstraints are present on some b-tagging and JES flavor related NPs, like FT_B and JET_FlavorComposition, which have been investigated and were found to have a different shape and normalisation variation between $t\bar{t}$ and W +jets processes, driving the fit to use this characteristics to better describe the fitted distributions. Constraints on these systematic sources are expected because of the selection applied in this analysis, which asks for exactly two b-tagged jets in the event. Such a requirement may result in the wrong selection of light flavor jets, which therefore enters as fakes in the analysis. JER NPs show some shift, which are likely due to small background sample fluctuation, like diboson processes, that are described by many small sliced samples dedicated to different diboson sub-processes. Uncertainties associated with lepton reconstruction and energy measurements, as well as MET related uncertainties have instead a negligible impact on the final results. Modelling uncertainties of $t\bar{t}$ and W +jets are present in Figure 4.45 as MODEL_ttbar_SCALE, MODEL_ttbar_PDF, MODEL_W_SCALE and MODEL_W_PDF. As it is visible, W +jets NPs are not pulled, while $t\bar{t}$ ones are pulled to better describe $t\bar{t}$ contribution in the different analysis regions, but are contained in the one sigma range. Finally, a NP is associated to the multijet estimate, QCD_FFstat. Up and down varied histograms are built for the multijet contribution using the procedure described in Section 6.2 and applying the FFs with their up and down statistical variation applied. These varied distributions are used as two-sided systematics of the multijet background. This NP is pulled to better fit the data, resulting in a smaller multijet normalisation factor with respect to the one obtained in Section 6.3.

After validating the fit results with the checks described above, the post-fit final d_{HH} distribution in dHHSR has been unblinded. The unblinded post-fit d_{HH} distribution of the DNN trained model of the analysis in dHHSR is shown in Figure 4.46.

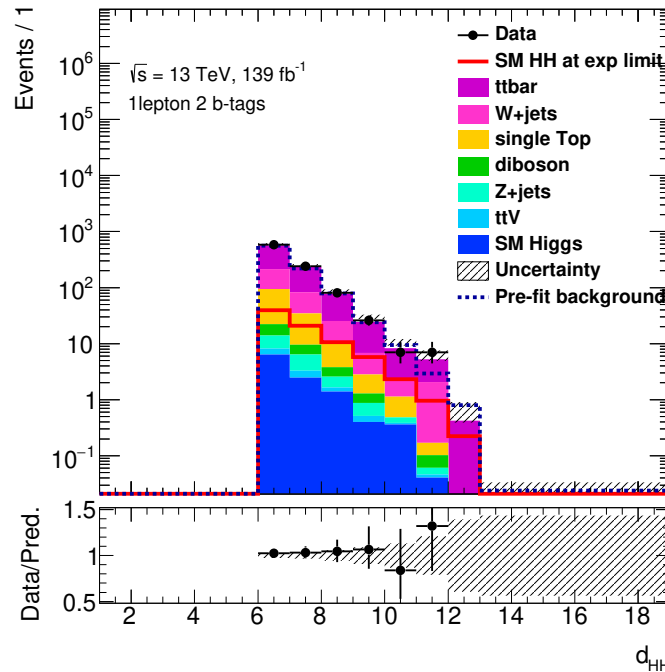


Figure 4.46: Post-fit distribution of the DNN final discriminant d_{HH} in dHHSR. All background components are scaled to their normalisations as determined in the fit. The pre-fit normalisation of the sum of all background components predicted by the simulated events is given by the dashed blue line. The SM $HH \rightarrow bbWW \rightarrow bbl\nu qq$ signal distribution is also shown with a normalisation scaled to the expected exclusion limit (see Section 9). The shaded bands represent the total post-fit uncertainty with NPs profiled to the data.

This distribution, used as final discriminant, is well described by the background-only prediction in the signal region. As it is visible, in fact, no significant excess over the expected background from SM processes was observed in the final tested d_{HH} distribution, meaning that no evidence for the presence of a signal is found. The data are thus used to set an upper limit on the SM di-Higgs production cross section by the use of an exclusion fit, which it is presented in the next Section.

9 Results

The observed data are found to be compatible with the background-only hypothesis, as visible from Figure 4.46; therefore, they are used to set an exclusion upper limit on the SM di-Higgs production cross section. The upper limit is computed at 95% CL performing a profile likelihood ratio test with the test statistics q_1 , defined in Equation 4.39, following the CLs prescription [83]. The same analysis regions, with the same distributions and the same binning used in the test of the background-only hypothesis, described in Section 8.3, are used in the limit setting. The result is summarised in Table 4.26, which reports the expected and the observed upper limit on the SM di-Higgs production cross section expressed in terms of HH production cross section times $bbWW$ single lepton BR and in terms of the signal strength μ . Upper limits on Higgs pair production cross section, and on $HH \rightarrow bbWW^*$ cross section, have been also evaluated and are reported in Table 4.26. The median expected upper limit on μ obtained in this analysis is 34.8, which is a factor 8.6 better than the previous expected value of 300 times the SM [79]. The increase of available luminosity with respect to the previous iteration of the analysis [79], from 36.1 to 139 fb^{-1} , contributes to almost a factor 2 in the improvement of the expected limit. The remaining factor 4.3 is therefore due to the development of the analysis done in this thesis work with the use of a multiclass DNN model for the classification of signal and main background sources, to the use of d_{HH} variable, coming from the output scores of the DNN model, as final discriminant for the statistical analysis, as well as to the addition of MET triggers and the other optimisations of the event selection.

	Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
$\mu = \sigma/\sigma_{SM}$	40.2	18.7	25.1	34.8	48.4	64.9
$\sigma(HH \rightarrow bbWW^* \rightarrow bbl\nu qq)$ [fb]	137.1	63.8	85.6	118.7	165.0	221.3
$\sigma(HH \rightarrow bbWW^*)$ [fb]	312.0	145.1	194.8	270.0	375.6	503.6
$\sigma(pp \rightarrow HH)$ [fb]	1248.2	580.6	779.4	1080.5	1502.8	2015.1

Table 4.26: Observed and expected upper limits on the SM di-Higgs production cross section times the $bbWW^*$ single lepton final state BR at 95% CL, and their ratios to the SM prediction. Upper limits on Higgs pair production cross section, and on $HH \rightarrow bbWW^*$ cross section have been also evaluated. The $\pm 1\sigma$ and $\pm 2\sigma$ variations from the expected limit are also reported.

This analysis strategy have improved significantly the result compared to the previous cut-based strategy [79], showing the power of the use of machine learning techniques in searches like the one performed in this work. The observed upper limit on the signal strength of 40.2 is in agreement with the median expected, and shows a small over-fluctuation of the observed data which is contained inside one sigma error of the expected value. This limit corresponds to a cross section for di-Higgs production processes times the BR of the single lepton $bbWW$ final state studied in this work of $\sigma(HH \rightarrow bbWW \rightarrow bbl\nu qq) < 137.1$ fb at 95% CL. The median expected upper limit on the signal strength obtained removing all the systematic uncertainties is 15.8,

meaning that the impact of these uncertainties on the result is of the order of 100% and thus that this result is limited by the systematic uncertainties related to this analysis. Therefore, a further improvement of the expected limit may be obtained with a better treatment of the systematic uncertainties, which it is planned in the prospects of the analysis.

The result obtained in this thesis work can be also compared with the results of the two searches in the di-lepton final state, previously described in Section 1. A summary of the new single-lepton $bbWW^*$ channel result reported here, and ATLAS and CMS $bb\ell\ell$ analyses upper limits is given in Table 4.27. The CMS $bb\ell\ell$ analysis [88] needs to be scaled to the luminosity of the full Run 2 to make a more precise comparison. The luminosity scaling provides a factor ~ 2 as the luminosity of the other analysis is about four times the CMS one. The scaled expected and observed limits are therefore 45 and 40 times the SM prediction, respectively, which are reported in brackets in Table 4.27.

Analysis	Luminosity [fb^{-1}]	Exp limit on μ	Observed limit on μ
$bbWW^*$ single-lepton (this work)	139	$34.8^{+13.6}_{-9.7}$	40.2
ATLAS $bb\ell\ell$ [90]	139	29^{+14}_{-9}	40
CMS $bb\ell\ell$ [88]	35.9 (139)	89^{+47}_{-28} (45)	79 (40)

Table 4.27: Comparison of the upper limits on the signal strength obtained by the analysis performed in this work and the two Higgs boson pair production searches in the $bb\ell\ell$ final state performed by ATLAS [90] and CMS [88] collaborations. Expected and observed upper limits are shown together with the integrated luminosity of the dataset used in each analysis. CMS analysis limits are also scaled to the integrated luminosity of 139 fb^{-1} of the other two analyses, and are reported in brackets.

