Search for the Standard Model Higgs boson produced in association with a pair of top quarks and decaying into a $b\bar{b}$-pair in the single lepton channel at $\sqrt{s}$ = 13 TeV with the ATLAS experiment at the LHC

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsprogramm ProPhys
der Georg-August University School of Science (GAUSS)

vorgelegt von

Matteo Mantoani

aus Friaul

Göttingen, 2017

Betreuungsausschuss

Prof. Dr. Arnulf Quadt

II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Ariane Frey

II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Kevin Kröninger

Lehrstuhl für Experimentelle Physik IV, Technische Universität Dortmund

Mitglieder der Prüfungskommission:

Referent: Prof. Dr. Arnulf Quadt

II. Physikalisches Institut, Georg-August-Universität Göttingen

Koreferent: Prof. Dr. Stanley Lai

II. Physikalisches Institut, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Ariane Frey

II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Wolfram Kollatschny

Institut für Astrophysik, Georg-August-Universität Göttingen

Prof. Dr. Karl-Henning Rehren

Institut für Theoretische Physik, Georg-August-Universität Göttingen

Prof. Dr. Steffen Schumann

II. Physikalisches Institut, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 28.06.2017

Referenz: II.Physik-UniGö-Diss-2017/02

Für Elisa, Unsterbliche Geliebte.

Search for the Standard Model Higgs boson produced in association with a pair of top quarks and decaying into a $b\bar{b}$-pair in the single lepton channel at $\sqrt{s}$ = 13 TeV with the ATLAS experiment at the LHC

Abstract

This thesis presents a search for Standard Model Higgs boson produced in association with top quark pairs, $t\bar{t}H$. The analysis uses 13.2 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV collected in 2015 and 2016 with the ATLAS detector at the LHC.

The considered decay mode for the Higgs boson is $H \rightarrow b\bar{b}$ and the single lepton decay channel (electron or muon) for the $t\bar{t}$ pair. The sensitivity of this channel is improved by an event categorisation according to the jet multiplicity and the number of jets containing a b-hadron decay. Multivariate techniques are used to distinguish the signal events from the background events, which are dominated by $t\bar{t}$ +jets production, in particular by the $t\bar{t} + b\bar{b}$ component.

The data are found to be consistent with both the background-only hypothesis and with the Standard Model $t\bar{t}H$ prediction. The ratio of the measured $t\bar{t}H$ signal cross-section to the Standard Model expectation is found to be $\mu = 1.6 \pm 1.1$, assuming a Higgs boson mass of 125 GeV. No significant excess of events above the background expectation is found and an observed (expected) upper limit on μ of 3.6 (2.2) is set at a 95% confidence level.

Post address:
Friedrich-Hund-Platz 1
37077 Göttingen
Germany

II. Physikalisches Institut
Georg-August-Universität Göttingen

May 2017

Contents

1 Introduction

From the beginning of its existence, humankind has asked a large number of questions about Nature. Ancient people from all civilisations looking at the sky started wondering about the mysterious appearance of the cosmos. Questions such as "How was the Universe made? How was it born? What are the fundamental building blocks of matter? How do they interact?" were asked since the time of Ancient Roman and Greek civilizations by philosophers such as Democritus and Lucretius. After two millennia, with the development of technology and knowledge, some of these questions have been answered, but many of them are still to be addressed.

Particle Physics is at present the most powerful tool to investigate these fundamental problems. Particle Physics is based on two large pillars. The first is a theoretical framework written in mathematical language. As Leibnitz said: "Quo facto, quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos computistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo (accito si placet amico) dicere: calculemus!"[1] This sentence, which summarises the essence of the philosopher's thoughts, expresses the idea that all possible scientific dilemmas can be addressed and solved just through the usage of mathematical methods. Unfortunately, this overly optimistic approach is far from sufficient when exploring Nature. This leads to the second pillar, being the approach to conduct experiments to prove the validity of theories, as stated by the scientific method. If the scientific method is one of the greatest conquests of humankind, the Standard Model (SM) can be considered one of the most important theories involved in the exploration of the fundamental questions. Its greatest success is perhaps the discovery of the Higgs boson, by both the ATLAS and

[1] if controversies were to arise, there would be no more need of disputation between two philosophers than between two calculators. For it would suffice for them to take their pencils in their hands and to sit down at the abacus, and say to each other (and if they so wish also to a friend called to help): Let us calculate!

CMS collaborations, at CERN in 2012 [1, 2]. The experimental setup employed to arrive to such a result consists of the largest and most complicated machines ever created in human history. If such a complexity is not enough, the analysis techniques employed in continuing this research journey, such as exploring the properties of the discovered Higgs boson, make use of techniques which are among the most sophisticated ever invented. These techniques were originally employed to implement software capable of simulating and reproducing the most complex object in the entire Universe: the human brain. Like Goethe said: "Den lieb ich, der Unmögliches begehrt"[2]. As an Italian scientist, the author cannot forget one of the greatest lessons of the greatest Italian poet: "fatti non foste a viver come bruti ma per seguir virtute e canoscenza"[3].
This thesis is organised as follows. In the second chapter, the basics of the first pillar are shown: the theoretical framework of the SM is described. Particular importance is given to the description of the Higgs boson and the $t\bar{t}H$ production channel, which is the main topic of the thesis. The third chapter is devoted to the description of the experimental setup of the LHC and the ATLAS experiment. Details on the data and on the modelling of the physics processes employed in the analysis are given in the fourth chapter. The definition and selection of the reconstructed objects is given in the fifth chapter. The analysis strategy is explained in great detail in the sixth chapter, as well as the theoretical framework and the implementation of the employed Multi Variate Analysis (MVA) techniques. The statistical tools used to get to the final result and the systematic uncertainties are described in the seventh chapter. Finally, the results are given and discussed in the eighth chapter.

[2] I love those who yearn for the impossible. *Faust*, Act II.
[3] you were not made to live as brutes, but to follow virtue and knowledge. Divina Commedia, Inferno, Canto XXVI

2 Theoretical Background

The discovery of the top quark in 1995 [3, 4] finally confirmed the quark model introduced for the first time in the 60s by Ne'eman, Gell-Mann and Zweig [5]. Several theories were developed since the 70s, concerning the identities and properties of the elementary constituents of matter, and also of the forces acting between them. This ensemble of theories forms what today is called *Standard Model* (SM) [6–9].

According to the SM, the elementary components that build ordinary matter, are quarks and leptons (and their anti-particles). They are spin-$\frac{1}{2}$ particles, so they are fermions, and are pointlike. The main difference between these kinds of particles is that quarks are ruled by the strong interaction, as described in the theory of Quantum-Chromo-Dynamics (QCD), while this interaction has no effect on leptons.

Quarks and leptons are paired and form three generations (see Table 2.1), which contain one pair of quarks and one pair of leptons. A quantum number called *colour* is introduced for every quark and corresponds to *blue, green, red*. Quarks are of six different *flavours* (u, d, c, s, t, b). The first generation includes the u and d quarks, together with the ν_e and e leptons. Since u and d quarks are the constituents of protons and neutrons, this generation represents all the main building blocks of ordinary matter. Particles that belong to the 2^{nd} and 3^{rd} generation appear only in cosmic events or are produced in high energy physics experiments. Interactions between members of the first generation can involve virtual particles from the 2^{nd} and 3^{rd} generations. The 2^{nd} and 3^{rd} generations are thus very important for understanding the first generation properties.

Within the SM, interactions between elementary particles are mediated by integer spin particles, that are called *gauge bosons* (See Table 2.2). QCD theory describes the strong interaction between quarks and the eight colours of the gluons, which mediate this interaction with quarks.

	Leptons		Quarks	
Electric charge [e]	q=0	q=-1	q=+2/3	q=-1/3
1st Generation	ν_e	e^-	u	d
Mass	<2 eV	0.51 MeV	2.3 MeV	4.8 MeV
2nd Generation	ν_μ	μ^-	c	s
Mass	<2 eV	105.66 MeV	1.275 GeV	95 MeV
3rd Generation	ν_τ	τ^-	t	b
Mass	<2 eV	1.77 GeV	173.5 GeV	4.65 GeV

Table 2.1.: The generations of quarks and leptons. Numbers are taken from [10].

Fundamental interaction	Carrying particle	Symbol	Mass (GeV)	Spin
Strong	8 gluons	g	0	1
Electromagnetic	photon	γ	0	1
Weak	W bosons	$W^\pm$	80.385	1
	Z boson	Z	91.1876	1
Gravitational	graviton	G	0	2

Table 2.2.: Fundamental interactions and gauge bosons. The graviton is currently only a hypothetical particle. Numbers are taken from [10].

At low energies the electromagnetic and weak interactions appear distinct from one another and the theory that describes the weak interaction is called Fermi V-A theory of the β-decay. The need to unify these two interactions comes from a problem in the Fermi theory of the β-decay: considering as an example a process such as the neutrino-electron scattering and computing the cross section, it is found that:

$$\sigma(\nu e^-) = G_F^2 \frac{s}{\pi}\,, \tag{2.1}$$

where G_F is the Fermi constant and $s = (p_e^\mu + p_\nu^\mu)^2$, where p_e^μ and p_ν^μ are the particles' four-momenta. From this equation, it follows that the cross section increases with energy. A general theorem of scattering theory implies that, at a given energy, the cross section must be:

$$\sigma \leq 4\pi \lambda\!\!\!^{-2}\,, \tag{2.2}$$

where $\lambda\!\!\!^{-} = \frac{\hbar}{|\vec{p}|}$ in the centre-of-mass frame. At high energies, where three-momentum, $\vec{p}$, of particles satisfies $p^2 \backsimeq s$. Looking at equation (2.1) it can be seen that using the inequality (2.2) for the cross section the unitarity bound is violated for $\sqrt{s} \geq 300$ GeV, called *Fermi Scale*. This issue would be avoided if there were intermediate bosons which mediate the weak interaction (W and Z bosons). This implies that, in other words, the Fermi theory is the low-energy limit of a more general theory that describes the phenomena occurring at the Fermi Scale. At present, we know that this theory is the *Unified ElectroWeak (EW)* theory, which together with QCD forms the Standard Model.
The gravitational interaction has not been included in the mathematical framework of the SM so far, however its effect is negligible at these scales.
The theoretical formalism on which the SM is based is called Quantum Field Theory (QFT), and makes use of quantum mechanics and special relativity. In QFT, $fields$ are mathematical quantities which have a value in every point of space and time, and *particles* are seen as excitations of the $fields$. Two ideas are fundamental in QFT: the *local gauge symmetry* and the *spontaneous symmetry breaking*. The first concept describes how fields behave under special transformations called *gauge transformations*, which operate in every point of the space-time. The second one is discussed at length in the following section (2.2).
The SM is a very successful theory and it has been used for making many computations for physics processes and, so far, it has not been refuted by any experiment. However, the evidence of the neutrino oscillations suggests that the SM is incomplete [11]. There

are other observations that prove that SM is not a complete theory. They are discussed in Sec. 2.3.

2.1. A Brief Theoretical Overview of the Standard Model

In the theory of weak interaction, left-handed particles form a doublet of a SU(2) symmetry, and right-handed particles are singlets. Starting from the first lepton-generation, the following notation for the left-handed doublet is introduced:

$$X_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L . \tag{2.3}$$

The singlet, instead, is denoted as e_R. Using this notation, the left-handed doublet transformation law in the SU(2) symmetry group, is:

$$X'_L = e^{i\overline{\epsilon(x)}\cdot\overline{T}} X_L \,, \tag{2.4}$$

where $T_i = \tau_i/2$ and τ_i are the Pauli matrices. The current triplet is defined as:

$$J^i_\mu(x) = \overline{X}_L \gamma_\mu \frac{1}{2} \tau_i X_L \,. \tag{2.5}$$

Defining:

$$J^+_\mu = \overline{X}_L \gamma_\mu \tau_+ X_L \tag{2.6}$$

and:

$$J^-_\mu = \overline{X}_L \gamma_\mu \tau_- X_L \tag{2.7}$$

with $\tau_\pm = \frac{1}{2}(\tau_1 \pm i\tau_2)$ the charged currents are obtained, and with:

$$J^3_\mu = \overline{X}_L \gamma_\mu \frac{\tau_3}{2} X_L \tag{2.8}$$

so are the neutral current. Neutral weak currents are observed and they use both e_L and e_R, but, in the definition of equation 2.8, they seem to interact only with left-handed particles. To solve this problem, it is necessary to combine them with some other currents that also use the right-handed fermions. Those are the electromagnetic currents, that have equation:

$$J_\mu^{em} = \overline{\psi}\gamma_\mu Q\psi \tag{2.9}$$

where Q is the electrical charge of the particle. To combine them, a new quantity is introduced, the hypercharge:

$$Y = 2(Q - T_3)\,. \tag{2.10}$$

This quantity is associated to the following current:

$$J_\mu^Y = 2(J_\mu^{em} - J_\mu^3) = -\overline{X}_L\gamma_\mu X_L - 2\overline{e}_R\gamma_\mu e_R\,. \tag{2.11}$$

The commutation relations between the new Y and old T matrices, are:

$$[T_+, T_-] = 2T_3 \tag{2.12}$$

$$[T_3, T_\pm] = \pm T_\pm \tag{2.13}$$

$$[Y, T_\pm] = [Y, T_3] = 0\,. \tag{2.14}$$

At this point, the hypercharge generates the $U(1)_Y$ group, for which phase transformations are:

$$X_L' = e^{i\beta(x)Y}X_L = e^{i\beta(x)y_L}X_L \tag{2.15}$$

$$e_R' = e^{i\beta(x)Y}e_R = e^{i\beta(x)y_R}e_R\,. \tag{2.16}$$

The $U(1)_Y$ group symmetry combines with the $SU(2)_L$ to create the $SU(2)_L \times U(1)_Y$ local group symmetry, which transformations are:

$$X_L' = e^{i\overline{\epsilon(x)}\cdot\overline{T} + i\beta(x)Y}X_L \tag{2.17}$$

$$\psi_R' = e^{i\beta(x)Y}\psi_R\,. \tag{2.18}$$

A new derivative is also defined to impose the Lagrangian invariance:

$$D_\mu X_L = (\partial_\mu + ig\frac{\overline{\tau}}{2} \cdot \overline{W}_\mu + i\frac{g'}{2} y_L B_\mu) X_L \tag{2.19}$$

$$D_\mu \psi_R = (\partial_\mu - ig' B_\mu) e_R \,. \tag{2.20}$$

Thus, the Lagrangian is:

$$\mathcal{L} = \overline{X}_L i\gamma_\mu \partial^\mu X_L + \overline{e_R} i\gamma_\mu \partial e_R - \frac{1}{4}\overline{W}_{\mu\nu}\overline{W}^{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \mathcal{L}_{int} \tag{2.21}$$

where, using $W^\pm_\mu = \frac{W^1_\mu \mp W^2_\mu}{\sqrt{2}}$ and the Weinberg angle θ_W:

$$\overline{W}_{\mu\nu} = \partial_\mu \overline{W}_\nu - \partial_\nu \overline{W}_\mu - g\overline{W}_\mu \times \overline{W}_\nu \tag{2.22}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.23}$$

$$\mathcal{L}_{int} = -\frac{g}{\sqrt{2}}(J^+_\mu W^{-\mu} + J^-_\mu W^{+\mu}) - \frac{g}{\cos\theta_W}(J^3_\mu - \sin^2\theta_W J^{em}_\mu) Z^\mu - e J^{em}_\mu A^\mu \,. \tag{2.24}$$

This Lagrangian, does not contain the mass terms of the particles, which for the moment are massless.

2.2. Spontaneous Symmetry Breaking

One of the most fundamental ideas of particle physics is that interactions among fundamental particles are described by symmetry principles. Using these symmetries, some conservation laws in nature are imposed. For example, electroweak interaction is based on a symmetry between gauge bosons of electromagnetic and weak interaction, but this symmetry cannot be seen in nature because W and Z bosons have mass and photons are massless. The explanation is that a short time after the Big Bang, electromagnetic and weak interactions were actually identical and carried by four massless particles. During the cooling down of the Universe, symmetry was spontaneously broken and these particles diversified. The concept of spontaneous symmetry breaking is borrowed from condensed matter physics, where it refers to phase transitions, for example the transition from water to ice. Water molecules are in a spherical symmetry situation because being free to move, they assume the same properties independently of their position. On the contrary, in ice, molecules are not free to move and they form a rigid structure. It is thus

possible to say that symmetry breaks in the transition between liquid water and ice. The breaking is due to inter-molecular interactions, that are negligible at higher temperature where molecules have greater kinetic energy, but they become important at lower temperatures. Between 1959 and 1966, several physicists (such as Nambu, Goldstone, Higgs and Weinberg) [12, 13] noticed that the principle of spontaneous symmetry breaking could be applied to particle physics to generate the mass of W and Z bosons and also for fermion masses.

2.2.1. Spontaneous Breaking of a Global Gauge Symmetry: Goldstone Model

Considering a complex electrically charged field, ϕ, the starting Lagrangian, involving the derivatives of the fields and a potential term expressed as a function of the fields themselves, can be:

$$\mathcal{L} = (\partial_\mu \phi^*)(\partial^\mu \phi) - V\,, \tag{2.25}$$

where V is the potential energy, defined as:

$$V = \mu^2 \phi^* \phi + \lambda(\phi^* \phi)^2\,. \tag{2.26}$$

Now, imposing:

$$\phi = \frac{\phi_1 + i\phi_2}{\sqrt{2}} \tag{2.27}$$

$$\phi^* = \frac{\phi_1 - i\phi_2}{\sqrt{2}} \tag{2.28}$$

one obtains:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi_1)^2 + \frac{1}{2}(\partial_\mu \phi_2)^2 - \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) - \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2\,, \tag{2.29}$$

The global gauge symmetry transformations are defined as:

$$\phi'(x) = e^{i\epsilon}\phi(x) \tag{2.30}$$

$$\phi^{*'}(x) = e^{-i\epsilon}\phi^*(x) \tag{2.31}$$

where ϵ is a real coordinates-independent parameter.
Minimising the potential energy V, one obtains two cases, Fig. 2.1:

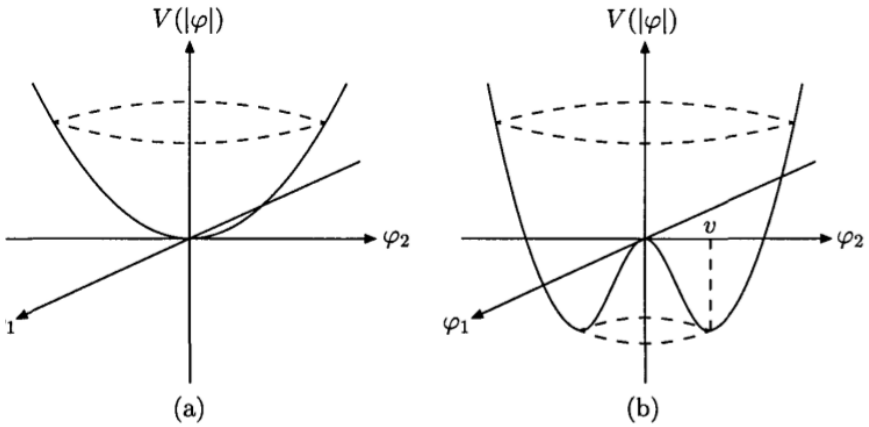


Figure 2.1.: V in the case where (a) $\mu^2 > 0$ and (b) $\mu^2 < 0$.

- $\mu^2 > 0$: in this case the stable equilibrium configuration is given by $|\phi| = 0$ and the vacuum $|0\rangle$ is unique and it respects the symmetry;

- $\mu^2 < 0$: in this case the equilibrium configuration is $|\phi|_0 = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}} \neq 0$ with $v = \sqrt{\frac{-\mu^2}{\lambda}} > 0$. This gives a condition on $|\phi|$, so every configuration $\phi = \frac{v}{\sqrt{2}} e^{i\alpha}$ where α is a real number, is a stable configuration. In this case there are infinite degenerate vacuum states and vacuum $|0\rangle$ is not symmetry invariant.

Considering the second case, setting $\alpha = 0$ and $\phi_1|_0 = v$, $\phi_2|_0 = 0$:

$$\phi(x) = \frac{\phi_1 + i\phi_2}{\sqrt{2}} = \frac{v + \sigma(x) + i\eta(x)}{\sqrt{2}} \,. \tag{2.32}$$

with $\sigma(x)$ and $\eta(x)$ real functions with zero vacuum expectation value. The new Lagrangian is:

$$\mathcal{L}(\sigma, \eta) = \frac{1}{2}(\partial_\mu \sigma)^2 - \frac{1}{2}(-2\mu^2)\sigma^2 + \frac{1}{2}(\partial_\mu \eta)^2 - \lambda v(\sigma^2 + \eta^2)\sigma - \frac{\lambda}{4}(\sigma^2 + \eta^2)^4 \tag{2.33}$$

The particles produced in this way are a chargeless scalar boson with mass $m_\sigma^2 = 2\lambda v^2 = -2\mu^2 > 0$ and a chargeless massless scalar one m_η^2 (the Goldstone boson).

2.2.2. Spontaneous Breaking of a Local Gauge Symmetry: Higgs Model

One now wants to verify the spontaneous breaking of a local gauge symmetry. This mechanism has two main effects:

- It generates the masses for vector gauge bosons;
- It eliminates the Goldstone bosons.

Different gauge transformations are defined:

$$\phi'(x) = e^{i\epsilon(x)}\phi(x) \tag{2.34}$$

$$\phi^{*'}(x) = e^{-i\epsilon(x)}\phi^*(x) \tag{2.35}$$

in which $\epsilon(x)$ now has a x dependence. In order to impose the invariance of the Lagrangian under these transformations, the definition of the covariant derivative is changed to:

$$D'_\mu = \partial_\mu - \frac{1}{e}A_\mu \tag{2.36}$$

with A_μ transforms as:

$$A'_\mu = A_\mu + \frac{1}{e}\partial_\mu\epsilon. \tag{2.37}$$

The final Lagrangian with these changes is:

$$\mathcal{L}(\phi, D_\mu; \phi^*, D^*_\mu) = (D_\mu\phi)^*(D^\mu\phi) - (\mu^2\phi^*\phi) - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \tag{2.38}$$

where:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{2.39}$$

As before, one looks for the configuration with minimal energy. There are two cases:

- $\mu^2 > 0$: in this case $\phi|_0 = 0$ is the solution with stable equilibrium and one has the electrodynamics of spinless particles with electrical charge;
- $\mu^2 < 0$: in this case one has spontaneous symmetry breaking. The minimal configuration is $\langle 0|\,\phi(x)\,|0\rangle = \frac{v}{\sqrt{2}}e^{i\alpha}$, so on obtains infinite states of degenerate minimal

energies.

Choosing the second case, and defining:

$$\phi(x) = \frac{v + \sigma(x) + i\eta(x)}{\sqrt{2}}\,, \tag{2.40}$$

the following Lagrangian is obtained:

$$\begin{aligned}\mathcal{L} = \frac{1}{2}(\partial_\mu\sigma)^2 + \frac{1}{2}(\partial_\mu\eta)^2 - \frac{1}{2}(2\lambda v^2)\sigma^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}e^2v^2A_\mu A^\mu + e(\eta\partial_\mu\sigma - \sigma\partial_\mu\eta)A^\mu + \\ + \frac{1}{2}e^2(\sigma^2+\eta^2)A_\mu A^\mu + e^2v\sigma A^\mu A_\mu - \lambda v\sigma(\sigma^2+\eta^2) - \frac{1}{4}\lambda(\sigma^2+\eta^2)^2 - evA^\mu\partial_\mu\eta.\end{aligned} \tag{2.41}$$

The generated particles are:

- A scalar particle (σ) with mass $m_\sigma^2 = 2\lambda v^2$;
- A massless Goldstone boson (η);
- A gauge boson with mass $M = ev$ and three different polarisations.

In total, one has five degrees of freedom, but having started only with four degrees of freedom. To eliminate the extra degree of freedom introduced with this formalism, it is possible to introduce the following gauge transformation:

$$\phi(x) = \frac{1}{\sqrt{2}}\varphi(x)e^{\frac{i\theta(x)}{v}} \tag{2.42}$$

where $\theta(x)$ is chosen in a way to have a real $\varphi(x)$. To have an invariant Lagrangian, one has to define another derivative:

$$D_\mu(B) = \partial_\mu - ieB_\mu \tag{2.43}$$

where $B_\mu = A_\mu - \frac{1}{ev}\partial_\mu\theta$. Taking into account that $F_{\mu\nu}(A) = F_{\mu\nu}(B)$, the new Lagrangian is obtained:

$$\begin{aligned}\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 - \frac{1}{2}m_h^2h^2 - \frac{1}{4}F_{\mu\nu}(B)F^{\mu\nu}(B) + \frac{1}{2}e^2v^2B_\mu B^\mu + e^2vhB_\mu B^\mu + \\ + \frac{1}{2}e^2h^2B^\mu B_\mu - \lambda vh^3 - \frac{\lambda}{4}h^4\,,\end{aligned} \tag{2.44}$$

where $\phi(x) = v + h(x)$. Thus, starting from a scalar complex field and a real massless vector boson, a massive chargeless scalar field and a massive gauge boson (for a total of four degrees of freedom) where λ and v are arbitrary parameters, are obtained. The particles masses are $m_h^2 = 2\lambda v^2$ and $M_B^2 = e^2 v^2$. The Goldstone boson has been eliminated and this gauge is called Unitary Gauge.

2.2.3. Spontaneous Breaking of a SU(2) Gauge Symmetry: Non-Abelian Higgs Model

The mechanism for the electroweak symmetry breaking considered in the SM is here described. Consider the following Lagrangian:

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \,, \tag{2.45}$$

where ϕ is a doublet of complex scalar fields:

$$\phi = \begin{pmatrix} \phi_\alpha \\ \phi_\beta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.46}$$

Requiring that $\mathcal{L}$ be invariant under local gauge transformations, for real arbitrary functions $\epsilon^k(x)$, one has:

$$\phi'(x) = e^{i \sum_{k=1}^{3} \epsilon^k(x) T^k} \phi(x) \,, \tag{2.47}$$

where $T^k = \frac{\tau^k}{2}$ and obeys the commutation relation $[T^i, T^j] = i\epsilon^{ijk} T^k$ for $i, j, k = 1, 2, 3$. One has to change the definition of the covariant derivative to have an invariant Lagrangian:

$$D_\mu = \partial_\mu + ig \frac{\overline{\tau}}{2} \cdot \overline{W}_\mu \tag{2.48}$$

where $W_\mu^{k=1,2,3}$ are the gauge fields. In the end, the following Lagrangian is obtained:

$$\mathcal{L} = (D_\mu^\dagger)(D^\mu) - V(\phi) - \frac{1}{4} \overline{W}_{\mu\nu} \cdot \overline{W}^{\mu\nu} \tag{2.49}$$

where $\overline{W}_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - g \overline{W}_\mu \times \overline{W}_\nu$ and $V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2$. In the case of spontaneous symmetry breaking (where $\mu^2 < 0$) the minimum of the potential is:

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{\lambda}. \tag{2.50}$$

Choosing to expand ϕ around the minimum, one gets:

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix}. \tag{2.51}$$

This is not invariant under symmetry transformations and so the symmetry is broken. Deviations of the Higgs field from the minimum are parametrised in this way:

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v+h(x)\end{pmatrix}. \tag{2.52}$$

This final model opens the way for a formulation of a theory of weak interactions based on spontaneous symmetry breaking for the generation of three massive bosons.

2.2.4. Gauge Boson Masses

The mechanism of generation of boson masses has to give mass to the $W^{\pm}$ and Z^0, while preserving $m_\gamma = 0$. This is achieved by the term:

$$\mathcal{L}_\phi = (D_\mu^\dagger)(D^\mu) - V(\phi) = \left\|(\partial_\mu + ig\overline{T}\cdot\overline{W}_\mu + i\frac{g'}{2}YB_\mu)\phi\right\|^2 - V(\phi) \tag{2.53}$$

to the initial Lagrangian, where:

$$V(\phi) = \mu^2\phi^+\phi + \lambda(\phi^+\phi)^2\,. \tag{2.54}$$

Choosing $\phi = \begin{pmatrix}0\\v+h\end{pmatrix}$, one obtains:

$$\begin{aligned}\mathcal{L}_\phi = \frac{1}{2}(\partial_\mu h)^2 + \frac{g^2}{8cos^2\theta_W}(v+h)^2Z^\mu Z_\mu + \frac{g^2}{4}(v+h)^2W_\mu^+W^{-\mu} - \\ -\frac{1}{2}(-2\mu^2)h^2 + \lambda v h^3 + \frac{\lambda}{4}h^4\,.\end{aligned} \tag{2.55}$$

Thus bosons masses have been generated with $M_W = \frac{gv}{2}$, $M_Z = \frac{gv}{2cos\theta_W}$, $M_\gamma = 0$ and $M_h = \sqrt{2v\lambda}$. The mass of the Higgs boson is not predicted by this model, since it depends on the free parameter λ.

2.2.5. Fermion Masses

The Yukawa coupling of leptons with the Higgs doublet is defined by:

$$\mathcal{L}_{l\phi} = -g_l^{ij}\overline{X}_L^i \phi l_R^j + h.c.\,, \tag{2.56}$$

where g_l^{ij} is a real diagonal matrix, l_R^j are right-handed lepton fields and X_L are the doublets for every lepton family L:

$$X_L = \begin{pmatrix} \nu_L \\ l_L \end{pmatrix}. \tag{2.57}$$

In this way, the masses of leptons are generated, $m_L = g_L \frac{v}{\sqrt{2}}$, and the couplings are defined by:

$$g_L = \frac{\sqrt{2}}{v} m_L\,. \tag{2.58}$$

They are proportional to the mass of the particles considered, such as in the case of the gauge bosons. For the quarks, in a similar way as before, it is possible to consider the following Yukawa coupling Lagrangian:

$$\mathcal{L}^{quark} = -g_d^{ij}\overline{Q}_L^i \phi D_R^j - g_u^{ij}\overline{Q}_L^i \tilde{\phi} U_R^j + h.c. \tag{2.59}$$

where g^{ij} are real constants, $\tilde{\phi} = -i(\phi^\dagger \tau_2)^T$, Q_L^i are the left-handed doublets for each quark generation and D_R^j are right-handed singlets. With these terms, after the symmetry breaking, the Lagrangian becomes:

$$\mathcal{L}^{quark} = -g_d^{ij}\frac{v}{\sqrt{2}}\overline{D}_L^i D_R^j - g_u^{ij}\frac{v}{\sqrt{2}}\overline{U}_L^i U_R^j - g_d^{ij}\frac{1}{\sqrt{2}}\overline{D}_L^i D_R^j h - g_u^{ij}\frac{1}{\sqrt{2}}\overline{U}_L^i U_R^j h + h.c. \tag{2.60}$$

which can be written as:

$$\mathcal{L}^{quark} = -\overline{D}_L M_d D_R - \overline{U}_L M_u U_R - \frac{1}{v}(\overline{D}_L M_d D_R h - \overline{U}_L M_u U_R)h + h.c. \tag{2.61}$$

where $M_{d/u} = \frac{v}{\sqrt{2}} g_{d/u}$ are complex, non-diagonal hermitian matrices that can be diagonalised using 4 unitary matrices U_u,U_d,V_u,V_d as follows:

$$M_u^{diag} = U_u^\dagger M_u V_u \tag{2.62}$$

$$M_d^{diag} = U_d^\dagger M_d V_d \tag{2.63}$$

Coupling	Strength
$hf\bar{f}$	$\frac{m_f}{v}$
hVV	$\frac{2}{v}m_V^2$
$hhVV$	$\frac{2}{v^2}m_V^2$
hhh	$\frac{3}{v}m_h^2$
$hhhh$	$\frac{3}{v^2}m_h^2$

Table 2.3.: Couplings of the Higgs field with fermions, vector gauge bosons and self couplings [10].

and M^{diag} elements are real and represent the physical masses of the quarks.

$$\mathcal{L}^{quark} = -\overline{D}'_L M_d^{diag} D'_R - \overline{U}'_L M_u^{diag} U'_R - \frac{1}{v}(\overline{D}'_L M_d^{diag} D'_R - \overline{U}'_L M_u^{diag} U'_R)h\,, \tag{2.64}$$

where $M_{d/u}^{diag} = \frac{v}{\sqrt{2}} U_{d/u}^{\dagger} g_{d/u} V_{d/u}$. As a result, defining the Cabibbo-Kobayashi-Maskawa (CKM) matrix, $V_{CKM} = U_{d/u}^{\dagger} V_{d/u}$, it is possible to describe the mixing between quark flavours via the charged weak interaction.
In general, the coupling of the Higgs field to a fermion, f, can be written as:

$$g_{fh} = \frac{\sqrt{2}}{v} m_f\,. \tag{2.65}$$

The CKM matrix is unitary and can be defined starting from 3 rotation angles and 1 complex phase.

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \times \begin{pmatrix} c_{13} & 0 & s_{13}e^{-1\delta'} \\ 0 & 1 & 0 \\ -s^{13}e^{i\delta'} & 0 & c_{13} \end{pmatrix} \times \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.66}$$

where $c_{ij} = \cos\phi_{ij}$ and $s_{ij} = \sin\phi_{ij}$.
Table 2.3 summarises the vertex factors for the interaction with the Higgs field.

2.3. Beyond the Standard Model

Although it is a very successful theory, the SM alone still cannot explain some open problems of particle physics. They can be summarised as follows:

Dark matter and dark energy: The fact that the velocity of rotation of galaxies does not decrease as a factor $r^{-1/2}$ gives an indication that their mass is not concentrated only in the central bulge [14] and $M(r) \propto r$. The significant non-luminous component of the galaxies is referred to as *dark matter*. More evidences of the existence of dark matter come from the precision measurements of the small fluctuations in the Cosmic Microwave Background (CMB). Theoretical models describe that only 5% of the total mass of the universe is visible, thus explainable in terms of the SM. While dark matter accounts for 23% of the universe, the majority, 72%, comes from *dark energy* [15], which can explain the acceleration of the expansion of our Universe. It is possible to extend the SM, for example with Supersymmetry (SUSY) [16], by introducing Weakly Interacting Massive Particles (WIMP), that could explain the existence of cold dark matter. At present, there is no robust explanation of dark energy.

Unification of forces: The strengths of the three forces described by SM change with energy scale, this behaviour is known as *running coupling*. In particular, g_W and g_S decrease with energy while g_{EM} increases. Thus, the running of the coupling constants could bring them together at a certain scale. The Grand Unified Theory (GUT) theory, which unifies the forces all together, can exist if the couplings converge at a certain scale. In the simplest symmetry group which includes the interactions all together, SU(5), the couplings do not converge at the same point, whereas this happens when including the particles predicted by SUSY, and the couplings meet at an energy scale of $\sim$ 1 TeV. Finally, gravity is not considered in any of these theories, but it is expected to become important at a scale known as the Planck Scale, $\Lambda_p \sim 10^{16}$ GeV. A theory in which all the four forces are unified, Theory of Everything (TOE), still does not exist.

Hierarchy problem: The mass of the Higgs boson is much smaller than the GUT mass scale, although it is expected that the large quantum contributions to the square of the Higgs boson mass would inevitably make the mass comparable to the scale at which new physics appears. So a very precise *fine-tuning* cancellation between the quadratic radiative corrections and the bare mass is required in order to keep the

mass at an EW scale. The SM alone cannot explain such corrections, while SUSY could do so.

Matter-antimatter asymmetry: There are indications of the existence of an asymmetry of particles and anti-particles in nature, which goes beyond the explanations that are possible to obtain through the observed Charge Parity (CP) violation in the flavour sector [17]. Thus, the SM alone cannot describe these additional CP violating effects.

Neutrino masses: Neutrino oscillations, first predicted by Bruno Pontecorvo in 1952 [18], had many experimental confirmations, and led to the Nobel Prize in Physics in 2015 [19]. Neutrino oscillations imply that neutrinos have mass, while in the SM they are massless. If neutrinos are *Majorana particles*, they can acquire mass through the *seesaw mechanism* [20]. In this case β decay without neutrinos can occur. At the moment much effort is dedicated to finding such a decay.

Other, more qualitative, motivations can be raised to support the idea that the SM is not a complete theory. In fact, the SM is based on the measurements of some parameters which are not predicted by the theory itself. These parameters are:

- v: Vacuum Expectation Values (VEV) of the Higgs potential; using the measured W boson mass and the coupling g_W, $v \simeq 246$ GeV;
- λ: quartic Higgs-self-coupling constant;
- θ_W: the Weinberg angle;
- g: one of the two coupling constants of the $SU(2)_L \times U(1)_Y$ symmetry (the other one is related to this through the Weinberg angle);
- quark and lepton masses;
- $(n-1)^2$ degrees of freedom from the CKM matrix, where n is the number of quark generations.

If one assumes that neutrinos are massless, nine mass parameters are obtained together with four parameters from the CKM matrix, for a total of 17. It is important to stress that fermion masses are free parameters of the SM, and have to be measured. At the moment, there is no theory that gives *a priori* a value to these free parameters and they have to be determined experimentally. SUSY is one of the most attractive extensions

of the SM, resolving many different issues such as described previously, but this theory predicts at least five physical Higgs bosons (A, h, H, H^+, H^-), together with a large number of additional superpartners and new mixing angles, increasing the number of free parameters to at least 115.

2.4. The Top Quark

Being the heaviest known elementary particle, the top quark plays an important role in the SM, especially in the sector of electroweak symmetry breaking. It was discovered in 1995 by the CDF and DØ collaborations at the Tevatron accelerator located at Fermilab [3, 4]. At present, the most recent result based on the combination of the top quark mass measurements performed at the Large Hadron Collider (LHC) and the Tevatron is m_t =173.21 ± 0.51 ± 0.71 GeV [10]. Even if this mass is not predicted by the SM, the other properties of the top quark are well described by the theory. A deviation from these predictions could lead to the discovery of new physics beyond the SM. For this reason, much effort is placed in measuring the electric charge, decay width, difference between top and anti-top quark masses, $t\bar{t}$ charge asymmetry and spin correlations, the helicity of the W boson from top quark decays, coupling to other particles and decay branching fractions. The top quark has a very short lifetime, ($\simeq 10^{-25}$s), thus it decays before hadronising and does not form any bound states as the other quarks do. The consequence is that its properties are directly propagated to its decay products, making them accessible by dedicated experiments [10]. At present, the measured properties of the top quark are compatible with the predictions of the SM. Special attention is given to the Yukawa coupling, which is $\simeq 1$. This could be a random feature of nature, but could also be an indication of some deeper unknown properties, which can be explored in new physics theories.