Both the expected and the observed upper limits of the analysis done in this thesis work are comparable with the results in literature obtained in other di-Higgs searches focusing on the $bbVV$ decay channel. The smaller cross section of the di-lepton final state is well compensated by the minor amount of background, which is instead large in the single-lepton final state due to the irreducible $t\bar{t}$ background and the larger contribution of the multijet background compared to the di-lepton final state. The development and the use of an optimised deep neural network based classifier done in this thesis work, providing a good reduction of the large background in the sensible region of the analysis, has therefore improved the previous non competitive result [79] to the level of the other $bbVV$ channels.

9.1 Impact of systematic uncertainties

In order to quantify the impact of the systematic uncertainties, a fit is performed where the estimated signal yield is multiplied by a scaling factor α_{sig} , which is treated as the parameter of interest in the fit. The fit is performed using pseudo-data and the contribution to the uncertainty in α_{sig} from several sources is determined. The contribution of the statistical and systematic uncertainty to the total uncertainty in α_{sig} , shown in Table 4.28, are evaluated for groups of systematic sources. Jet and MET uncertainty sources have been grouped in one category as well as electron and muon systematics; then b-tagging, tau, pileup reweighting and luminosity related systematics are treated separately. Modelling systematics are evaluated separately for $t\bar{t}$ and W +jets, while multijet FFs systematic is considered alone. Finally, theoretical systematics on cross sections and branching fractions of all the processes are grouped together. The dominant systematic uncertainties are due to b-tagging systematic uncertainties, and those

related to jet measurements and MET resolution. Other contributions come from lepton, tau and modelling systematic uncertainties. An important impact is given also by the normalisations of the four main backgrounds (see Table 4.24), which are correlated with modelling uncertainties for $t\bar{t}$ and W +jets. Pileup and luminosity systematic uncertainties as well as theoretical ones, have instead a negligible contribution to the total uncertainty. Another large uncertainty source is related to the statistics of the MC samples. This large uncertainty was also noted in the sensitive range of the analysis signal region, where the single-top background presence was found to be not negligible, determining the addition of a dedicated normalisation factor in the fit, and in the statistical fluctuations of other backgrounds in this region. Statistical uncertainty has a contribution which is a factor two smaller than the one related to the total systematic uncertainties, confirming that the current analysis is systematically limited.

Systematic source	Contribution [%]
Total Stat.	± 28.5
Total Syst.	± 69.9
All normalisations	± 19.9
Jet/MET	± 30.0
B-tagging	± 14.2
Electron/Muon	± 7.8
Tau	± 12.6
Pileup reweighting	± 0.3
Luminosity	± 0.3
MC statistics	± 24.0
Multijet estimate	± 0.4
$t\bar{t}$ modelling	± 5.1
W +jets modelling	± 10.0
Theoretical systematics	± 0.6

Table 4.28: Statistical and systematic contributions (in percentage) to the total error in the scaling factor α_{sig} for the signal hypothesis. The first column quotes the source of the systematic uncertainty. The contribution is obtained by calculating the difference in quadrature between the total error in α_{sig} and that obtained by setting constant the nuisance parameter(s) relative to the contribution(s) under study.

Conclusions

In this thesis a search for the Standard Model Higgs boson pair production in the single-lepton $WWb\bar{b}$ channel is presented. The search uses the full Run 2 pp collisions data, corresponding to 139 fb^{-1} , at a centre-of-mass energy of 13 TeV recorded by the ATLAS experiment between 2015 and 2018. The main challenge of this channel, which was already studied with 2015 and 2016 data by the ATLAS Collaboration, is due to the huge presence of $t\bar{t}$ background that has the same final state of the process under study. A new complete analysis of this di-Higgs channel has been developed in this thesis work, with focus on the optimisation of the event selection with the use of machine learning techniques, increasing the signal-to-background separation to improve the sensitivity to the small signal of the analysed process. A deep neural network (DNN) based multi-output classifier has been optimised to correctly classify $t\bar{t}$, W +jets and signal HH events, exploiting the different kinematics of the three processes. The DNN output distributions have been used to define analysis signal and control regions for the estimation and normalisation of background contribution and for the evaluation of the sensitivity of the analysis. The final result has been obtained by performing a simultaneous binned maximum likelihood fit on the analysis signal and control regions, exploiting the discriminating power of the combined DNN output distributions. No significant excess over the Standard Model background prediction was seen, and the result of this analysis was interpreted in the context of the Standard Model di-Higgs production as an upper limit on the production cross section times the branching fraction of the single-lepton $WWb\bar{b}$ final state using the CLs method. The observed (expected) upper limit at 95% confidence level is 40.2 (34.8) times the Standard Model expectation, improving the previous ATLAS result on the same final state by a factor 4.3, after normalising it to the same integrated luminosity of 139 fb^{-1} . This big improvement is the result of the new analysis strategy and of the optimisation of the selection performed in this thesis. In particular the development of a DNN multi-output classifier done in this work has provided a very effective background reduction in the sensible phase space of the analysis. These optimisations have allowed to improve the sensitivity of the analysis to the level of the other machine learning based di-Higgs searches performed on the $WWb\bar{b}$ channel. Systematic uncertainties have been observed to have a large impact on the final result, and a more refined treatment of some of the uncertainty sources may further improve the result obtained in this work. Finally, a small presence of single-top background in the sensible region of the analysis has been observed. Including this background in the training of the DNN, may be remove it leading to a possible improvement in the sensitivity of this analysis.

Bibliography

- [1] Y. Fukuda and others (Super-Kamiokande Collaboration), “Evidence for oscillation of atmospheric neutrinos,” *Phys. Rev. Lett.*, vol. 81, pp. 1562–1567, Aug 1998.
- [2] Q. R. Ahmad and others (SNO Collaboration), “Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the sudbury neutrino observatory,” *Phys. Rev. Lett.*, vol. 87, p. 071301, Jul 2001.
- [3] M. Tanabashi and others (Particle Data Group), “Review of particle physics,” *Phys. Rev. D*, vol. 98, p. 030001, 2018 and 2019 update.
- [4] E. Noether, “Invariante Variationsprobleme,” *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1918)* 235, 1918.
- [5] E. Fermi, “An attempt of a theory of beta radiation. 1.,” *Z. Phys.*, vol. 88, pp. 161–177, 1934.
- [6] E. C. G. Sudarshan and R. E. Marshak, “Chirality invariance and the universal fermi interaction,” *Phys. Rev.*, vol. 109, pp. 1860–1862, Mar 1958.
- [7] R. P. Feynman and M. Gell-Mann, “Theory of the fermi interaction,” *Phys. Rev.*, vol. 109, pp. 193–198, Jan 1958.
- [8] S. Weinberg, “A model of leptons,” *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.
- [9] A. Salam, “Weak and Electromagnetic Interactions,” *Conf. Proc. C*, vol. 680519, pp. 367–377, 1968.
- [10] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964.
- [11] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.*, vol. 13, pp. 321–323, Aug 1964.
- [12] J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, *A Historical Profile of the Higgs Boson*, pp. 255–274. edited by L. Maiani and L. Rolandi, 2016.
- [13] J. Goldstone, A. Salam, and S. Weinberg, “Broken symmetries,” *Phys. Rev.*, vol. 127, pp. 965–970, Aug 1962.
- [14] D. J. Gross and F. Wilczek, “Ultraviolet behavior of non-abelian gauge theories,” *Phys. Rev. Lett.*, vol. 30, pp. 1343–1346, Jun 1973.
- [15] H. D. Politzer, “Reliable perturbative results for strong interactions?,” *Phys. Rev. Lett.*, vol. 30, pp. 1346–1349, Jun 1973.

- [16] R. P. Feynman, “Very high-energy collisions of hadrons,” *Phys. Rev. Lett.*, vol. 23, pp. 1415–1417, 1969.
- [17] 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*, vol. 46, pp. 641–653, 1977.
- [18] V. Gribov and L. Lipatov, “Deep inelastic $e p$ scattering in perturbation theory,” *Sov. J. Nucl. Phys.*, vol. 15, pp. 438–450, 1972.
- [19] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language,” *Nucl. Phys. B*, vol. 126, pp. 298–318, 1977.
- [20] L. Del Debbio, “Parton distributions in the LHC era,” *EPJ Web Conf.*, vol. 175, p. 01006, 2018.
- [21] B. Richter, “Very High-Energy electron-Positron Colliding Beams for the Study of the Weak Interactions,” *Nucl. Instrum. Meth.*, vol. 136, pp. 47–60, 1976.
- [22] H. T. Edwards, “The tevatron energy doubler: A superconducting accelerator,” *Annual Review of Nuclear and Particle Science*, vol. 35, no. 1, pp. 605–660, 1985.
- [23] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012.
- [24] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012.
- [25] G. Aad *et al.*, “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.*, vol. 114, p. 191803, 2015.
- [26] D. de Florian *et al.*, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector,” vol. 2/2017, 10 2016.
- [27] J. R. Andersen *et al.*, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties,” 7 2013.
- [28] T. Plehn and M. Rauch, “The quartic higgs coupling at hadron colliders,” *Phys. Rev. D*, vol. 72, p. 053008, 2005.
- [29] J. Alison *et al.*, “Higgs Boson Pair Production at Colliders: Status and Perspectives,” in *Double Higgs Production at Colliders* (B. Di Micco, M. Gouzevitch, J. Mazzitelli, and C. Vernieri, eds.), 9 2019.
- [30] D. O’Connell, M. J. Ramsey-Musolf, and M. B. Wise, “Minimal Extension of the Standard Model Scalar Sector,” *Phys. Rev. D*, vol. 75, p. 037701, 2007.
- [31] V. Barger, P. Langacker, M. McCaskey, M. J. Ramsey-Musolf, and G. Shaughnessy, “LHC Phenomenology of an Extended Standard Model with a Real Scalar Singlet,” *Phys. Rev. D*, vol. 77, p. 035005, 2008.
- [32] L. Evans and P. Bryant, “LHC machine,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08001, 2008.

- [33] G. Aad *et al.*, “The ATLAS experiment at the CERN Large Hadron Collider,” *Journal of Instrumentation*, vol. 3, no. 08, p. S08003, 2008.
- [34] CMS Collaboration, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, pp. S08004–S08004, aug 2008.
- [35] J. Alves, A. Augusto *et al.*, “The LHCb Detector at the LHC,” *JINST*, vol. 3, p. S08005, 2008.
- [36] K. Aamodt *et al.*, “The ALICE experiment at the CERN LHC,” *JINST*, vol. 3, p. S08002, 2008.
- [37] E. Mobs, “The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019,” Jul 2019. General Photo.
- [38] ATLAS Collaboration, “Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC,” Tech. Rep. ATLAS-CONF-2019-021, CERN, Geneva, Jun 2019.
- [39] ATLAS Luminosity group, “[Luminosity public results Run 2.](#)”
- [40] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1.* Technical Design Report ATLAS, Geneva: CERN, 1997. <https://cds.cern.ch/record/338080>.
- [41] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1.* Technical Design Report ATLAS, Geneva: CERN, 1997.
- [42] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 2.* Technical Design Report ATLAS, Geneva: CERN, 1997.
- [43] J. Pequeno, “Computer generated image of the ATLAS inner detector.” General Photo, Mar 2008.
- [44] ATLAS Collaboration, “Experiment Briefing: Keeping the ATLAS Inner Detector in perfect alignment.” General Photo, Jul 2020.
- [45] ATLAS Collaboration, *ATLAS calorimeter performance: Technical Design Report.* Technical Design Report ATLAS, Geneva: CERN, 1996.
- [46] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report.* Technical Design Report ATLAS, Geneva: CERN, 1996.
- [47] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report.* Technical Design Report ATLAS, Geneva: CERN, 1996.
- [48] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report.* Technical Design Report ATLAS, Geneva: CERN, 1997. <https://cds.cern.ch/record/331068>.
- [49] G. Aad *et al.*, “Operation of the ATLAS trigger system in Run 2,” 7 2020.
- [50] M. Aaboud *et al.*, “Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2,” *Eur. Phys. J. C*, vol. 77, no. 10, p. 673, 2017.
- [51] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting,” *Nucl. Instrum. Methods Phys. Res., A*, vol. 262, p. 444. 19 p, Jun 1987.

- [52] M. Aaboud *et al.*, “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC,” *Eur. Phys. J. C*, vol. 77, no. 5, p. 332, 2017.
- [53] R. Frühwirth, W. Waltenberger, and P. Vanlaer, “Adaptive Vertex Fitting,” Tech. Rep. CMS-NOTE-2007-008, CERN, Geneva, Mar 2007.
- [54] G. Aad *et al.*, “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data,” *JINST*, vol. 14, no. 12, p. P12006, 2019.
- [55] J. Illingworth and J. Kittler, “A survey of the hough transform,” *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87 – 116, 1988.
- [56] G. Aad *et al.*, “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}=13$ TeV,” *Eur. Phys. J. C*, vol. 76, no. 5, p. 292, 2016.
- [57] S. Tarem, Z. Tarem, N. Panikashvili, and O. Belkind, “Mugirl – muon identification in the atlas detector from the inside out,” in *2006 IEEE Nuclear Science Symposium Conference Record*, vol. 1, pp. 617–621, 2006.
- [58] ATLAS Collaboration, “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV,” 8 2020.
- [59] ATLAS Collaboration, “Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC,” Tech. Rep. ATL-PHYS-PUB-2015-045, CERN, Geneva, Nov 2015.
- [60] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008.
- [61] ATLAS Collaboration, “Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment,” Tech. Rep. ATL-PHYS-PUB-2019-033, CERN, Geneva, Aug 2019.
- [62] M. Aaboud *et al.*, “Jet reconstruction and performance using particle flow with the ATLAS Detector,” *Eur. Phys. J. C*, vol. 77, no. 7, p. 466, 2017.
- [63] G. Aad *et al.*, “Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector,” *Eur. Phys. J. C*, vol. 76, no. 11, p. 581, 2016.
- [64] G. Aad *et al.*, “ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV,” *Eur. Phys. J. C*, vol. 79, no. 11, p. 970, 2019.
- [65] M. Aaboud *et al.*, “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV,” *Eur. Phys. J. C*, vol. 78, no. 11, p. 903, 2018.
- [66] M. Aaboud *et al.*, “Constraints on off-shell Higgs boson production and the Higgs boson total width in $ZZ \rightarrow 4\ell$ and $ZZ \rightarrow 2\ell 2\nu$ final states with the ATLAS detector,” *Phys. Lett. B*, vol. 786, pp. 223–244, 2018.

- [67] A. M. Sirunyan *et al.*, “Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state,” *Phys. Rev. D*, vol. 99, no. 11, p. 112003, 2019.
- [68] ATLAS Collaboration, “A combination of measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13 \text{ TeV}$ collected with the ATLAS experiment,” 8 2020.
- [69] A. M. Sirunyan *et al.*, “Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$,” *Eur. Phys. J. C*, vol. 79, no. 5, p. 421, 2019.
- [70] G. Aad *et al.*, “Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8 \text{ TeV}$ from the ATLAS Detector,” *Phys. Rev. Lett.*, vol. 114, no. 8, p. 081802, 2015.
- [71] G. Aad *et al.*, “Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8 \text{ TeV}$ with the ATLAS detector,” *Eur. Phys. J. C*, vol. 75, no. 9, p. 412, 2015.
- [72] G. Aad *et al.*, “Searches for Higgs boson pair production in the $hh \rightarrow b\bar{b}\tau\tau, \gamma\gamma WW^*, \gamma\gamma b\bar{b}, b\bar{b}b\bar{b}$ channels with the ATLAS detector,” *Phys. Rev. D*, vol. 92, p. 092004, 2015.
- [73] A. M. Sirunyan *et al.*, “Search for Higgs boson pair production in the $b\bar{b}\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 8 \text{ TeV}$,” *Phys. Rev. D*, vol. 96, no. 7, p. 072004, 2017.
- [74] V. Khachatryan *et al.*, “Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV ,” *Phys. Rev. D*, vol. 94, no. 5, p. 052012, 2016.
- [75] M. Grazzini, G. Heinrich, S. Jones, S. Kallweit, M. Kerner, J. M. Lindert, and J. Mazzitelli, “Higgs boson pair production at NNLO with top quark mass effects,” *JHEP*, vol. 05, p. 059, 2018.
- [76] M. Aaboud *et al.*, “Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” *JHEP*, vol. 01, p. 030, 2019.
- [77] M. Aaboud *et al.*, “Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” *Phys. Rev. Lett.*, vol. 121, no. 19, p. 191801, 2018. [Erratum: *Phys.Rev.Lett.* 122, 089901 (2019)].
- [78] M. Aaboud *et al.*, “Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment,” *JHEP*, vol. 11, p. 040, 2018.
- [79] M. Aaboud *et al.*, “Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” *JHEP*, vol. 04, p. 092, 2019.
- [80] M. Aaboud *et al.*, “Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,” *Eur. Phys. J. C*, vol. 78, no. 12, p. 1007, 2018.