2.4.1. Top Quark Production and Decay

The top quark production at hadron colliders primarily occurs through strong interactions and in association with an anti-top quark. Single top production can also occur through electroweak interactions. The four Leading Order (LO) Feynman diagrams for $t\bar{t}$ production via strong interaction are shown in Figure 2.2.

At the Tevatron, a $p\bar{p}$ collider, $q\bar{q}$ annihilation was the most important production channel for the discovery of the top quark, while at the LHC, the gluon fusion process dominates, accounting for 80-90% of the cross section, depending on $\sqrt{s}$. The theoretical cross sec-

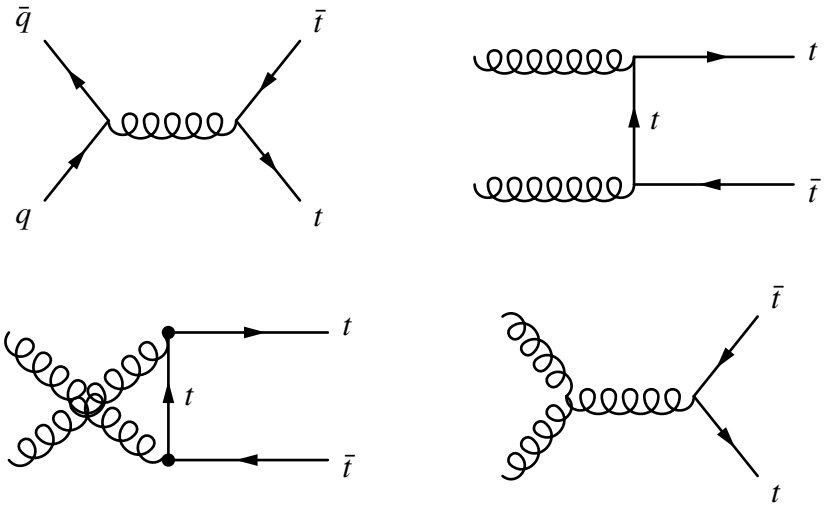


Figure 2.2.: The four LO Feynman diagrams for $t\bar{t}$ production at hadron colliders.

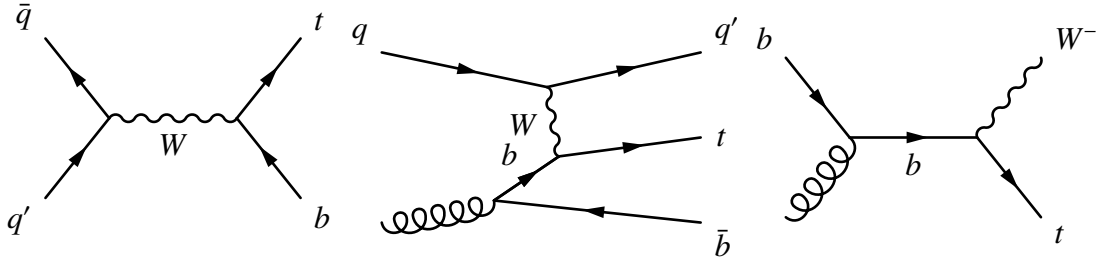


Figure 2.3.: The Feynman diagrams for single top production at hadron colliders.

tion of 832^{+46}_{-51} pb is calculated at $\sqrt{s}$ =13 TeV at Next-to-Next-to-Leading-Order (NNLO) in QCD and includes resummation of Next-to-Next-to-Leading-Logarithmic (NNLL) soft gluon terms [21–25]. Theoretical uncertainties on this cross section result from variations of the factorisation and renormalisation scales, as well as from uncertainties on the Parton Distribution Functions (PDFs) and α_s.

The Feynman diagrams related to the single top quark production are shown in Fig. 2.3. The large amount of background and the low cross section make the studies of single top quark challenging. Its predicted cross section is also lower than the one of $t\bar{t}$ production [26]. Measurements of single top quark production allow the study of the properties of the Wtb vertex, giving the possibility to directly measure the CKM matrix element $|V_{tb}|$.

Top quarks decay almost exclusively through $t \to Wb$. As a consequence, the W boson decay modes define the signature of the $t\bar{t}$ final states. For $t\bar{t}$ decays, the final states are categorised according to the number of charged leptons, shown in Fig. 2.4. Since τ leptons can both decay leptonically and hadronically, final states containing this lepton are treated separately and not included in this classification. In summary, the possible

final states from the decay of the $t\bar{t}$ system are:

Dilepton channel: Both W bosons decay into leptons and neutrinos. This channel has the lowest branching ratio, but it permits a very high $t\bar{t}$ purity.

Single lepton channel: One W boson decays leptonically and the other hadronically. The signature of the final state is characterised by the presence of one lepton, a neutrino, two b-jets and two light-jets.

All hadronic channel: Both W bosons decay hadronically, so the final state contains only jets, two of which are b-jets. The branching ratio is $\simeq$ 46% and its main background is coming from multijet production.

τ leptons that decay leptonically give final states that are experimentally included in dilepton and single lepton channels.

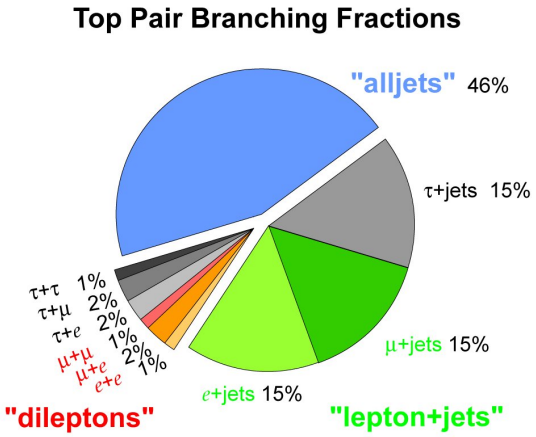


Figure 2.4.: The possible $t\bar{t}$ decay topologies.

2.5. The Higgs Boson

The discovery of a particle compatible with the Higgs Boson has been announced by both of the two experiments at CERN: A Toroidal LHC ApparatuS (ATLAS) and Compact Muon Solenoid (CMS). The announcement was given on July 4$^{\text{th}}$, 2012 [1, 2]. For the ATLAS collaboration, this observation was performed using 5.85 fb^{-1} of pp collision data recorded during April to June 2012 at a centre-of-mass energy of 8 TeV combined with the 4.7 fb^{-1} recorded in 2011 at a centre-of-mass energy of 7 TeV.
An excess of events with significance 5.9σ in the mass range of 122-131 GeV was observed. The excess was driven by the two channels with the highest mass resolution and lowest background, $H \rightarrow ZZ^{(*)} \rightarrow 4l$ and $H \rightarrow \gamma\gamma$, and by the $H \rightarrow WW^{(*)} \rightarrow lvlv$ channel, characterised by a lower mass resolution. Taking into account the entire mass range of the search, 110-660 GeV, the global significance of the excess was 5.1σ. These results provided conclusive evidence for the discovery of a new particle with mass $126.0 \pm 0.4\,(stat) \pm 0.4\,(sys)$ GeV. Until now, the measurements of the properties of this particle have supported the assumption that it is indeed the SM Higgs boson. However, to confirm the SM hypothesis, all of the accessible production and decay rates need to be measured and finally compared to the SM predictions.

2.5.1. Higgs Boson Production and Decay

As already stated in the previous sections, the SM Higgs boson couples primarily to W and Z bosons, t and b quarks. Hence, at the LHC, Higgs boson production is achieved through the processes explained below. The theoretical cross sections refer to a mass of 125 GeV for the Higgs boson and a centre-of-mass energy of 13 TeV [27].

Gluon-gluon fusion, ggF: This is the dominant process where $gg \rightarrow H$, Fig. 2.5a. Here the internal loop is dominated by top quarks. Its cross section is $48.58^{+4.6\%}_{-6.7\%}$ pb.

Vector boson fusion, VBF: Is the second most important production channel, where $qq \rightarrow qqH$ and consists of the annihilation of two virtual vector bosons (W or Z) which create a Higgs boson, Fig. 2.5b. The cross section for this process is $3.78 \pm 2.7\%$ pb.

Associated production with a vector boson, VH: The considered process is $q\bar{q} \rightarrow VH$, Fig. 2.5c. In case of the association with a W boson the cross section is $1.37 \pm 1.9\%$ pb and in case of a Z boson is $0.88^{+3.8\%}_{-3.1\%}$ pb.

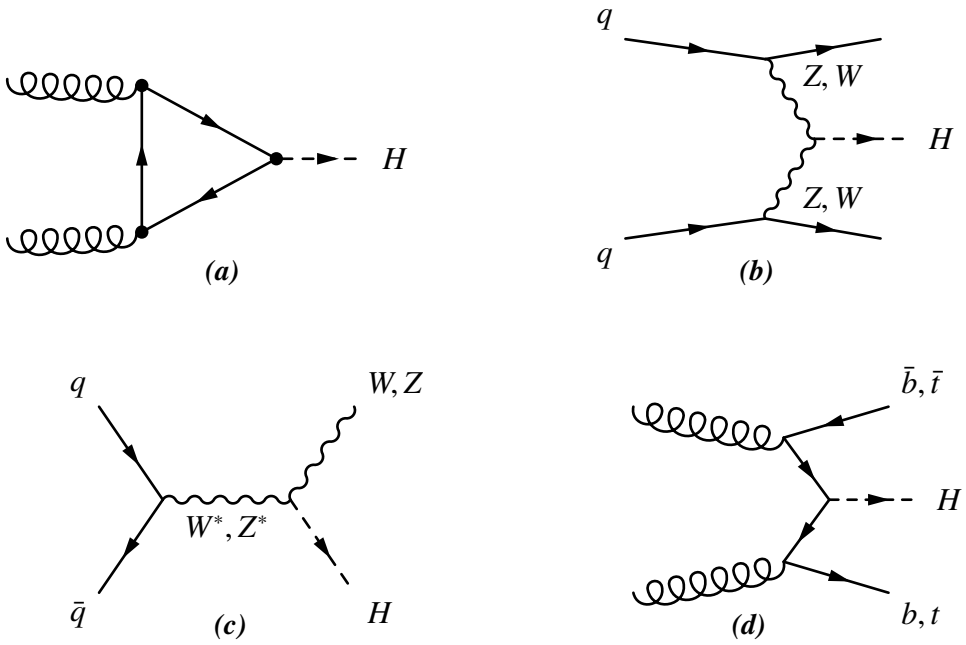


Figure 2.5.: Feynman diagrams for Higgs production at the LHC.

Associated production with two heavy quarks, qqH: Here the Higgs boson is produced in association with a heavy quark pair, and the process is dominated by the $t\bar{t}H$ process, Fig. 2.5d. The cross section for $t\bar{t}H$ is $0.507^{+5.8\%}_{-9.2\%}$ pb.

Fig. 2.6a shows the different production cross sections as a function of $\sqrt{s}$ at the LHC. The branching ratios of the possible Higgs boson decay modes in the SM depend on

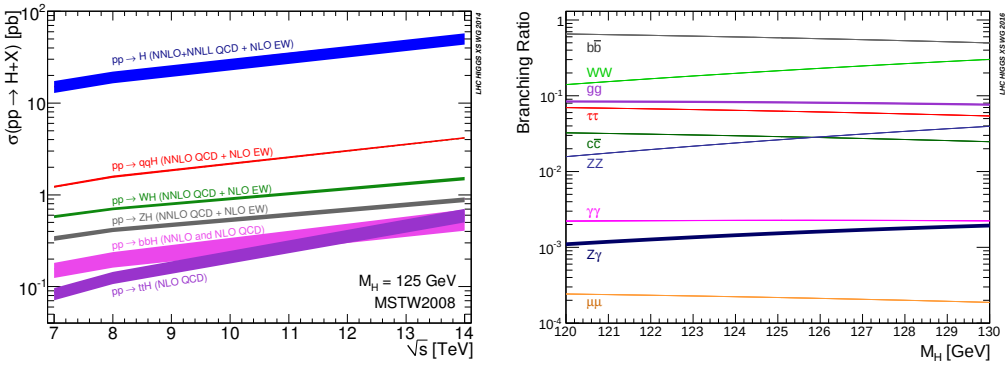
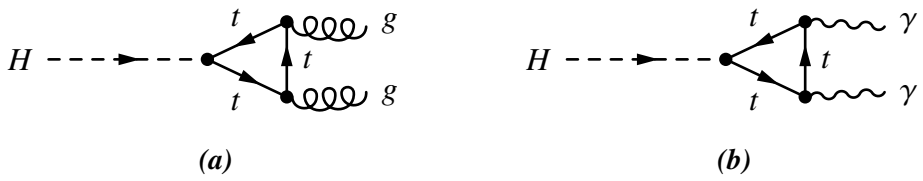


Figure 2.6.: (a) Higgs boson production cross section as a function of $\sqrt{s}$ for the different production channels at the LHC. (b) Higgs boson decays BRs near its mass [27].

its mass. The total width of the Higgs boson is predicted to be $\Gamma_H \simeq 4$ MeV and the

corresponding lifetime is $\tau_H \simeq 10^{-22}$ s [10]. Thus, the Higgs boson can only be detected through its decay products. The partial width of each of the possible decays is related to the couplings of the Higgs boson to the decay products. For this reason, the decays into massive gauge bosons (W,Z) or fermions (b,τ) are the preferred decay channels and account alone for over the 99% of the total width, Fig. 2.6b. Since the Higgs boson only couples to massive particles, decays into two gluons or two photons occur through Next-to-Leading-Order (NLO) loops of heavy particles in the Feynman diagrams [28], Fig. 2.7. With a mass of 125 GeV, the Higgs boson's preferred decay mode is $H \rightarrow b\bar{b}$, directly followed by $H \rightarrow WW^*$.



(a) *(b)*

Figure 2.7.: NLO decay modes of the Higgs boson into (a) gg and (b) $\gamma\gamma$.

The $H \rightarrow b\bar{b}$ channel is very important and challenging and accounts alone for almost half of the Higgs boson total decays. The $t\bar{t}H$ production channel is very important because it allows the study of Higgs-top quark Yukawa coupling constant. The $t\bar{t}H$ with $H \rightarrow b\bar{b}$ allows the study of the Higgs boson decay in $b\bar{b}$, otherwise prevented because of the presence of a large multijet background.

2.6. Summary of Run 1 $t\bar{t}H(H\rightarrow b\bar{b})$ searches

The searches of the $t\bar{t}H$ process, with $H \rightarrow b\bar{b}$, are divided in different analyses according to the considered decay channel for the $t\bar{t}$ system. This can be all hadronic, single lepton or dilepton.
Results were published both by the ATLAS and CMS collaborations based on the Run 1 data, at $\sqrt{s}$ = 7 TeV and $\sqrt{s}$ = 8 TeV. The CMS collaboration published the following results:

- *Search for the associated production of the Higgs boson with a top-quark pair* [29]. This analysis presented a search of the $t\bar{t}H$ channel in different H boson decay modes, using the dataset from $\sqrt{s}$ = 7 TeV and $\sqrt{s}$ = 8 TeV run periods. Regarding the $t\bar{t}H(H \rightarrow b\bar{b})$ analysis, both the dilepton and single lepton channel were

considered and a BDT (see Sec. 6.4.4) was used to discriminate the signal from the background.

- *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method* [30]. This analysis used the data from the $\sqrt{s}$ = 8 TeV period and investigated the $t\bar{t}H(H \to b\bar{b})$ in both the single lepton and dilepton channels using a particular reconstruction techniques named *matrix element method*. This technique combines both theoretical and experimental information in order to perform a probability calculation that a certain measured event is consistent with background or signal hypothesis. The usage of this technique improved the final separation and the results from the previous analysis.

The results of the ATLAS collaboration using the Run 1 data are listed as follows:

- *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector* [31]. This analysis used an ANN to discriminate signal from background and the final state reconstruction involved the matrix element method in the single lepton channel. The reconstruction information was used in the ANN to enhance its final separation.

- *Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector* [32]. This analysis considered the all hadronic channel and used BDT techniques to discriminate the signal from the multijet background.

Both the ATLAS and CMS collaborations published combined results of their searches [30, 32]. The ATLAS collaboration measured a signal strength μ (where $\mu = \frac{\sigma_{\text{obs}}}{\sigma_{\text{SM}}}$) of 1.4±1.0, while a signal strength of $1.2^{+1.6}_{-1.5}$ was measured by the CMS collaboration. The ATLAS and CMS results have been combined and the resulting signal strength is 1.09 ± 0.11 [33].
The measurements of the $t\bar{t}H$, with $H \to b\bar{b}$ process is one of the main goals of Run 2 at the LHC. At $\sqrt{s}$ = 13 TeV the $t\bar{t}H$ cross section increases by a factor of 3.9 compared to the Run 1 period, while the cross section of the main $t\bar{t}$ +jets background increases by a factor of 3.3 [27]. Results based on Run 2 are compared in Sec. 9.2.

3 CERN, LHC and the ATLAS Experiment

3.1. CERN and LHC

The Conseil Européen pour la Recherche Nucléaire (CERN), is a complex of laboratories for physics research among the most important of the world. Founded in 1954, it is located on the Franco-Swiss border near Geneva, in Switzerland. It was one of the first European joint projects and at present day (2017) it has 22 member nations. Its goal is fundamental physics, researching what the Universe is made of and how it works. CERN hosts the largest and most complex scientific instrumentation, in order to study the building blocks of matter. These instruments are particle accelerators and detectors: accelerators boost beams of particles at very high energy or collide them at stationary targets; detectors collect and observe the results of these collisions.

The LHC [34], Fig. 3.1, is a particle collider built in a circular tunnel 27 km in circumference, which had been previously built for the Large Electron-Positron Collider (LEP) accelerator. The tunnel is located approximately 100 m underground and it crosses the Swiss and French borders near Geneva, in Switzerland.

The LHC is designed to collide two counter rotating beams of protons or heavy Pb ions. Proton-proton collisions are at present at a centre-of-mass energy of 13 TeV per beam, but this energy will be increased to 14 TeV in the future. Protons reach this energy using a complex chain of accelerators, of which the final step is the LHC: they are first accelerated to 50 MeV using the LINear ACcelerator (LINAC), then they are injected in the Proton-Syncrotron (PS) which increases the energy to 26 GeV, where they enter in the Super-Proton-Synchrotron (SPS) to reach 450 GeV of energy. At this point, they are injected in the LHC, grouped in bunches, se Fig. 3.2. The beams move around the LHC ring inside a continuous vacuum guided by superconducting magnets which provide a 8.3

T magnetic field, and are cooled by a huge cryogenics system, which operates at 1.9 K. The beams are stored at high energy for hours and, during this time, collisions take place inside the four main LHC experiments (see Sect.3.2). The main technical design parameters of LHC are summarised in Table 3.1.

LHC Parameter	Nominal Value
Beam Energy	7 TeV
Injection Energy	450 GeV
Dipole Magnetic Field	8.33 T
Space Between Bunches	7.5 m
Time Between Bunches	25 ns
Protons for Bunch	10^{11}
Collision Angle	300 μrad
Beam Lifetime	10 h
Energy Loss per Loop	7 KeV
Power Radiated	3.7 KW

Table 3.1.: Design parameters (1995) of the LHC collider [35]

.

The LHC has been designed to reach the highest luminosity ever achieved, where the luminosity $\mathcal{L}$ is a parameter defined as:

$$\mathcal{R} = \mathcal{L} \cdot \sigma, \tag{3.1}$$

$\mathcal{R}$ being the event rate and σ the interaction cross section. Luminosity is given by the accelerator parameters as:

$$\mathcal{L} = \frac{n_p^2 f k}{4\pi\rho^2}, \tag{3.2}$$

where n_p is the number of protons per bunch, f is the frequency of circulation of the bunches, k is the number of bunches, ρ is the mean square radius of the spacial distributions of protons in the orthogonal plane of the beam direction.
The expected peak luminosity is $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$, so given the inelastic cross section for protons at $\sqrt{s}$ =14 TeV of σ = 70 mb, the expected event rate is given by:

$$\mathcal{R} = 70\text{ mb} \cdot 10^{34}\text{cm}^{-2}\text{s}^{-1} \approx 1\text{GHz}\,. \tag{3.3}$$

The number of events for a given process characterised by a cross section σ produced by

the collisions at the LHC can be obtained with the time-integrated luminosity:

$$N = \sigma \int \mathcal{L} dt. \tag{3.4}$$

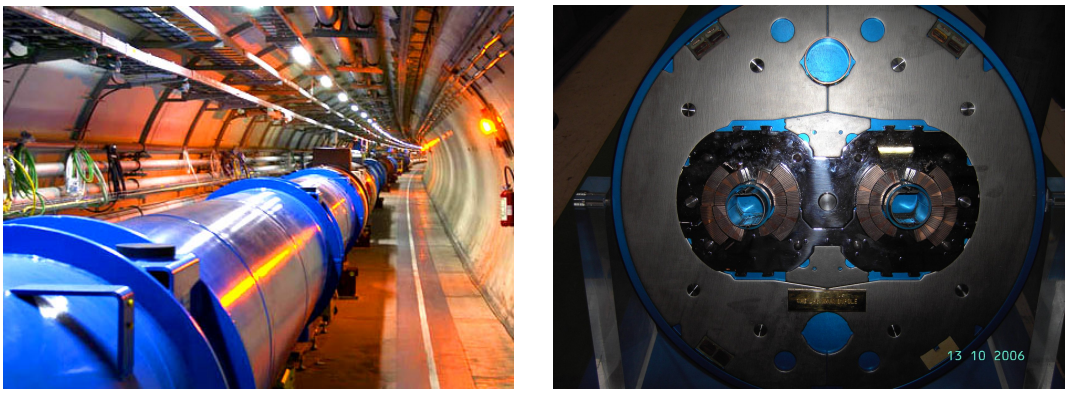


Figure 3.1.: View of the LHC tunnel (a) and of one of the magnets (b).

3.2. Experiments at the LHC

At present day, there are seven experiments working at the LHC, all of them being large international collaborations, unifying scientists from many institutes around the world. The two largest ATLAS and CMS experiments, are devoted to the study of the particles produced in the proton-proton collisions to search for new physics beyond the SM. They are therefore designed in such a way to be able to investigate the largest possible range of physics. They have independently designed detectors, to double confirm every potential discovery.

The two medium-sized detectors A Large Ion Collider Experiment (ALICE) and LHCb experiments have detectors specialised in the study of the LHC collisions related to specific phenomena. In particular, ALICE studies Pb-Pb collisions in order to investigate quark-gluon-plasma, and LHCb has the main goal to study CP violation in B mesons.

The TOTEM and LHCf experiments are smaller in size. They are designed to focus on forward produced particles (protons or heavy ions), and also measure the LHC luminosity.

Finally, MoEDAL has the goal to look for the existence of magnetic monopoles and other ionising stable massive particles. ATLAS, CMS, ALICE and LHCb are located around the LHC ring inside huge underground caverns. TOTEM, LHCf and MoEdal are posi-

tioned near CMS, ATLAS and LHCb respectively. A view of the four main experiments positions is shown in Fig. 3.2

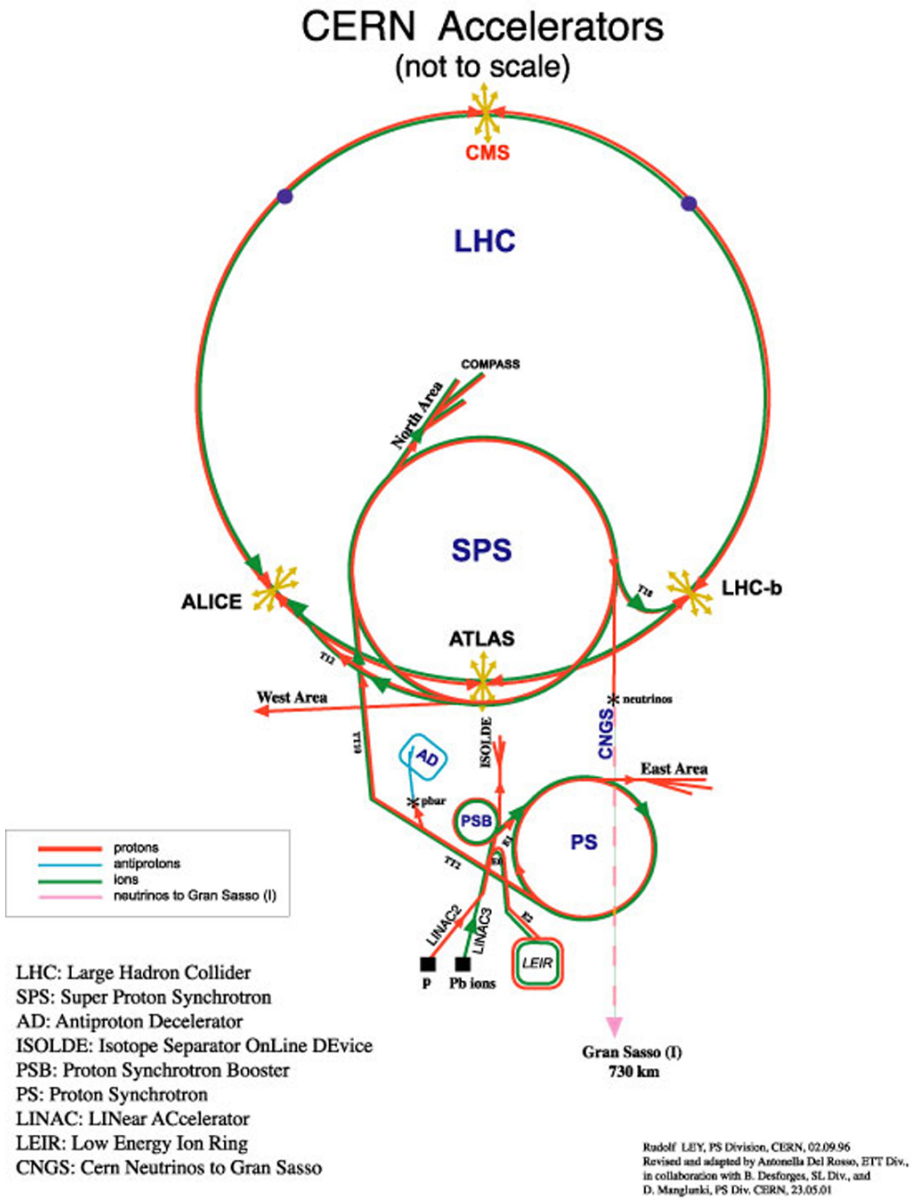


Figure 3.2.: A view of the accelerators and detectors at CERN, © by CERN.

3.3. ATLAS Experiment Overview

ATLAS is a particle physics experiment at the LHC at CERN, Fig. 3.3. The ATLAS detector searches for new physics in the head-on collisions of protons at the LHC, and in 2012 announced the discovery of a new particle which is believed to be the Higgs boson. The detector has to be able to detect as many signatures as possible in the very harsh LHC environment, in order to obtain measurements of electrons, photons, muons, hadronic jets, missing transverse momentum and also recognise the b-quark jets.

3.3.1. Coordinate System

In the description of the detectors, a cylindrical coordinate system is used. The LHC beam direction defines the z-axis, and so the x-y plane is the plane transverse to the beam direction. The x-axis points to the centre of the LHC ring and the positive direction of the y-axis points upwards. The azimuthal angle ϕ is defined around the beam axis and the polar angle θ is the angle taken from the beam axis. At high energies, it is convenient to define the pseudorapidity as $\eta = -\log\left(\tan\left(\frac{\theta}{2}\right)\right)$ and use it instead of the angle θ. The transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ and all other transverse variables, are defined in the transverse plane x-y.
The main design criteria of the ATLAS detector, were defined according to the following requests:

- have very good electromagnetic calorimetry for electron and photon detection;
- have a full hadronic calorimetric-coverage for the jet and missing transverse momentum measurements;
- perform high-precision muon momentum measurements;
- guarantee an efficient tracking for leptons at high p_T at high luminosity;
- perform τ-lepton and heavy flavour quark identification;
- have a good acceptance in η with a complete coverage for the azimuthal angle (ϕ).

The overall dimensions of ATLAS are defined by the muon spectrometer. The detector has a cylindrical shape, with subdetectors arranged as concentric cylinders around the beam axis. The outer chambers have a radius of 11 m and the length of the cylinder around 45 m. The weight of the whole detector is approximately 7000 tons, see Fig. 3.3.

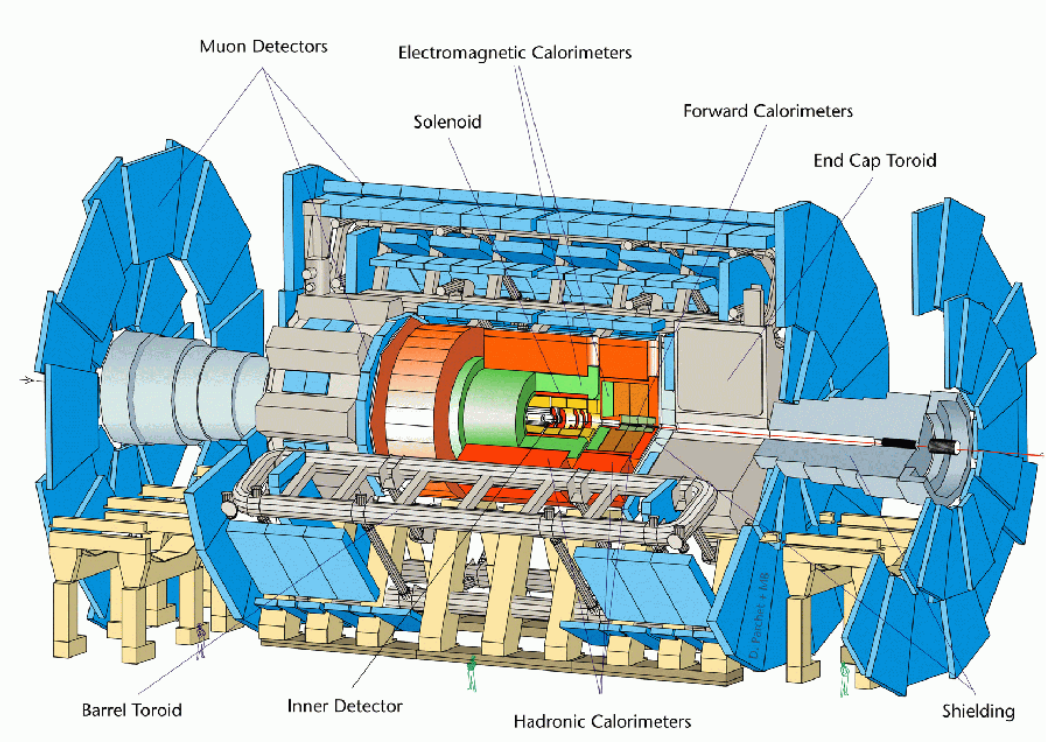


Figure 3.3.: ATLAS experiment overview. The human figure in the bottom gives an idea of the overall size of the detector.

3.3.2. Inner Detector

The Inner Detector (ID) is contained in a cylinder of radius 1.15 m and length of 7 m, embedded in a 2 T solenoid magnetic field.
The required momentum and vertex resolutions need high-precision measurements to be performed using fine-granularity detectors, given the very large track density of LHC. These detectors also have to be radiation-hard in order to work for at least ten years. The highest granularity is obtained using Semi-Conductor Tracker (SCT) and pixel detectors. The SCT Barrel region comprises four cylindrical microstrip layers of silicon microstrips modules and nine disks in the End-Cap, while the pixel detectors are made up of four cylindrical layers of silicon pixels in the Barrel region and of three disks in the End-Cap. The total number of precision layers must be limited because of the quantity of additional material they introduce, which can perturb the measurements of the properties (energy and momentum) of the particles produced in the collisions. Another reason is related to the cost of such layers. In association with this high precision instrumentation, the Transition Radiation Tracker (TRT) provides a larger number of tracking points, requiring much less material per point and a lower cost.

The ID combines discrete high-resolution semiconductor pixel and strip detectors in the inner part, and continuous straw-tube tracking detector with transition radiation capability in the outer part. This layout can be seen in Fig. 3.4.
In summary, in the ID the following technologies are present:

- The pixel detectors determine the primary vertex of the collisions and allow measuring secondary vertices coming from the *long* lived particles such as B hadrons and τ leptons. The system contains a total of 140 million detector elements, giving a very high-granularity and precision resolution.

- The SCT consists of layers of silicon microstrip detectors designed to contribute to the measurements of momentum, impact parameter and vertex position, providing also a good pattern recognition using a high granularity.

- The TRT is based on the use of straw tube detectors, which can operate at very high rates thanks to their thin size and isolation of the sensitive wires within individual volumes filled with gas. The electron identification is enhanced employing Xenon gas to detect transition radiation photons created in a radiator between the straws. TRT provides a good discrimination between electron and hadron signals.

The ID layout provides full tracking coverage over $|\eta| \leq 2.5$ and provide an experimental resolution of 10 x 115 μm for the particle position. The ID reaches a designed resolution of the track momentum of:

$$\frac{\sigma_{p_T}}{p_T} = 0.05\% \times p_T(\text{GeV}) \oplus 0.1\% \,. \tag{3.5}$$

A fourth pixel layer has been installed in the ID in 2014 to recover the loss of sensitivity of the Pixel Detector due to radiation damage. This Insertable B-Layer (IBL) has been installed between the beam pipe and the Pixel Detector. The internal radius of IBL is 31 mm and the outer one is 38.2 mm, the sensors are present at a radius of 33.4 mm and face the beam pipe on the range of $|\eta| < 2.5$. With its 50 x 250 μm pixels, the IBL adds additional 12 million pixels to the overall Pixel Detector. The physics performance of the ATLAS detector highly depends on the capabilities of the IBL, which provides an improved vertexing and a better b-tagging [36].

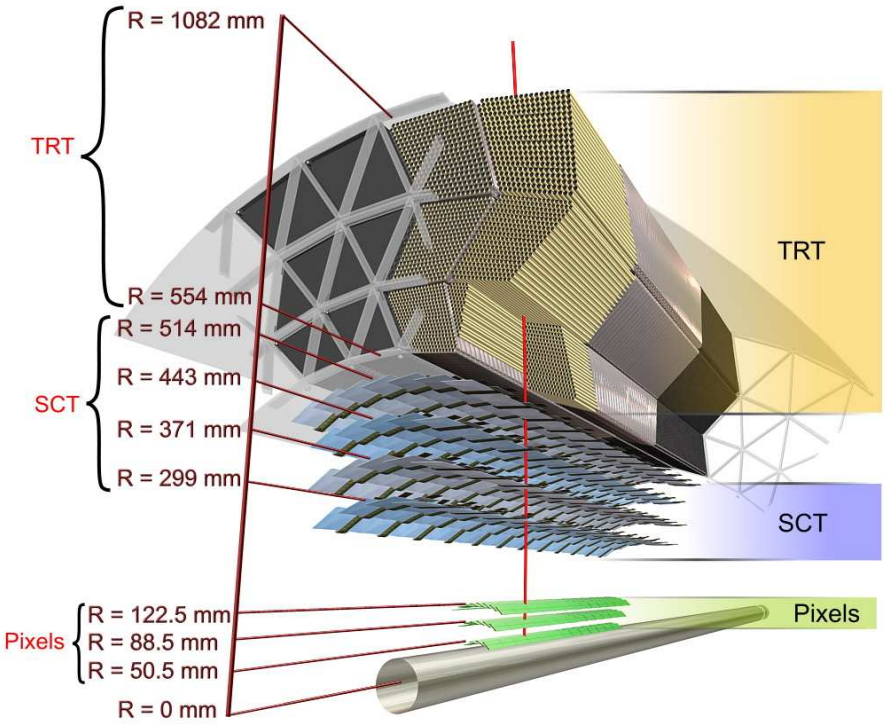


Figure 3.4.: A view of the Inner Detector layers.

3.3.3. Calorimeters

The ATLAS calorimetry system is designed to serve in a very harsh environment of proton-proton collisions, in particular it has to be efficient at the high luminosity of LHC. The overall structure of the ATLAS calorimeters is shown in Fig. 3.5.

The barrel Electromagnetic (EM) calorimeter is a highly granular Lead/Liquid-Argon (LAr) sampling calorimeter, see Fig. 3.6. It has a good energy and position resolution and covers the pseudorapidity region of $|\eta| < 3.2$. The EM calorimeter is housed in a barrel cryostat, it surrounds the ID, in front of the solenoid which generates the 2 T magnetic field. The calorimeter is also very important for particle identification and hadronic-electromagnetic separation ($\gamma/\pi^0, e/\pi$ separation, etc.). It also provides a precise position measurements in η through high granularity. The design of the EM calorimeter is an arrangement of absorber layers and active layers in a characteristic accordion geometry, see Fig. 3.6b.

The ATLAS hadronic calorimeters cover the range $|\eta| < 4.9$ using different techniques suited for the varying requirements and radiation environment over the large η-range. The bulk of hadronic calorimetry is given by the Iron-scintillator tile calorimeter, which is

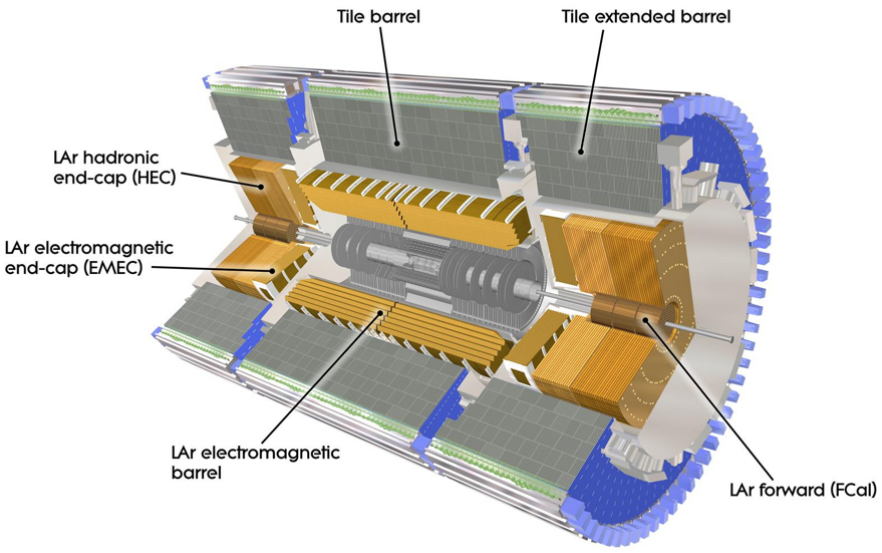


Figure 3.5.: A view of the ATLAS calorimeters.

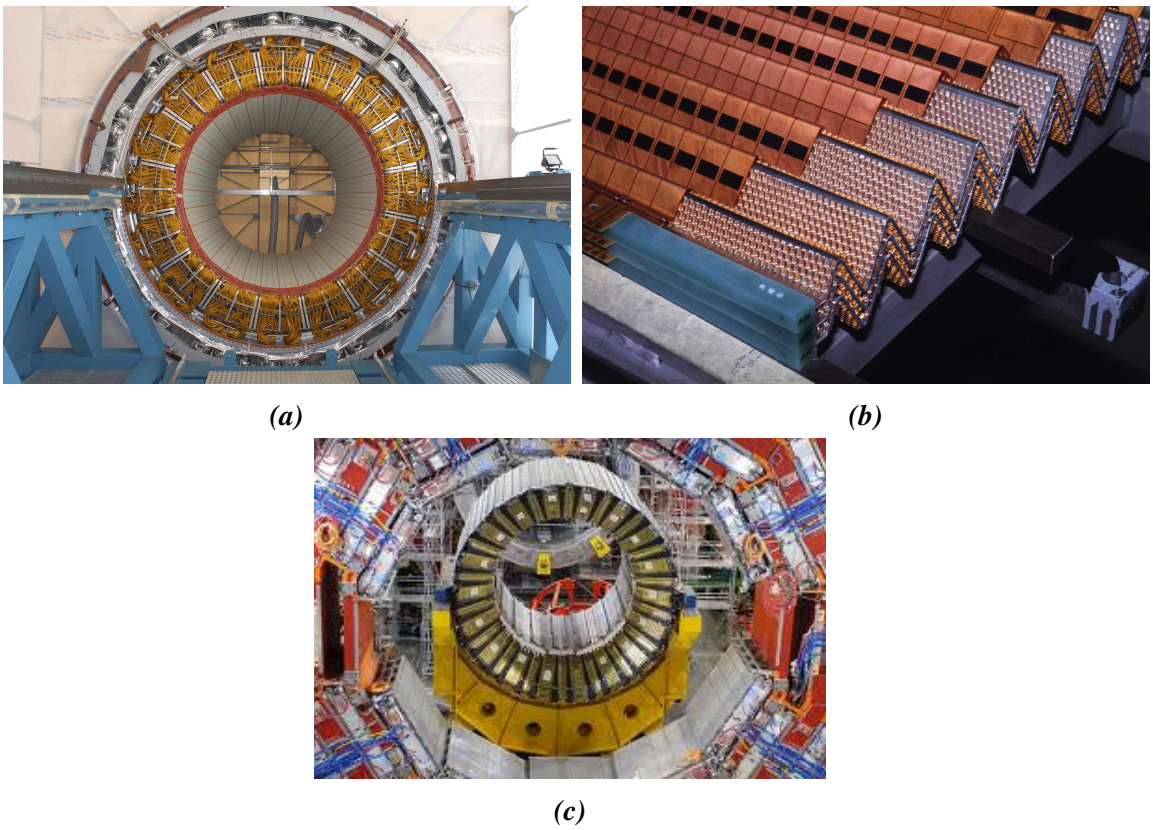
separated in a large barrel and two smaller extended barrel cylinders, one for each side of the barrel, as shown in Fig. 3.6c. The Hadronic End-Cap (HEC) calorimeter and the high density Forward CALorimeter (FCAL) share the LAr technology and are integrated in the same cryostat, which houses the EM end-cap, see Fig. 3.7.
The coverage of the hadronic calorimeter guarantees a good missing transverse momentum measurement, which is very important for many physics signatures and also for the detection of SUSY particles. The energy resolution of the calorimetry system is summarised in Table 3.2.

Detector Component	Energy resolution
EM Calorimetry	$10\%/\sqrt{E} \oplus 0.7\%$
Hadronic Calorimetry	
Barrel & End-Cap	$50\%/\sqrt{E} \oplus 3\%$
Forward	$100\%/\sqrt{E} \oplus 3.1\%$

Table 3.2.: Nominal detector performance goals for the ATLAS calorimetry system.

3.3.4. Muon Spectrometer

The Muon Spectrometer (MS) uses the magnetic deflection of muon tracks inside the superconducting air-core toroid magnets. It consists of separate trigger and high-precision



(a) *(b)*

(c)

Figure 3.6.: EM and hadronic calorimeters. (a) EM calorimeter. (b) Accordion geometry of the EM calorimeter. (c) Hadronic calorimeter.

tracking chambers. There are a Barrel and two End-Cap magnet regions, which can together cover the pseudorapidity region of $|\eta| \leq 2.7$. This magnet configuration provides a field that is mostly orthogonal to the muon trajectories and minimises the resolution degradation due to the multiple scattering.

In the Barrel region, a precision measurement of the track coordinates along the principal bending direction of the magnetic field is performed. Optical alignment systems have been designed to satisfy the strict requirements on the mechanics and the survey of the precision chambers.

In both the Barrel and the two End-Cap magnet regions, two different kinds of detectors are installed. The first one is a trigger chamber system which provides a fast response. It consists of Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). The second type are precision tracking chambers which provide more accurate measurements. They consist of Monitored Drift Tubes (MDT) and Cathode Strip Chamber (CSC).

The trigger chambers system can provide a time resolution better than the LHC spacing of 25 ns, providing a trigger with a well-defined p_T cut-off to reject low momentum particles

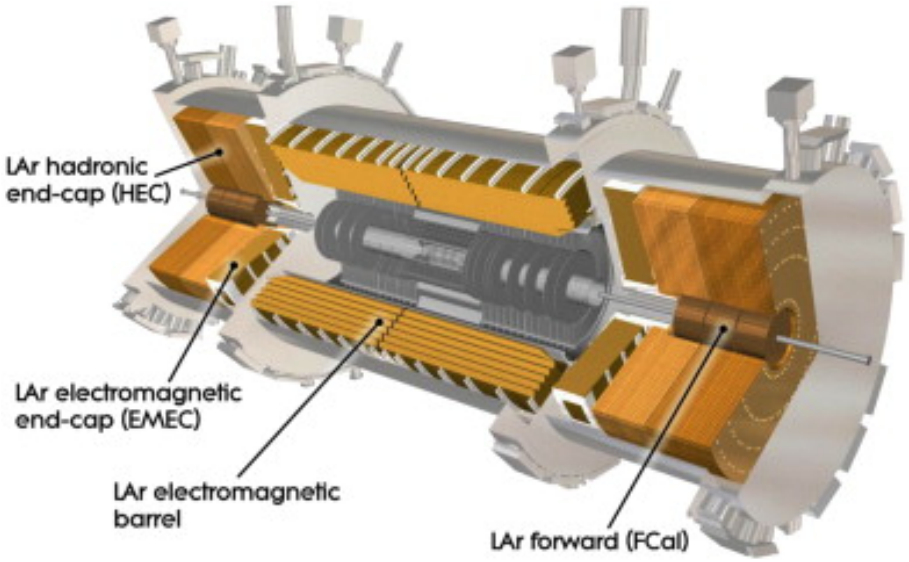


Figure 3.7.: A view of the LAr calorimeter system, which includes the Barrel, Forward and End-Caps EM calorimeters and the HEC and FCAL hadronic calorimeters.

originated in hadronic showers.
The requirements on the momentum resolution of the spectrometer are met by constantly monitoring possible chamber deformations and positions, thanks to optical alignment systems. The momentum resolution is $\simeq$10% at transverse momentum of 1 TeV.
The muon spectrometer defines the overall dimensions of the ATLAS detector, with a length of 45 m and a radius of 11 m, see Fig. 3.8.

3.3.5. Magnet System

The ATLAS superconducting magnet system (whose elements are shown in Fig. 3.9), consists of a Central Solenoid (CS) which provides the magnetic field to the ID, and a system of three external large air-core toroids generating the magnetic field for the muon spectrometer. These air core toroids are the two End-Cap Toroids (ECT) at both sides and in line with the CS, and the Barrel Toroid (BT). The overall dimension of the external large air-core toroids magnet system is 26 m in length and 20 m in diameter. The CS provides a central magnetic field of 2 T, parallel to the beam axis. It is positioned in front of the EM calorimeter, sharing the same vacuum vessel, eliminating in this way two vacuum walls.
Each of the three external toroids consists of eight coils assembled radially and symmetrically around the beam axis, providing a field between 0.15 T and 2.5 T, with an average

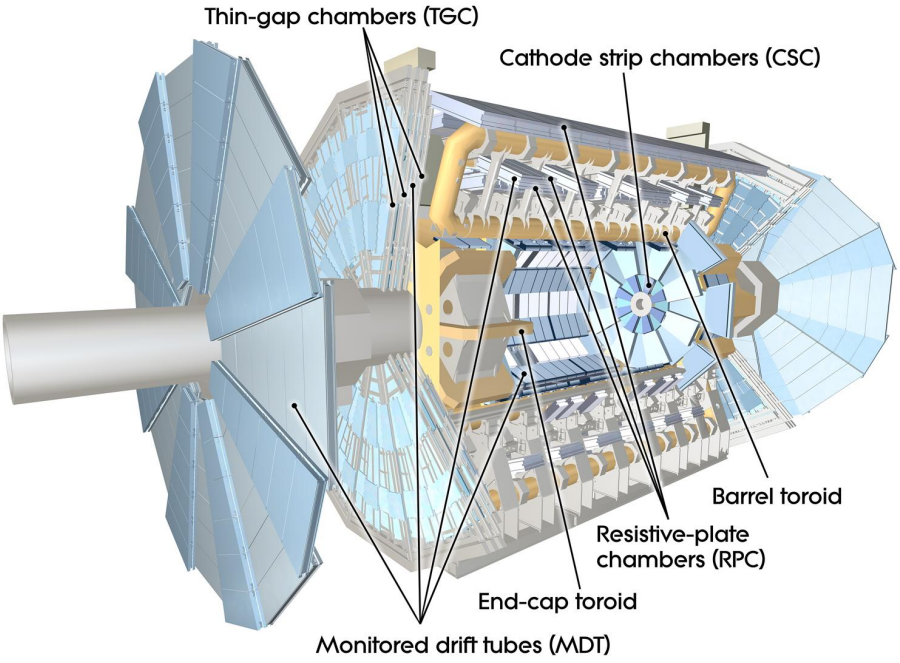
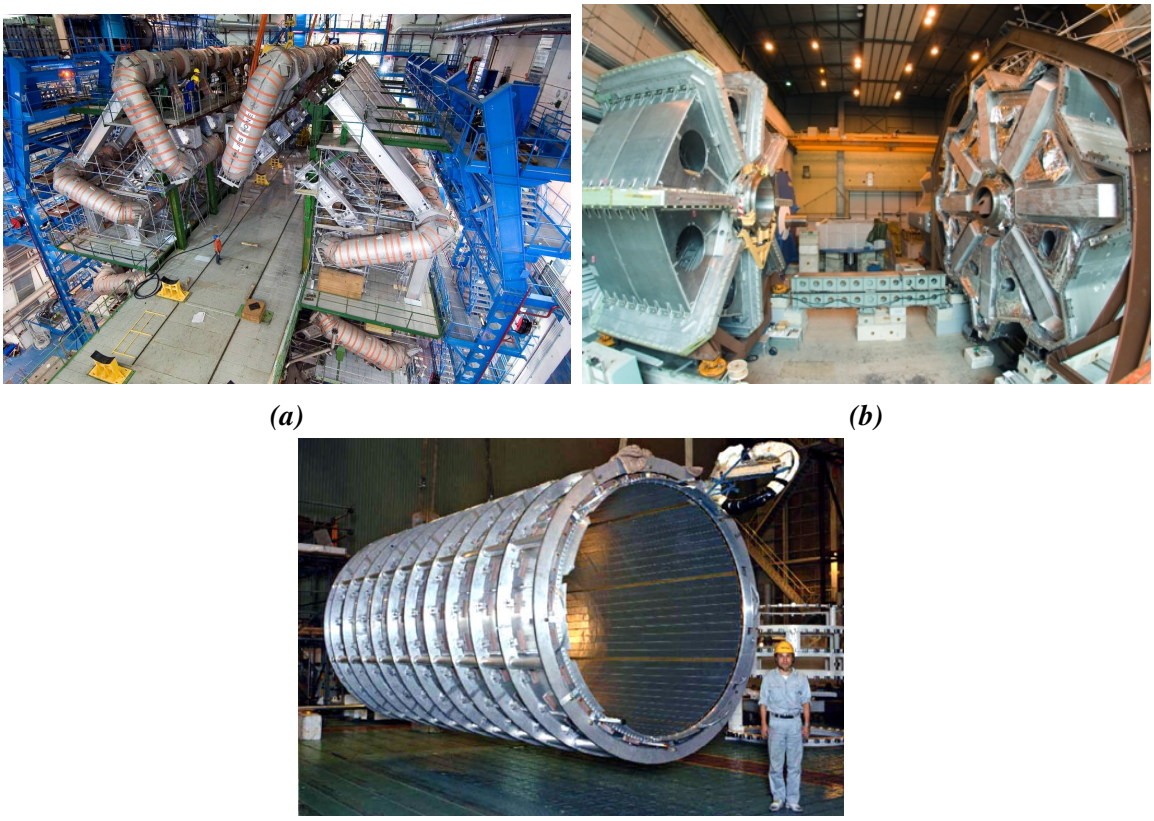


Figure 3.8.: The muon spectrometer, with its barrel and end-cap regions.

of 0.5 T, while the end-cap magnetic field covers the range between 0.2 and 3.5 T. The barrel and end-cap regions can cover together the pseudorapidity region of $|\eta| < 3$.

3.3.6. Trigger System

The trigger system of ATLAS has been successfully developed and used during the period of Run 1 [37], however the increased instantaneous luminosity, beam energies and frequency of the collisions in Run 2 implied the production of higher background rates compared to Run 1. For this reason, during Long Shutdown 1 (LS1), many important design changes and additions to the trigger and Trigger and Data Acquisition (TDAQ) systems were performed [38]. In Run 2 the trigger system consists of a hardware and a single software based trigger level, Level-1 (L1) and High-level trigger (HLT), see Fig. 3.10. The reduction of the recording rate achieved is from 40 MHz of the bunch-crossing rate to 100 kHz at the L1, with a final output rate of 1 kHz on average at the HLT [39].
In Ref. [40], a strategy for the trigger menu in Run2 is reported.



(a) *(b)*

(c)

Figure 3.9.: Magnetic system of ATLAS detector. (a) The eight coils of the barrel toroid. (b) The end-cap toroids. (c) The central solenoid.

Level-1 Trigger

The primary event selection is performed using the information coming from calorimeters and muon detectors by the L1 trigger level. The L1 trigger system in Run 2 consists of different components: the Level-1 calorimeter trigger (L1Calo), Level-1 topological trigger (L1Topo) [41], Level-1 muon trigger (L1Muon) and Central Trigger Processors (CTP) [42]. L1Calo and L1Muon sublevels process the topological properties of the event, such as angles between the objects. The L1Topo trigger makes some more complex decisions performing some geometrical cuts. The CTP makes the final decision.

High Level Trigger

In Run 1, L1 trigger was followed by a Level-2 trigger which performed another selection based on detailed physics properties. At this stage the rate was reduced to 2-3 kHz. A final third-Level trigger finally used the full available information to decide which events to

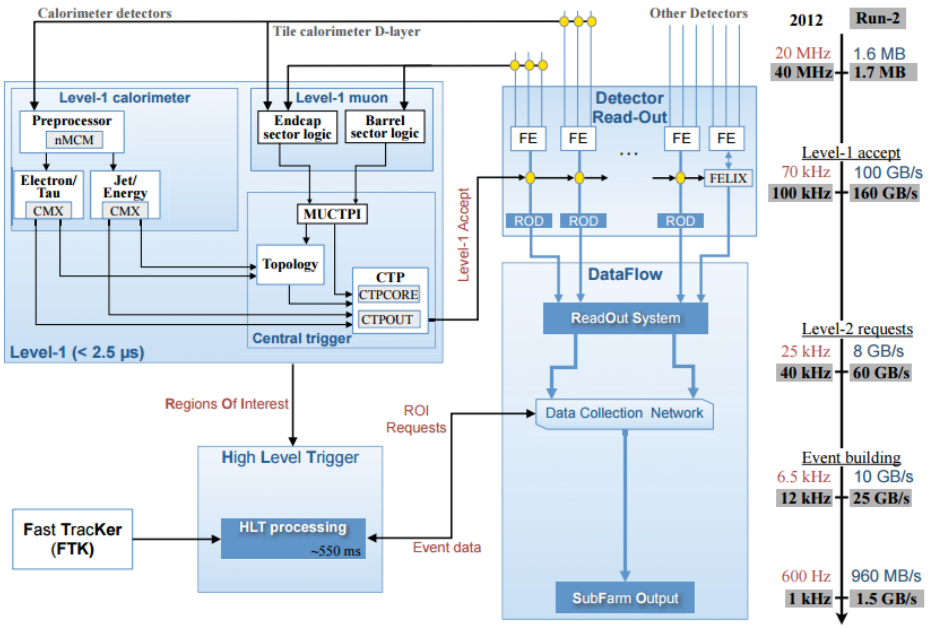


Figure 3.10.: Schematic overview of the Run-2 configuration of the Trigger and TDAQ system.

save for an offline analysis and the final rate was here reduced to 300-400 Hz. In Run 2 the Level-2 and Level-3 were merged in a single HLT farm to provide dynamic and simplified resource sharing. With this configuration, the read-out system (ROS) computers save network bandwidth and decrease the read-out request rate. For Run 2, many multivariate analysis techniques have been developed and used at the HLT stage, as for example, in electron and photon trigger systems [43]. The final rate for data is about 1 kHz at peak luminosity at the HLT.

4 Experimental Data and MC Modelling

The proton beams accelerated by the LHC collide at certain specific points along the circumference where detectors are installed. The ATLAS detector is located at one of these collision points, being devoted to the detection of the particles produced in these collisions. The resulting experimental data is analysed by scientists. The data consists of information about electrical voltages and currents that are recorded in all the different active parts of the detector. This information is converted into tracks and energy deposits, which are the building blocks for the reconstruction of the different physics objects. Some modelling is also required to be compared with the real data, which is achieved through MC simulation or from estimation of the experimental data.

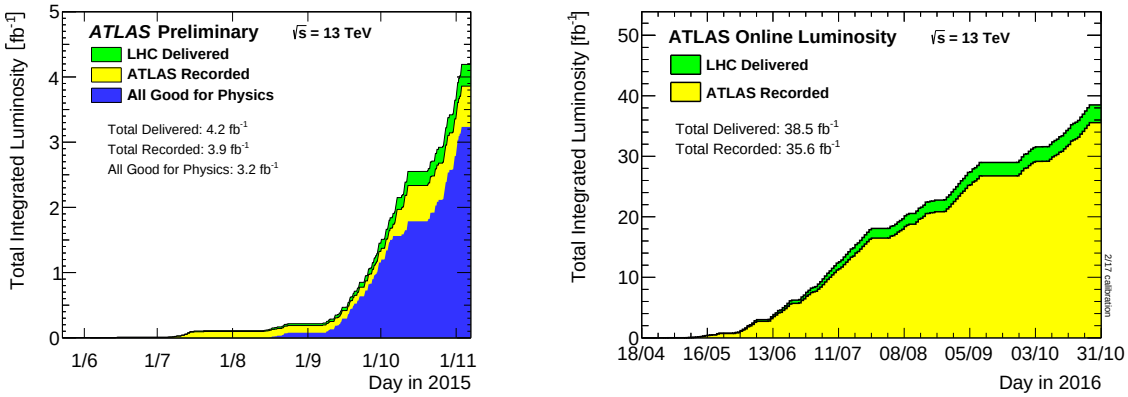
4.1. Experimental Data

The first proton-proton beams were successfully circulated on 10$^{\text{th}}$ of September 2008 but on 19$^{\text{th}}$ of September a serious fault damaged several superconducting magnets and the machine had to stop. The repair required a long technical intervention and the LHC did not work again until November 2009. First collisions at $\sqrt{s} = 7$ TeV took place on 30$^{\text{th}}$ of March 2010 with the rest of the year mainly devoted to commissioning. During 2011, over 5 fb^{-1} were delivered to both ATLAS and CMS experiments at $\sqrt{s} = 7$ TeV. Between April and June 2012, over 5 fb^{-1} at $\sqrt{s} = 8$ TeV were delivered. These data allowed the first observation of a particle compatible with the Higgs boson, as announced on 4$^{\text{th}}$ July 2012. The integrated delivered luminosity up to December 2012 was 23.3 fb^{-1}. The LHC paused its activity on December 2013 (Run 1) and it resumed in April 2015 at the centre-of-mass energy of 13 TeV, very close to the nominal value. In the period of 2015-2016 over 43 fb^{-1} were delivered by the LHC. In Fig. 4.1, the corresponding luminosity dis-

tribution over time is shown for 2015 and 2016, the colours refer to luminosity delivered by the LHC (green), recorded by the ATLAS detector (yellow), and classified as good for physics analyses after data quality checks (blue). At present, a $\mathcal{L} = 13.8 \cdot 10^{33}\text{cm}^{-2}\text{s}^{-1}$ peak luminosity has been reached, Fig. 4.2*a*.

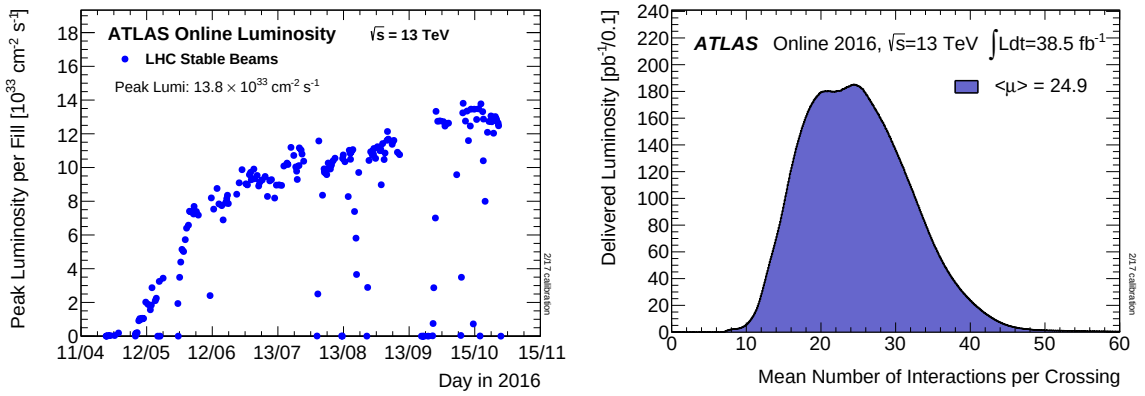
An important effect that can cause a significant degradation of the object reconstruction, is the *pile-up* noise. Two different types of pile-up are defined: the *in-time pile-up*, originated by additional interactions in the same bunch crossing with respect to the hard scattering process of interest, and the *out-of-time pile-up*, which is the effect of multiple interactions from events prior or posterior to the analysed one. At 13 TeV and during the data taking period of 2016, the mean number of interactions per crossing was $\mu = 24.9$, see Fig. 4.2*b*. This value is useful to quantify the overall pile-up conditions. The algorithms used for the object reconstruction are designed to compensate for these degradation effects and recover the detector performance.

The analysis described in this thesis, is performed using data recorded at $\sqrt{s} = 13$ TeV by the ATLAS experiment during 2015 and between April and July 2016. The corresponding time integrated luminosity labelled as good for physics is 13.2 fb^{-1}.



(a) ***(b)***

Figure 4.1.: Integrated luminosity vs day up to December 2015 (a), and up to September 2016 (b), delivered to (green) and recorded by the ATLAS experiment (yellow) during stable beams and for pp collisions at $\sqrt{s} = 13$ TeV.



(a) *(b)*

Figure 4.2.: (a) Peak luminosity vs day in 2016. (b) Mean number of interactions per crossing for the 2016 pp collision data.

4.2. MC Modelling of Physics Processes

The ATLAS simulation chain consists of four different steps [44]: generation of the event (event generation), simulation of the detector geometry and physics interactions with material (detector simulation), digitisation of the signals (voltages and currents) in the sensitive regions of the detector. The output of the simulation chain is presented in a format identical to the output of the ATLAS data acquisition system in a way that both the simulated and real data can be run through the same trigger and reconstruction packages. This final step is devoted to the reconstruction of physical quantities and is discussed in Chapter 5. The simulation chain is depicted in Fig. 4.3

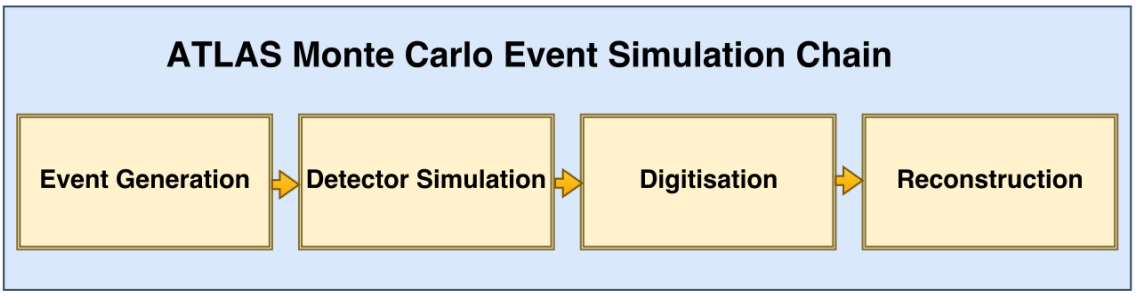


Figure 4.3.: ATLAS simulation chain.

4.2.1. Event Generation

Given that protons are composite particles made of partons, the modelling of pp collisions at the LHC demands a deep understanding of high energy (short distance) deep-inelastic interactions (that is possible to describe through a perturbative QCD approach) and low energy (long distance) structure of the proton and the interactions of its composing partons (non perturbative QCD). The separation between these two regimes is defined through an arbitrary *factorisation scale*, μ_F, that sets a limit on the energy regime where the running of α_S becomes too large to permit a convergence of the perturbation series. This splitting leads to the factorisation theorem [45], whose main idea is explained as follows. In the centre-of-mass frame, the hard interactions between partons occur very quickly relative to the time for them to interact. As a result, the hadronic collision can be factorised into a parton collision weighted by PDFs, $f_i(x_i)$, which expresses the probability for the parton of flavour i to carry the momentum fraction, x_i, of its parent hadron. A specific process production cross section, denoted as X, is then calculated as:

$$\sigma(pp \to X) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, Q^2, \mu_F^2)\,, \tag{4.1}$$

where the sum runs over gluons and quarks in the colliding protons, and $\hat{\sigma}_{ij}$ is the perturbative cross section for collisions of partons i and j, and can be calculated via the Feynman diagrams according to the Fermi Golden Rule and to the perturbation theory. The *hard scale* Q^2 is typically chosen at the invariant mass^2 of the final state of the considered process. The factorisation scale, μ_F^2, is usually taken to be equal to the *renormalisation scale* μ_R^2 which is also $\simeq Q^2$. In the case of $t\bar{t}H$, μ_F could be chosen equal to $m_{top} + m_H/2$.
Since QCD cannot predict the actual form of the PDFs, they need to be experimentally evaluated and parametrised at a starting scale Q_0^2. Historically the data from experiments for the PDFs determination come from deep inelastic scattering experiments performed mainly at HERA electron-proton collider of DESY. The functions are parametrised with respect of x at defined a starting scale Q_0^2 from fits to the experimental data and then extended to higher scales Q^2 using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations. These equations are computed for the quarks and gluons respectively and take the form:

$$\frac{\partial q(x,Q^2)}{\partial \log Q^2} = \frac{\alpha_s}{2\pi}(P_{qq} \otimes q + P_{qg} \otimes g) \tag{4.2}$$

$$\frac{\partial g(x,Q^2)}{\partial \log Q^2} = \frac{\alpha_s}{2\pi}\left(\sum_i P_{gq} \otimes (q_i + \bar{q}_i) + P_{gg} \otimes g\right), \tag{4.3}$$

where $q \equiv f_q$ is the quark density and $g \equiv f_g$ is the gluon density, $\otimes$ indicates the convolution integral:

$$P \otimes f \equiv \int_x^1 \frac{dy}{y} f_q(y) P\left(\frac{x}{y}\right) \tag{4.4}$$

and $P(z)_{ab}$ are the universal parton splitting functions that are defined as the distribution of the fraction z of the energy of the parton a carried by the parton b in the parton splitting process $a \rightarrow b$. At present day, several collaborations provide PDF sets to be used for simulation purposes, some of them are CTEQ [46], NNPDF [47] and MSTW [48], the latter displayed in Fig. 4.4.

At high energy collisions, high momentum transfers between partons occur, thus they are drastically accelerated. As in the case of the electrically charged particles, which emit photons when accelerated as described by Quantum-Electro-Dynamics (QED), coloured partons emit gluons following QCD. Unlike photons, that do not carry any electrical charge, the gluons do carry colour charge, so they can emit further QCD radiation themselves. This effect leads to the formation of *parton showers*. Each parton shower emission is relative to a phase space region where emissions are collinear or soft (low energy) and can be described as high order corrections to the hard process. For practical reasons an approach based on a particular approximation scheme is used. In this approach, only the dominant contribution is included for each order.

Considering a collinear splitting of a parton i in $j + k$, $q \rightarrow q + g$, and assuming that the differential cross section for n partons before the parton splitting is $d\sigma_n$, the cross section for the next emission becomes:

$$d\sigma_{n+1} \approx d\sigma_n \frac{\alpha_S}{2\pi} \frac{d\theta^2}{\theta^2} dz d\phi P_{ij}(z,\phi)\,, \tag{4.5}$$

where, θ and ϕ are the transverse and azimuthal angle of the splitting, and P_{ij} are the splitting functions mentioned above. The parton shower develops iteratively starting from each of the partons which take a role in the hard process. The probability of not splitting during the evolution between two scales q_1^2 and q_2^2 is given by the *Sudakov form factor*:

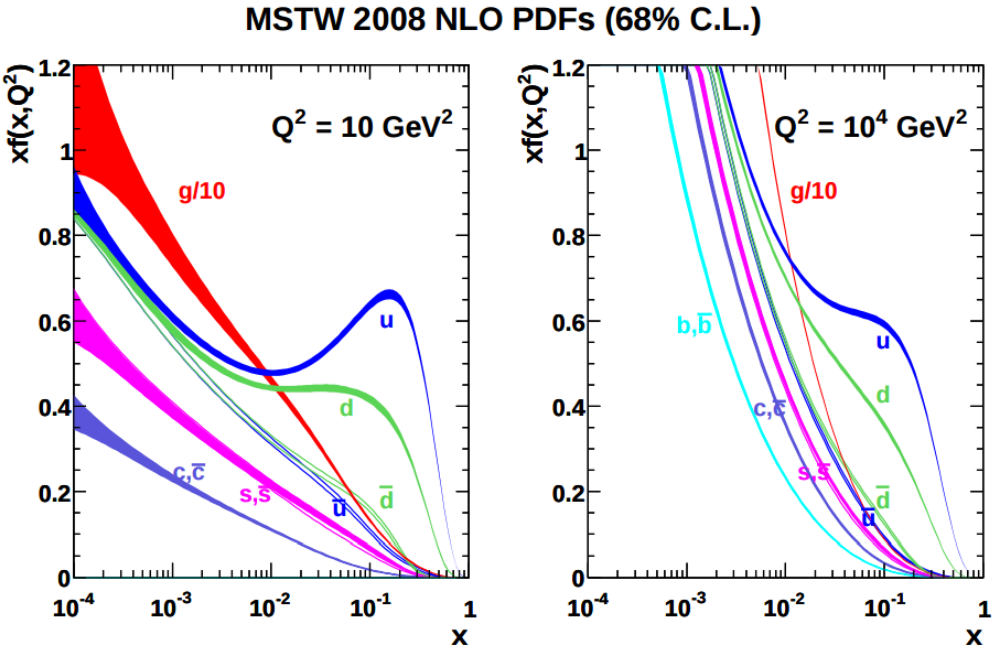


Figure 4.4.: Parton density functions and their dependence on $xf(x, Q^2)$ at $Q^2 = 10\,\mathrm{GeV}^2$ and $Q^2 = 10^4\,\mathrm{GeV}^2$, obtained by the MSTW 2008 NLO group. It is possible to notice the dominance of the gluon for small x ranges and of the valence quarks for large x ranges [48, 49]

.

$$\Delta_i(q_1^2, q_2^2) = \exp\left(-\int_{q_1^2}^{q_2^2} \frac{dq^2}{q^2}\frac{\alpha_S}{2\pi}\int_{z_{min}}^{z_{max}} dz \int_0^{2\pi} d\phi P_{i,j}(z,\phi)\right). \tag{4.6}$$

In this expression, the equality $\frac{d\theta^2}{\theta^2} = \frac{dq^2}{q^2}$ is used. The range $[z_{min}, z_{max}]$ specifies the range in z in which the emissions are resolvable, outside they cannot be detected and they are not included in the parton shower. To better understand the Sudakov form factors, it is possible to consider the analogy in the radioactive decay, where having a decay probability per unit of time λ, the probability for a decay not to occur in the time interval Δt is given by $e^{-\lambda\Delta t}$.

Fixing the scale beyond which the parton shower cannot develop any more (*hadronisation scale*) to Q_0^2, whose value is typically 1 GeV^2, the parton shower is generated as follows. Given an initial scale Q^2 (which is referred to as *resummation scale*) one needs to solve the equation $\Delta_i(Q^2, q_1^2) = R_1$ for a scale q_1^2 where R_1 is a random number thrown in the range [0,1]. If the resulting q_1^2 is below the hadronisation scale then the splitting is unresolved and the shower is terminated, otherwise the splitting occurs and the procedure is

repeated for the resulting partons iteratively taking as initial scale q_1^2, until all the resulting splittings fall below the scale Q_0^2. At each step, the z and ϕ for the produced emissions are computed according to the splitting functions $P_{ij}(z, \phi)$ using Monte Carlo methods.
Once the parton shower emissions fall below the hadronisation scale, the *hadronisation* process starts to develop. In this energy regime the strong coupling constant α_S is large enough to confine the partons in colourless hadrons. The hadronisation process occurs at a level where the perturbative approximations of QCD are not reliable any more. For this reason, the formation of hadrons is described through phenomenological models. Two important examples of such models are the *Lund String Model* [50] and the *Cluster Model* [51]. The first comes from the observation that the potential energy of colour sources, such as quark-antiquark pairs, increases linearly with their separation if their distance is large enough. This linear increase of the potential corresponds to a force of attraction which is independent on the distances of the colour charges. Due to the gluons self interactions the field lines are attracted to each other and collapse into a string (this is not the case for electromagnetic field lines, which are spread away because the photon has no self interaction). Since the quark and the antiquark move rapidly apart, the string is stretched and the potential energy grows. Once this potential reaches the order of the hadron masses, then the string can break along its length by creating a new quark-antiquark pair. The new quark and the new antiquark are connected by the two remaining string segments to the original antiquark and quark respectively. The two pairs created in this way continue to move apart, so the two strings stretch and the process starts again iteratively until all the energy is converted in hadrons. A visual representation of such a process is shown in Fig. 4.5. A typical MC generator involving the Lund String Model for the hadronisation is Pythia [52].

The Cluster Model uses a property of the parton shower which is known as *colour preconfinement* that was found by Amati and Veneziano [54]. The property basically implies that the partons created during the parton shower form colour singlet structures, called clusters, which have an asymptotic invariant mass distribution. These clusters are identified as proto-hadron structures that decay into the hadrons observed in the final state. MC generators that use the cluster model are Herwig [55] and Sherpa [56]. Fig. 4.6 shows a visual representation of this model.

Fig. 4.7 summarises the different generation stages taking as an example the $t\bar{t}H$ process. At the beginning the momenta of the partons participating in the hard scattering are ob-

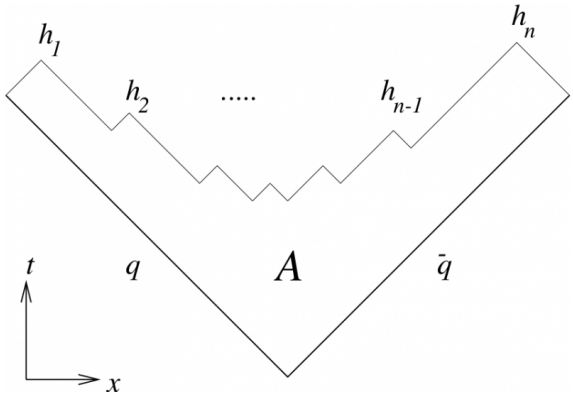


Figure 4.5.: A visual representation of the Lund String Model for the hadronisation process. A represents the world-sheet of the string; $h_1, h_2, ..., h_n$ represent the hadrons produced by the string breaking [53]

.

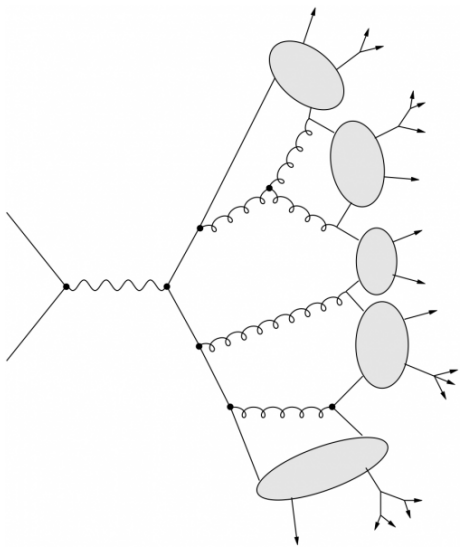


Figure 4.6.: A visual representation of the Cluster Model for the hadronisation process [53]

.

tained using the PDFs parametrised at the energy scale of the considered process. Then, the red portion of the figure represents the first stage in the event generation, being the hard-scattering followed by the decays of the created Higgs boson and top quarks, depicted as red blobs. Scientists in the ATLAS collaboration use several different matrix element generators for the modelling of the hard process. Some of them are Sherpa [56], MadGraph [57] and Powheg [58]. The accelerated coloured partons radiate gluons which create the parton shower (depicted in blue). This radiation corresponds to high order corrections to the hard process and simulates the Initial State Radiation (ISR) and the

Final State Radiation (FSR). The next stage of the simulation consists in the transition of the partons produced in the parton showers into colourless hadrons. This process is called *hadronisation* and it is depicted with the green circles. The decay products of these hadrons, are then shown with the green lines. MC generators used for the parton showers and hadronisation are typically Herwig [55] or Pythia [52].
The final step is the simulation of the Underlying Event (UE) (violet), which describes possible interactions between the proton remnants, contributions of ISR and FSR, Multi Parton Interaction (MPI) or in-time pile-up.