- [81] M. Aaboud *et al.*, “Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$ decay channel using ATLAS data recorded at $\sqrt{s} = 13$ TeV,” *JHEP*, vol. 05, p. 124, 2019.
- [82] G. Aad *et al.*, “Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” *Phys. Lett. B*, vol. 800, p. 135103, 2020.
- [83] A. L. Read, “Presentation of search results: The CL(s) technique,” *J. Phys.*, vol. G28, pp. 2693–2704, 2002. [11(2002)].
- [84] G. Aad *et al.*, “Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,” *JHEP*, vol. 07, p. 108, 2020.
- [85] A. M. Sirunyan *et al.*, “Search for nonresonant Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state at $\sqrt{s} = 13$ TeV,” *JHEP*, vol. 04, p. 112, 2019.
- [86] A. M. Sirunyan *et al.*, “Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV,” *Phys. Lett. B*, vol. 778, pp. 101–127, 2018.
- [87] A. M. Sirunyan *et al.*, “Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV,” *Phys. Lett. B*, vol. 788, pp. 7–36, 2019.
- [88] A. M. Sirunyan *et al.*, “Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JHEP*, vol. 01, p. 054, 2018.
- [89] A. M. Sirunyan *et al.*, “Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *Phys. Rev. Lett.*, vol. 122, no. 12, p. 121803, 2019.
- [90] G. Aad *et al.*, “Search for non-resonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV,” *Phys. Lett. B*, vol. 801, p. 135145, 2020.
- [91] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual,” *JHEP*, vol. 05, p. 026, 2006.
- [92] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1,” *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.
- [93] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An introduction to pythia 8.2,” *Computer Physics Communications*, vol. 191, pp. 159 – 177, 2015.
- [94] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter, “Event generation with SHERPA 1.1,” *Journal of High Energy Physics*, vol. 2009, pp. 007–007, feb 2009.
- [95] M. Bahr *et al.*, “Herwig++ Physics and Manual,” *Eur. Phys. J. C*, vol. 58, pp. 639–707, 2008.
- [96] J. Bellm *et al.*, “Herwig 7.0/Herwig++ 3.0 release note,” *Eur. Phys. J. C*, vol. 76, no. 4, p. 196, 2016.

- [97] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method,” *Journal of High Energy Physics*, vol. 2007, pp. 070–070, nov 2007.
- [98] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations,” *JHEP*, vol. 07, p. 079, 2014.
- [99] ATLAS Collaboration, “The ATLAS Simulation Infrastructure,” *Eur.Phys.J.*, vol. C70, pp. 823–874, 2010.
- [100] S. Agostinelli *et al.*, “Geant4 - a simulation toolkit,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.
- [101] W. Lukas, “Fast simulation for ATLAS: Atfast-II and ISF,” *Journal of Physics: Conference Series*, vol. 396, p. 022031, dec 2012.
- [102] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C.-P. Yuan, “CT10 next-to-next-to-leading order global analysis of QCD,” *Phys. Rev. D*, vol. 89, no. 3, p. 033009, 2014.
- [103] S. Borowka, N. Greiner, G. Heinrich, S. Jones, M. Kerner, J. Schlenk, U. Schubert, and T. Zirke, “Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass Dependence,” *Phys. Rev. Lett.*, vol. 117, no. 1, p. 012001, 2016. [Erratum: *Phys.Rev.Lett.* 117, 079901 (2016)].
- [104] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, M. Spira, and J. Streicher, “Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme,” *Eur. Phys. J. C*, vol. 79, no. 6, p. 459, 2019.
- [105] D. de Florian and J. Mazzitelli, “Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD,” *Phys. Rev. Lett.*, vol. 111, p. 201801, 2013.
- [106] D. Y. Shao, C. S. Li, H. T. Li, and J. Wang, “Threshold resummation effects in Higgs boson pair production at the LHC,” *JHEP*, vol. 07, p. 169, 2013.
- [107] D. de Florian and J. Mazzitelli, “Higgs pair production at next-to-next-to-leading logarithmic accuracy at the LHC,” *JHEP*, vol. 09, p. 053, 2015.
- [108] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box,” *Journal of High Energy Physics*, vol. 2010, Jun 2010.
- [109] R. D. Ball *et al.*, “Parton distributions from high-precision collider data,” *Eur. Phys. J. C*, vol. 77, no. 10, p. 663, 2017.
- [110] ATLAS Collaboration, “ATLAS Pythia 8 tunes to 7 TeV data,” Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov 2014.
- [111] D. Lange, “The EvtGen particle decay simulation package,” *Nucl. Instrum. Meth. A*, vol. 462, pp. 152–155, 2001.
- [112] ATLAS Collaboration, “Comparison of Monte Carlo generator predictions for gap fraction and jet multiplicity observables in top-antitop events,” Tech. Rep. ATL-PHYS-PUB-2014-005, CERN, Geneva, May 2014.

- [113] M. Czakon, P. Fiedler, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$,” *Phys. Rev. Lett.*, vol. 110, p. 252004, 2013.
- [114] P. Kant, O. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer, “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions,” *Comput. Phys. Commun.*, vol. 191, pp. 74–89, 2015.
- [115] M. Aliev, H. Lacker, U. Langenfeld, S. Moch, P. Uwer, and M. Wiedermann, “HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR,” *Comput. Phys. Commun.*, vol. 182, pp. 1034–1046, 2011.
- [116] N. Kidonakis, “Top Quark Production,” in *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*, pp. 139–168, 2014.
- [117] N. Kidonakis, “Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-,” *Phys. Rev. D*, vol. 82, p. 054018, 2010.
- [118] C. Anastasiou, L. Dixon, K. Melnikov, and F. Petriello, “High-precision qcd at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order,” *Phys. Rev. D*, vol. 69, p. 094008, May 2004.
- [119] J. M. Campbell, R. Ellis, and C. Williams, “Vector boson pair production at the LHC,” *JHEP*, vol. 07, p. 018, 2011.
- [120] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C.-P. Yuan, “New parton distributions for collider physics,” *Phys. Rev. D*, vol. 82, p. 074024, 2010.
- [121] K. Hamilton, P. Nason, and G. Zanderighi, “MINLO: Multi-Scale Improved NLO,” *JHEP*, vol. 10, p. 155, 2012.
- [122] G. Aad *et al.*, “Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector,” *JHEP*, vol. 09, p. 145, 2014.
- [123] J. H. Kim, K. Kong, K. T. Matchev, and M. Park, “Probing the Triple Higgs Self-Interaction at the Large Hadron Collider,” *Phys. Rev. Lett.*, vol. 122, no. 9, p. 091801, 2019.
- [124] ATLAS Collaboration, “Higgsness and Topness plots for non-resonant HH production in the single lepton $bbWW$ decay channel,” Tech. Rep. ATL-PHYS-PUB-2019-040, CERN, Geneva, Oct 2019.
- [125] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification; 2nd ed.* New York, NY: Wiley, 2001.
- [126] A. R. Webb and K. D. Copsey, *Statistical Pattern Recognition; 3rd ed.* Hoboken, NJ: John Wiley & Sons, 2011.
- [127] C. M. Bishop, *Pattern recognition and machine learning*. Information science and statistics, New York, NY: Springer, 2006. Softcover published in 2016.
- [128] A. Hocker *et al.*, “TMVA - Toolkit for Multivariate Data Analysis,” 3 2007.

- [129] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Adaptive computation and machine learning series, Cambridge, MA: The MIT Press, 2016.
- [130] F. Chollet *et al.*, “Keras.” <https://keras.io>, 2015.
- [131] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, “Tensorflow: A system for large-scale machine learning,” in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265–283, 2016.
- [132] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [133] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014.
- [134] Buttinger, W., “Background Estimation with the ABCD Method.” [document link](#).
- [135] J. Butterworth *et al.*, “PDF4LHC recommendations for LHC Run II,” *J. Phys. G*, vol. 43, p. 023001, 2016.
- [136] R. D. Ball *et al.*, “Parton distributions for the LHC Run II,” *JHEP*, vol. 04, p. 040, 2015.
- [137] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011. [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)].