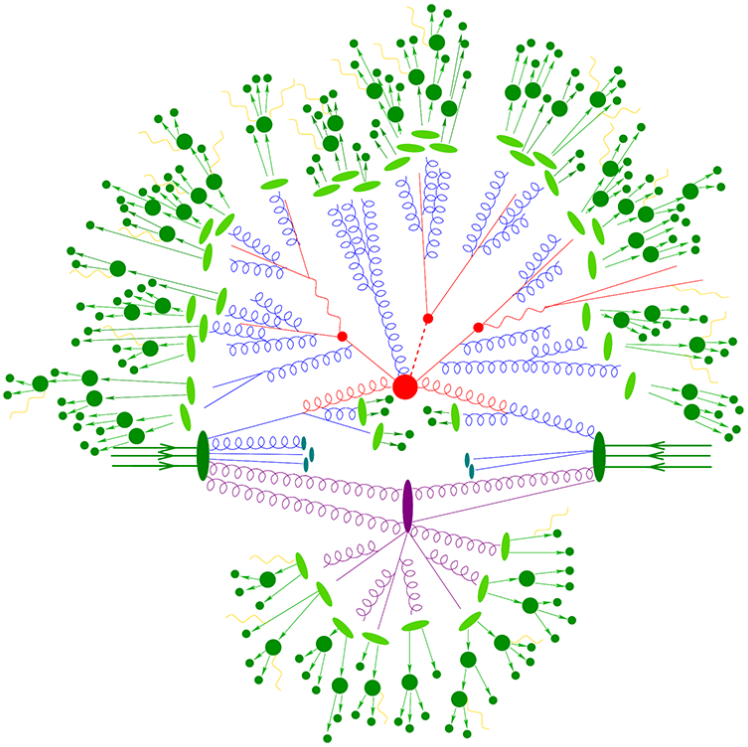


Figure 4.7.: A schematic illustration of the different steps of a MC event generation.

4.2.2. Detector Simulation and Digitisation

All the simulated particles that have a lifetime large enough to travel from the interaction point through the detector, are processed using an ATLAS detector simulation using Geant4 [59], which is a framework that provides a simulation of interactions of particles with the matter that builds the detector components. Since the computing time needed to perform this step is very large, often a simplified simulation model for the calorimeter

Sample	Generator	PDF	Shower	Normalisation
$t\bar{t}H$	MG5_aMC	NNPDF3.0NLO	Pythia 8.2	(N)NLO
$t\bar{t}$ + jets	PowHeg	CTEQ6L1	Pythia 6.428	NNLO+NNLL
W + jets	Sherpa	CT10	Sherpa 2.1.1	NNLO
Z + jets	Sherpa	CT10	Sherpa 2.1.1	NNLO
Single top (s-channel, Wt)	PowHeg	CT10	Pythia 6.428	aNNLO
Single top (t-channel)	PowHeg	CT10f4	Pythia 6.428	aNNLO
$t\bar{t}V$	MG5_aMC	NNPDF3.0NLO	Pythia 8.2	NLO
Diboson	Sherpa	CT10	Sherpa 2.1.1	NLO

Table 4.1.: A summary of the various physics samples and the used hard processes generators, PDFs, parton shower and the normalisation precision considered for the calculation of the cross section [61].

response is used. Such simulation is called *fast* simulation and labelled as AFII, while the complete one is referred to as *full* simulation, or FS. After this, a digitisation software is used to transform the output obtained with Geant4 in the default output format of the detector. Finally, after the digitisation step, the events are reconstructed as it is explained in Chapter 5. In the following sections, the simulation and data-driven techniques used to model the $t\bar{t}H$ signal and the background processes are described. All samples use a top quark mass of 172.5 GeV. Decays of b and c hadrons are simulated using EvtGen v1.2.0 [60], except in samples simulated by Sherpa. The Table 4.1 summarises the different MC samples considered for the analysis presented in this thesis.

4.3. Monte Carlo Samples

4.3.1. Signal Modelling

The $t\bar{t}H$ signal modelling is performed using MadGraph5_aMC@NLO version 2.3.2 [62] (referred to in the following as MG5_aMC) for the matrix element (ME) calculation, interfaced to Pythia 8.210 [63] parton shower (PS) generator using the A14 tune [64] for the tunable parameters used to model the UE. The used PDF setting is NNPDF3.0NLO [47], while the factorisation and renormalisation scales are both set to $\mu_F = \mu_R = H_T/2$, where H_T is the scalar sum of the transverse masses $\sqrt{p_T^2 + m^2}$ of all the particles appearing in the final state. The decay of the top quarks is simulated by MadSpin [65], which preserves all the spin correlations. The mass of the Higgs boson is set to 125 GeV and all its possible decay modes are included. The $t\bar{t}H$ cross section is computed at NLO [66–70]. The branching fractions for the Higgs decays are calculated using HDECAY [71].

4.3.2. $t\bar{t}$ +jets Background Modelling

The dominant $t\bar{t}$ background is modelled using the Powheg-Box v2 NLO generator [58, 72–74] using the CT10 PDF. The simulation is done setting the *hdamp* parameter, which controls the p_T of the first emission beyond the Born level, equal to the top quark mass. To model the parton shower and the hadronisation, Pythia 6.428 [52] is used together with the CTEQ6L1 PDF set [46] and the Perugia2012 [75] UE tune. The obtained sample is normalised to the Top++2.0 [76] theoretical cross section of 832^{+46}_{-51} pb, which is calculated at NNLO in QCD and includes resummation of NNLL soft gluon terms [77–81]. In addition, some alternative $t\bar{t}$ samples are used to derive systematic uncertainties. They are described in Sec. 7.1.3.

For both, the nominal and the alternative $t\bar{t}$ samples, a correction is computed for the top quark p_T and the p_T of the $t\bar{t}$ system in order to match predictions at NNLO accuracy in QCD [82, 83]. This method is referred to as a *reweighting* procedure. The correction defined in this way is not applied to the $t\bar{t}$ events with additional b-jets. Those events have instead a dedicated reweighting which is described below.

A categorisation is defined for the $t\bar{t}$ background according to the flavour of the additional jets produced in an event, this procedure is the same as in Ref. [84]. Such as it is explained in Chapter 5, jets are reconstructed starting from stable particles using the anti-k_t algorithm. The flavour of the jets is determined by matching within a radius of $\Delta R < 0.4$ to b or c hadrons. Jets matched to exactly one b hadron, with p_T above 5 GeV, are labelled b-jets, while those matched to two or more b hadrons are labelled B-jets (without p_T requirement on the second hadron). For c and C jets the definition is analogous. Events which have at least a b or B-jet, excluding jets from top or W decays, are labelled as $t\bar{t}+ \geq 1b$, while the events without any b or B-jet but with at least one c or C-jet are labelled as $t\bar{t}+ \geq 1c$. These two contributions together are referred to as $t\bar{t}$ +HF. The events with no $t\bar{t}$ +HF jets are labelled as $t\bar{t}$ +light.

It is possible to define a more detailed classification: the events which have at least three b or B-jets are labelled as $t\bar{t}+ \geq 3b$, those having exactly two b or B-jets are labelled as $t\bar{t} + b\bar{b}$, those having only one B-jet are labelled as $t\bar{t} + B$, and finally those having only one b-jet are labelled as $t\bar{t} + b$. Events with c jets or C-jets can be divided analogously. The latter classification is particularly useful for modelling studies, i.e. to compare the produced events among generators and to derive corrections or estimate uncertainties. Since the $t\bar{t}+ \geq 1b$ is the main background, it is important to model it with the best possible precision. Thus the nominal Powheg+Pythia 6 sample and all the

other alternative $t\bar{t}$ samples are corrected in order to match the predictions of a NLO $t\bar{t} + b\bar{b}$ sample generated using Sherpa+OpenLoops [56, 85]. This sample uses CT10 four-flavour scheme PDF set. For this sample, the renormalisation scale is set to the CMMPS [86] value of $\mu_{\text{CMMPS}} = \prod_{i=t,\bar{t},b,\bar{b}} E_{T,i}^{1/4}$, while for the factorisation scale, the value is set to $H_{\text{T}}/2 = \frac{1}{2}\sum_{i\in FS} E_{T,i}$. The resummation scale, which sets an upper bound for the hardness of the parton shower emissions, is set to $H_{\text{T}}/2$. The correction is performed by applying a kinematic reweighting separately in all of the $t\bar{t}+ \geq 1b$ sub-categories ($t\bar{t} + b\bar{b}$, $t\bar{t} + B$, $t\bar{t} + b$, $t\bar{t}+ \geq 3b$), in such a way that at the end the relative normalisation of the sub-categories and the kinematic distributions match the ones obtained with Sherpa+OpenLoops. In each sub-category, a reweighting is applied using the p_T of the top quark and of the $t\bar{t}$ system. This is followed in the $t\bar{t}+ \geq 3b$ and $t\bar{t} + b\bar{b}$ sub-categories by a reweighting of the ΔR between the b-jets and the p_T of the b-jet system. In the $t\bar{t} + B$ and $t\bar{t} + b$ sub-categories, the B or b-jet p_T and η are used instead. Some topologies included in the NLO calculations and labelled as $t\bar{t}+ \geq 1b$ are not reweighted, these include events with b-jets from MPI and from FSR. The predicted cross section for all the sub-categories and for the different generators considered, is shown in Figure 4.8.

4.3.3. Other Backgrounds

Other background samples used in this analysis consist of single top production, W/Z+jets, diboson production in association with jets, $t\bar{t}V$ ($V = W, Z$) events. The Wt and s-channel single top quark backgrounds are obtained using the Powheg-Box 2.0 generator and the CT10 PDF set [87, 88]. To handle the overlap between the $t\bar{t}$ and Wt, the *diagram removal* scheme [89] is used. The t-channel single top-quark events are generated with the Powheg-Box v1 generator and the CT10f4 PDF. All these samples are interfaced to Pythia 6.428 with the Perugia 2012 UE tune. The t- and s-channel samples are normalised to the NNLO cross sections predictions of [90–92].

The W/Z+jets events and diboson production in association with jets samples are generated with Sherpa 2.1. For W/Z+jets samples, matrix elements are calculated for up to two partons at NLO and four partons at LO using the Comix [93] and OpenLoops generators with the Sherpa parton shower [94] using the ME+PS@NLO prescription [95]. The events are normalised to the NNLO cross section as [96]. The diboson+jets samples are generated following the same approach but with up to one additional parton at NLO and up to three additional partons at LO. They are normalised to their respective NLO cross sections.

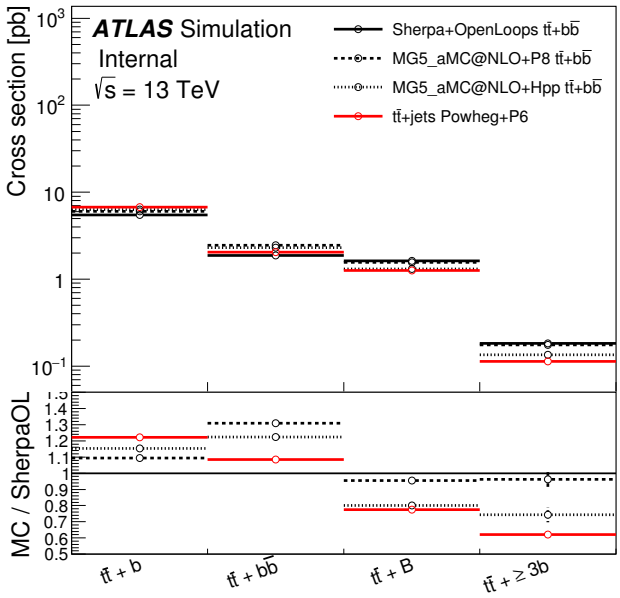


Figure 4.8.: The predicted cross sections for each of the $t\bar{t}+ \geq 1b$ sub-categories [61]. The inclusive prediction obtained with POWHEG+Pythia 6 is compared to the four-flavour calculations from Sherpa+OpenLoops and from MG5_aMC with different parton showers. The reweighting to Sherpa+OpenLoops has not been applied.

Events for $t\bar{t}V$ are generated using a NLO matrix element with MG5_aMC interfaced to Pythia 8.210 with the `NNPDF3.0NLO` PDF and A14 UE tune.

5 Object reconstruction at ATLAS

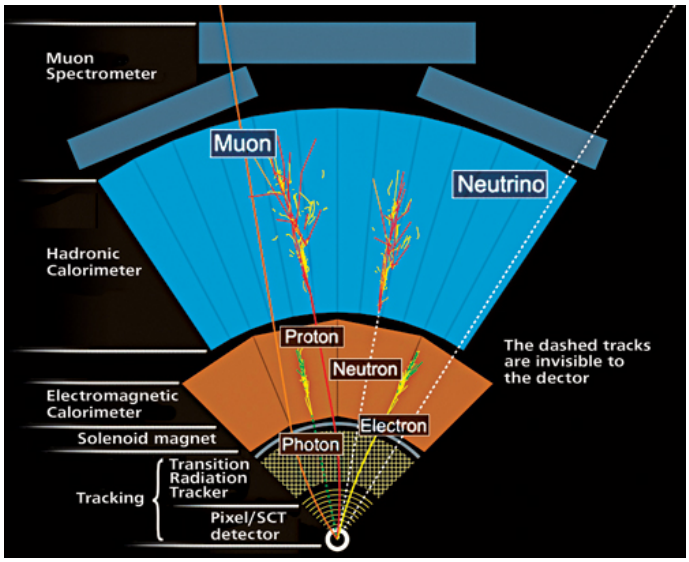
The ATLAS detector provides information for reconstruction of the objects in the final state of proton-proton collisions. The detector has an *onion-like* structure being composed by several cylinders one inside another. Each sub-detector is designed to identify and measure a particular set of particles.

Particles pass first through the tracking system, which is the innermost layer, able to reconstruct the tracks of the electrically charged particles. These tracks are bent by the solenoidal magnetic field so that the sign of their electric charge and momentum can be measured.

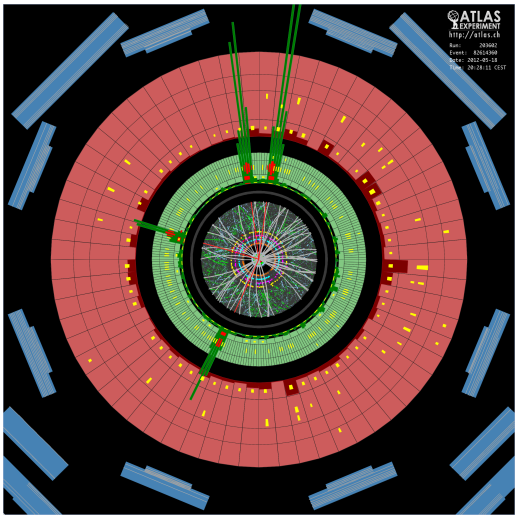
Particles pass then through the electromagnetic and hadronic calorimeters, which measure their energy. Calorimeters can interact with particles also if they are not electrically charged, as for example is the case for photons. Analysing the amount of particle energy loss inside the calorimeters, it is also possible to discriminate between particles which can or cannot interact strongly. In this way, for example, hadrons are distinguished from electrons and photons.

Particles which are not stopped by the calorimeter system, can be detected using the external muon tracking system. A sketch of the various particle signals in the different layers of the ATLAS detector, is shown in Fig. 5.1a.

Using this information, it is possible to identify particles and reconstruct their momentum and energy, Fig. 5.1b shows an example of event reconstruction in ATLAS. It is impossible though to directly detect neutrinos, which interact only weakly with matter. Some indirect information can be deduced from the so-called missing transverse momentum.



(a)



(b)

Figure 5.1.: Fig. 5.1a: Onion-like structure of ATLAS and particle identification for each layer, Fig. 5.1b: A display of a collision event at the ATLAS experiment. It shows a good candidate for a Higgs boson decaying into four leptons with a total mass of about 124.5 GeV ($H \rightarrow ZZ \rightarrow 4e$). The electrons are shown by red tracks inside the tracker and green ones in the calorimeters.

5.1. Particle tracks and primary vertices

The passage of charged particles through the detector leaves some tracks, whose detection and reconstruction is crucial to identify electrons, muons and reconstruct interaction vertices. Track reconstruction happens in the ID, where it is performed following different steps each one using a different pattern recognition algorithm [97]. The algorithms employed use a so called *inside-out* pattern recognition technique. Its main idea is to build the track *seeds* starting from space points located inside the silicon detector and then extending the track candidate out towards the TRT. The hits not used by this algorithm, are then taken into account by a *back-tracing* (also called *outside-in*) algorithm. This algorithm does the opposite to the inside-out one, starting to seed in the TRT and then extending the obtained track candidate towards the silicon detectors. After the back-tracing another inside-out sequence is employed but with a looser pattern requirement for the recognition of the tracks. The goal of this last step is to collect the missed low p_T tracks. For this reason this is called *low p_T tracking*.
The parameters that are considered in the reconstruction of the tracks are: the minimum distances to the interaction point defined in the transverse plane, d_0, and in the longitudinal direction, z_0; ϕ and θ angles defined in the transverse and in the longitudinal direction respectively; $\frac{q}{p}$ being the charge of the track divided by the momentum. A visual representation of these parameters is shown in Fig. 5.2. In order to improve the performance of the algorithms, some quality cuts are defined and used at the different stages of the track reconstruction.
Since a great number of protons collide at each bunch crossing, for each event, several interaction vertices can be reconstructed. Adaptive vertex fitting algorithms [98] are used to reconstruct primary vertices from a combination of tracks. Some boundary conditions are employed to keep the vertices inside the estimated location of the spatial region defined around the interaction point and in which the beams' profiles overlap (*beam spot*). To improve the resolution on the vertices' spatial position, the vertices are required to have at least two associated tracks with $p_T > 400$ MeV. The vertices are ordered by sum of the p_T^2 of the associated tracks of the considered vertex. The one with the highest sum is considered to be the primary vertex, while the others are labelled as pile-up vertices.

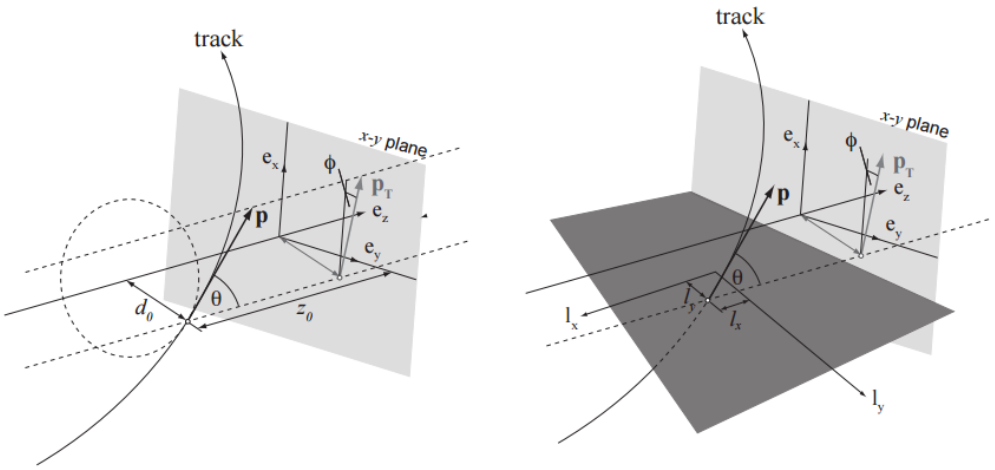


Figure 5.2.: A track reconstructed with respect to two different surfaces: with a Perigee representation of the track (left), with an intersection with a planar surface (right) [99].

5.2. Leptons

The design of the ATLAS detector permits a very efficient detection of charged leptons through the reconstruction of charge, momentum and direction of the tracks. Since the analysis described in this thesis does not make explicit use of the tau-leptons, their reconstruction is not discussed here. Depending on the decay mode, the tau-leptons are reconstructed as isolated electrons or muons or as jets.

5.2.1. Muons

To reconstruct the muons, the information coming from different subdetectors (inner detector, calorimeters and muon spectrometer) is combined [100]. Depending on the reconstruction method, muons can be defined within four different categories:

- *Stand-alone* muons: The muon trajectory is only reconstructed from hits in the MS. The tracks are built starting from the segments found in the muon stations, which are combined together, and extrapolated back to the beam line, taking into account the energy loss in the calorimeters.
- *Segment-tagged* muons: This classification applies if a track of the ID is matched to at least one track segment coming from the MDT or CSC chambers.

- *Calorimeter-tagged* muons: The muon is reconstructed by a combination of tracks in the ID and energy deposits in the calorimeters that are consistent with an energy deposit by minimum ionising particles.

- *Combined* muons: A fit algorithm combines the independent reconstruction obtained in the MS and in the ID. This reconstruction algorithm has the best p_T resolution and rejection efficiency for fake muons.

In order to fulfil the requirements of the different physics analyses performed in ATLAS, four different identification selections are provided (Medium, Loose, Tight, High-p_T). The Medium identification criteria is used as default selection for muons in ATLAS and it minimises the reconstruction and calibration uncertainties. Some corrections are applied to MC simulations in form of efficiency scale factors, in order to match isolation and trigger efficiency in data. These corrections are obtained by comparing MC predictions to large samples of $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ data events using the tag-and-probe method. In Fig. 5.3 the efficiency of the reconstruction for Medium muons as a function of their p_T is shown. As can be seen, it is very close to 99%.

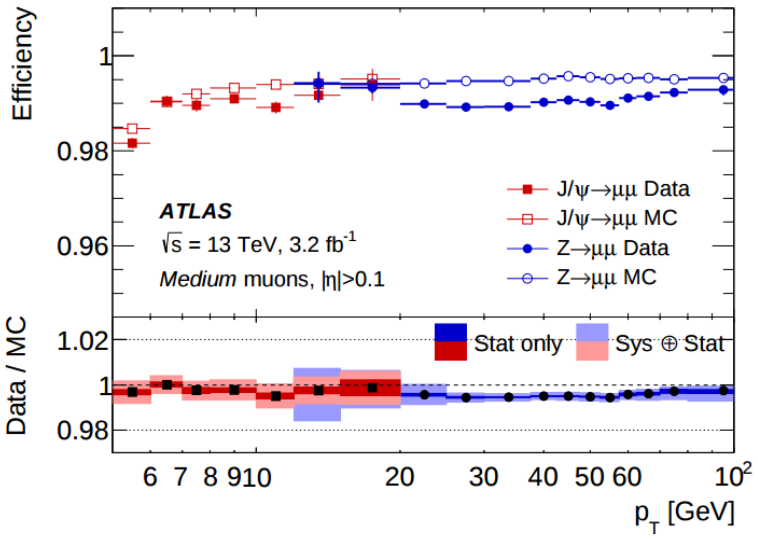


Figure 5.3.: Reconstruction efficiency for the Medium muon selection as a function of the p_T of the muon, in the region $0.1 < |\eta| < 2.5$ as obtained with $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events. The studies are performed using 3.2 fb^{-1} of pp collisions data at $\sqrt{s}$=13 TeV collected in 2015 [100].

5.2.2. Electrons

To reconstruct electrons, information coming from the ID and from EM calorimeter is combined [101]. The reconstruction is achieved in two steps. The first one, referred to as *cluster reconstruction*, creates an electron candidate using the EM calorimeter clusters that are matched to the corresponding tracks in the ID. The clustering in the EM calorimeter is done using the *sliding window* algorithm [102]. In this algorithm the cells within a rectangular window with the size of 3x5 towers size (a tower size is 0.025x0.025 in the $[\eta, \phi]$ phase space) are summed and the position of the window is chosen to maximise the energy deposit inside it.

The second step is called *electron identification*. In this step, a discrimination between signal-like or background-like electrons (these can come from converted photons or hadronic showers in the calorimeters) is achieved through a likelihood method [103]. This method is an MVA algorithm that uses probability density functions associated to signal and background built from discriminating variables based on electron track and cluster measurements, such as calorimeter shower shape, quality of the matching between track and calorimeter cluster, track quality, energy lost to bremsstrahlung, and hadronic leakage. Starting from these probability density functions, a likelihood is defined under the signal and background hypothesis.

$$\mathcal{L}_{S/B}(\boldsymbol{x}) = \prod_{i=1}^{N} P_{S/B,i}(x_i) \tag{5.1}$$

where $P_{S/B,i}(x_i)$ is the signal/background probability density function of the i-th variable evaluated at x_i in the vector of variable $\boldsymbol{x}$. From the likelihoods, the $d_{\mathcal{L}}$ discriminant is defined as follows:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}. \tag{5.2}$$

Cutting on this discriminant, five different operating points for the electron reconstruction are defined: Very Tight, Tight, Medium, Loose, Very Loose, each one corresponding to different levels of electron efficiency and background rejection.

As for the muons, some corrections are applied to MC simulation in form of efficiency scale factors, in order to match isolation and trigger efficiency in data. These corrections are obtained by comparing MC predictions to $J/\psi \to ee$ and $Z \to ee$ data samples using the tag-and-probe method. In Fig. 5.4 the combined electron reconstruction and identification efficiencies in $Z \to ee$ events as a function of the transverse energy E_T, and as a function of pseudorapidity η are shown [104].

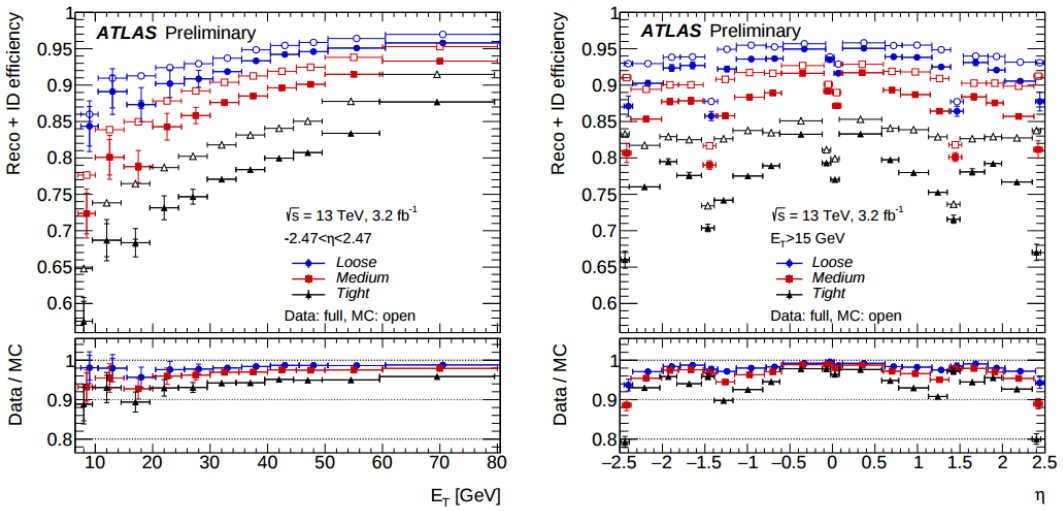


Figure 5.4.: Combined electron reconstruction and identification efficiencies in $Z \to ee$ events as a function of the transverse energy E_T, integrated over the full pseudorapidity range (left), and as a function of pseudorapidity η, integrated over the full E_T range (right). The data efficiencies are obtained from the data-to-MC efficiency ratios measured using $J/\psi \to ee$ and $Z \to ee$ tag-and-probe method, multiplied by the MC prediction for electrons from $Z \to ee$ decays [104].

5.3. Jets

After a pp collision, charged and neutral hadrons are produced and appear in form of *jets* of particles. These ensembles of particles are interpreted as coming from a two-stage process involving the production and subsequent *fragmentation* of gluons or quarks. Jets are also referred to as a spray of collimated particles. As already explained in Sec. 4.2.1, the process of generation of colourless hadrons from the partons and the following formation of the jets is called hadronisation. A jet is identified by the detector, which reconstructs its momentum and direction, giving information about the parton that generated it. Jets used to develop the analysis presented in this thesis are reconstructed with the anti-k_T algorithm [105], which is a particular sequential recombination jet algorithm. These algorithms are designed to provide a jet reconstruction that is insensitive to the so called infrared and collinear (IRC) emissions. In fact, soft (infrared) gluon emissions can lead to an incorrect reconstruction of jets. These algorithms use the idea of defining a jet as the cone region around some distances between clusters of dominant energy flow. The distance between

clusters i and j is defined as:

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta_{ij}^2}{R^2}\,, \tag{5.3}$$

where $\Delta_{ij}^2 = (\phi_i - \phi_j)^2 + (y_i - y_j)^2$, and $p_{T,i}$, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuthal angle of particle i. For $p = 1$ one recovers the inclusive k_T algorithm, the case where $p = 0$ corresponds to the inclusive Cambridge/Aachen algorithm. The case of $p = -1$ corresponds to the anti-k_T algorithm. The latter is an IRC safe algorithm that gives as an output circular hard jets. R is a radius parameter that defines the size of the jet. For the analysis presented in this thesis, jets with $R = 0.4$ are used. Fig. 5.5, shows the reconstruction of jets using different algorithms.

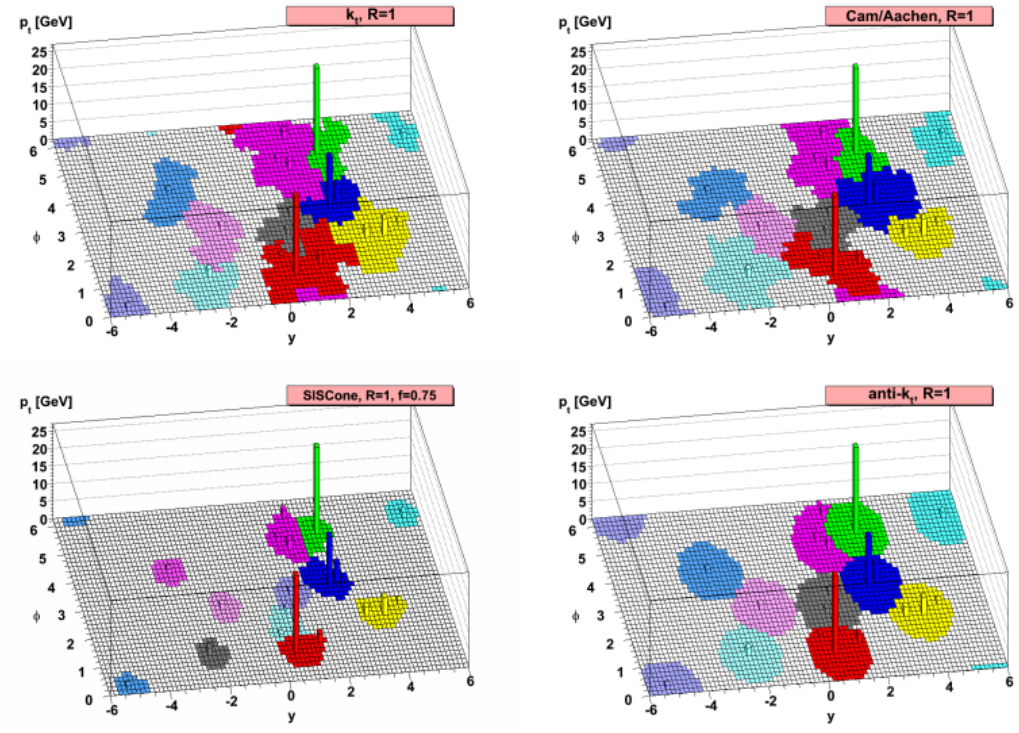


Figure 5.5.: A sample parton-level event clustered with four different jet algorithms [105].

As inputs to the jet algorithm topological calorimeter clusters (topo-clusters) are chosen [102]. The clusters are obtained from the topological algorithm, which starts with a seed cell and iteratively adds to the cluster the neighbour of a cell already in the cluster, provided that the energy in the new cell is above a threshold defined as a function of the

expected noise. This algorithm is efficient at suppressing noise on clusters with large number of cells. It is the best algorithm for jet and missing transverse momentum reconstruction.
After a jet is reconstructed, some corrections are applied [106, 107]:

- pile-up corrections, whose aim is to subtract the energy deposit due to pile-up processes from the reconstructed jet energy;
- correction to the jet directions;
- calibration of the energy of the jet based on MC generated samples;
- a final correction to data derived using in-situ measurements.

In order to reduce a pile-up component of the reconstructed jets, a quantity referred to as Jet Vertex Fraction (JVF) is defined. This quantity is defined as the scalar sum of transverse momenta of all the tracks matched to the considered jet divided by the sum of transverse momenta of tracks within the jet and originated from the hard-scatter vertex. This can also be interpreted as the fraction of the energy of the jet that is associated to the hard interaction [108]. During Run 1, a cut on the JVF of 0.5 was applied. This removed the pile-up contribution on the jet reconstruction, but led to hard-scatter jet efficiencies that depended on the number of reconstructed primary vertices in the event. For this reason, for Run 2, new track-based variables were developed in order to obtain a hard-scatter jet efficiency stable as a function of the number of primary vertices. These variables, are combined in an MVA algorithm called Jet Vertex Tagger (JVT) [109]. Fig. 5.6 shows efficiency curves for JVF and JVT and their dependence on the number of primary vertices.

5.4. *b*-tagging

It is crucial for the analysis described in this thesis to be able to distinguish jets originating from light quarks from jets from heavy quarks, as for example b-quarks. To identify jets produced from a b-quark, sophisticated techniques are used, which are grouped together in a process called b-tagging. The goal of each b-tagging algorithm is to identify the b-jets with the highest efficiency possible and rejecting the background arising from jets originated from *light*-quarks, c-quarks or gluons. The idea behind the b-tagging is to take advantage of the fact that, when hadronising, b-jets contain b-hadrons such as $B^{\pm}$, B^0 and B_s. These hadrons are typically long lived particles, with the mean lifetime of the order

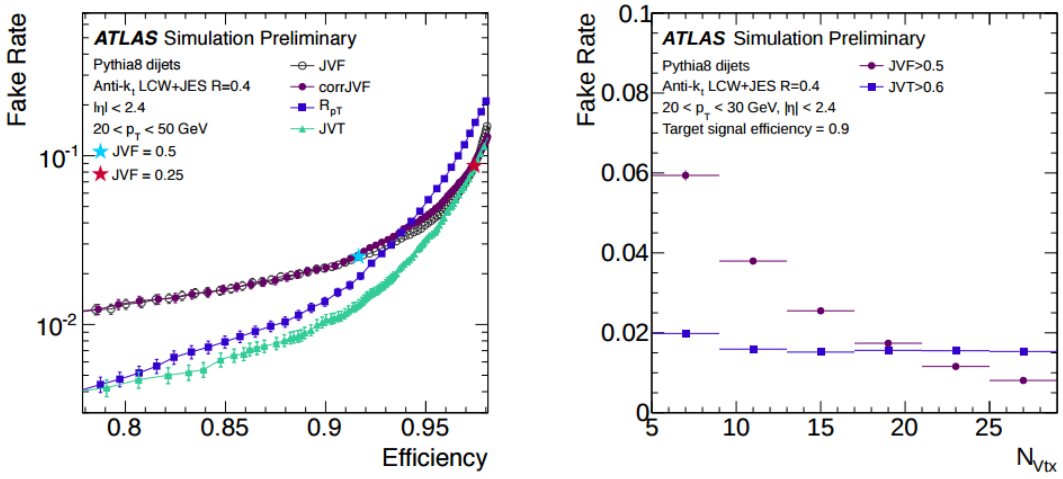


Figure 5.6.: (a) Fake rate from pile-up jets versus hard-scatter jet efficiency curves for JVF and JVT. (b) Number of primary vertices dependence of the pile-up jet fake rate when imposing cuts on JVT (blue) and JVF (violet) such that the inclusive hard-scatter jet efficiency is 90% [109].

of $\tau \simeq 1.5$ ns. This means that these particles (having energy of the order of tens of GeV) would travel for on average 3 mm before decaying. This behaviour can lead to two effects that can be used for *b*-jet tagging:

- Presence of secondary vertices in the primary jet;
- Presence of soft leptons inside the jet originating form the semileptonic decay of the *b*-hadron.

A schematic depiction of this mechanism is shown in Fig. 5.7. Using ID information, it is possible to detect a presence of a secondary vertex in the event and use it to identify the *b*-quark jets.

There are several *b*-tagging algorithms in ATLAS, which are sometimes combined using dedicated MVA techniques, which provide the best performance in the separation of *b*-jets and other flavour jets [110]. The MV1 tagger, widely used during Run 1, was a neural network discriminant which combined several MVA based taggers. In Run 2, a new BDT based tagger called MV2, combines together 24 input variables. Three different versions of the MV2 tagger are provided, each one distinguished by the contribution of *c*-jets used in the training phase for the background discrimination. The taggers used in this analysis

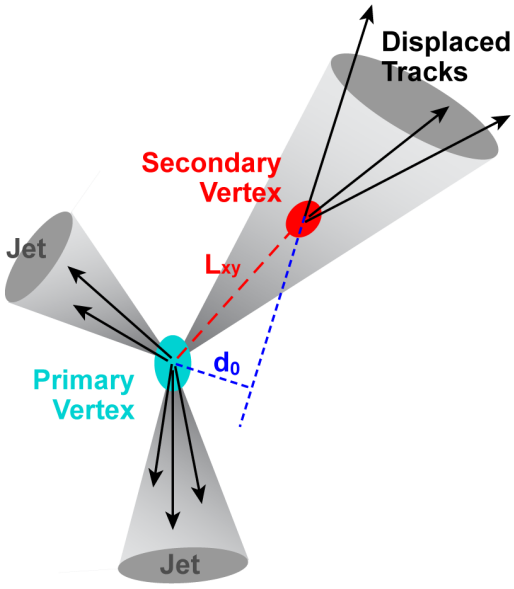


Figure 5.7.: Visual representation of a *b*-hadron decay within a jet resulting in a secondary vertex displaced to the primary vertex

is the MV2c10 which corresponds to a 15% *c*-jet fraction. A cut on the output of the MV2 distribution defines an operating point which corresponds to a specific *b*-jet tagging efficiency and background rejection. Data/simulation efficiency scale factors are defined to correct the *b*-tagging efficiency in simulation to data. Table 5.1 shows operating points for the MV2c10 *b*-tagging algorithm, including benchmark numbers for the efficiency and rejections rates.

BDT Cut Value	*b*-jet Efficiency [%]	*c*-jet Rejection	*light*-jet Rejection
0.9349	60	34	1538
0.8244	70	12	381
0.6459	77	6	334
0.1758	85	3	33

Table 5.1.: Operating points for the MV2c10 *b*-tagging algorithm, including benchmark numbers for the efficiency and rejections rates [110].

5.4.1. Tag Rate Function

Requiring three or four b-tagged jets for a background dominated by light jets, reduces significantly the number of selected MC events, making the modelling of such background a challenging task. The reduced statistics together with the large cross sections of some backgrounds can produce large statistical fluctuations in the kinematic distributions, which lead to instabilities in the likelihood fit, used to extrapolate a parameter of interest, due to spiky templates. To mitigate this effect the Tag Rate Function (TRF) method is used [111]. According to this method, no event is rejected based on the number of b-tagged jets, but a per-event weighting is applied. The event weight is obtained through the jet tagging efficiency, ε, which is a function of p_T, η, and flavour of the jet. In this way, for a given event with N_{jets} jets, the probability to contain exactly N_b b-tagged jets is defined as:

$$P(N_{tag} = N_b|N_{jets}) = \sum_{m+n=N_{jets}} \left(\prod_{i \in T_m} \varepsilon_i \prod_{j \in U_n} \left(1 - \varepsilon_j\right) \right),$$

where the sum is calculated for all the permutations in which T_m (U_n) designates the subset of m (n) jets considered (un)tagged.
As a consequence, the probability that a considered event consists of at least N_b b-tagged jets is:

$$P(N_{tag} \geq N_b|N_{jets}) = 1 - \sum_{N_{b'} < N_b} P(N_{tag} = N_{b'}|N_{jets}) \,.$$

With this method, a permutation is selected among all the possible ones for N jets and a given number of b-tags. The permutation is chosen according to the procedure described in the following. At a first step, the sum of the TRF weights defined above, S, for the permutations corresponding to a precise number of b-tagged jets is computed, and each partial sum, S_X is calculated and recorded. At a second step, a random number in the uniform range between 0 and 1 is selected. The permutation i corresponding to the partial sum up to i, which value is greater or equal to the random number, is chosen. A scheme showing this procedure is shown in Fig. 5.8
In the analysis described in this thesis, TRF is used to evaluate some systematic uncertainties for some MC samples with low statistics, and also for increasing the effective statistics of the samples used for training the ANN.

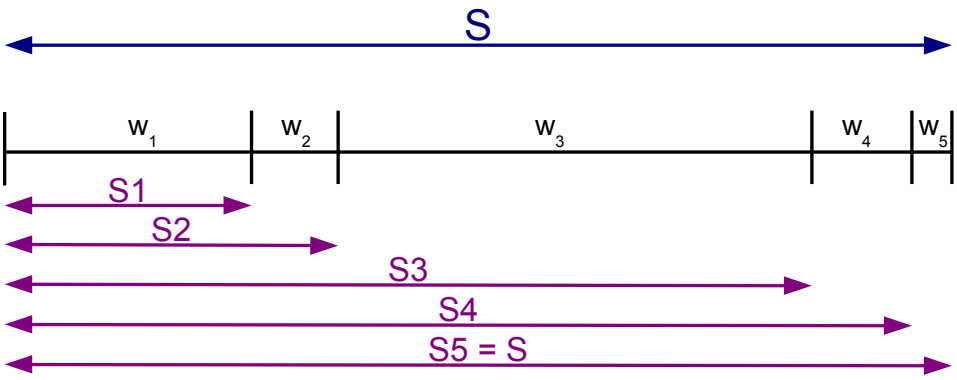


Figure 5.8.: A visual representation of TRF [61]. In this scheme, the choice of the permutation in the case where there are 5 possible permutations (for example 1 b-tag among 5 jets) is illustrated. S defines the total sum, SX are the partial sums, where $S5 = S$. If the random number falls, as an example, between $S2$ and $S3$ then the permutation 3 is picked.

5.5. Missing Transverse Momentum

Although neutrinos do not leave any direct sign of their passage through the detector, some information about their kinematics is possible to obtain. Partons carry all the 4-momentum of the proton, which collide longitudinally inside the detector. Although it is not possible to have information about the partons longitudinal 4-momenta, it is possible to assume that the sum of the component of the partons 4-momenta in the transverse plane is zero. In this way, a variable called *missing transverse momentum* $\vec{p}_T^{\,\text{miss}}$ is defined, whose magnitude is E_T^{miss}:

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}\,. \tag{5.4}$$

The $x(y)$ components of E_T^{miss} are obtained by the negative sum of the reconstructed and calibrated physics objects, as well as detector signal objects which are not associated to any of the reconstructed ones. To avoid double counting, a specific order for the physics object is generally followed [112]: electrons (e), photons (γ), hadronically decaying taus (τ), jets and finally muons (μ):

$$E^{\text{miss}}_{x(y)} = E^{\text{miss},e}_{x(y)} + E^{\text{miss},\gamma}_{x(y)} + E^{\text{miss},\tau}_{x(y)} + E^{\text{miss},jets}_{x(y)} + E^{\text{miss},\mu}_{x(y)} + E^{\text{miss,soft}}_{x(y)}, \tag{5.5}$$

where the $E^{\text{miss,soft}}_{x(y)}$ term takes into account soft calorimeter topo-cluster and tracks not associated to any reconstructed object.

5.6. Data Driven Background Determination

The background from events with a mis-identified lepton arises from the identification of *fake* or *non-prompt* leptons instead of the *real* ones.
The dominant sources for the mis-identified leptons, are semileptonic b-jet decays, long lived weakly decaying states (such as $\pi^{\pm}$ or K mesons) and, in case of electrons, also photons conversions or direct photons and π^0 showers reconstructed as an electron.
This kind of background is strongly influenced by the detector characteristics, so data-driven methods are used to evaluate it.
The Matrix Method [113], is a data-driven method which is based on the selection of two categories of events using loose and tight lepton selection requirements. The tight lepton selection is the same lepton selection as used in the analysis, while the loose selection is obtained reducing some lepton selection requirements. In this way, it is possible to write the following equations for the number of events passing the loose selection (N^{loose}) and for the number of events passing the tight selection (N^{tight}):

$$N^{\text{loose}} = N^{\text{loose}}_{\text{fake}} + N^{\text{loose}}_{\text{real}} \tag{5.6}$$

$$N^{\text{tight}} = N^{\text{tight}}_{\text{fake}} + N^{\text{tight}}_{\text{real}}, \tag{5.7}$$

where $N^{\text{loose}}_{\text{fake(real)}}$ is the number of events which pass the loose selection containing a fake (real) lepton and $N^{\text{tight}}_{\text{fake(real)}}$ is the number of events which pass the tight selection containing a fake (real) one. The efficiency $\epsilon_{\text{fake(real)}}$ for a fake(real) lepton that passed the loose selection to pass also tight selection is:

$$\epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}} \tag{5.8}$$

$$\epsilon_{\text{real}} = \frac{N_{\text{real}}^{\text{tight}}}{N_{\text{real}}^{\text{loose}}} \,. \tag{5.9}$$

Using the equations in 5.7, it is possible to obtain:

$$N^{\text{loose}} = N_{\text{fake}}^{\text{loose}} + N_{\text{real}}^{\text{loose}} \tag{5.10}$$

$$N^{\text{tight}} = \epsilon_{\text{fake}} N_{\text{fake}}^{\text{loose}} + \epsilon_{\text{real}} N_{\text{real}}^{\text{loose}} \,. \tag{5.11}$$

The efficiency ϵ_{real} is measured using $Z \rightarrow ll$ events dominated by real leptons and counting the tight events that pass the selection. ϵ_{fake} is instead obtained selecting a sample of events enriched in fake leptons. Having the two efficiencies $\epsilon_{\text{real(fake)}}$, a system of two equations and two unknown is obtained, which can be solved to obtain $N_{\text{fake}}^{\text{tight}}$:

$$N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{fake}} + \epsilon_{\text{real}}} \left(\epsilon_{\text{real}} N^{\text{loose}} - N^{\text{tight}} \right) \,. \tag{5.12}$$

6 Analysis Strategy and MVA Techniques

The purpose of this thesis is to give an overview of the search of $t\bar{t}H(H \rightarrow b\bar{b})$ signal in the single lepton channel produced using pp collisions at $\sqrt{s} = 13$ TeV at the LHC. The signal events include the production of a Higgs boson in association with a $t\bar{t}$ pair, whose decay products give rise to six non-overlapping high-p_T jets, of which two are light jets coming from the decay of the W boson, and four are b-jets coming from the decay of the $t\bar{t}$ pair and the Higgs boson. A charged lepton is also produced from the leptonic decay of the remaining W boson. In order to check the background modelling and simultaneously constrain the systematic uncertainties, events are classified into several categories. In the signal rich regions, an MVA variable is developed in order to give the best possible separation between signal and background.

6.1. Event Selection

The objects considered in this analysis are jets, b-jets, electrons, muons. The reconstruction of these objects is described in Chapter 5. In the following, additional requirements on these objects in the current analysis are explained.

Events were recorded using dedicated single-lepton triggers [114]: one with a low p_T threshold and isolation requirements and another with a high p_T threshold and no isolation requirement. These two triggers were combined by a logical OR in order to achieve the maximum efficiency. For muons, the lowest p_T threshold is 20 GeV for 2015 data period and 24 GeV for 2016, while the higher thresholds are 40 GeV for 2015 and 50 GeV 2015. For the electrons, isolated triggers with a p_T threshold of 24 GeV are used with non-isolated triggers at 60 GeV in both years, along with a 120(140) GeV trigger which also uses looser identification criteria.

For electrons, a tight likelihood identification criterion is imposed [115]. Electron tracks are required to match the primary vertex of the event. This is achieved by imposing the longitudinal impact parameter condition $|z_0 \sin\theta| < 0.5$ mm and the transverse impact parameter significance condition $\frac{d_0}{\sigma(d_0)} < 5$. Only electrons with $p_T > 25$ GeV and $|\eta| < 2.5$ are considered. Isolation criteria are also employed to further suppress background coming from non-prompt electrons [116].
For muons, a medium quality requirement is imposed [100] together with a *gradient* isolation criteria [117]. Additional requirements are imposed on the impact parameters: $\frac{d_0}{\sigma(d_0)} < 3$ and $|z_0 \sin\theta| < 0.5$ mm. Further requirements on muons are $p_T > 25$ GeV and $|\eta| < 2.5$.
Jets are required to have $p_T > 25$ GeV and $|\eta| < 2.5$ after calibration. Jet cleaning criteria are used to remove jets coming from noisy sources [118]. To suppress pile-up, an additional selection of JVT > 0.59 is applied to jets with $p_T < 60$ GeV and $|\eta| < 2.4$ [109].
Jets originating from the hadronisation of a b-quark are identified using the MV2c10 algorithm at the 70% working point, which corresponds to a b-tagging efficiency of 70% for b-jets in $t\bar{t}$ events, while the light-jet rejection is 381 and the charm jet rejection is 12. Scale factors are used to correct the MC efficiency to the data one. Events are required to have at least four jets, with at least two of them b-tagged.
An overlap-removal procedure is implemented to avoid double counting of individual detector responses. To prevent an electron energy deposit from also being reconstructed as jet, the jet closest to the electron within the range of $\Delta R < 0.2$ is removed. If any jet surviving the selection above is within $\Delta R < 0.4$ of the electron, the electron is discarded instead. Muons within $\Delta R < 0.4$ of a selected jet are also removed to reject background muons coming from decays of heavy-flavour hadrons of the jet. On the contrary, if the jet has fewer than three associated tracks, the muon is kept while the jet is removed. This is done to avoid any inefficiency in the reconstruction of high-energy muons that lose a significant amount of energy in the calorimeter.
No selection is made on missing transverse momentum. However, it is used for the event reconstruction purposes.

6.2. Event Categorisation

After the preselection defined above, events are further categorised according to the number of jets and number of b-jets and corresponding event regions are defined. The regions

are referred to in the following as (mj,nb), where m is the number of selected jets and n is the number of b-jets. In this way, nine independent regions are defined. If a region has S/B>1%, it is referred to as signal-enriched region (S and B indicate respectively the expected signal yield for a Higgs boson mass of 125 GeV and the expected background yield). For these regions, the signal significance is $S/\sqrt{B} > 0.5$ and these regions provide most of the sensitivity of the signal. The signal regions, are the three regions (5j,≥4b), (≥6j,3b), (≥6j,≥4b). The other six regions, (4j,2b), (4j,3b), (4j,4b), (5j,2b), (5j,3b) and (≥6j,2b), are referred to as control regions as they are almost purely background-only regions. These regions are useful to constrain the systematic uncertainties and, in this way, to improve the background prediction in the signal regions. For the regions with a large fraction of $t\bar{t}$ +HF events, a discrepancy between data and prediction is observed and it is discussed in Sec. 7.1.3. Tables 6.1-6.3 show the event yields for all the regions. Control plots of basic distributions for the inclusive multiplicity selection region (≥4j, ≥2b) are shown in Appendix A. A comparison between the predicted yields and data in each of the considered regions is shown in Fig. 6.1. Fig. 6.2a shows the signal and control regions and their S/B and $S/\sqrt{B}$. The background composition in different regions is shown in Fig. 6.2b. Here, the $t\bar{t}$ background is split in different categories as described previously. In the most sensitive regions, the Higgs boson decay into $b\bar{b}$ pairs constitutes approximately 90% of the total signal. In Fig. 6.2c the different contributions to the Higgs boson decays are shown for each region.

	4j,2b	4j,3b	4j,≥4b
$t\bar{t}$ + light	160 000 ± 30 000	5290 ± 1540	17.3 ± 10.7
$t\bar{t}$ + ≥1c	10 800 ± 2350	882 ± 297	11.7 ± 5.44
$t\bar{t}$ + ≥1b	4580 ± 925	1570 ± 473	76.2 ± 24.4
$t\bar{t}$ + W	99.0 ± 16.7	4.30 ± 0.972	0.035 ± 0.029
$t\bar{t}$ + Z	113 ± 21.2	14.1 ± 3.69	1.57 ± 0.418
Wt channel	5980 ± 1610	239 ± 81.9	3.46 ± 2.79
t and s channel	4330 ± 612	150 ± 28.3	6.12 ± 2.15
Diboson	416 ± 216	15.2 ± 10.2	3.88 ± 3.30
W + jets	5250 ± 2370	181 ± 97.7	1.64 ± 1.10
Z + jets	1210 ± 584	41.3 ± 22.6	0.460 ± 0.379
Fakes & non-prompt	9150 ± 4150	771 ± 364	29.1 ± 27.2
$t\bar{t}H$ ($H \to b\bar{b}$)	42.4 ± 5.63	23.3 ± 4.05	3.30 ± 0.868
$t\bar{t}H$ ($H \to WW$)	9.73 ± 1.71	0.504 ± 0.143	0.0045 ± 0.0046
$t\bar{t}H$ ($H \to$ other)	11.7 ± 1.89	0.820 ± 0.165	0.022 ± 0.0060
Total	202 000 ± 32 400	9190 ± 1920	155 ± 44.7
Data	208239	11686	218

Table 6.1.: Yields before the fit in the exclusive four jet regions.

	5j,2b	5j,3b	5j,≥4b
$t\bar{t}$ + light	90 800 ± 17 400	3640 ± 877	23.9 ± 14.9
$t\bar{t}$ + ≥1c	10 800 ± 2100	1170 ± 332	30.1 ± 11.6
$t\bar{t}$ + ≥1b	4440 ± 533	2230 ± 459	224 ± 61.7
$t\bar{t}$ + W	130 ± 21.5	8.27 ± 1.79	0.19 ± 0.073
$t\bar{t}$ + Z	147 ± 25.0	27.1 ± 5.86	4.76 ± 1.53
Wt channel	3470 ± 1140	218 ± 85.0	8.08 ± 4.97
t and s channel	1480 ± 283	87.1 ± 17.3	6.26 ± 2.50
Diboson	195 ± 108	15.7 ± 9.66	0.387 ± 0.280
W + jets	2300 ± 1090	165 ± 99.7	2.33 ± 3.13
Z + jets	413 ± 211	37.0 ± 26.8	0.717 ± 0.646
Fakes & non-prompt	3270 ± 1500	298 ± 152	20.0 ± 17.2
$t\bar{t}H$ ($H \to b\bar{b}$)	59.7 ± 6.64	46.8 ± 6.86	11.7 ± 2.58
$t\bar{t}H$ ($H \to WW$)	18.3 ± 2.70	1.25 ± 0.347	0.025 ± 0.013
$t\bar{t}H$ ($H \to$ other)	18.5 ± 2.73	1.69 ± 0.333	0.140 ± 0.101
Total	118 000 ± 19 500	7940 ± 1430	333 ± 79.4
Data	124688	10755	418

Table 6.2.: Yields before the fit in the exclusive five jet regions.

	≥6j,2b	≥6j,3b	≥6j,≥4b
$t\bar{t}$ + light	54 400 ± 24 300	2590 ± 1070	33.8 ± 21.9
$t\bar{t}$ + ≥1c	11 500 ± 3740	1550 ± 557	70.7 ± 36.8
$t\bar{t}$ + ≥1b	4840 ± 1190	3240 ± 796	674 ± 190
$t\bar{t}+W$	204 ± 39.9	20.8 ± 4.59	1.24 ± 0.391
$t\bar{t}+Z$	270 ± 46.1	65.7 ± 11.6	17.9 ± 4.23
Wt channel	2060 ± 821	212 ± 103	20.6 ± 13.7
t and s channel	632 ± 159	65.7 ± 15.7	8.28 ± 2.34
Diboson	164 ± 88.1	14.4 ± 8.33	2.03 ± 1.34
W + jets	1350 ± 650	106 ± 54.4	10.4 ± 6.67
Z + jets	264 ± 132	15.2 ± 8.11	1.47 ± 0.874
Fakes & non-prompt	1220 ± 556	272 ± 151	1.17 ± 1.17
$t\bar{t}H$ ($H \to b\bar{b}$)	104 ± 14.8	108 ± 16.3	43.5 ± 9.37
$t\bar{t}H$ ($H \to WW$)	53.0 ± 8.43	5.75 ± 1.19	0.496 ± 0.184
$t\bar{t}H$ ($H \to$ other)	40.9 ± 6.88	5.72 ± 1.06	0.822 ± 0.244
Total	77 100 ± 26 200	8270 ± 1940	888 ± 233
Data	84556	11561	1285

Table 6.3.: Yields before the fit in the inclusive six jet regions.

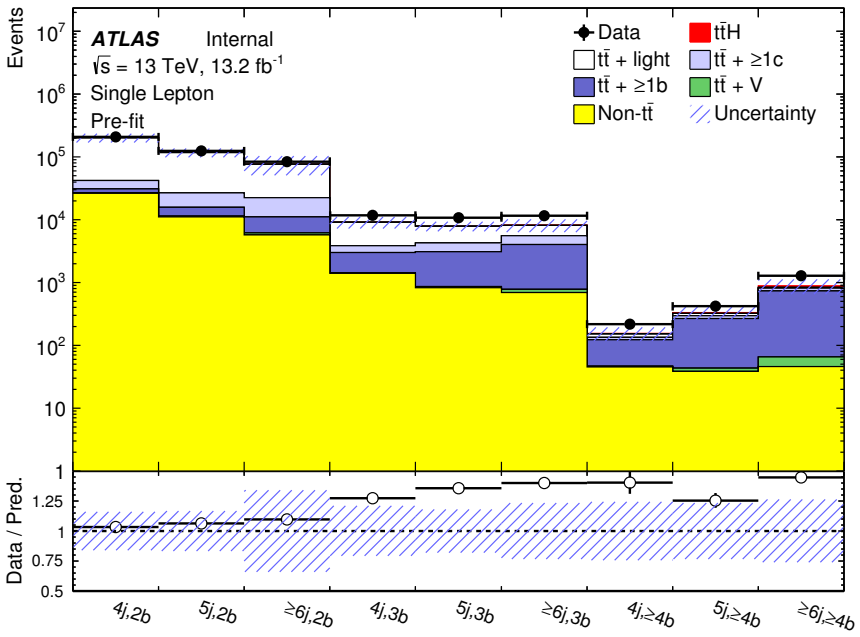


Figure 6.1.: Comparison between predicted yields and data for each region before the fit to data.

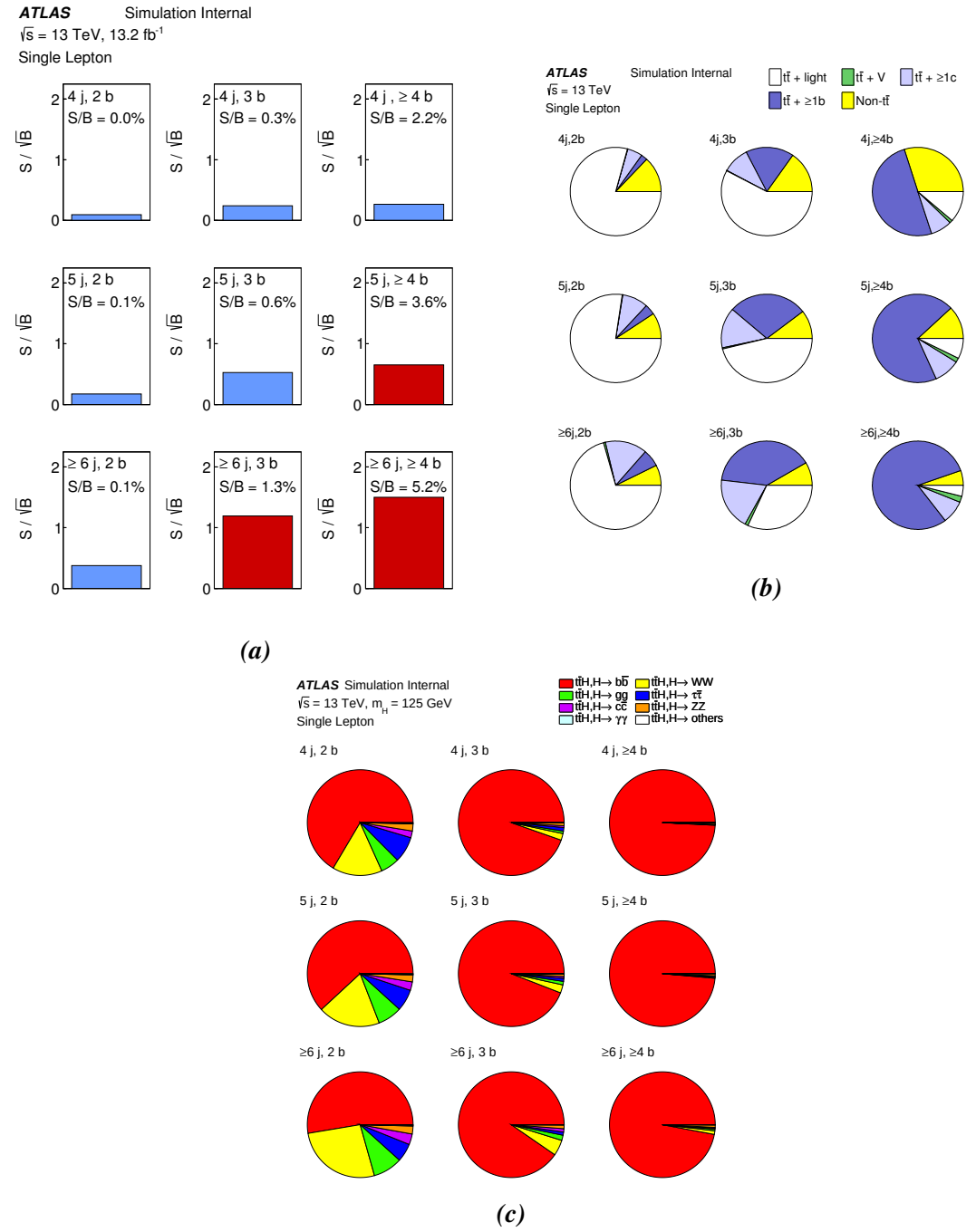


Figure 6.2.: (a) Analysis regions in the single lepton channel. Each row corresponds to a different jet multiplicity and each column corresponds to a different *b*-jet multiplicity. (b) Fractional contributions of the different backgrounds to the total background prediction in each of the regions. (c) Relative contributions of various Higgs decays in each of the regions.

6.3. Multivariate Analysis Strategy

Even with a dedicated event categorisation, it is not possible to define regions that have a high signal fraction. In the most important signal region the S/B is 5.2% and $S/\sqrt{B} \simeq 1.4$. For these reasons a *cut-and-count* analysis to extract the signal from the background is not possible. To discriminate signal from background, it is necessary to search for variables that show a good separation. These variables can be constructed from the objects in the final state. However, there is no single variable with a significant separation power. The analysis of $t\bar{t}H$ production requires sophisticated MVA techniques. Such an approach is used to obtain a good separation between $t\bar{t}H$ signal and $t\bar{t}$+jets background in the most signal enriched regions (5j,≥4b), (≥6j,3b), (≥6j,≥4b). The MVA discriminants are trained independently in the three signal regions. The shapes of the output distributions are used as templates in the final fit to data. For the control regions, the scalar sum of the transverse momenta of jets in the final state (H_T^{had}) is used instead for the templates. Plots of H_T^{had} in the considered control regions are shown in Fig. 6.3. This variable helps to constrain the different sources of systematic uncertainties. The discriminating variables used in each region are summarised in Table 6.4.

The considered MVA techniques are either an ANN, or a BDT. The latter consists of two steps, the first aimed to kinematically reconstruct the final state. This approach is discussed in Sec. 6.7. Two different likelihood fits are considered for the two cases and results are explained in Chapter 8.

region	2 b-tags	3 b-tags	≥4 b-tags
4 jets	H_T^{had}	H_T^{had}	H_T^{had}
5 jets	H_T^{had}	H_T^{had}	MVA
≥ 6 jets	H_T^{had}	MVA	MVA

Table 6.4.: Summary of the discriminants used in the analysis regions. In the signal regions, an MVA technique is employed. In all the other regions, H_T^{had} is used.

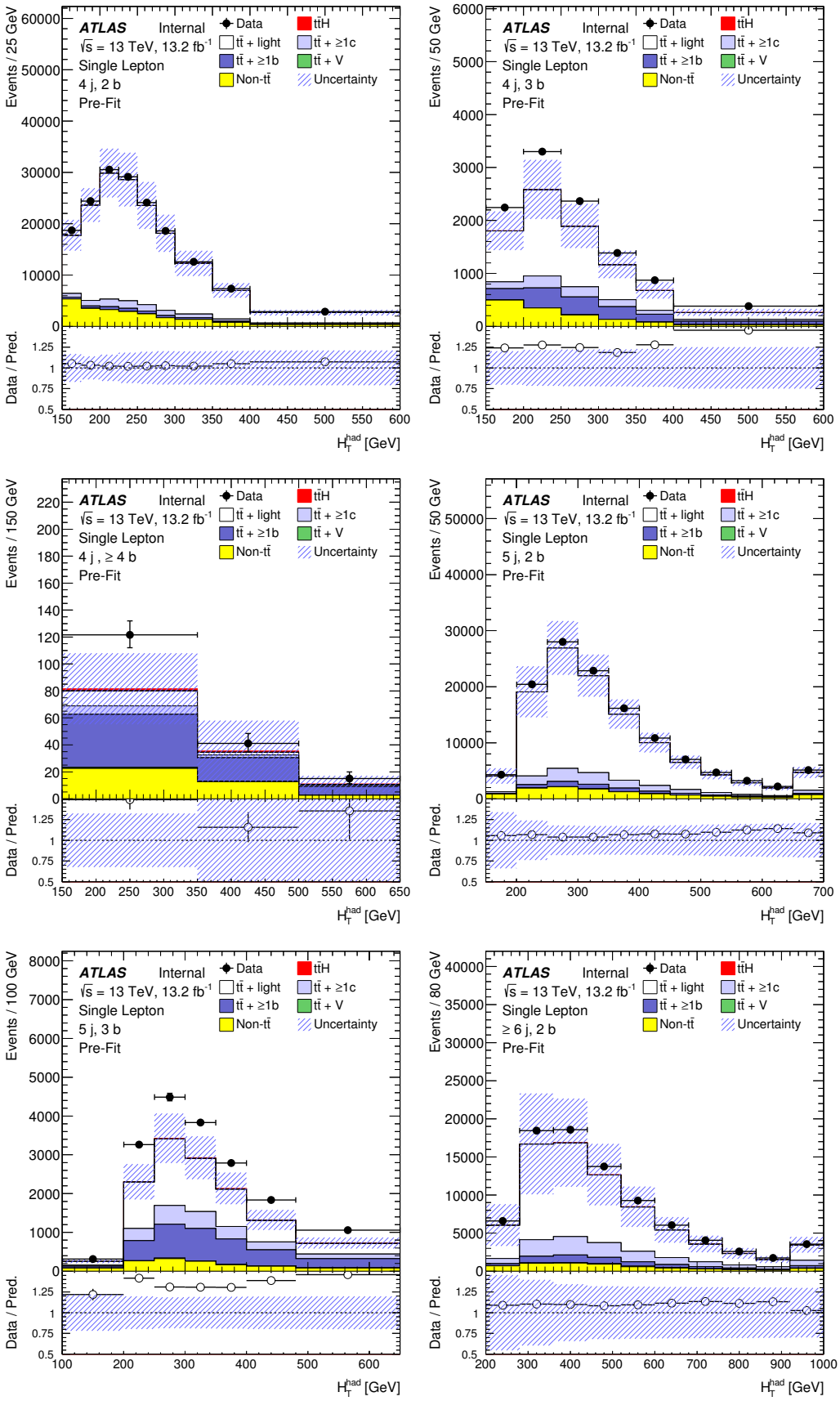


Figure 6.3.: The H_T^{had} distribution in the control regions.

6.4. Machine Learning Algorithms

New generation high energy physics experiments are growing both in terms of their dimensions and also in complexity of the structure of their basic components. In most cases, discoveries are made from a very complex analysis of a large amount of data, where the rare signal is extracted from a large amount of background processes. Thus, analysis techniques that have been developed for such purposes involve advanced *machine learning algorithms*, whose goal is to carry out accurate predictions, based on mathematical assumptions, out of a sample of recorded data or MC simulated datasets. Examples of such algorithms are already given in this thesis in Sec. 5.4. Given that the analysis presented in this thesis makes use of an ANN and of a BDT that are both machine learning algorithms, a dedicated section is devoted to the description of the basics of both methods [119], starting from a brief introduction on the human brain system.

6.4.1. Human Central Nervous System

The structure of the human central nervous system has been studied since the Middle Ages, while its actual structure was confirmed only after the observations of the Spanish neuroanatomist Santiago Ramon y Cajal in 1888 based on the revolutionary microscopy silver staining technique discovered by the Italian physician Camillo Golgi in 1873. For their discoveries, the two shared the Nobel Prize in medicine in 1906. The structure of the human central nervous system is based on elementary interconnected cellular units, called *neurons*, Fig. 6.4.

The investigation of the neuronal structure revealed that all the neurons are built of the same basic parts, independently of their shape or size. The bulbous body of the neuron is called *soma*; the root extensions departing from the soma are called *dendrites*; the single tubular fibre originating from the soma is called *axon*, which also ramifies into small branches. The joint between the end of the axon and another neuron is called *synapse*. The overall size of the neuron in the human body varies from 0.01 mm in the brain to 1 m in the limbs. The tiny gap at the synapses is just 200 nm wide.

Nervous signal impulses are transmitted both electrically and chemically. Electrical transmissions are typical of the interior of the neurons, while chemical mechanisms prevail at the synapses through the exchange of chemical substances called *neurotransmitters*. The aim of those is to either excite or inhibit the activation of the neuron, as stated by the Dale's law. The body of the neuron acts like a summing device that adds all the contributions of its various input signals. If this signal exceeds a critical threshold then the

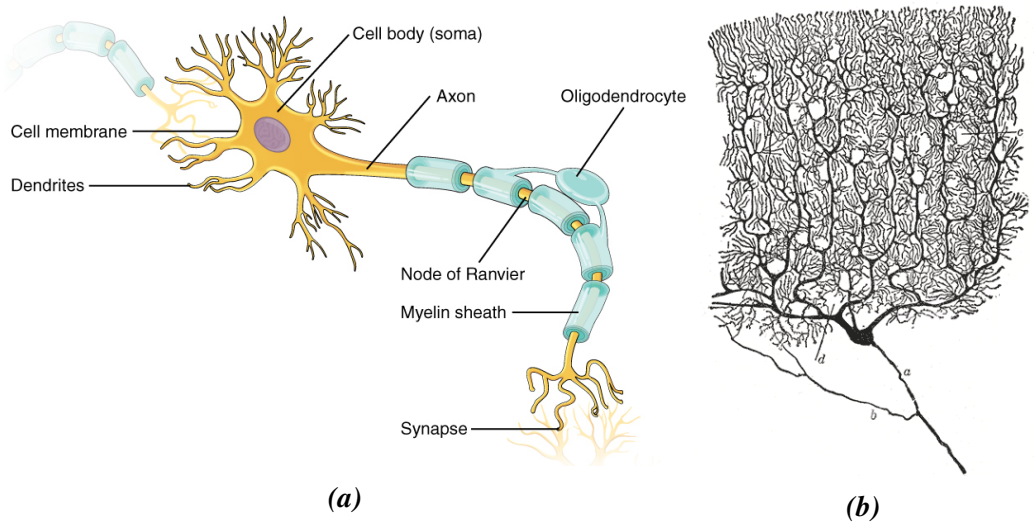


Figure 6.4.: (a) A visual representation of the human neuron cells. (b) Original drawing by Ramon y Cajal showing a complex neuron net structure.

neuron is activated. The speed of propagation of the electrical impulses through the body of the neuron is 0.5-2 m/s, which gives an unacceptably long reaction time between central neurons and neurons of the limb. To overcome this problem, the axons of the neurons are made of individual segments that are covered by am insulating substance, the *myelin*, which is interrupted in the *Ranvier nodes*. This allows the electric signal to be propagated almost instantaneously from one Ranvier node to the next. Thanks to this the overall speed of propagation is enhanced up to 100 m/s.
The human brain is one of the most complex structures in biological systems, although its complexity cannot be ascribed to the single neuron cells, but to the huge number of its constituent neurons and as a consequence the incredible number of possible mutual connections between them. An estimation of the total number of neurons in the human cortex, considering a density of 150000 neurons per mm^3, is 3×10^{10} neurons, for a total of about 10^{15} synaptic connections.

6.4.2. Generalities on ANN

ANN models are algorithms employed for cognitive tasks and are based on concepts derived from the study of the human brain. By definition, a neural network model is mathematically described by a *directed graph* that fulfils the following properties:

- To each node i an associated state variable n_i is defined.

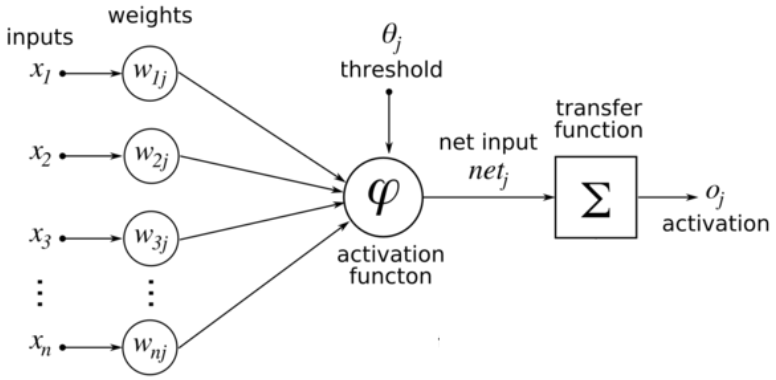


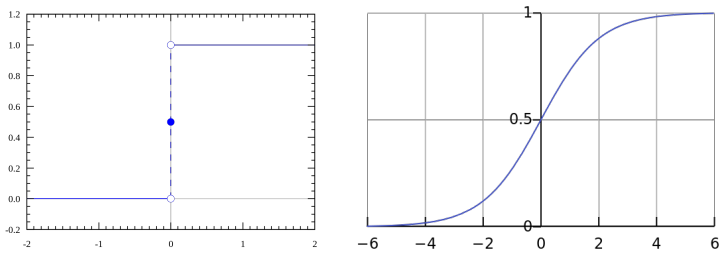
Figure 6.5.: A visual representation of an ANN. Nodes are referred to as neurons and they are connected to each other by weighted links. The activation function, φ, sets the state for the neuron j after the evaluation of a certain transfer function $\Sigma(\varphi)$.

- A real weight value, w_{ik}, is associated with each connection between two different nodes i and k.
- A real bias value, θ_i, is associated with the node i.
- A transfer function, $f_i\,(n_k, w_{ik}, \theta_i)$ for $k \neq i$, which describes the state of the node k as a function of θ_k and w_{ik} associated with its links with other states, is defined for each node i.