Appendix A

Cut-based and BDT truth level studies

A BDT selection have been performed with TMVA [128]. With this classification, using the CLs method we have estimated the minimum achievable upper limit μ^* , scanning over the possible working points of the selection. Signal and $t\bar{t}$ background truth level kinematic variables were used in the training. Cuts similar to the ones used in the object selection of the previous analysis [79] have been applied: events with taus are removed, jets are required to have $p_T > 20$ GeV, leptons are required to have $p_T > 28$ GeV. Only full-resolved events have been used, checking the ΔR distance between the particles to be $\Delta R > 0.4$. Jet pairing is performed among the three leading jets to reconstruct the hadronic decaying W boson. The features used in the BDT classifier, whose distributions are shown in Figure A.2, are:

- $\Delta R(b, \bar{b})$
- $\Delta\theta(b, \bar{b})$ in C.O.M. frame of $WWbb$ system
- $p_T^{b\bar{b}}$
- $m(b\bar{b})$
- m_{WW}
- p_T^{WW}
- $\Delta\phi(W, W)$
- $\Delta\eta(W, W)$
- $\Delta R(W, W)$
- m_{W+b} or $m_{W-\bar{b}}$, using the two jets of the hadronic W
- $\Delta R_{min}(l, q)$
- $\Delta R_{min}(l, q)$
- $m_{jj} = m_{W_{had}}$
- Higgsness
- Topness

A comparison was done with the cut-based approach used in the previous 36 fb^{-1} analysis [79]. The cuts used in the previous analysis were optimised, keeping the same set of kinematic variables, whose distributions are shown in Figure A.1. Here the new

cut values are listed, with the original value in brackets: $m_{WW} < 141.8$ GeV (130), $p_T^{WW} > 248.4$ GeV (250), $m_{bb} \in (71.8, 176)$ GeV (105, 135), $p_T^{bb} > 287$ GeV (300), $E_T^{miss} > 20$ GeV (20). After the optimisation of both BDT and cut-based models, the BDT was found to be a factor ~ 4 more sensible than the cut-based approach. The results are shown in Table A.1.

Analysis	comments	ε full-res (+ cuts)		Selection efficiency		# expected events		$s/\sqrt{b}$	upper limit μ^*
		signal	bkg	ε_{sig}	ε_{bkg}	signal	bkg		
BDT	FR + JetCuts + JetPair + noTaus + LepCuts*	0.138	0.323	0.220	$2.61 \cdot 10^{-6}$	5.96	16.33	1.48	1.60
Cut-based ~ old analysis	FR + JetCuts + JetPair + noTaus + LepCuts*	0.138	0.323	0.091	$7.76 \cdot 10^{-6}$	2.47	48.48	0.36	6.15

Table A.1: Analysis sensitivity comparison between BDT and cut-based selection approach.

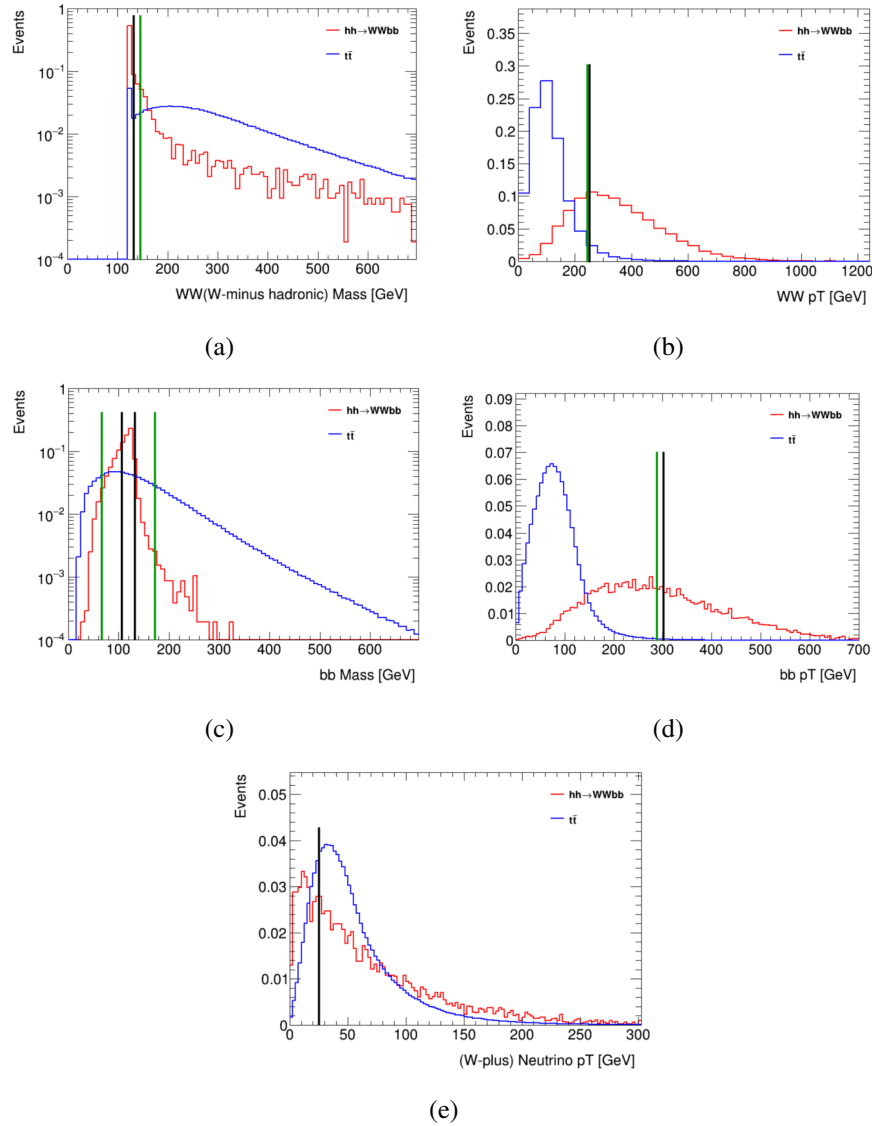


Figure A.1: Distributions used in the previous cut-based analysis: m_{WW} (a), p_T^{WW} (b), m_{bb} (c), p_T^{bb} (d) and p_T^V (e) taken as the E_T^{miss} . In black are reported the exact cuts used in the previous analysis, while in green are indicated the values obtained by TMVA optimisation of the cut-based selection.

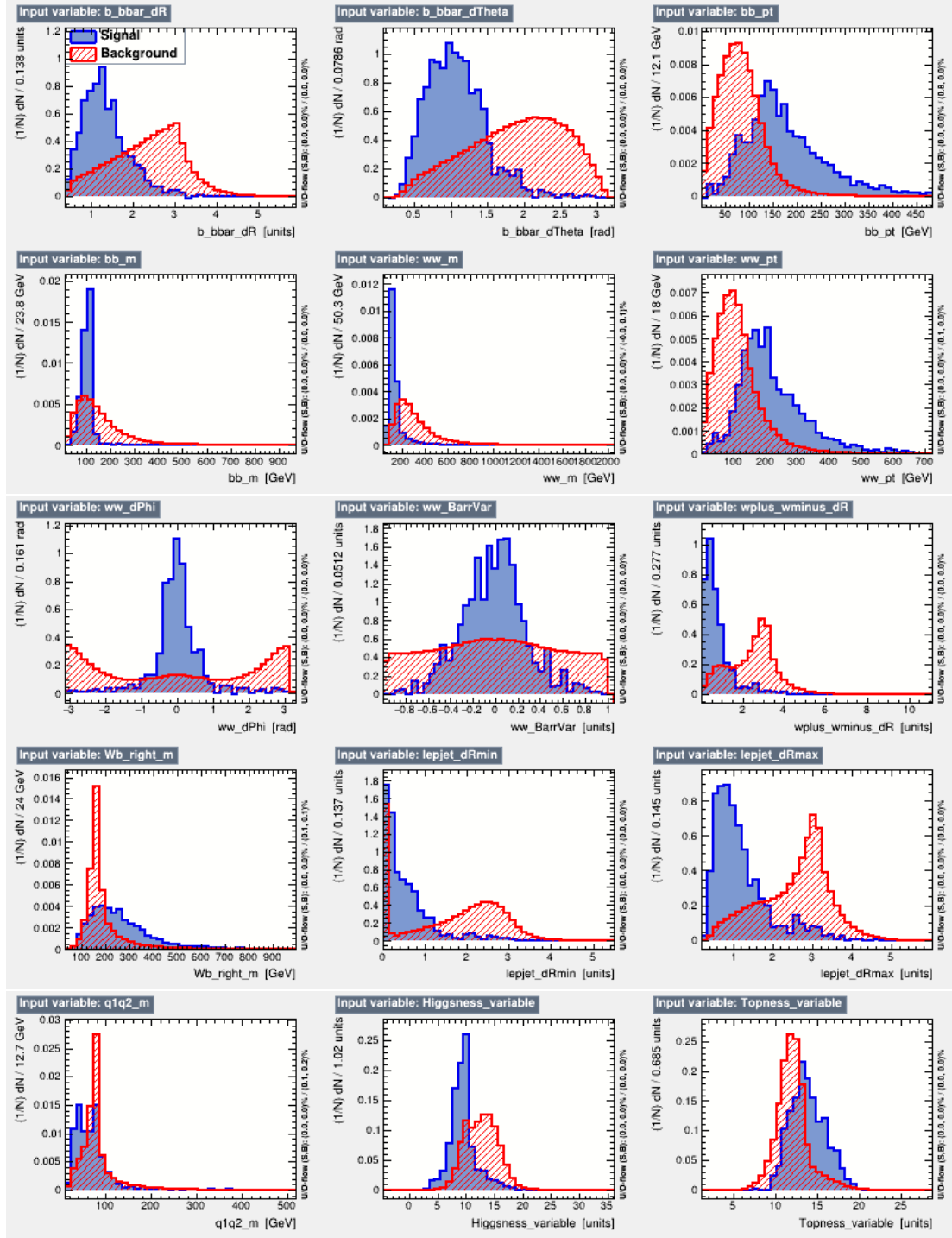


Figure A.2: Distributions of the input BDT variables for HH signal, in blue, and $t\bar{t}$ background, in red, truth-level events. In order from the upper left pad, from the left to the right and from the top to the bottom: ΔR^{bb} , $\Delta\theta^{bb}$, p_T^{bb} , m_{bb} , m_{WW} , p_T^{WW} , $\Delta\phi(W, W)$, $\Delta\eta(WW)$, ΔR^{WW} , m_{Wb} , $\Delta R_{min}(l, j)$, $\Delta R_{max}(l, j)$, m_{jj} , $\log(H)$, $\log(T)$.