From this definition, the parallel to the biological systems is straightforward, see Fig. 6.5. The nodes are individual elementary constituents connected to each other by weighted connections. Thus, they are also called *neurons*, while the weighted links are called *synapses*. The electrical activation threshold of biological neurons is reproduced here by the bias, which is therefore called *activation threshold*. The state of a certain neural unit i is described as a linear combination of its input connections, so the *activation function* for the node i is defined as $\varphi = \sum_k w_{ik} n_k - \theta_i$. An important feature is the form of the transfer function, Fig. 6.6, which can be a step function in the most elementary example, like in the first theory of ANN by McCulloch and Pitts [120], but it has the form of a sigmoid function in real cases. In the case of a sigmoid shape, the activation threshold is defined as:

$$f(\varphi) = \frac{2}{1 + e^{-\varphi}} - 1 \tag{6.1}$$

Nodes with no connection into them are called *input nodes*, while the nodes with no exiting connections are called *output nodes*. The ANN is called *feed-forward* if its link



(a) *(b)*

Figure 6.6.: A visual representation of (a) a step transfer function and (b) a sigmoid transfer function.

structure does not admit closed paths. Feed forward ANN are often referred as *perceptrons*. In the simplest case [121], the perceptron consists of two separate layers of neurons, representing layers of input and output neurons. The neurons of the output layer receive signals only from the input layer, but not vice versa, and the neurons in the same layer are not connected to each other. A special implementation of perceptrons, the *multilayer* perceptron, is at present widely used. These perceptrons are not made up of only two layers, but have one or more intermediate layers of neurons known as *hidden layers*, Fig. 6.7. The design of an ANN is devoted to the choice of w_{ij} and θ_j such that a precise cognitive

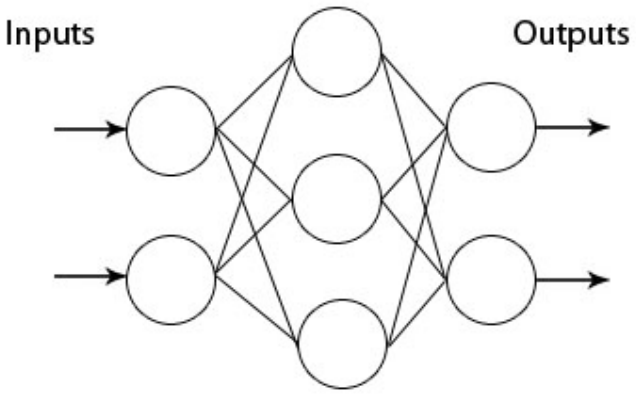


Figure 6.7.: A visual representation of a multilayer perceptron with three layers made of: two neurons in the input layers, three in the hidden layer and two in the output layer.

task can be performed by a machine. The process of finding the best weights is referred to as *learning* or also as *training*. Multilayer perceptrons have a structure which gives the possibility of employing powerful learning methods, one of which was introduced in

1974 by Werbos [122] and is known as *error back-propagation*. This method uses the assumption that the desired values for the output neurons are known, so the weights are corrected iteratively until the difference between the actual value and the desired value is as small as possible. The error back-propagation, makes use of another more basic algorithm known as *gradient descent* method, which implies the evaluation of the output after changing the weights by a small amount δw_{ij} and $\delta\theta_j$. Both of these methods are briefly discussed in the next sections. Error back-propagation, is part of a wider range of learning algorithms known as *supervised learning algorithms*, since at each step the ANN is adjusted after the comparison of its actual output and the desired output. These methods, although designed to mimic the human brain structure, are implausibly implemented in biological systems. Biological systems are also not fully feed-forward, but most likely *recurrent*, since an output layer of neurons can be connected with its own inputs. Even with those limitations, multilayer perceptrons are simplified models that offer a very powerful tool in signal-background discrimination and thus are employed in high energy physics analyses.

The Gradient Descent Method

In the case of a perceptron made of only two layers, considering a transfer function $f(x)$ and denoting the states of the input layer neurons as σ_k for $k = 1, 2, ..., N_i$ and the output states as S_i for $i = 1, 2, ..., N_o$, the activation of each output neuron is determined by:

$$S_i = f\left(\sum_k w_{ik}\sigma_k\right) \tag{6.2}$$

where, for simplicity, no biases, θ_k, are considered. Learning implies finding of the best values for the synaptic connections, w_{ik}, such that a certain input σ_k leads to desired output, defined as the correct states for the neurons in the output layer, ζ_i. Ideally, one should obtain $S_i = \zeta_i$. Considering all the possible input cases, indicated with the superscripts μ, the desired relation is found to be: $S_i^\mu = \zeta_i^\mu$. At present, an explicit function that achieves this goal is not known, but it is possible to get it through the usage of iterative procedures. The idea is to increment w_{ik} by a small quantity, δw_{ik}, until w_{ik} converges to the desired values for the synaptic connections. To define how to choose δw_{ik}, the error function (or loss function) is introduced:

$$D = \frac{1}{2}\sum_\mu \sum_i \left(S_i^\mu - \zeta_i^\mu\right)^2 = \sum_{\mu,i}\left(1 - S_i^\mu \zeta_i^\mu\right), \tag{6.3}$$

which represents the deviation, or error, between the correct output of each node and its actual value. Using equation 6.2, it is possible to rewrite the definition of the deviation as follows:

$$D\left(w_{ik}\right) = \frac{1}{2}\sum_{\mu}\sum_{i}\left[\zeta_i^{\mu} - f\left(\sum_k w_{ik}\sigma_k^{\mu}\right)\right]^2 \equiv \frac{1}{2}\sum_{\mu}\sum_{i}\left[\zeta_i^{\mu} - f\left(h_i^{\mu}\right)\right]^2 . \tag{6.4}$$

Ideally, such a deviation should be zero, while what is instead achievable is to reach a minimum for it. For this, the gradient of D with respect to the synaptic weights is defined:

$$\frac{\partial D}{\partial w_{ik}} = -\sum_{\mu}\left[\zeta_i^{\mu} - f\left(h_i^{\mu}\right)\right] f'\left(h_i^{\mu}\right)\frac{\partial h_i^{\mu}}{\partial w_{ik}} = -\sum_{\mu}\Delta_i^{\mu}\sigma_k^{\mu} . \tag{6.5}$$

Here the abbreviation

$$\Delta_i^{\mu} = \left[\zeta_i^{\mu} - f\left(h_i^{\mu}\right)\right] f'\left(h_i^{\mu}\right) \tag{6.6}$$

is used. By definition of a function gradient, the direction in which the gradient increases follows the direction of the primary function increase. Hence, the direction of the negative gradient points towards the steepest descent of the primary function. Using this property δw_{ik} is defined as:

$$\delta w_{ik} = -\eta\frac{\partial D}{\partial w_{ik}} = \eta\sum_{\mu}\Delta_i^{\mu}\sigma_k^{\mu} . \tag{6.7}$$

The parameter η, known as *learning rate*, is a positive real value which is usually taken as $\eta << 1$ to obtain an optimal convergence to the minimum at each iteration. The number of iterations may be adjusted as a function of the chosen η, Fig. 6.8. This leads to the limitations of the gradient descent method, since a wrong combination of the number of iterations and the learning rate causes wrong convergence to a local minimum or a very slow convergence to the global minimum.

The Error Back-Propagation Method

The error back-propagation method is a generalisation of the gradient descent method for the multilayer perceptrons. A three layer perceptron is considered. The definitions of the labels used in the previous section are completed by introducing the missing correspondent quantities for the intermediate, hidden, layer of neurons. The synaptic connections between output and hidden layer are denoted by w_{ij} and the states are called s_j, where the

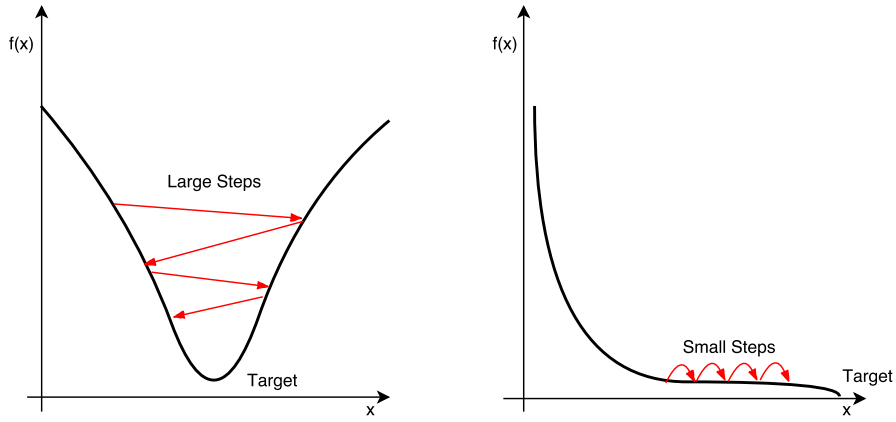


Figure 6.8.: Different choices for the learning rate parameter. In the left case the learning rate is too large to permit the method to converge to the minimum of the error function. In the right case the learning rate is small enough but the number of iterations is not sufficient to reach the minimum.

index j labels all the nodes in the input layer, $j = 1, 2, ..., N_h$. The threshold potentials are also denoted as θ_{ij}. Similar definitions are applied for the connections between hidden and input layer, whose synaptic connections are indicated by $\bar{w}_{jk}$ and the corresponding biases as $\bar{\theta}_k$. Using these definitions, it is possible to introduce the equations that govern the state of the ANN:

$$\begin{aligned} S_i &= f\left(\sum_j w_{ij} s_j - \theta_i\right) \equiv f\left(h_i\right) , \\ s_j &= f\left(\sum_k \bar{w}_{jk} \sigma_j - \bar{\theta}_j\right) \equiv f\left(\bar{h}_j\right) . \end{aligned} \tag{6.8}$$

Using the same ideas as in the previous section, the variations for the synaptic connections and for the activation thresholds are defined. The connections between the hidden layer and the output layer are:

$$\begin{aligned} \delta w_{ij} &= -\eta \frac{\partial D}{\partial w_{ij}} = \eta \sum_\mu \left[\zeta_i^\mu - f\left(h_i^\mu\right)\right] f'\left(h_i^\mu\right) \frac{\partial h_i^\mu}{\partial w_{ij}} = \eta \sum_\mu \Delta_i^\mu s_j^\mu , \\ \delta \theta_i &= -\eta \frac{\partial D}{\partial \theta_i} = \eta \sum_\mu \left[\zeta_i^\mu - f\left(h_i^\mu\right)\right] f'\left(h_i^\mu\right) \frac{\partial h_i^\mu}{\partial \theta_i} = -\eta \sum_\mu \Delta_i^\mu . \end{aligned} \tag{6.9}$$

The connections between the input and the hidden layer are similarly achieved:

$$
\begin{aligned}
\delta\bar{w}_{jk} &= -\eta\frac{\partial D}{\partial\bar{w}_{jk}} = \eta\sum_{\mu,i}\left[\zeta_i^\mu - f\left(h_i^\mu\right)\right]f'\left(h_i^\mu\right)\frac{\partial h_i^\mu}{\partial s_j}\frac{\partial s_j}{\partial\bar{w}_{jk}} = \\
&= \eta\sum_{\mu,i}\Delta_i^\mu w_{ij}f'\left(\bar{h}_j^\mu\right)\frac{\partial\bar{h}_j}{\partial\bar{w}_{jk}} = \eta\sum_{\mu}\bar{\Delta}_j^\mu\sigma_k^\mu\,, \\
\delta\bar{\theta}_j &= -\eta\frac{\partial D}{\partial\bar{\theta}_j} = \eta\sum_{\mu,i}\left[\zeta_i^\mu - f\left(h_i^\mu\right)\right]f'\left(h_i^\mu\right)\frac{\partial h_i^\mu}{\partial s_j}\frac{\partial s_j}{\partial\bar{\theta}_j} = \\
&= \eta\sum_{\mu,i}\Delta_i^\mu w_{ij}f'\left(\bar{h}_j^\mu\right)\frac{\partial\bar{h}_j}{\partial\bar{\theta}_j} = -\eta\sum_{\mu}\bar{\Delta}_i^\mu\,.
\end{aligned}
\tag{6.10}
$$

Here a new quantity is introduced:

$$
\bar{\Delta}_j^\mu = \left(\sum_i \Delta_i^\mu w_{ij}\right)f'\left(\bar{h}_j^\mu\right)\,. \tag{6.11}
$$

These equations obtained for the adjustments of the synaptic connections are very similar to the case discussed in the previous section, with the only difference that the quantity $\bar{\Delta}_j^\mu$ is obtained recursively from Δ_i^μ. This recursion relation gives the name of the method, since the corrections propagate from the output layer backwards to the input layer. It is possible to generalise the method to more hidden layers, by defining other Δ parameters always obtained for a certain layer as a function of the same Δ obtained in the previous (in backward propagation direction) layer.

Alternative Approaches For the Loss Function Minimisation

As already discussed, back propagation is based on gradient descent method, which is an effective method and offers sufficient solutions for the implementation of an ANN, but suffers of some problems related to the convergence to the global minimum of the error function. Some variations of this method use alternative minimisation algorithms. An important example is based on the so called classical *method of Newton* [123]. Given a non-linear function $D(x)$, it is possible to find a set of iteratively defined positions x_i, which converges to the minimum position $x_1, x_2, ... \rightarrow x_{min}$. Each step is iteratively defined as:

$$
x_{n+1} = x_n - \mathscr{H}^{-1}\nabla D(x_n)\,, \tag{6.12}
$$

where $\mathscr{H}^{-1}$ is the inverse of the Hessian matrix $\mathscr{H}$. Since the Hessian matrix contains the

second derivatives of $D(x)$, it is computationally expensive to obtain. The *quasi-Newton methods* make use of the same idea, but use approximated versions of the Hessian matrix (or its inverse), B, which are built-up iteratively after evaluations of the gradient of the function. Since B is built in steps starting from a randomly defined B_0 initial state, many numerical algorithms try to optimise the updating step $\Delta B = B_{k+1} - B_k$, the most popular being the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [123]. Such approach is also considered in the analysis presented in this thesis.

Weight Decay Regularisation

A common problem of ANN is given by the loss of generalisation power, which occurs when the trained ANN loses performance (i.e. separation power) when introducing a different input data set of the same nature as the training set. This is given mainly by two factors, the first is the amount of information in the input dataset, and second the complexity of the network. If the complexity does not match the quantity of information, then the ANN loses generalisation. If the network is very complex and there is too little information, an *overtraining* effect occurs, while the opposite situation is known as *undertraining*. Undertraining can be recovered by adding information to the input dataset, i.e. defining more input training variables, while overtraining is more difficult to avoid. One way to constrain the network and reduce its complexity is to introduce a mechanism which limits the growth of the weights [124]. One way to do that is to add a decay term to the error function:

$$D = \sum_{\mu,i} \left(1 - S_i^\mu \zeta_i^\mu\right) + \frac{1}{2}\alpha \sum_j w_j^2 \,. \tag{6.13}$$

This new term is governed by the *regularisation parameter* α which determines how strongly the weights with a large value are penalised. The choice of this term is arbitrary, but there are several methods to find the best values for α.

6.4.3. The NeuroBayes ANN Implementation

For this analysis an ANN was implemented through the usage of the NeuroBayes commercial package [125, 126]. This implementation considers a three layer feed-forward ANN and provides three distinctive advantages: first, there is an important input variable pre-processing step; second, there is an internal method to provide a ranking of the variables used in the training phase according to their separation power; third, a Bayesian regularisation procedure is used to optimise the training. A very useful feature of Neu-

roBayes is the possibility to use weights associated to each input event during the training phase. This is particularly useful in physics analyses, since the generated MC events are often associated with MC weights. The weights are used to adjust the minimisation of the loss function to give an optimised result. In the definition of the ANN architecture, NeuroBayes gives the possibility to define *input bias nodes*. These nodes have a constant value, and help to avoid saturation of the nodes and improve the achieved training.
The input variable pre-processing plays a very important role in the NeuroBayes framework, since it has been shown [126] that such a procedure greatly helps in providing an optimisation in speed and robustness of the training procedure and also in avoiding neuron saturation. The pre-processing is done in different steps. In the first step, the input variables are transformed into flat distribution. This is achieved through an integral transformation of the initial variable $y = f(x)$:

$$g(x) = \int_{-\infty}^{x} f(x')dx' \,, \tag{6.14}$$

where the obtained $g(x)$ is the cumulative distribution of $f(x)$. Thus, the variable defined as $Y = g(y)$ has a uniform distribution. The variable is then rescaled in order to be in the [-1,1] interval. This first step is important to prevent extreme values of an input variable that saturate the neurons and thus bias the net output. A second pre-processing step is transforming the obtained flat distribution into a distribution with mean value centred at zero and unitary standard deviation. This is achieved with the transformation defined as:

$$\tilde{Y} = \frac{Y - \bar{Y}}{\sigma_Y} \,, \tag{6.15}$$

where $\bar{Y}$ and σ_Y are the mean and standard deviation of the distribution of Y. As a last step the input variables are decorrelated. This also helps to achieve an optimal training. The decorrelation is obtained by calculating the covariance matrix for the input variables and then diagonalising it. The diagonalisation method employed in NeuroBayes makes use of the iterative method of the Jacobian rotations [127]. The idea of this method is to perform several two-dimensional rotations of the matrix until the obtained matrix is as close to diagonal as possible.
One of the most useful features of NeuroBayes is the ranking of the input variables based on their significance in the training. After the variables have been reprocessed their correlation with the output of the ANN is calculated. After the correlation matrix for all the variables is computed, the correlation coefficients are recomputed after removing one

variable at a time from the input set. At the end the variables are sorted according to the loss of correlation to the final discriminant caused by their exclusion from the calculation. The variable which gives the least loss of correlation is removed, and the procedure is repeated for the N-1 remaining variables, until only a single variable remains. This variable is the most significant. A quantity given by the loss of correlation caused by its removal multiplied by $\sqrt{n}$, with n the number of events in the training dataset, is associated to each variable. The ranking of each variable is defined by this quantity. Since the ranking depends on the statistics and on the other variables in the input dataset, it is quite difficult to define which are the most important variables in different training processes. In any case, this functionality is particularly useful when the user has to deal with many possible input variables, since it gives a clear method to choose the best variables for each training procedure. This feature of NeuroBayes allows to consider only the highest ranked variables for the actual training, rejecting the others.

NeuroBayes offers an important regularisation method for improving the generalisation of the ANN and thus avoiding overtraining. It is based on a Bayesian approach, [128]. A general idea of regularisation is explained in the Sec. 6.4.2. In the NeuroBayes package, it is achieved using a particular form of loss function defined as:

$$D = \sum_j w_j \sum_i \log\left[\frac{1}{2}(1 + S_i\zeta_i + \epsilon)\right] + \frac{1}{2}\sum_c \alpha_c \sum_j w_j^2 \tag{6.16}$$

The first term is a modified version of the error function (*entropy error function*) and the second term is the weight decay term. The entropy error function has the advantage of assigning infinite values to a wrong classification (in which $\zeta = 1$ and $S_i = -1$ or vice versa). The ϵ parameter is called *regularisation constant*. It is a number which is introduced at the beginning of the training and it is zero after few iterations. This form of the loss function is also important since it gives the possibility to implement the Bayesian regularisation method [128], devoted to find the best value for α_c. In this case α_c is not just a single parameter, but it is split in three components (c = 1, 2, 3) [128], each governing the decay of the weights between input and hidden layer, the bias node and the hidden layer, and the hidden layer and the output layer. With this construction, a NeuroBayes based ANN is very robust against overtraining.

A detailed list of all the options used in the NeuroBayes configuration is listed in Table 6.5.

# of variables	N_I
# nodes in hidden layer	$N_I + 2$
Update weights interval	50 events
Learning speed factor	1
Maximum learning speed	0.01
# of iterations	100
Minimisation approach	BFGS

Table 6.5.: Full detail of the NeuroBayes settings.

6.4.4. Decision Trees

A popular and commonly used MVA method is based on the concept of Decision Tree (DT). First introduced by Breiman [129] in 1984, DTs are employed to define sequential cuts in order to perform a classification task. Iteratively, DT searches the best possible set of cuts to split the input dataset. The process is continued recursively in the obtained partition of the original input dataset until a user defined criterion is satisfied. At the beginning of the training, the DT considers the whole set of events for both signal and background samples, this step is depicted as a *root node*, Fig. 6.9. At each iteration, the algorithm splits the training dataset according to the best cut on each of the training variables, forming two different branches at each *decision node*. The best cut is decided calculating the associated impurity, according to the *Gini* index [130]:

$$Gini = (s + b)P(1 - P) = \frac{sb}{s + b}\,, \tag{6.17}$$

where $P = s/(s + b)$ defines the purity of signal events and s and b are the number of signal and background events at each selection step. The cut which minimises the *Gini* index is chosen as the best cut for each decision node. The procedure is repeated until the impurity for the considered split does no longer reduce the impurity any more (or when the achieved reduction is considered too small). In this case the node becomes a terminal node, or a *leaf*.

DT algorithms have some advantages [130]: they are more transparent than other methods in terms of the procedure and the interpretation of the achieved results; they are strong against missing information in the training and testing samples; they are insensitive to input training variables with low separation power. DTs have also some limitations. First of all, the algorithms are usually not generalised, and they greatly suffer of overtraining since small changes in the input dataset translate into drastically different trees. Another

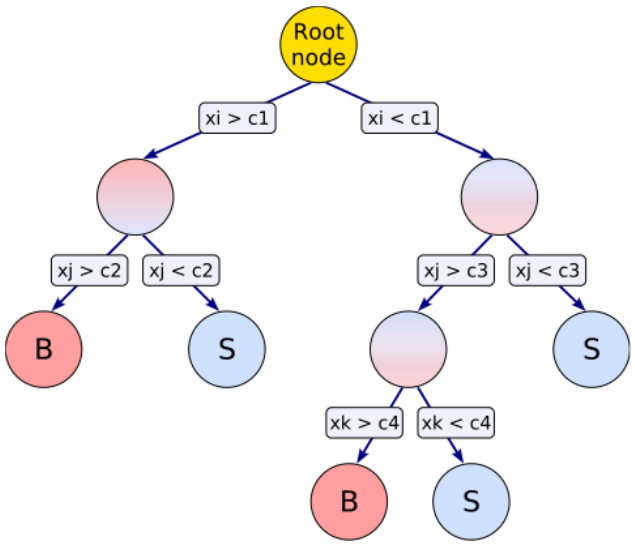


Figure 6.9.: A visual representation of a DT

drawback is that for each leaf the prediction is constant, so the prediction is binned and discontinuous at the edges of each bin. Most of the limitations are overcome by introducing *ensemble learning* techniques, such as *boosting*, *bagging*, *random forests*. The idea behind these techniques is to define a set of different discriminants and to make an average over their different predictions. This gives a generalisation error that is usually smaller compared to the error of each single discriminant, thus providing good discriminants starting from a large set of modest ones.

Ensemble Learning Techniques

The idea behind *boosting* techniques is to create an ensemble of discriminants that all together sum up into a boosted high-performance, even if the performance of each of the single discriminant is weak. For a set of N discriminants, its combined performance is defined as:

$$\tilde{y} = \sum_{n=1}^{N} \alpha_n y_n \,, \tag{6.18}$$

where y_n is the output of the n-discriminant and α_n is a set of N weights associated to each discriminant, which are defined differently according to the considered algorithm. In the case of the AdaBoost algorithm [131], the underlying discriminants, y_n, are DT and the

obtained $\tilde{y}$ is called BDT. In this algorithm, the α_n are defined as:

$$\alpha_n = \log\left(\frac{1-\epsilon_n}{\epsilon_n}\right), \tag{6.19}$$

where ϵ_n is defined as the misclassification error for the n-th DT.
Bagging is a technique in which the available N discriminants are simply defined over a different randomly selected dataset and then averaged. In this case, looking at equation 6.18, $\alpha_n = 1/N$.
Finally, in the *random forests*, many different DT are trained, but for each of them only a random set of input training variables is considered for defining the best cut in each decision node. At the end, all the obtained outputs are averaged to get the final classification decision.

6.4.5. The TMVA Package

In the ROOT environment [132], the integrated Toolkit for Multivariate Analysis (TMVA) [133] package, provides a processing, parallel evaluation and application of MVA techniques. All the MVA techniques provided by TMVA are *supervised learning* algorithms. The training samples in TMVA are defined both for signal and background regions and are employed in order to define a precise mapping function that describes a decision boundary. A typical TMVA analysis consists of two independent phases: the training phase, where the multivariate discriminants are trained, and an application phase, where the discriminants are applied to the concrete classification problem. For each TMVA method, the configuration and training results are written in a result (weight) file, that is used for the application of the training results to the analysis. The software developed in the TMVA, consists of abstract, object-oriented implementations in C++/ROOT of several MVA techniques. TMVA provides also training, testing and performance evaluation algorithms through visualisation scripts.
The package is used for training and testing of a BDT configured with the AdaBoost algorithm. Training and testing is performed through user-supplied datasets, that contain an individual weight for each of the events. The true sample composition and the desired target value needs to be supplied by the user for each event. Once the discriminant is trained, its signal efficiency and the background rejection performance are stored in a ROOT output file. These results can be displayed using built-in macros, that are executed through dedicated graphical user interfaces.

6.5. Variables Used in the ANN Analysis

The large variety of physics objects present in the final state of the $t\bar{t}H$ production, gives the possibility to define a large number of variables with the goal to find the best possible discrimination of signal and background. These variables can be grouped in different categories:

- Object kinematics: η and p_T of the jets, b-jets and leptons in the final state;
- Global event variables: Scalar sum of the p_T of the jets in the final state (H_T^{had}), the mass of the 4-vector resulting from the combinations of the 4-vectors of the objects in the final state, number of jets beyond a fixed p_T threshold (40, 60, 80 GeV).
- Event shape variables: Include several combinations of the eigenvalues of the linear momentum tensor (such as Aplanarity), Centrality and Fox-Wolfram moments (which describe any geometrical correlation among the objects in the final state in terms of spherical harmonics [134]). The momentum tensor is defined as:

$$M_{ij} = \frac{\sum_k p_k^i p_k^j}{\sum_k (p_k)^2} \,. \tag{6.20}$$

- Object pair properties: Invariant mass, p_T, and ΔR of the jet pair that has the largest p_T sum ($\max p_T$), smallest ΔR, the combinations of pairs of jets differentiated according to the number of b-jets considered; if no b-tagged jets are considered, the corresponding pair of jets with the minimum angular distance is defined to identify the W boson candidate decaying in a pair of light jets.

All variables are defined considering at most seven jets in the final state. In case more than seven jets are found, the b-tagged jets are considered first and then the missing jets are chosen among the non b-tagged jets with highest p_T. This allows avoiding using soft jets, which are not well modelled. For each of the variables, a separation is defined as follows:

$$S = \frac{1}{2} \cdot \sum_i^{\text{bin}} \frac{(N_i^{\text{signal}} - N_i^{\text{background}})^2}{(N_i^{\text{signal}} + N_i^{\text{background}})} \tag{6.21}$$

where N_i^{signal} and $N_i^{\text{background}}$ are the entries in each bin after histograms have been normalised to the same unit area.
The final obtained separation between the signal and the background is originated from the nature of the b-quarks produced in the events as well as by the different mechanisms

involved in the production of the $t\bar{t}$ pair, that are reflected in the kinematic properties of the final state objects. Since the energy required to produce the $t\bar{t}H$ signal is higher than the one needed to produce the $t\bar{t}$ background, signal events are expected to be on average more energetic and more central in the detector than background ones.
The ranking procedure of NeuroBayes is used to rank the variables according to their separation power. In general, the best 12 variables are found to be enough in order to achieve a signal-to-background separation close to the maximal one. The choice of the small number of input variables for each region is the result of a compromise between the complexity of the analysis and the best achievable performance. In order to enhance the robustness against overtraining, TRF (defined in Sec. 5.4.1) is used in the training phase for increasing the available MC statistics. A complete list of the used input variables is available in Table 6.6. The NeuroBayes variables ranking for each analysis region is given in Table 6.7. Plots of the input variables are available in Appendix B.

Variable	Definition
$\text{Centrality}_{\text{all}}$	Sum of the p_T divided by sum of the E for all jets and the lepton
$\text{M}_{bb}^{\text{min}\Delta R}$	Mass of the combination of two b-tagged jets with the smallest ΔR
$\Delta R_{bb}^{\text{avg}}$	Average ΔR for all b-tagged jet pairs
$H1_{\text{all}}$	Second Fox-Wolfram moment computed using all jets and the lepton
$\text{Aplan}_{\text{jets}}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with all the jets
$\text{Aplan}_{b\text{-jets}}$	Same as $\text{Aplan}_{\text{jets}}$ but only for b-jets
H_T^{Had}	Scalar sum of the p_T of the jets in the final state
H_T	Scalar sum of final state objects p_T
$\text{M}_{jj}^{\text{minM}}$	Mass of the combination of any two jets with the smallest invariant mass
$\Delta R_{l,bb}^{\text{min}\Delta R}$	ΔR between the lepton and the combination of two b-tagged jets with the smallest ΔR
$\Delta\eta_{jj}^{\text{max}\Delta\eta}$	the maximum $\Delta\eta$ between any pair of jets
$\text{M}_{bj}^{\text{max}p_T}$	Mass of the combination of a b-tagged jet and any jet with the largest vector sum p_T
$p_T^{\text{jet}_5}$	Fifth leading jet p_T
$\Delta R_{bb}^{\text{max}p_T}$	ΔR between two b-tagged jets with the largest vector sum p_T
$\Delta R_{uu}^{\text{min}\Delta R}$	Minimum ΔR between two untagged jets
$\text{M}_{uu}^{\text{min}\Delta R}$	Invariant mass of the pair of untagged jets with the smallest ΔR
N_{30}^{Higgs}	Number of b-jets pairs with a mass of within 30 GeV of the defined Higgs mass (125 GeV)
$\text{M}_{jj}^{\text{max}p_T}$	Mass of the combination of two b-tagged jets with the largest vector sum p_T

Table 6.6.: List of variables uses in the ANN analysis in the single lepton channel.

Variable	≥6j,≥4b	≥6j,3b	5j,≥ 4b
$\Delta R_{bb}^{\text{avg}}$	1	6	1
$\text{M}_{bb}^{\text{min}\Delta R}$	2	10	2
$\text{Centrality}_{\text{all}}$	3	2	3
$H1_{\text{all}}$	4	3	4
$p_T^{\text{jet}_5}$	5	1	5
N_{30}^{Higgs}	6	7	6
H_T	7	-	7
$\Delta\eta_{jj}^{\text{max}\Delta\eta}$	8	-	8
$\text{M}_{jj}^{\text{minM}}$	9	-	9
$\Delta R_{l,bb}^{\text{min}\Delta R}$	10	-	10
$\text{Aplan}_{b\text{-jets}}$	11	-	11
$\Delta R_{bb}^{\text{max}p_T}$	12	-	12
H_T^{Had}	-	4	-
$\text{Aplan}_{\text{jets}}$	-	5	.
$\text{M}_{bj}^{\text{max}p_T}$	-	8	-
$\Delta R_{uu}^{\text{min}\Delta R}$	-	9	-
$\text{M}_{uu}^{\text{min}\Delta R}$	-	11	-
$\text{M}_{jj}^{\text{max}p_T}$	-	12	-

Table 6.7.: The lists and rankings of the variables in each of the regions in the single lepton channel.

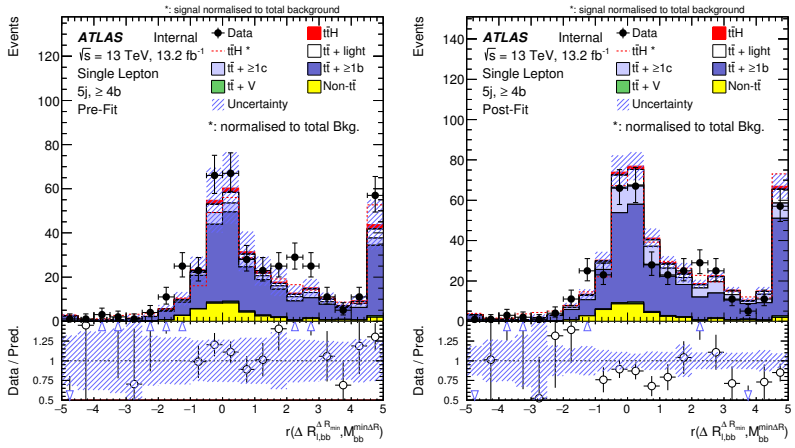
6.5.1. Validation of Input Variables

Input variables used in the training of an MVA are required to be well modelled and the correlations among them have to be studied. Since for the regions with a large fraction of $t\bar{t}$ +HF events a discrepancy between data and prediction is observed, the validation implies also a check of the post-fit distributions (the discussion of the fit technique is a part of the next chapter). The validation is thus performed in different steps:

1. Check the agreement between data and prediction for the considered variables before and after the fit.

2. Compare the linear correlation coefficients, ρ, among input variables, between data and prediction.

3. Check the agreement between data and prediction for the 1D correlation of the pairs of variables which present a correlation discrepancy $|\Delta\rho| > 0.1$ in the previous step and also for all the pairs of variables with a significant correlation.



(a) ***(b)***

Figure 6.10.: 1D correlation for the two variables $\Delta R_{l,bb}^{\min\Delta R}$ and $\mathrm{M}_{bb}^{\min\Delta R}$ in the (5j,≥4b) region. (a) Shows pre-fit distribution and (b) shows post-fit distribution.

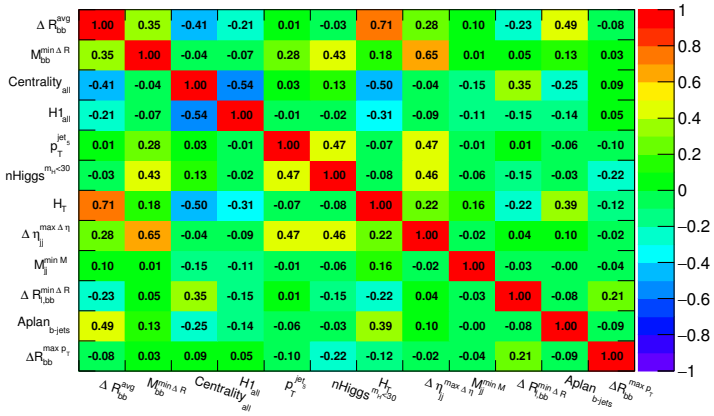
The 1D correlation between two variables x and y is calculated event by event as:

$$r(x,y) = \frac{(x-\bar{x})(y-\bar{y})}{\sqrt{\sum_i (x_i-\bar{x})(y_i-\bar{y})}}\,, \tag{6.22}$$

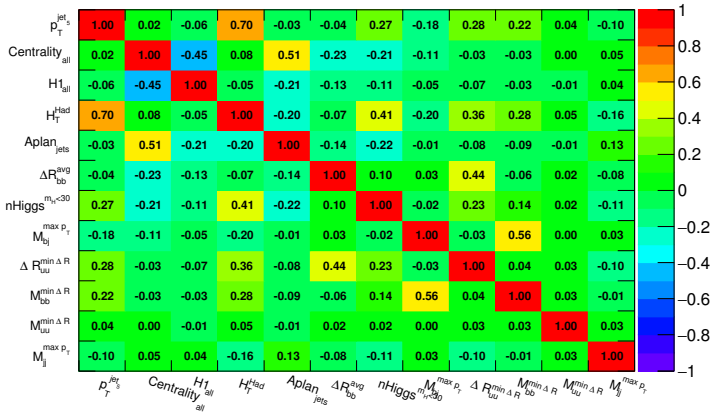
where $\bar{x}(\bar{y})$ is the mean value of the $x(y)$ variable. The plots for the step 1 are shown in Appendix B. No large disagreement between data and prediction is found in the pre-fit distributions apart from the normalisation of the $t\bar{t}$ +HF component of the background. Post-fit distributions show a good agreement between data and prediction, apart from some statistical fluctuations in the regions with the lowest statistics.

Plots for the step 3 are shown in Appendix C for variables with $|\rho|> 0.4$. The conclusions are similar as the previous case. The plots for the 1D correlation of the two variables with $|\Delta\rho|>0.1$ are shown in Fig. 6.10. The comparison between data and prediction shows some statistical fluctuations that are within uncertainty bands.

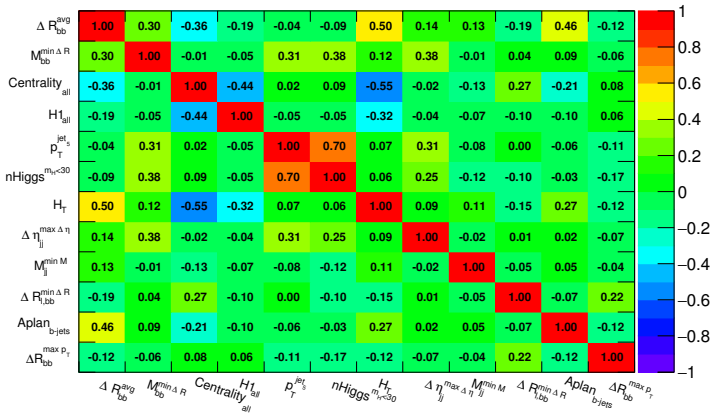
The correlation coefficients for each pair of variables in the data and in the prediction are shown in Figs. 6.11 and 6.12. No significant discrepancy is found, Fig. 6.13, apart from the two variables $\Delta R_{l,bb}^{\min\Delta R}$ and $\mathrm{M}_{bb}^{\min\Delta R}$ which show $|\Delta\rho|$=0.17 in the (5j,≥4b) region. This difference is due to statistical fluctuations that are significant in this region because of limited statistics.



(a)

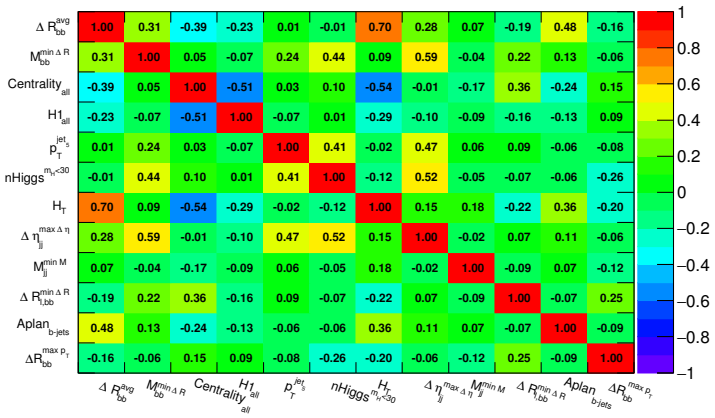


(b)

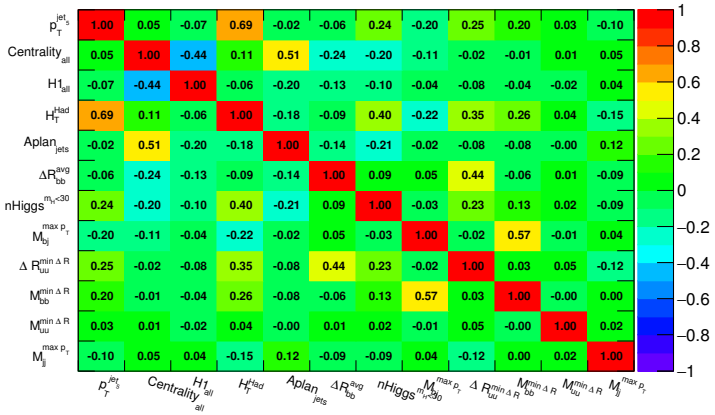


(c)

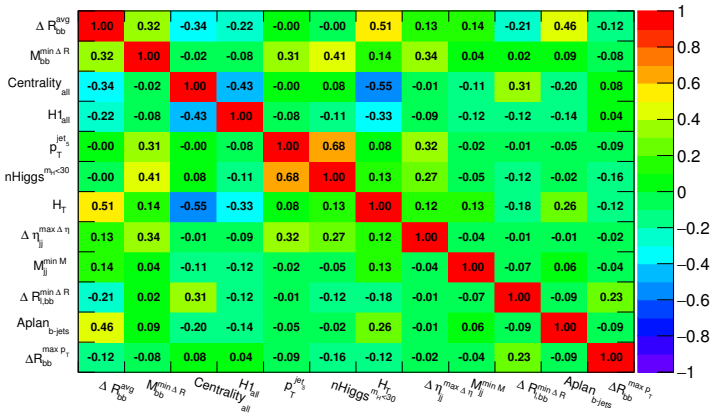
Figure 6.11.: Validation of ANN input variables. Matries of predicted correlation coefficients. (a) (5j,≥4b). (b) (≥6j,3b). (c) (≥6j,≥4b).



(a)

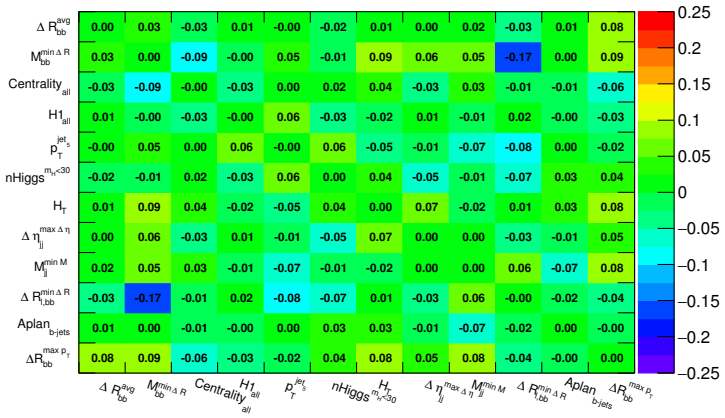


(b)

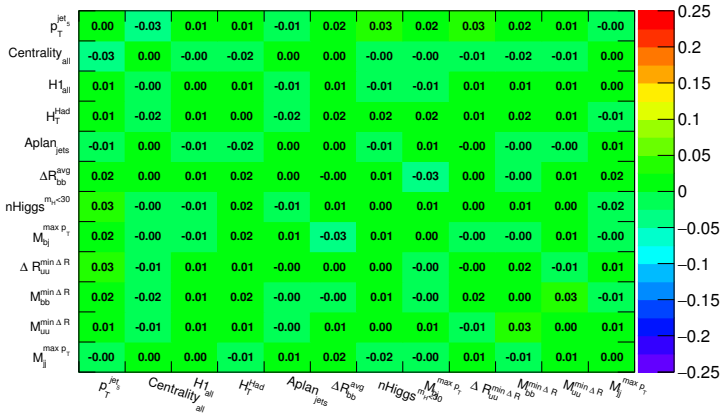


(c)

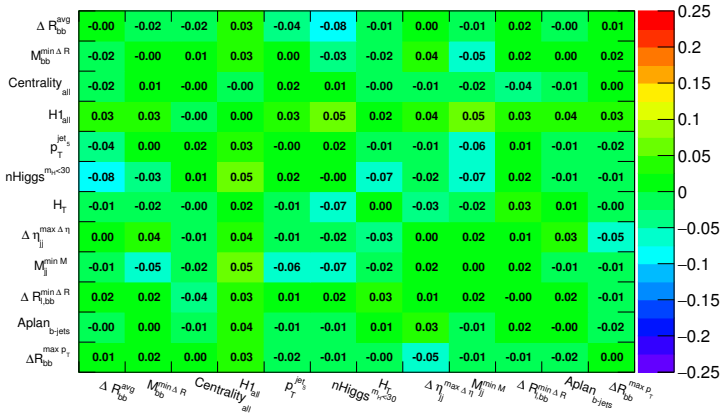
Figure 6.12.: Validation of ANN input variables. Matrices of correlation coefficients extracted from data. (a) (5j,≥4b). (b) (≥6j,3b). (c) (≥6j,≥4b).



(a)



(b)



(c)

Figure 6.13.: Validation of ANN input variables. Differences in the correlation between prediction and data. (a) (5j,≥4b). b (≥6j,3b). (c) (≥6j,≥4b).

6.6. ANN Training and Final Discriminant

For the training of the ANN, a global weight for all events is used in order to normalise signal yield to the background yield. Additional weights are applied in the MC background samples according to the expected yield. Since $t\bar{t}$+jets accounts for more than 95% of the total background in the signal regions, only this source of background is used for the training. The training is performed independently in each signal region.
The achieved separation for the ANN can be evaluated in different ways. One way is to use the same method as defined for the input variables. However, using the Receiver Operating Characteristic (ROC) curve is preferred, since the ROC curve definition is independent of the considered binning. To define the ROC curve, the efficiency is first introduced as follows:

$$\varepsilon(x) = \frac{\int_x^1 \text{Discriminant}}{\int_{-1}^1 \text{Discriminant}}, \tag{6.23}$$

where:

$$\varepsilon(x) \, \forall x \in [\text{-}1, 1]\,. \tag{6.24}$$

The ROC curve is obtained plotting ε_{sig} versus $1 - \varepsilon_{\text{bkg}}$, where ε_{sig} and ε_{bkg} are calculated using the output discriminant obtained from signal events or background events.
To evaluate the performance of the ANN and the quality of the achieved training, two different tests are considered. These tests require splitting the sample for both signal and background in subsets. The number of subsets can be arbitrary, but given the limited statistics for both signal and background, splitting in more than two subsets is not recommended. Thus, splitting is done according to the even/odd event number. Having two equivalent sets of samples for both signal and background, which are referred to as the training and the testing sample respectively, the following tests are defined:

- The overtraining test: This test is performed to evaluate a given ANN, obtained from the training sample, in the training and in the testing sample. The resulting two classification plots and the corresponding ROC curves are overlayed and compared. The test is passed if the ROC curves are equivalent for these two evaluations, otherwise the ANN is overfitted.

- The two-fold validation test: Having a training on the training (testing) sample, the performance of each training is evaluated on the testing (training) sample. The obtained two output classification plots and the corresponding ROC curves are overlayed and compared. The test is passed if the ROC curves are equivalent for these

two evaluations.

The overtraining test is performed, to check if the obtained ANN pattern recognition is biased towards the sample used for the training. The cause of overtraining is usually insufficient statistics of the sample used for the training, together with a number of input variables too large to train. This is another reason to use a limited number of input variables to train, especially for regions with lower statistics.
Being forced to split the samples in two subsets in order to perform the overtraining test, two trainings can be made, using both the training and the testing samples, with the advantage of considering all the available statistics. The two different trainings need to be equivalent, and this is exploited through the two-fold validation test. If this test is passed, the two ANN can be used when evaluating the ANN in the analysis, taking into account the even/odd splitting.
Using the ROC curves, a characteristic of the achieved separation power of the ANN is defined. By definition, the ROC curve integral is proportional to the latter, so the definition of the indicator is straightforward:

$$Index = 2(AUC) - 1\,, \tag{6.25}$$

where AUC is the area under the curve. The $Index$ can be expressed in percentage and is referred to as the *Frico-Gini Index*. The maximum integral value is 1, which corresponds to Frico-Gini Index of 100%. In the worst case, if the integral is 0.5, the Frico-Gini Index is 0%. The trainings are required to pass the two tests above and maximise the Frico-Gini index at the same time.
The plots for the overtraining and for the two-fold validation tests are shown in Figs. 6.14, 6.15 and 6.16. Plots of the final discriminants and their S/B separation are shown in Fig. 6.17. The number of events available for the training of the ANN is shown in Table 6.8.

	≥6j,≥4b	≥6j,3b	5j,≥ 4b
Signal	2M	2M	1M
Background	1.5M	1.5M	2M

Table 6.8.: Number of events used in the training of the ANN.

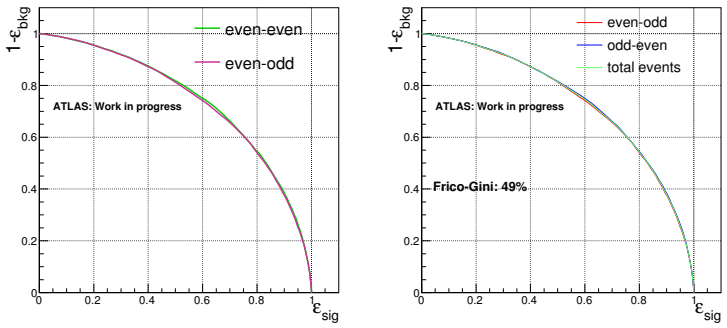


Figure 6.14.: Overtraining and two-fold validation tests for the (5j,≥4b) region.

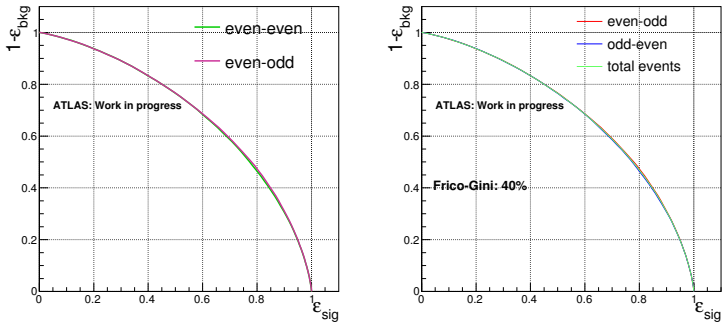


Figure 6.15.: Overtraining and two-fold validation tests for the (≥6j,3b) region.

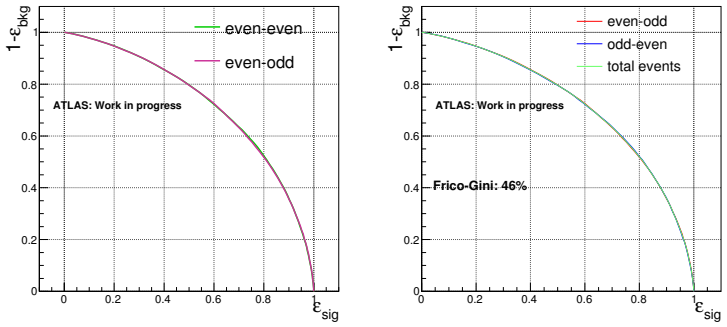
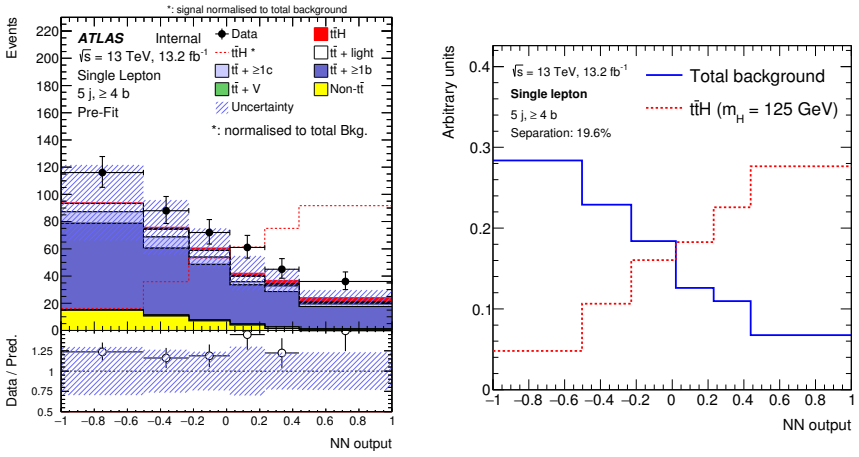
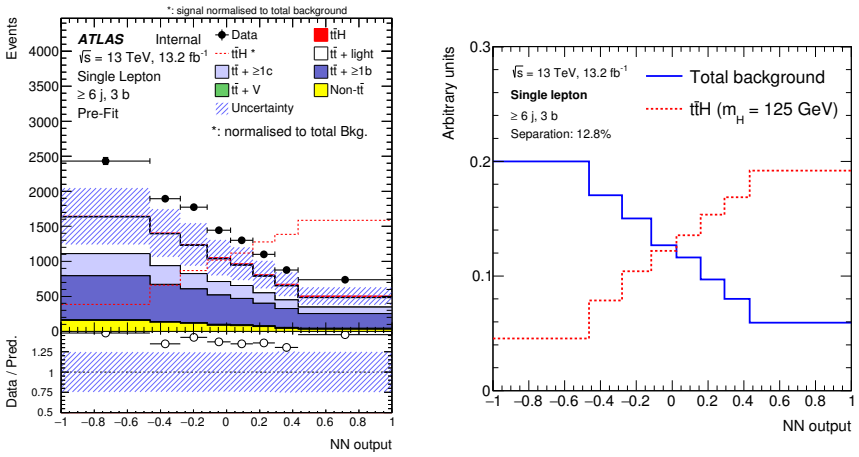


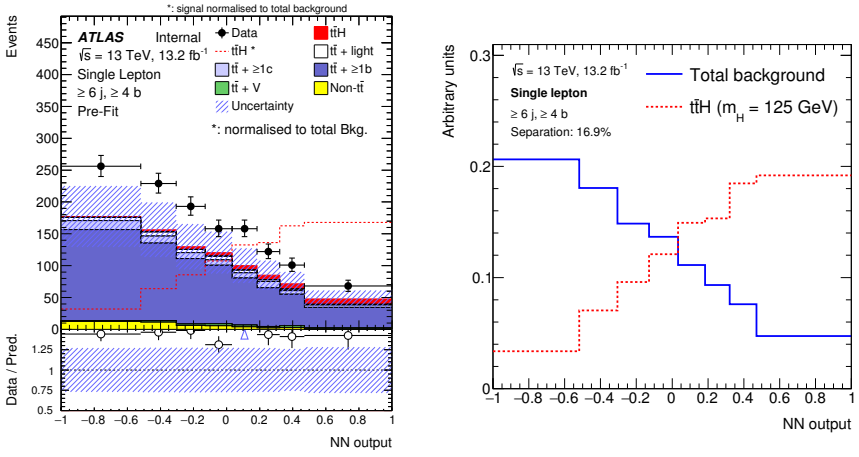
Figure 6.16.: Overtraining and two-fold validation tests for the (≥6j,≥4b) region.



(a)



(b)



(c)

Figure 6.17.: Final ANN discriminants. For each training region both the data-prediction comparison (left) and separation of the normalised plots (right) are shown. The binning choice is done in order to achieve the best possible separation. (a) (5j,≥4b). (b) (≥6j,3b). (c) (≥6j,≥4b)

6.7. BDT Based Event Reconstruction

Another MVA technique has been developed independently to perform an event reconstruction. In the final state, the identification of the correct $b\bar{b}$ pair coming from the decay of the Higgs boson is quite difficult since the choice of such pair has to be made among the possible $\binom{4}{2} = 6$ combinations, for the events with 4 b-jets. The possibility of picking up a wrong pair gives rise to a special background, called *combinatorial background*. A BDT implemented through the TMVA package is employed to reconstruct the $t\bar{t}H$ signal, discriminating the correct jets assignment from the combinatorial background.

For training purposes, a truth matching technique is employed to match each of the final state jets to the corresponding quarks from the hard scattering process. This is achieved by requiring that ΔR between the considered jet and the final state parton is less than 0.3. In this way, variables that take into account the characteristics of the truth-matched objects are defined, such as the invariant mass of the quark pair that matches the Higgs-boson, as well as angular distances between them. In the most signal sensitive region, (≥6j,≥4b), only in the 42% of the available events all the final state jets can be matched to the corresponding partons. The reason is that not all the decay products of the signal process are present, both due to the detector acceptance and the selection requirements for the physics objects.

The training of the reconstruction BDT is performed in $t\bar{t}H$ events. After the training, all the possible combinations of jets in the event are constructed and the BDT is evaluated for each combination, both for $t\bar{t}H$ and for $t\bar{t}$ samples. The combination that maximises the BDT output is selected as the final state reconstruction. The best reconstruction efficiency is achieved by using variables that contain information about the properties of the Higgs boson. However, this has a disvantage when trying to reconstruct $t\bar{t}$ events, since the chosen jet combination biases the distribution of the Higgs mass variable towards the $t\bar{t}H$ case. This results in the reduction of its discrimination power. To avoid this effect, two different reconstruction BDT are trained, with and without making use of the variables based on Higgs-boson information.

In the (≥6j,≥4b) region the reconstruction efficiency of 16% is achieved using Higgs-related variables in the training. Since the maximum achievable matching efficiency is about 42%, the relative all partons matching efficiency is about 38%. The efficiency is defined as the fraction of events in which each of the partons coming from the decay of the top quark or the Higgs boson are matched correctly to the final state jets within ΔR<0.3. For Higgs boson reconstruction, the relative efficiency is about 50% for the training using

the Higgs-related variables, while it is 30% otherwise, Fig. 6.18.
Given the reconstruction difficulties, a final classification BDT is trained to discriminate signal from the background as follows. A scheme showing the training method of the classification BDT is shown in Fig. 6.19.
Two different sets of variables, related to the quantities of reconstructed objects, are obtained from the best combinations of jets according to the choices of both reconstruction BDTs. Additionally, a third set of global event variables (as explained in Sec. 6.5), is considered for the training of the classification BDT, whose goal is to distinguish signal $t\bar{t}H$ events from $t\bar{t}$ background ones. Plots of the final classification BDT for each of the signal regions and the corresponding separation power are shown in Fig. 6.20.

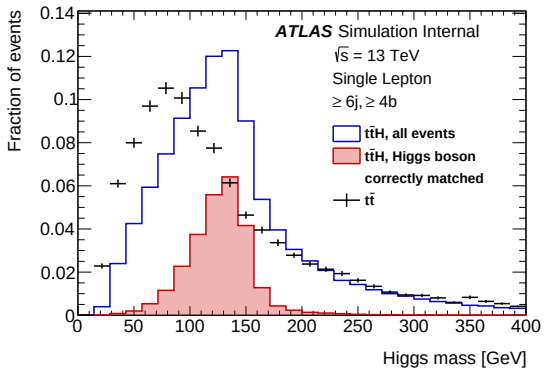


Figure 6.18.: The reconstructed Higgs boson invariant mass, from the reconstruction BDT that does not use Higgs-related input variables.

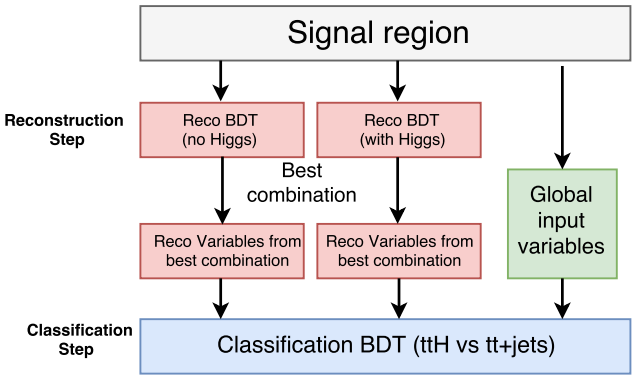
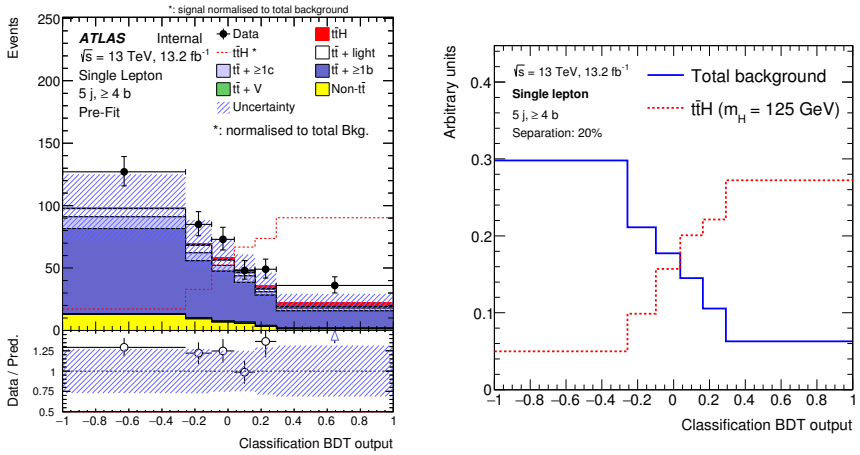
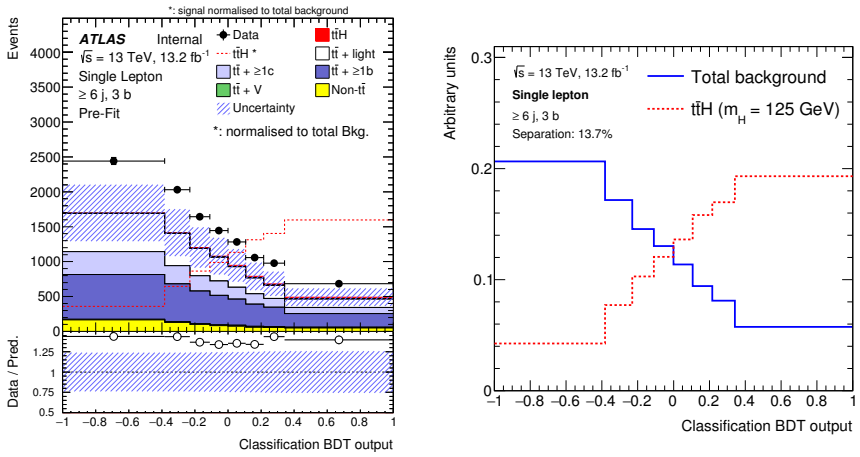


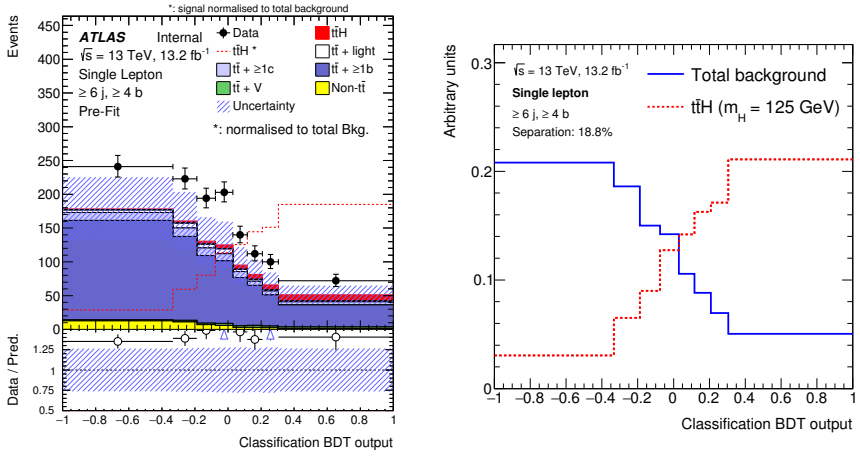
Figure 6.19.: Summary of the classification BDT training method.



(a)



(b)



(c)

Figure 6.20.: Final BDT discriminants. For each training region both the data-prediction comparison (left) and separation of the normalised plots (right) are shown. The binning choice is done in order to achieve the best separation power as possible. (a) (5j,≥4b). (b) (≥6j,3b). (c) (≥6j,≥4b)

7 Systematic Uncertainties and Statistical Tools

The systematic uncertainties arise from uncertainties in the calibrations of resolutions and efficiencies used to correct the simulations, or from an incomplete modelling of the physics of the MC processes that are compared to data. While statistical uncertainties are reduced by increasing the amount of the analysed data, the systematic uncertainties need dedicated studies. For this reason, the analysis presented in this thesis needs to be performed for each of the different set of objects for each systematic variation. Some systematic uncertainties affect only the number of the events for a certain physics process or shapes of certain distributions. In these cases, the analysis just needs to be adjusted by applying dedicated weights to each event without the need of repeating the full analysis.
In the case of the analysis presented in this thesis, since the considered signal is expected to be much smaller than the background, and thus a very sophisticated fit model is needed to extract it and thus many systematic uncertainties are considered. For a more accurate treatment, some systematic uncertainties are split into several components. In Table 7.1 the systematic uncertainties considered in this analysis are listed and details are given in Sec. 7.1
The distributions of the discriminants considered in each of the analysis regions are combined to test the presence of the $t\bar{t}H$ signal by performing a likelihood fit to data. This approach helps to reduce the impact of the systematic uncertainties. The statistical procedure is discussed in Sec. 7.2.

Systematic uncertainty	Type	Components
Luminosity	N	1
Reconstructed Objects		
Electron trigger+reco+ID+isolation	SN	5
Electron energy scale+resolution	SN	2
Muon trigger+reco+ID+isolation	SN	6
Muon momentum scale+resolution	SN	3
Pile-up modelling	SN	1
Jet vertex tagger	SN	1
Jet energy scale	SN	19
Jet energy resolution	SN	1
Missing transverse momentum	SN	3
b-tagging efficiency	SN	5
c-tagging efficiency	SN	4
Light-jet tagging efficiency	SN	14
High-p_T tagging	SN	2
Background and Signal Model		
$t\bar{t}$ cross section	N	1
$t\bar{t}$+HF: normalisation	N	2
$t\bar{t}+\geq 1b$: NLO Shape	SN	10
$t\bar{t}+\geq 1c$: NLO Shape	SN	1
$t\bar{t}$ modelling: residual Radiation	SN	3
$t\bar{t}$ modelling: residual NLO generator	SN	3
$t\bar{t}$ modelling: residual parton shower+hadronisation	SN	3
$t\bar{t}$ NNLO reweighting	SN	4
W+jets normalisation	N	6
Z+jets normalisation	N	6
Single top cross section	N	2
Single top model	SN	2
Diboson normalisation	N	1
Fakes normalisation	SN	7
$t\bar{t}V$ cross section	N	4
$t\bar{t}V$ modelling	SN	2
$t\bar{t}H$ cross section	N	2
$t\bar{t}H$ branching ratios	N	4
$t\bar{t}H$ modelling	SN	2

Table 7.1.: The list of systematic uncertainties considered. An "N" means that the uncertainty affects the normalisation of the discriminators for all processes and channels affected, whereas "SN" means that the uncertainty affects both shape and normalisation. Some systematic uncertainties are split into several components for a more accurate treatment.

7.1. Systematic Uncertainties

7.1.1. Luminosity

The uncertainty on the combined 2015 and 2016 run period integrated luminosity is estimated to 2.9% and affects the overall normalisation of all the processes estimated through a MC simulation. The methodology for estimating this uncertainty is based on a preliminary calibration of the luminosity scale using x-y beam-separation scans performed in August 2015 [135]. The luminosity systematic uncertainty is considered to be fully correlated for the 2015 and the 2016 datasets.

7.1.2. Reconstructed Objects

Jets

The systematic uncertainties associated with the jets are originated from the efficiency of the jet reconstruction, identification, the JVT, uncertainties on the Jet Energy Scale (JES) corrections and on the Jet Energy Resolution (JER).
The JES calibration and its uncertainty are derived by combining test-beam and LHC collision data information, and simulation [106]. Although more than 50 different sources of systematic uncertainties are considered, most of them are combined resulting in 19 uncorrelated sources that come from the different components of the JES calibrations:

- Six uncertainties related to the different in-situ techniques and divided according to their origin in different categories (statistical, modelling, detector and mixed).
- Four uncertainties related to the pile-up corrections, due to potential mismodelling of the number of primary vertices and the average number of interactions per bunch crossing.
- Three uncertainties related to the η-intercalibration techniques.
- Three uncertainties related to jet-flavour, which take into account the different response of the calorimeter to jets originated from quarks or gluons.
- One high-p_T uncertainty, which is derived from the single-particle response measurements for high-p_T jets (>1 TeV).
- One uncertainty related to the corrections applied for the global sequential calibration.

- One uncertainty related to the correction applied to the simulated jets that use a parametrised simulation of the calorimeter.

The JES uncertainty has different impact depending on the p_T of the jets. It is about 5.5% for jets with $p_T = 25$ GeV and it is below 1.5% for jets with p_T in the range of 100 GeV-1.5 TeV, see Fig. 7.1. JES represents one of the most important sources of uncertainty since this analysis uses many jets, playing a role in the normalisation of signal and backgrounds in multiple regions.

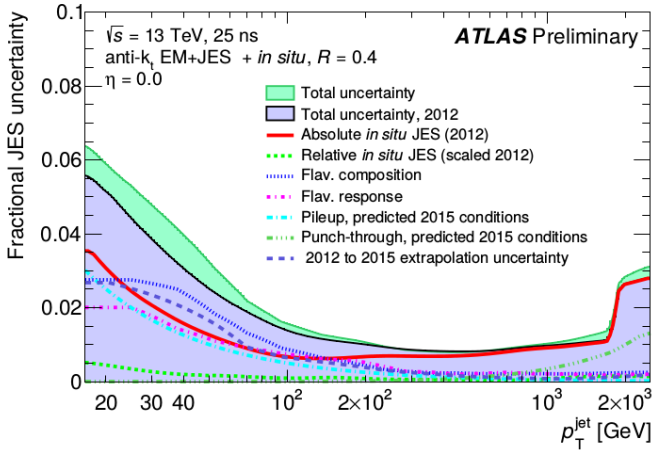


Figure 7.1.: JES estimated for 2015 data as a function of jet p_T for jets of $\eta = 0$ [106].

The reconstructed jet has an energy whose measurement is affected by resolution effects, due to stochastic fluctuations in the calorimeters, electronic noise and fluctuations due to detector calibration measurements. For this reason, the measurements of the energy for jets having the same true energy follow a Gaussian distribution, whose standard deviation is the JER. The JER measurement during Run 1 was compared with a simulation of dijet events and it was found to agree within 10% [136]. These uncertainties have been combined with the ones obtained from extrapolation from Run 1 to Run 2 conditions [106] and have the effect of smearing the jet p_T in the MC simulations.

Heavy and Light Flavour Tagging

The efficiencies associated to b- and c-tagging in simulation are corrected making use of dedicated scale factors to achieve the best possible agreement with the efficiencies in data. The scale factors are p_T dependent for the case of b- and c-quarks, while they depend also on η for the light-jets.
The efficiencies and corrections associated to b-jets are measured on data using $t\bar{t}$ events,

while the mis-tag efficiencies for c-jets are derived from measurements of the D^* meson decays [137]. Uncertainties on the scale factors are factorised into a total of five independent components associated to b-jets and four independent components associated to c-jets. Fourteen uncertainty components are considered for the light-jet tagging. These uncertainties are uncorrelated.
Since the c- and light-jet measurements made use of data collected during the Run 1 period, an extrapolation to the Run 2 uncertainty is taken into account.
An additional uncertainty is finally included to take into account the extrapolation of the different scale factors for jets beyond the kinematic range of the samples used for the data calibration. This uncertainty is taken as correlated among the different jet flavours.

Leptons

The uncertainties associated to the lepton reconstruction are due to uncertainties coming from lepton momentum scale and lepton resolution as well as from trigger, identification, reconstruction and isolation efficiencies. The uncertainties have been measured in data from $Z \to \ell^+\ell^-$, $J/\psi \to \ell^+\ell^-$ events, and also from the energy and momentum of the electrons in $W \to e\nu$ events. These uncertainties have a minimal impact on the analysis presented in this thesis.

Missing Transverse Momentum

The uncertainties associated to reconstruction of E_T^{miss} are propagated from leptons, JES and JER. These uncertainties have a small impact in the analysis since E_T^{miss} is not used in the event selection but only in the event reconstruction.

7.1.3. Uncertainties on Background Modelling

$t\bar{t}$ + jets Modelling

$t\bar{t}$ +jets process is the main source of background for the analysis presented in this thesis. For this reason, a large number of systematic uncertainties affecting this process are considered. These are uncertainties associated to the theoretical prediction for the inclusive cross section, as well as overall normalisation of the $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ components, uncertainties associated to the modelling of the $t\bar{t}+ \geq 1b$ event production, such as choice of the matrix element generator, modelling of extra radiation, choice of the model for the

parton shower and hadronisation. A summary of these uncertainties is presented in Table 7.2.

The uncertainty considered for the inclusive $t\bar{t}$ production cross section is ±6% [76] and includes contributions from variations of PDFs, α_S, renormalisation and factorisation scale and from top quark mass uncertainty. To take into account the different choices for the NLO generator, an uncertainty is derived by comparing two different predictions using Powheg-Box and MG5_aMC, each of them making use of Herwig++ for the parton shower simulation. The uncertainty associated to the choice of the modelling of initial and final state radiation is derived with two different Powheg-Box+Pythia6 samples, one with increased and another with decreased radiation. The first is obtained decreasing the renormalisation and factorisation scale by a factor of two and with the hdamp parameter (defined in Sec. 4.3.2) doubled, while the sample with radiation decreased has the scales increased by a factor of two [138]. A systematic uncertainty is finally derived for the top quark and $t\bar{t}$ transverse momentum prediction by considering the largest disagreement between the NNLO prediction and the uncorrected prediction obtained with any of the already mentioned alternative samples. This uncertainty is considered to be uncorrelated among $t\bar{t}+ \geq 1b$, $t\bar{t}+ \geq 1c$ and $t\bar{t}$ light components.

For the $t\bar{t}+ \geq 1b$ component, all the samples have been reweighted to the NLO SherpaOL prediction. The differences which remain after the reweighting procedure are referred to as *residual uncertainties*. The uncertainties on the reweighting procedure are evaluated by lowering and increasing the renormalisation scale by a factor of two, changing the resummation scale definition to μ_{CMMPS}, setting a global value for the scale, and fixing them to the same μ_{CMMPS}. Regarding the PDFs choice, two alternative PDFs are considered: MSTW [49] and NNPDF [47]. The uncertainty associated to the choice of the generator is derived by a comparison between the predictions of SherpaOL and MG5_aMC+Pythia8, while the uncertainty from the parton shower and hadronisation model is taken from the differences between the samples generated with MG5_aMC and showered either with Pythia8 or Herwig++. One last systematic uncertainty is associated to the events not included in the original NLO calculation but coming from MPI or FSR. A 50% uncertainty is assigned for the contribution of MPI while the uncertainty on FSR is derived from the radiation samples introduced before.

Since the charm jets are produced by a parton shower, a prediction obtained with the $t\bar{t} + c\bar{c}$ matrix element calculation, has been studied. An NLO matrix element calculation for the $t\bar{t} + c\bar{c}$ process has been obtained with MG5_aMC and Herwig++ as described in [139]. This sample has been compared to an inclusive $t\bar{t}$ sample simulated with the same

generator, and the differences have been taken as systematic uncertainty for the $t\bar{t}+ \geq 1c$ prediction.
As already mentioned, the data overshoot the prediction in all the multiplicity regions with significant $t\bar{t}$ +HF contributions, including all the considered signal regions. Even if the excess is consistent with the prediction due to large uncertainties associated to the $t\bar{t}$ +HF predictions [140], to avoid a potential bias that can affect the MVA discriminant shape, the normalisation of the $t\bar{t}+ \geq 1b$, $t\bar{t}+ \geq 1c$ components are allowed to float freely in the fit.

Multijet Modelling

The uncertainties on the data-driven multijet background estimate mainly originate from the limited sample size in data, especially at high jet and b-tag multiplicities, and also from the uncertainties on the rate of fake leptons. Following the approach used in Run 1, a normalisation of 50% is assigned and it is taken as uncorrelated across the lepton flavours and b-tag multiplicities.

Other Simulated Backgrounds

For the W/Z+jets background normalisation, an uncertainty of 30% is used and taken as uncorrelated among the jet multiplicities. For the W+HF jets process, an additional 30% uncertainty based on variation of the scales and matching parameters in the Sherpa MC simulation is taken into account.
Concerning the single top production, an uncertainty of $^{+5\%}_{-4\%}$ is applied on the total cross section prediction [90–92]. Similarly to the $t\bar{t}$ case, an additional systematic uncertainty associated to final state radiation is considered. An additional uncertainty is considered to account for the interference between the $t\bar{t}$ and Wt processes at NLO. This is achieved by comparing an alternative sample generated with the diagram subtraction scheme to the default sample produced with the diagram removal scheme [89].
A normalisation uncertainty of 50% is considered for the diboson background. This takes into account both the prediction for the inclusive cross section and the additional jet production [141]. An uncertainty of 15% is used for the $t\bar{t} + V$ NLO cross section prediction [142].

Systematic source	How evaluated	$t\bar{t}$ categories
$t\bar{t}$ cross-section	±6%	All, correlated
NLO generator (*residual*)	Powheg-Box + Herwig++ vs. MG5_aMC + Herwig++	All, uncorrelated
Radiation (*residual*)	Variations of μ_R, μ_F, and *hdamp*	All, uncorrelated
PS & hadronisation (*residual*)	Powheg-Box +Pythia 6 vs. Powheg-Box + Herwig++	All, uncorrelated
NNLO top & $t\bar{t}$ p_T	Maximum variation from any NLO prediction	$t\bar{t}+ \geq 1c$, $t\bar{t}$ +light, uncorr.
$t\bar{t} + b\bar{b}$ NLO generator *reweighting*	Sherpa+OpenLoops vs. MG5_aMC +Pythia8	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ PS & hadronis. *reweighting*	MG5_aMC +Pythia8 vs. MG5_aMC + Herwig++	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ renorm. scale *reweighting*	Up or down a by factor of two	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ resumm. scale *reweighting*	Vary μ_Q from $H_T/2$ to μ_{CMMPS}	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ global scales *reweighting*	Set μ_Q, μ_R, and μ_F to μ_{CMMPS}	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ shower recoil *reweighting*	Alternative model scheme	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ PDF *reweighting*	CT10 vs. MSTW or NNPDF	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ MPI	Up or down by 50%	$t\bar{t}+ \geq 1b$
$t\bar{t} + b\bar{b}$ FSR	Radiation variation samples	$t\bar{t}+ \geq 1b$
$t\bar{t} + c\bar{c}$ ME calculation	MG5_aMC + Herwig++ inclusive vs. ME prediction	$t\bar{t}+ \geq 1c$

Table 7.2.: A summary of the systematic uncertainties on the $t\bar{t}$+jets modelling. For the $t\bar{t}+ \geq 1b$ background, the inclusive $t\bar{t}$ sample is reweighted to an NLO $t\bar{t} + b\bar{b}$ prediction; uncertainties on the inclusive sample are labelled *residual*, while those on the NLO prediction are labelled *reweighting*.

7.1.4. Signal Modelling

An uncertainty of +10%/-13% is assigned to the NLO cross section prediction of the $t\bar{t}H$ process [66–68]. This is split into scale and PDFs uncertainties, which are considered as uncorrelated. Additionally, uncertainties are derived to take into account the differences between the parton shower and hadronisation models. The uncertainty due to the parton shower model choice is derived by comparison of the samples obtained with MG5_aMC interfaced to either Pythia8 or Herwig++. The effect of the variation of the renormalisation and factorisation scales is taken into account by a dedicated systematic uncertainty. Finally, uncertainties on the Higgs boson branching ratios to $b\bar{b}$, WW, and other decays are also considered [27].