Appendix B

Fake factors for multijet background estimation

Tag and probe FF p_T distributions in the two η regions are reported in Figure B.1 and B.2 respectively. Tag FF distributions exhibit a step in lepton p_T at 61 GeV for the electrons and at 50 GeV for the muons. This is due to transition from low p_T to high p_T lepton trigger. Probe lepton FFs do not have this kind of feature because they are trigger independent.

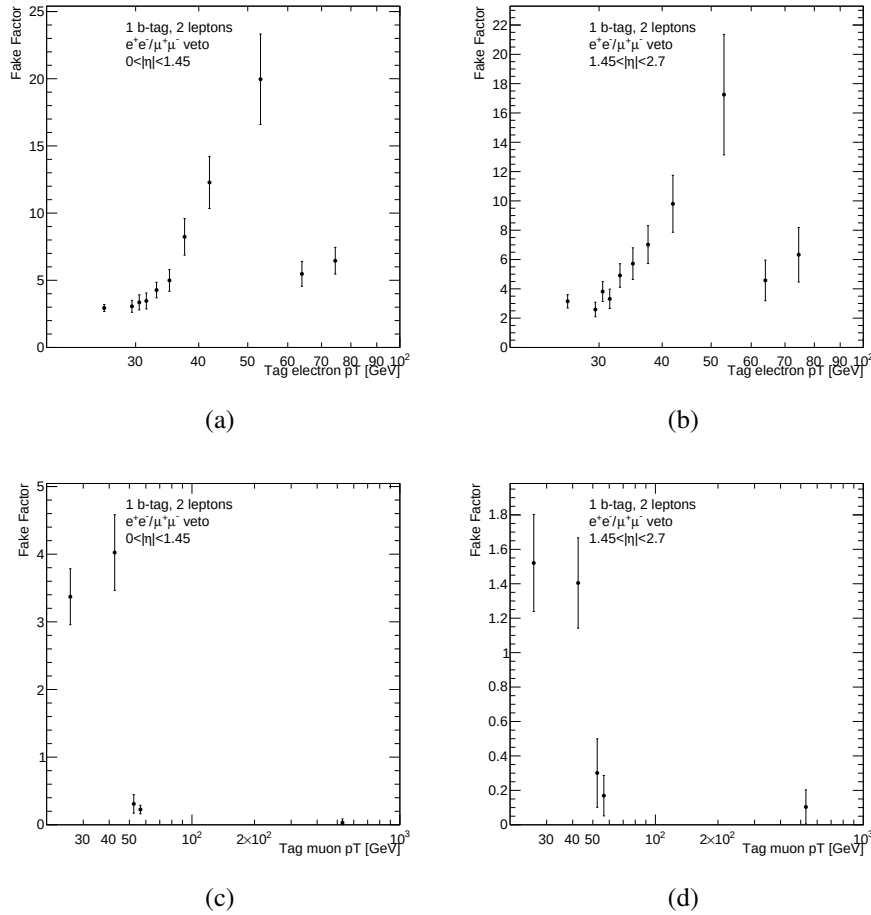


Figure B.1: Tag electron and muon FF p_T dependency in barrel and end-cap region. Barrel (a) and end-cap (b) electron, and barrel (c) and end-cap (d) muon.

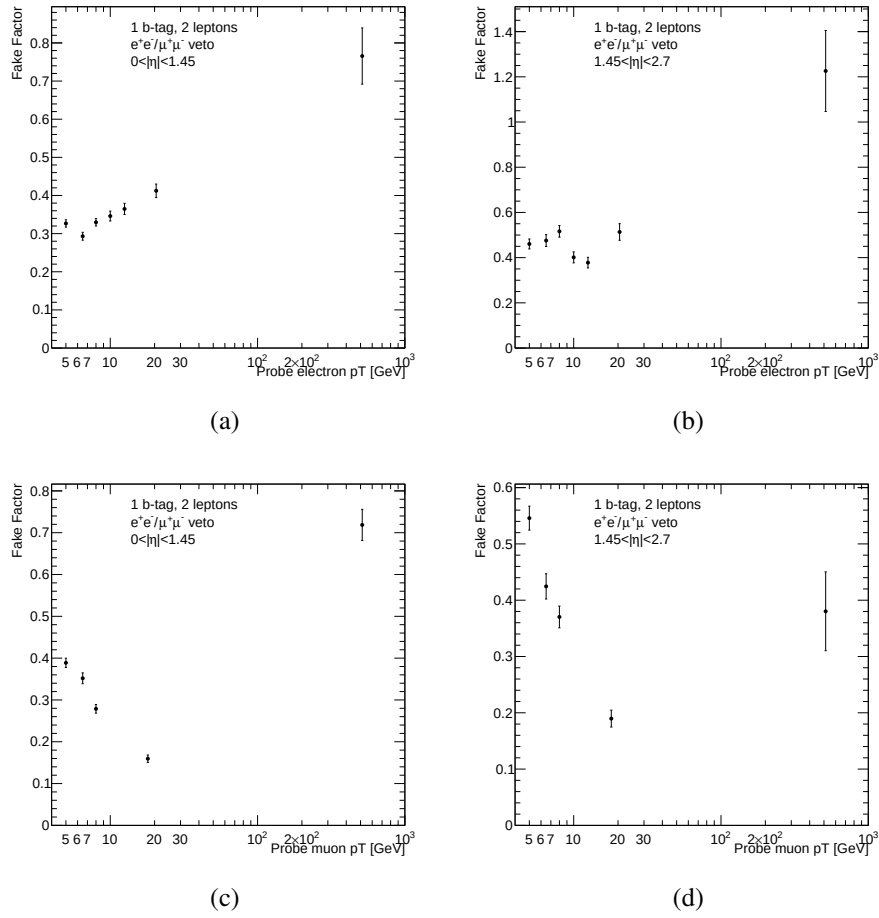


Figure B.2: Probe electron and muon FF p_T dependency in barrel and end-cap region. Barrel (a) and end-cap (b) electron, and barrel (c) and end-cap (d) muon.

Appendix C

MET trigger scale factors

MET trigger scale factors are calculated in the TopCR comparing $t\bar{t}$ background and collected data as a function of the E_T^{miss} of the event. Scale factors are derived separately for periods with the same combination of trigger algorithms, a summary is given in Table C.1. If multiple triggers are used in some periods, a logical OR is performed in the computation of trigger scale factors. The binning has been chosen in order to have a similar statistics in all the bins, keeping the statistical error under control for each SF. Efficiency SFs are calculated as:

$$\varepsilon = \frac{\text{number of events passing trigger(s) and selection}}{\text{total number of events passing selection}} \quad (\text{C.1})$$

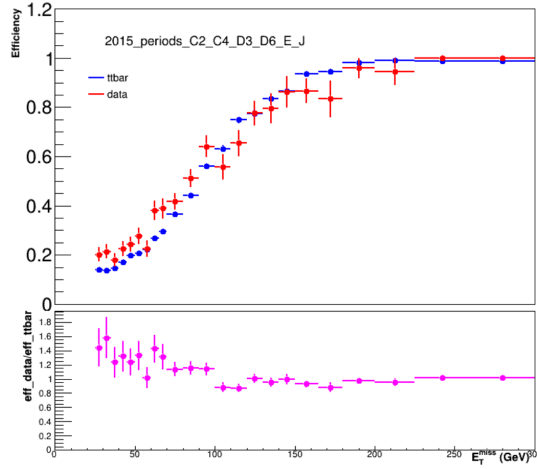


Figure C.1: MET trigger efficiency as a function of E_T^{miss} for data and $t\bar{t}$ background in TopCR (upper pad), and MET trigger scale factors (lower pad) for 2015 HLT trigger algorithms.

Year	Periods	Trigger
2015	all	HLT_xe70_mht
2016	A to C	HLT_xe80_tc_lcw_L1XE50
		HLT_xe90_mht_L1XE50
		HLT_xe90_mht_wEFMu_L1XE50
		HLT_xe90_tc_lcw_wEFMu_L1XE50
		HLT_xe100_L1XE50
		HLT_xe100_tc_em_L1XE50
		HLT_xe120_pueta
		HLT_xe120_pufit
	D1 to D3	HLT_xe90_mht_L1XE50
		HLT_xe90_mht_wEFMu_L1XE50
		HLT_xe100_tc_em_L1XE50
		HLT_xe110_pueta_L1XE50
		HLT_xe120_tc_lcw_L1XE50
	D4 to L	HLT_xe110_mht_L1XE50
2017	B to D5	HLT_xe110_pufit_L1XE55
	D6 to K	HLT_xe110_pufit_L1XE50
2018	B to C	HLT_xe110_pufit_xe70_L1XE50
		HLT_xe120_pufit_L1XE50
	D to J	HLT_xe110_pufit_xe65_L1XE50
		HLT_xe110_pufit_xe70_L1XE50
		HLT_xe120_pufit_L1XE50
	K to R	HLT_xe110_pufit_xe65_L1XE50
		HLT_xe110_pufit_xe70_L1XE50

Table C.1: Summary of MET trigger used in each data taking period, for which a logical OR in the selection requirement is performed.

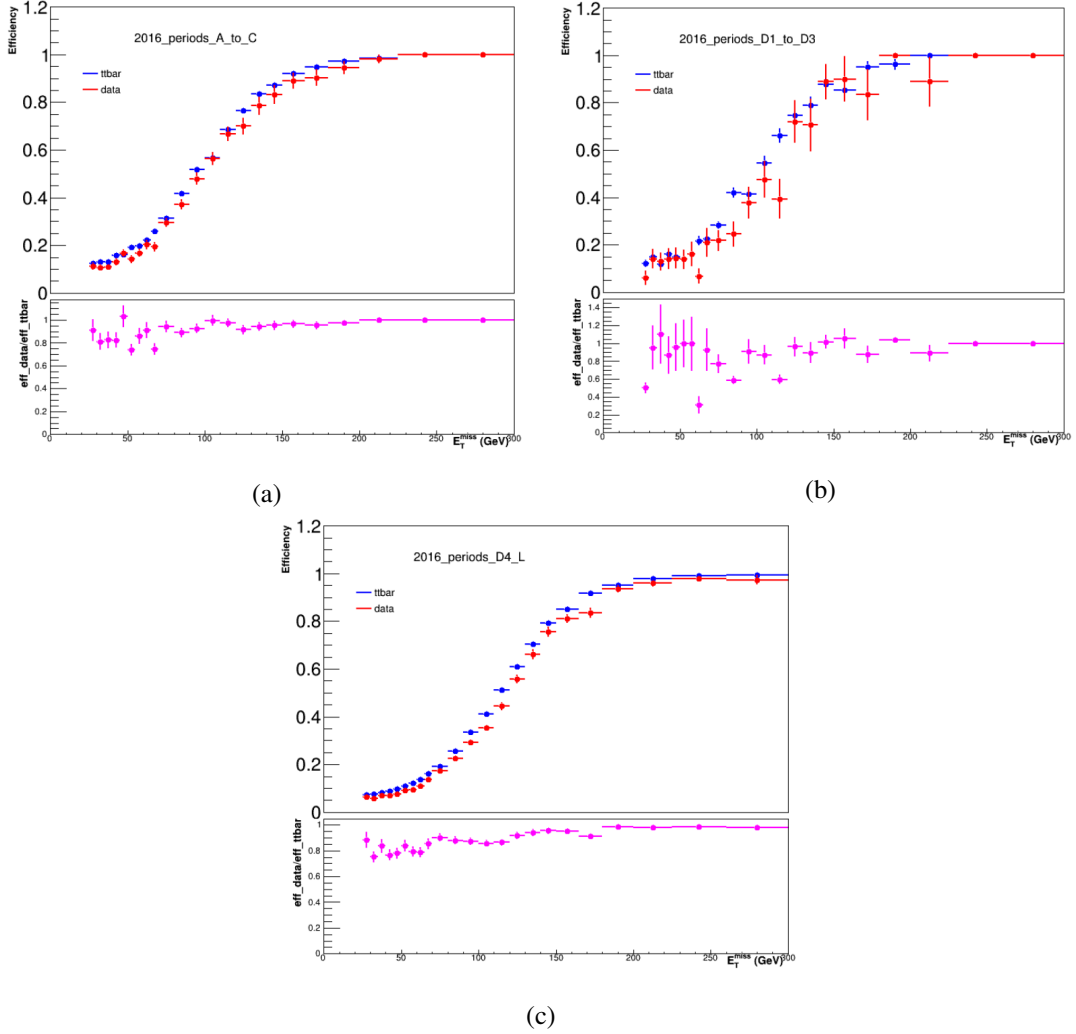


Figure C.2: MET trigger efficiency as a function of E_T^{miss} for data and $t\bar{t}$ background in TopCR (upper pad), and MET trigger scale factors (lower pad) for 2016 HLT trigger algorithms. Distributions are shown for three data taking periods.

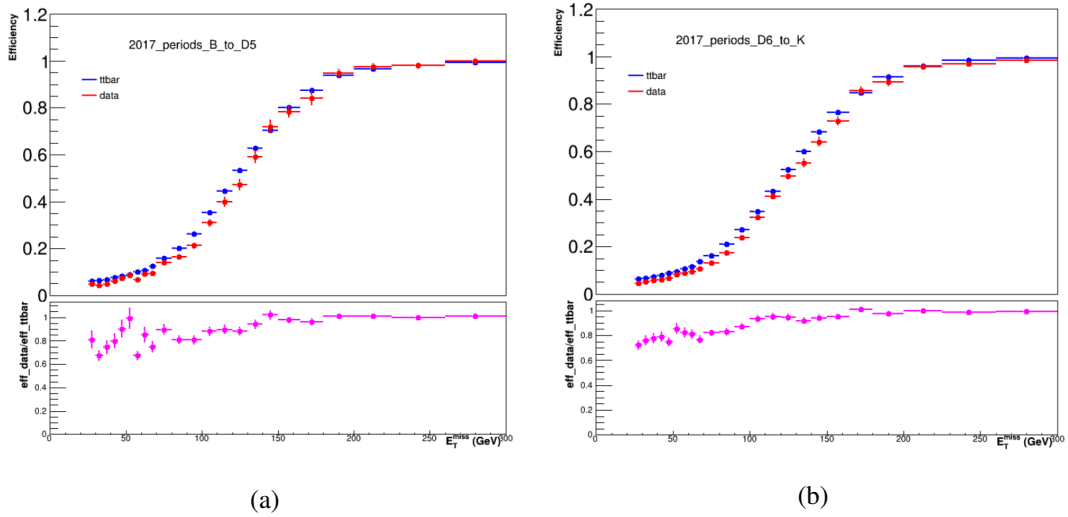


Figure C.3: MET trigger efficiency as a function of E_T^{miss} for data and $t\bar{t}$ background in TopCR (upper pad), and MET trigger scale factors (lower pad) for 2017 HLT trigger algorithms. Distributions are shown for two data taking periods.

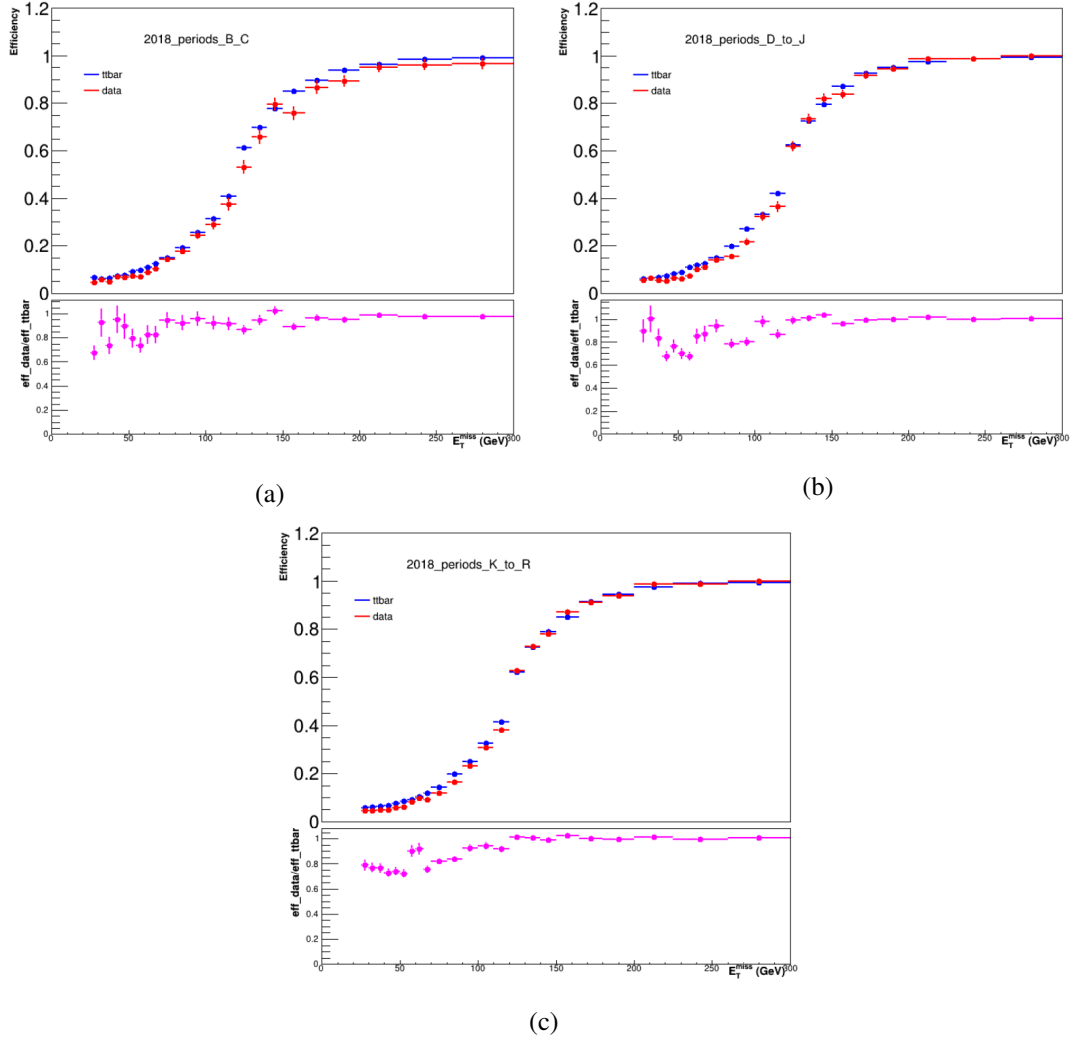


Figure C.4: MET trigger efficiency as a function of E_T^{miss} for data and $t\bar{t}$ background in TopCR (upper pad), and MET trigger scale factors (lower pad) for 2018 HLT trigger algorithms. Distributions are shown for three data taking periods.

Appendix D

Systematic uncertainties list

Object	Nuisance parameter	Type
Electron	EG_RESOLUTION_ALL	NS
	EG_SCALE_ALL	NS
	EG_SCALE_AF2	NS
	EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	N
	EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	N
	EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	N
	EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR	N
Muon	MUON_SCALE	NS
	MUON_ID	NS
	MUON_MS	NS
	MUON_SAGITTA_RHO	NS
	MUON_SAGITTA_RESBIAS	NS
	MUON_EFF_ISO_STAT	N
	MUON_EFF_ISO_SYS	N
	MUON_EFF_RECO_STAT	N
	MUON_EFF_RECO_STAT_LOWPT	N
	MUON_EFF_RECO_SYS	N
	MUON_EFF_RECO_SYS_LOWPT	N
	MUON_EFF_TTVA_STAT	N
	MUON_EFF_TTVA_SYS	N
Tau	TAUS_TRUEHADTAU_SME_TES_PHYSICSLIST	NS
	TAUS_TRUEHADTAU_SME_TES_INSITUEXP	NS
	TAUS_TRUEHADTAU_SME_TES_INSITUFIT	NS
	TAUS_TRUEHADTAU_SME_TES_MODEL_CLOSURE	NS
	TAUS_TRUEHADTAU_SME_TES_AFII	NS
	TAUS_TRUEHADTAU_EFF_RECO_TOTAL	N
	TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPT2025	N
	TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPT2530	N
	TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPT3040	N
	TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPTGE40	N
	TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPT2025	N
	TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPT2530	N
	TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPT3040	N
	TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPTGE40	N
	TAUS_TRUEHADTAU_EFF_RNNID_SYST	N
	TAUS_TRUEHADTAU_EFF_RNNID_HIGHPT	N
	TAUS_TRUEHADTAU_EFF_ELEOLR_TOTAL	N
	TAUS_TRUEELECTRON_EFF_ELEBDT_STAT	N
	TAUS_TRUEELECTRON_EFF_ELEBDT_SYST	N

Table D.1: Lepton related nuisance parameter list. Type is labelled "N" when the systematic is taken as a normalisation, or "NS" when it is taken both as normalisation and shape.

Object	Nuisance parameter	Type
Jet	JET_BJES_Response	NS
	JET_EffectiveNP_Detector1	NS
	JET_EffectiveNP_Detector2	NS
	JET_EffectiveNP_Mixed1	NS
	JET_EffectiveNP_Mixed2	NS
	JET_EffectiveNP_Mixed3	NS
	JET_EffectiveNP_Modelling1	NS
	JET_EffectiveNP_Modelling2	NS
	JET_EffectiveNP_Modelling3	NS
	JET_EffectiveNP_Modelling4	NS
	JET_EffectiveNP_Statistical1	NS
	JET_EffectiveNP_Statistical2	NS
	JET_EffectiveNP_Statistical3	NS
	JET_EffectiveNP_Statistical4	NS
	JET_EffectiveNP_Statistical5	NS
	JET_EffectiveNP_Statistical6	NS
	JET_EtaIntercalibration_Modelling	NS
	JET_EtaIntercalibration_NonClosure_2018data	NS
	JET_EtaIntercalibration_NonClosure_highE	NS
	JET_EtaIntercalibration_NonClosure_negEta	NS
	JET_EtaIntercalibration_NonClosure_posEta	NS
	JET_EtaIntercalibration_TotalStat	NS
	JET_Flavor_Composition	NS
	JET_Flavor_Response	NS
	JET_JER_DataVsMC_AFII	NS
	JET_JER_DataVsMC_MC16	NS
	JET_JER_EffectiveNP_1	NS
	JET_JER_EffectiveNP_2	NS
	JET_JER_EffectiveNP_3	NS
	JET_JER_EffectiveNP_4	NS
	JET_JER_EffectiveNP_5	NS
	JET_JER_EffectiveNP_6	NS
	JET_JER_EffectiveNP_7	NS
	JET_JER_EffectiveNP_8	NS
	JET_JER_EffectiveNP_9	NS
	JET_JER_EffectiveNP_10	NS
	JET_JER_EffectiveNP_11	NS
	JET_JER_EffectiveNP_12restTerm	NS
	JET_Pileup_OffsetMu	NS
	JET_Pileup_OffsetNPV	NS
	JET_Pileup_PtTerm	NS
	JET_Pileup_RhoTopology	NS
	JET_PunchThrough_AFII	NS
	JET_PunchThrough_MC16	NS
	JET_RelativeNonClosure_AFII	NS
	JET_RelativeNonClosure_MC16	NS
	JET_SingleParticle_HighPt	NS
	JET_JvtEfficiency	N
	JET_fJvtEfficiency	N

Table D.2: Jet related nuisance parameter list. Type is labelled "N" when the systematic is taken as a normalisation, or "NS" when it is taken both as normalisation and shape.

Object	Nuisance parameter	Type
B-tagging	FT_EFF_Eigen_B_0_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_B_1_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_B_2_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_C_0_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_C_1_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_C_2_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_C_3_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_Light_0_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_Light_1_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_Light_2_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_Eigen_Light_3_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_extrapolation_AntiKt4EMPFlowJets_BTagging201903	NS
	FT_EFF_extrapolation_from_charm_AntiKt4EMPFlowJets_BTagging201903	NS
Track	TRK_BIAS_DO_WM	NS
	TRK_BIAS_ZO_WM	NS
	TRK_BIAS_QOVERP_SAGITTA_WM	NS
	TRK_FAKE_RATE_LOOSE	NS
	TRK_FAKE_RATE_LOOSE_TIDE	NS
	TRK_EFF_LOOSE_GLOBAL	N
	TRK_EFF_LOOSE_TIDE	N
Pile-up re-weight	PRW_DATASF	N
E_T^{miss}	MET_SoftTrk_ScaleDown	One-sided
	MET_SoftTrk_ScaleUp	One-sided
	MET_SoftTrk_ResoPara	One-sided
	MET_SoftTrk_ResoPerp	One-sided

Table D.3: B-tagging, track, pile-up and MET related nuisance parameter list. Type is labelled "N" when the systematic is taken as a normalisation, or "NS" when it is taken both as normalisation and shape.