7.2. Statistical Tools

In the analysis presented in this thesis, as well as many other analysis, the object of the search is a physical process that is predicted by the SM but still not observed. To quantify the agreement between theoretical predictions and experimental data, some statistical concepts are used, such as the frequentist statistical test. In the context of this test, a null hypothesis is defined, H_0, describing all the processes that are already known, together with an alternative hypothesis, H_1, that includes also other not yet discovered processes. For the analysis presented in this thesis, the null hypothesis consists of a background hypothesis (where the SM does not predict $t\bar{t}H$), while the H_1 considers both background and signal. The two different hypothesis can be represented by a multiplicative factor μ to the $t\bar{t}H$ predicted cross section. Defining $\mu = \frac{\sigma_{\text{obs}}}{\sigma_{\text{SM}}}$ as *signal strength*, μ is equal to 0 for H_0 and to 1 for H_1.

Having a hypothesis associated to a certain μ, one needs to quantify the level of agreement of the experimental data with it. If the null hypothesis can be rejected, i.e. if data are incompatible with the background-only hypothesis, a discovery can be claimed. To quantify the agreement, one needs to assign a probability (p-value), referred to as p_μ, where μ refers to the signal strenght associated to the hypothesis under test, of finding data incompatible with the predictions of the hypothesis under the assumption that it is true. The hypothesis can be discarded if the p-value is observed to be below a certain threshold. Usually, the concept of p-value is converted into an equivalent parameter Z, being the number of standard deviations above which a Gaussian distributed variable has an upper-tail probability equal to p, Fig. 7.2. This is defined as:

$$Z = \Phi^{-1}(1 - p)\,, \tag{7.1}$$

where Φ^{-1} is the inverse of the cumulative distribution of the standard Gaussian.

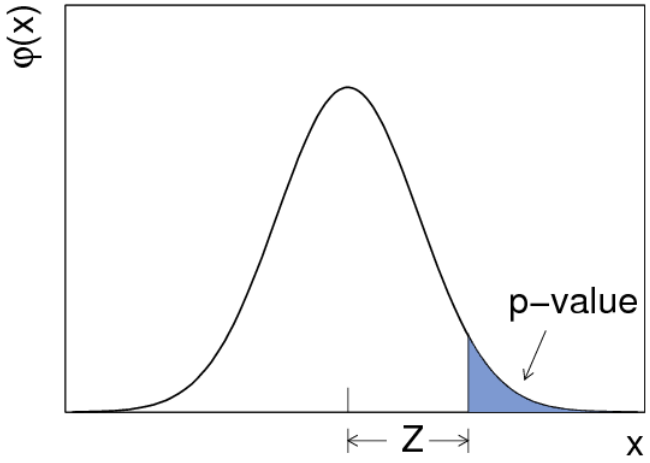


Figure 7.2.: The standard normal distribution $\varphi(x)$ showing the relation between the significance Z and the p-value [143].

By convention, a p-value greater than 5% (Z<1.64) is considered to exclude a new process which corresponds to a $\mu = 1$ hypothesis at 95% CL, while a Z>5 is needed to claim a discovery, and it corresponds to a p-value less than $2.9 \cdot 10^{-7}$.

7.2.1. Profile Likelihood Ratio

In order to test the presence of the $t\bar{t}H$ signal, a binned profile likelihood fit is performed simultaneously in all the analysis regions considered using the distributions of H_T^{had} in the control regions and of the MVA in the signal regions.
Each bin of each distribution in each of the regions has an expected number of events given by the equation:

$$E_{ij} = \mu s_{ij} + b_{ij} \tag{7.2}$$

where s_{ij} and b_{ij} represent the number of expected events associated to either signal or background processes in the i bin of the j histogram. Since the data follow a Poisson distribution around the number of expected events, it is possible to define a merely statistical *binned likelihood function* $\mathcal{L}(\mu)$ as the product of the Poisson probability terms over each bin of each of the considered distributions:

$$\mathcal{L}(\mu) = \prod_{j}^{9} \prod_{i}^{n_{\text{bins}(j)}} \frac{(E_{ij})^{n_{ij}}}{n_{ij}!} e^{-E_{ij}}, \tag{7.3}$$

where $n_{\text{bins}(j)}$ is the number of bins for the j histogram and n_{ij} is the observed number of events for the bin i and for the histogram j. The free parameter μ is estimated by maximising the $\mathcal{L}(\mu)$ or minimising its logarithm by a *fit* procedure. The error on the achieved best estimate of μ is obtained through a scan of the values of the likelihood as a function of μ. The 1-σ band is set finding the points in which the logarithm of the likelihood decreases by a factor of two with respect to its maximum.
In real cases, the expected number of events for signal and background processes is affected by both statistical and systematic uncertainties. The k systematic uncertainties are considered directly in the definition of the likelihood, through a collection of k continuous parameters θ_k, referred to as Nuisance Parameters (NPs). By varying the values of the NPs, one changes both the shape and the normalisation of the predictions, so the s_{ij} and b_{ij} are then also dependent on k NPs, referred to in the following as $\boldsymbol{\theta}$. By maximising the likelihood, the best values for the $\boldsymbol{\theta}$ in order to improve the agreement between expected and observed number of events are found. The NPs are inserted in the definition of the likelihood through their probability distribution functions $\rho(\boldsymbol{\theta})$.

$$\mathcal{L}(\mu, \theta) = \mathcal{L}(\mu) \prod_{k} \rho(\theta_k) \tag{7.4}$$

The $\rho(\boldsymbol{\theta})$ are also referred to as *penalty* terms or *prior distributions* on $\boldsymbol{\theta}$. The assumed functional form of the priors depends on the considered nuisance parameter. Three different types are used in this analysis [144]:

- Gaussian prior distribution: this is the assumed shape for most of the NPs. The associated function is:

$$\rho(\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{ -\frac{(\theta_k - \bar{\theta}_k)^2}{2\sigma_k^2} \right\} \tag{7.5}$$

  where the central value $\bar{\theta}_k$ is the measured value of a certain systematic variation and σ_k is the uncertainty associated to it. The usage of a Gaussian distribution prevents the fit to prefer very large deviations from the measured value in the minimisation procedure.

- Log-normal prior distribution: this shape is used for those NPs associated to quanti-

ties that always need to be positive defined, such as normalisations. The associated function is:

$$\rho(\theta_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{ -\frac{\left(\log\left(\theta_k/\bar{\theta}_k\right)\right)^2}{2(\log(\sigma_k))^2} \right\} \frac{1}{\theta_k} \tag{7.6}$$

- Gamma prior distribution: the gamma distribution is associated to the NPs which are introduced to take into account the statistical uncertainty on the number of the selected MC events. This takes the form:

$$\rho(\theta_k) = \frac{A}{\Gamma(B)} (A\theta_k)^B e^{-A\theta_k} \tag{7.7}$$

where $A = (1/\sigma_k^{rel})^2$, σ_k^{rel} is the relative statistical uncertainty of the considered bin, and $B = N - 1$ with N the bin content, rounded to the nearest integer.

By convention, the NPs are defined such that the value of $\boldsymbol{\theta} = 0$ refers to the nominal value of the prediction while the values of ± 1 refer to $\pm 1\sigma$ variations of the systematic uncertainty associated to the considered θ:

$$\theta' = \frac{\theta - \bar{\theta}}{\sigma} \,. \tag{7.8}$$

After the maximisation procedure is concluded, the values of $\hat{\boldsymbol{\theta}}$ and $\hat{\mu}$ are defined as the ones which maximise the likelihood. If the observed data are not sensitive to a given source of systematic uncertainty, the best value of the corresponding θ_k stays at 0 and its error is consistent with the input uncertainty. In the opposite case, the fit can shift (pull) the best value for a given NPs to achieve a better data/MC agreement or produce a reduction (constraint) of the error associated to a nuisance parameter. The latter case happens when the large effects of a given systematic uncertainty are not supported by the available data. Constraints provided by data can help to increase the sensitivity of the measurement. Statistical fluctuations can produce additional shape differences in the considered distributions, changing the result of the fit. To avoid this, a *smoothing* procedure is applied to merge the bins until the shape differences are significant compared to the statistical fluctuations. To neglect those uncertainties which do not play a role in the fit, all the systematic uncertainties affecting the total normalisation or the total shape by less than 0.5% are dropped. This procedure is referred to as *pruning*. Pruning does not affect the result of the fit. The tool used to implement the profile likelihood fit is the RooFit framework [145].

The likelihood definition gives the possibility to define confidence intervals as well. The p_μ are defined as profile likelihood ratios:

$$q_\mu = -2\log(\lambda(\mu)) = -2\log\left(\frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}\right), \tag{7.9}$$

where $\hat{\hat{\boldsymbol{\theta}}}$ are the values of the NPs that maximise the likelihood for a given value of μ, with the constraint $0 \leq \hat{\mu} \leq \mu$ since physics only allows the $\hat{\mu}$ to be positive and the exclusion limit μ needs to be greater than the best estimator. By using the Wilk and Wald theorems [146] which hold for sufficiently large dataset statistics, the asymptotic approximation is obtained:

$$q_\mu = -2\log(\lambda(\mu)) \simeq \frac{(\mu - \hat{\mu})^2}{\sigma^2}, \tag{7.10}$$

where σ represents the variance of the likelihood estimate of μ. Such a parameter is calculated making use of the so called Asimov dataset [143], an artificial dataset in which all observed quantities are set equal to their expected values[1]. A dataset defined in this way is such that when it is used to evaluate the estimators for all parameters, the true parameter values are obtained. The Asimov dataset has the particularity that all the pulls for the NPs are zero by definition.

7.2.2. Limit Setting

When a significant excess above the background-only hypothesis is not found, the q_μ can be used to obtain an upper limit on the production cross section of the $t\bar{t}H$ process. This is achieved by a frequentist Confidence Level (CL) approach [143]. The definition of the CL limit is based on the q_μ, which is used to construct two different distributions of the test statistics: the one for the background-only hypothesis indicated as $f(q_\mu|b)$ or $f(q_\mu|0)$ and the one for the signal plus background hypothesis indicated as $f(q_\mu|s+b)$ or $f(q_\mu|1)$. Introducing the test statistic indicated as q_{obs}, which refers to the value of μ under test, the compatibility of the result with the signal plus background hypothesis is given by the p-value:

$$p(\mu = 1) = f(q \geq q_{\text{obs}}|1) = \int_{q_{\text{obs}}}^{\infty} f(q_\mu|1)dq_\mu, \tag{7.11}$$

[1]The name of this dataset is given to remind its connection to the tale *Franchise* by the great science fiction writer Isaac Asimov (1920-1992). This tale describes a society where elections are conducted by replacing the whole electorate by the most representative voter.

which indicates the probability of finding a result less compatible to the signal plus background hypothesis than the considered μ under test. An analogous definition is given for the compatibility with the background-only hypothesis:

$$p(\mu = 0) = f(q \geq q_{\text{obs}}|0) = \int_{-\infty}^{q_{\text{obs}}} f(q_\mu|0) dq_\mu \,. \tag{7.12}$$

A visual representation of both is given in Fig. 7.3. With these definitions it is possible to define the confidence level for the signal hypothesis as [147]:

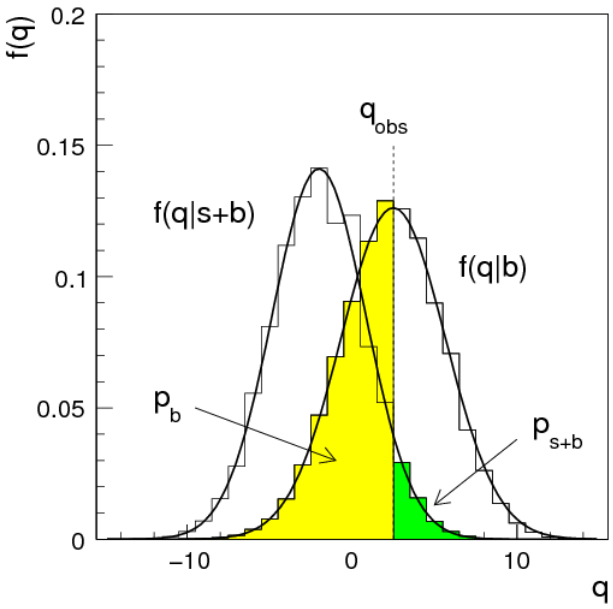


Figure 7.3.: Example distribution of the test statistics for background-only and signal plus background hypothesis [143].

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p(\mu = 1)}{1 - p(\mu = 0)} \,. \tag{7.13}$$

This definition allows to avoid excluding models for which low sensitivity is expected. The CL of α is set by adjusting the μ until the value of $CL_s = 1 - \alpha$ is reached. Values of μ for which the CL_s is smaller than 0.05 are excluded at 95% confidence level

8 Results and Limits on $t\bar{t}H(H\rightarrow b\bar{b})$ Production

In this chapter the final fit result is presented for the cases where the ANN or the BDT are used as signal/background discriminants for the signal regions. The two results were found to be consistent. The ANN final result is the main contribution to the analysis by the author of this thesis and was considered as an internal cross check, while the one obtained with the BDT was used as the public result of the analysis presented at the ICHEP conference 2016 [61]. This choice has been made for a number of reasons, the most important one being a slightly better signal/background separation obtained by the BDT.
The limits on the $t\bar{t}H$ production and combination with the dileptonic channel were derived from the BDT analysis, as discussed in Sec. 8.2.2
The fit result is obtained by performing a simultaneous likelihood fit to data using all the distributions of the discriminants in the nine analysis regions. The hypothesis considered for the fit is background plus signal, where the signal strength, μ, is the parameter of interest of the fit. μ is allowed to float freely in the fit, but with the condition to be the same in all the considered regions. The normalisation factors for the $t\bar{t}$ +HF background are similarly free parameters of the fit. The normalisation of each of the other backgrounds is constrained by the uncertainties of the respective theoretical cross section calculations, as well as by instrumental uncertainties. The MC statistical uncertainties, are taken into account in the fit by dedicated NPs. The analysis regions are sensitive to different sets of systematic uncertainties due to the differences in the event composition, allowing changing the values of the corresponding NPs to best fit the data. The effects of the several sources of systematic uncertainties are constrained by the large number of events selected in the control regions. The total background uncertainty after the fit is sensibly reduced compared to the one before the fit, not only for the control regions, but also for the signal regions. The regions with only two $b-$jets are almost pure in $t\bar{t}$ +light background and for

this reason they provide an important constraint on the $t\bar{t}$ modelling uncertainties in both shape and normalisation. Uncertainties on c-tagging are constrained mainly in the (4j,3b) region due to the large contribution of $W \to cs$ events in the $t\bar{t}$ +light background.
To study the expected performance of the fit, an Asimov dataset (defined in Sec. 7.2.1) assuming the signal plus background hypothesis is generated both for the ANN and the BDT case. Since the pulls obtained after the fit to data in the two cases are very similar, as well as the constraints of the NPs, they are discussed only for the ANN case. The explanation for the BDT case is equivalent.
The binning of the two MVA discriminants is chosen in order to maximise the signal/background separation, resulting in different binning choices for the two cases, which plays a main role in the smoothing procedure for the systematic uncertainties and affects the final result.

8.1. ANN Results

8.1.1. Expected Fit Results

A fit to an Asimov dataset is performed to evaluate the sensitivity of the analysis. The error on the fitted NPs is expected to be close to 1 if the data do not provide any improvement of that uncertainty. Plots of the fitted NPs are available in Fig. 8.1
Among the NPs related to modelling systematic uncertainties, the ones related to the modelling of the $t\bar{t}$ +jets events are constrained. This means that the large model variations are not compatible with the available data precision. The most notable constraint is obtained by the $t\bar{t}$ +PS systematic uncertainty, which changes the predicted yields by more than 10%, while the data statistical uncertainty in the regions with two b-tagged jets is only a few %.
NPs related to instrumental systematic uncertainties also show some minor constraints, in particular for some components of the JES uncertainty, as well as JER, JVT and jet flavour tagging ones. The first three uncertainties have a large effect on the low jet p_T region, where large data statistics is available.

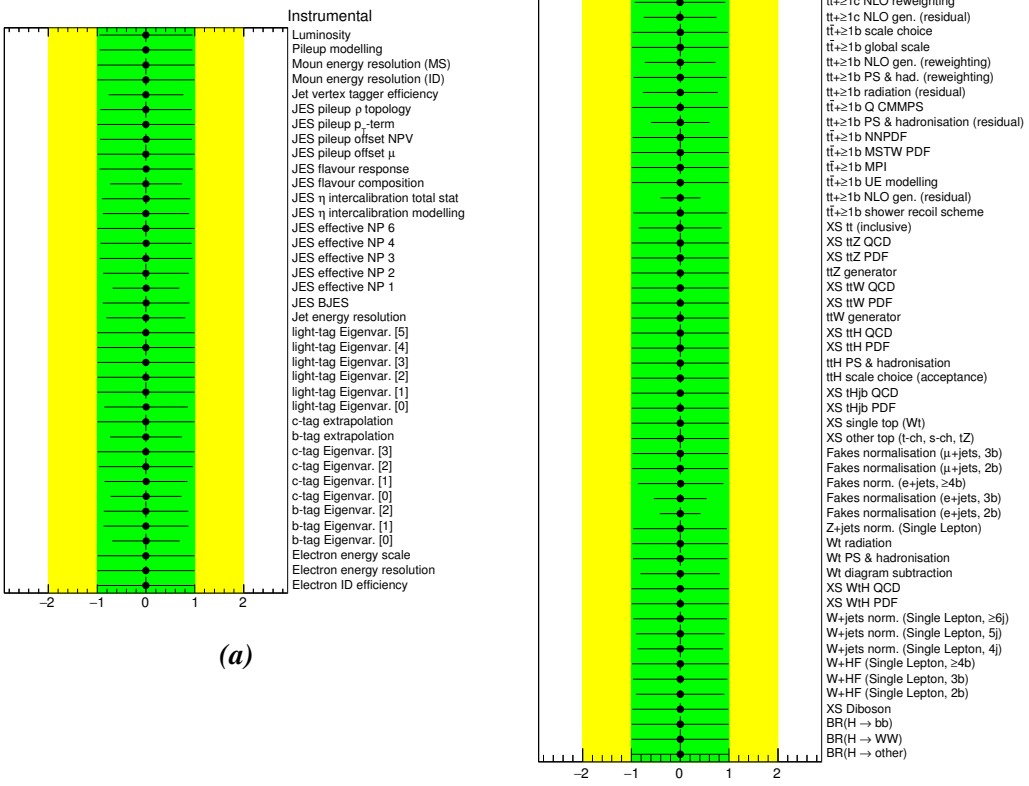


Figure 8.1.: Fitted NPs from a fit to the Asimov dataset using the ANN. (a) NPs associated to instrumental systematic uncertainties; (b) NPs associated to modelling systematic uncertainties.

8.1.2. Fit to Data

The result of the fit to data yields a signal strength of $\mu = 0.9^{+1.2}_{-1.1}$, a normalisation factor $k_{t\bar{t}+b\bar{b}} = 1.27^{+0.24}_{-0.21}$ for the $t\bar{t}+ \geq 1b$ background and $k_{t\bar{t}+c\bar{c}} = 1.40^{+0.74}_{-0.62}$ for the $t\bar{t}+ \geq 1c$. A comparison between the analysis yields and data in each of the fitted regions is available in Fig. 8.2. The fitted NPs are shown in Fig. 8.3.

The observed constraints on the NPs are compatible with the expected ones obtained from the fit to the Asimov dataset. Different pulls of the fitted values of the NPs are visible. The most notable ones are:

- Heavy and light flavour tagging: These pulls result in increased values of the scale factors applied to MC associated to c-tag and mis-tag. The largest component of the c-tagging systematic uncertainty is pulled to -1.03. This is due to the used calibrations for the scale factors, which are based on data from Run 1 which might not be valid for the different tagger choice of Run 2. This is enhanced by the large statistics available for the control regions. The most important component of the light flavour tagging systematic uncertainty is pulled by the regions with large component of $t\bar{t}$ +light jets events.

- $t\bar{t}$ modelling: These pulls are due to the differences on the predictions of the different used MC generators, especially the most visible pull on the $t\bar{t} + c\bar{c}$ generator.

A notable feature of the likelihood fit is that correlations between the NPs are introduced. Fig. 8.4 shows the correlation coefficients obtained. Even if the uncertainties are taken as uncorrelated before the fit, the correlations are derived by the fit procedure. As expected, the two normalisation factors in the fit for the $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ components of the $t\bar{t}$ +jets background are anti-correlated. An anti-correlation between the $t\bar{t}+ \geq 1b$ uncertainty and the signal strength uncertainty is found. This correlation is related to the ability of separating signal from background in the signal regions. The correlation becomes larger when using a variable with lower discrimination power.

To study the dependence of the fit result on the different sources of systematic uncertainty, the fit procedure is repeated for each of the considered NPs keeping its value fixed at different values. All the other NPs are allowed to vary in the new fit. Fig. 8.5 shows the variation of the signal strength, defined as *impact on* μ, when a considered NPs is fixed at the two values of $\hat{\theta} \pm \sigma_\theta$, where σ_θ is the considered pre-fit or post-fit error on the parameter. Only top twenty NPs with the largest impact on μ are shown. The sum of the impact on μ of the single systematic uncertainties in quadrature is not the total error on

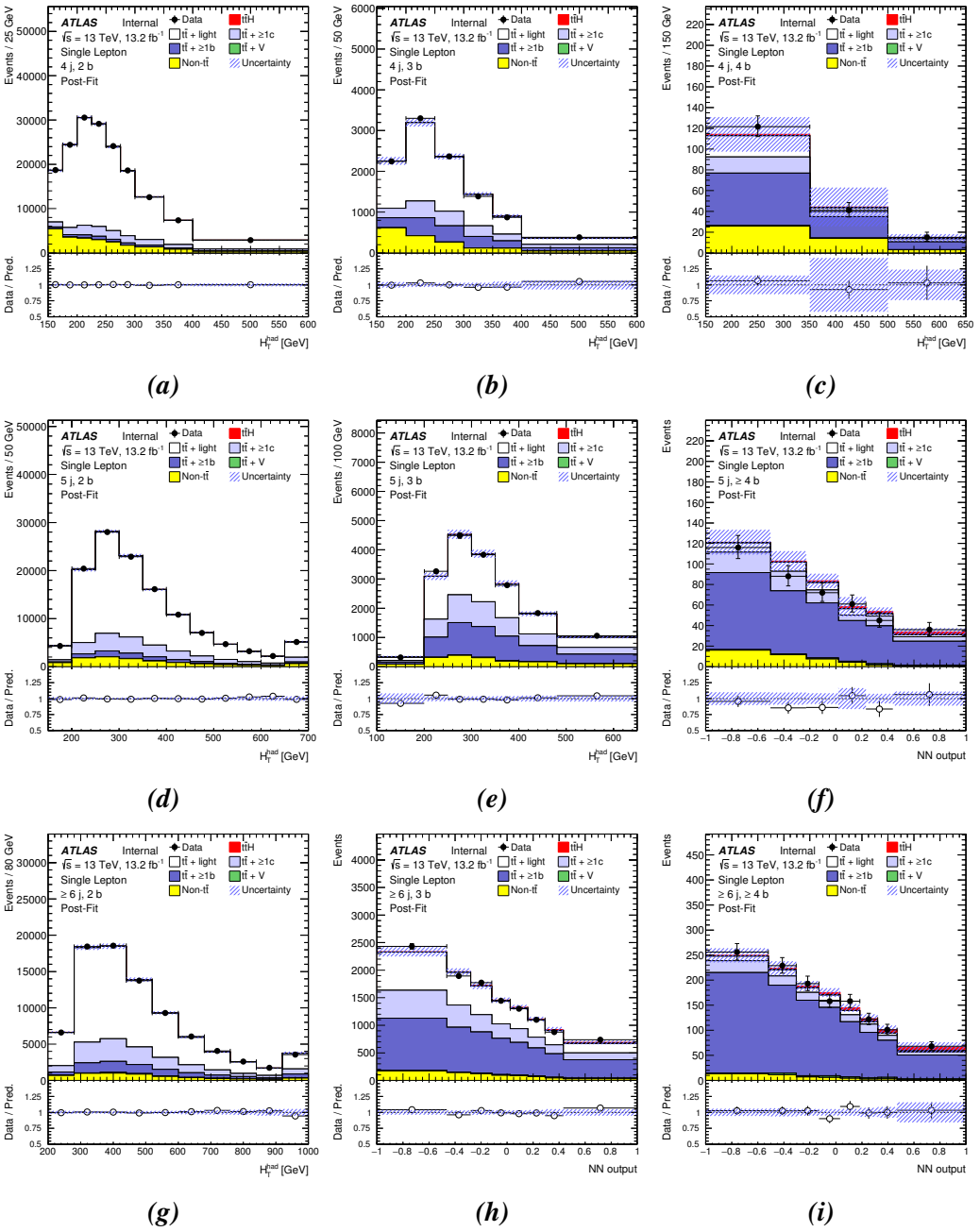


Figure 8.2.: (a) to (i), post-fit distributions of H_T^{had} and ANN in the regions used for the fit to data using the ANN in the signal regions.

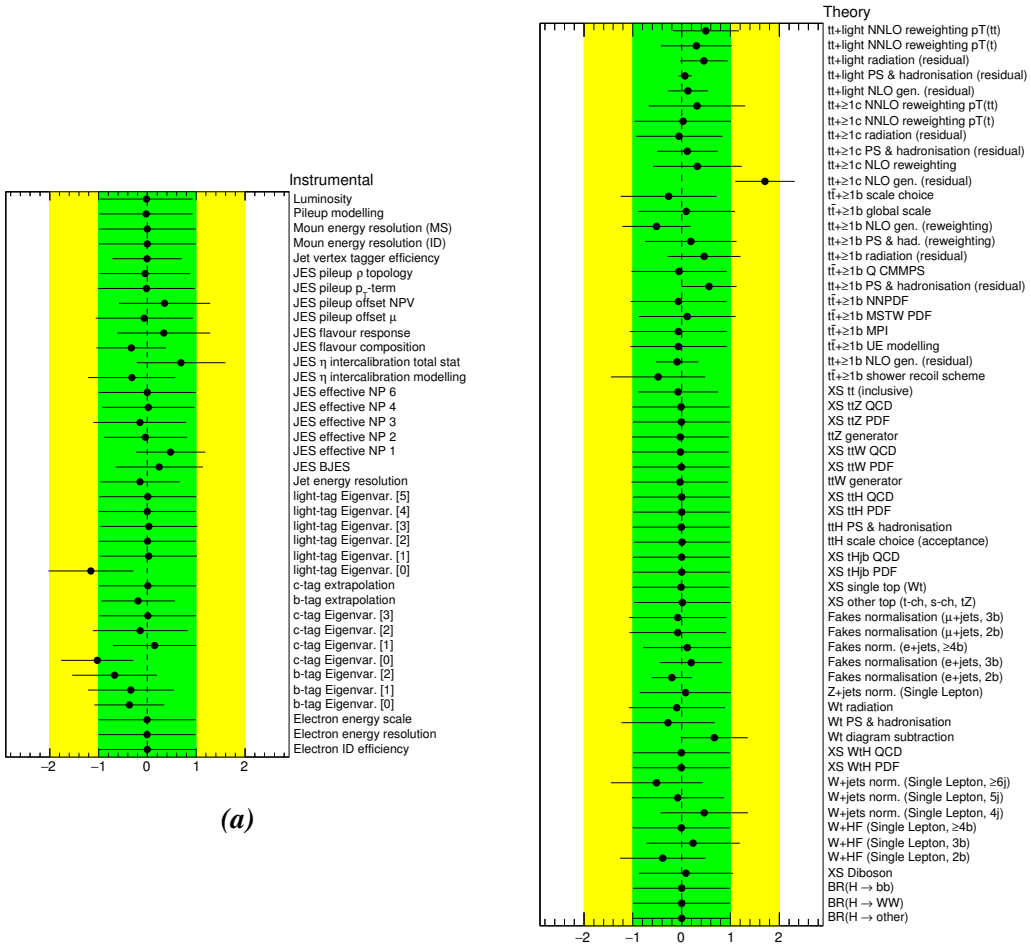


Figure 8.3.: Fitted NPs from a fit to the measured data sample using the ANN. (a) NPs associated to instrumental systematic uncertainties; (b) NPs associated to modelling systematic uncertainties.

μ due to the correlations among the systematic uncertainties. The largest impact comes from $t\bar{t}+ \geq 1b$ related systematic uncertainties.

Fig. 8.6 shows comparisons between data and prediction for the fitted distributions after the fit to data.

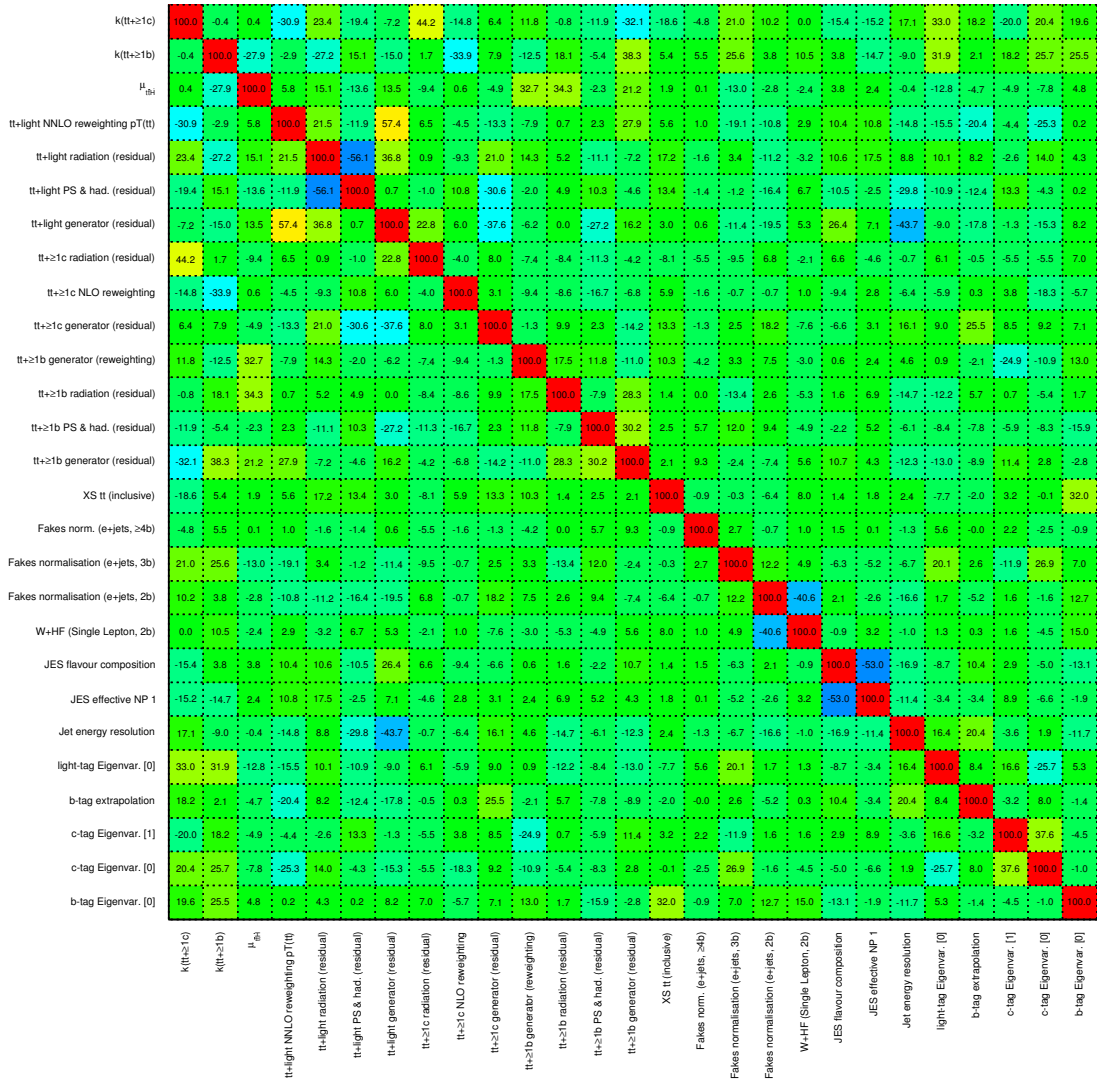


Figure 8.4.: The correlation matrix for the NPs obtained from the fit to data in the single lepton channel under the signal-plus-background hypothesis. Only NPs with a correlation coefficient of at least 30% with any other parameter are displayed.

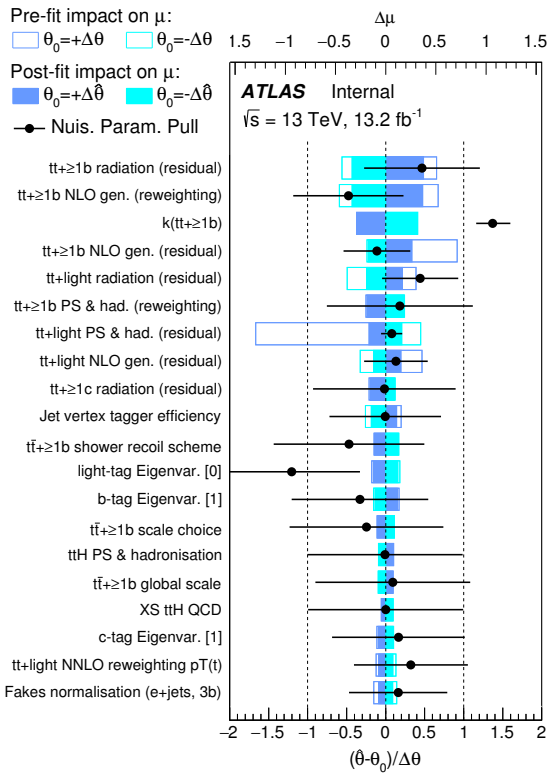


Figure 8.5.: Ranking of the NPs used in the fit according to their effect on the best fitted μ for the single lepton channel ANN fit to data. The top twenty parameters are shown. The empty blue rectangles correspond to the pre-fit impact while the filled blue ones to post-fit impact. The k is the normalisation factor (with respect to prediction) for the $t\bar{t}+\geq 1b$ component.

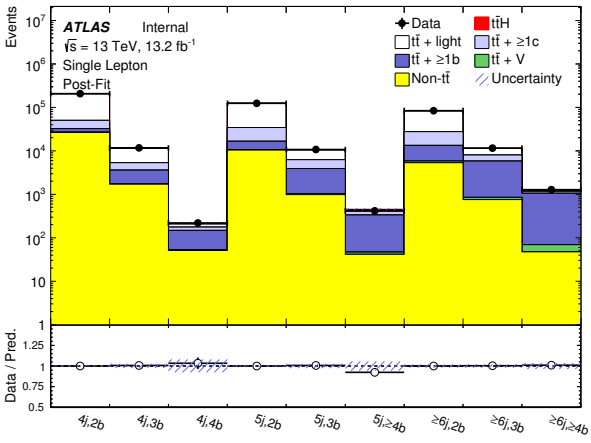


Figure 8.6.: Yield comparison between data and prediction after the fit to the measured data sample using the ANN.

8.2. BDT Results

As already discussed, both fits to the Asimov dataset and to data for the signal plus background hypothesis show similar constraints and pulls of the NPs parameters for BDT as shown for the ANN, therefore those are not shown. Differences in the fit to the data result is expected mainly due to the difference in shape and separation power. The shape comparison of the most important source of systematic variations for the two discriminants is presented in Appendix D.

8.2.1. Fit to Data and Limit on the $t\bar{t}H(H \to b\bar{b})$ Production

The result on data using the BDT is a signal strength of $\mu = 1.6^{+1.1}_{-1.1}$, a normalisation factor $k_{t\bar{t}+b\bar{b}} = 1.24^{+0.23}_{-0.21}$ for the $t\bar{t}+ \geq 1b$ background and $k_{t\bar{t}+c\bar{c}} = 1.37^{+0.70}_{-0.60}$ for the $t\bar{t}+ \geq 1c$. These results are similar to the one obtained with the ANN fit, but the errors are slighly reduced. A ranking plot of the twenty most important systematic uncertainties is available in Fig. 8.7.
A comparison between the fitted yields and data in each of the fitted regions is available in Fig. 8.8.
Fig. 8.9 shows the comparison between data and prediction for the distributions after the fit to data.
Since no significant excess of events above the background is visible for the SM Higgs boson with a mass of 125 GeV, a 95% CL upper limit on the signal strength is calculated. From the measurement, a signal cross section larger than 3.6 times the SM prediction is excluded. The expected exclusion limit in the hypothesis of absence of the $t\bar{t}H$ process is 2.2 times the SM prediction.

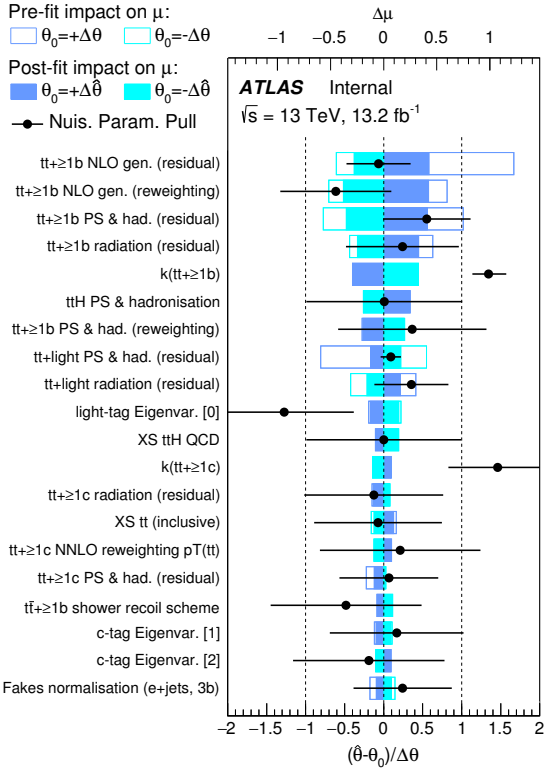


Figure 8.7.: Ranking of the NPs used in the fit according to their effect on the best fitted μ for the single lepton channel BDT fit to data. The top twenty parameters are shown. The empty blue rectangles correspond to the pre-fit impact while the filled blue ones to post-fit impact. The k are the normalisation factors (with respect to prediction) for the $t\bar{t}$ +HF (b and c) components.

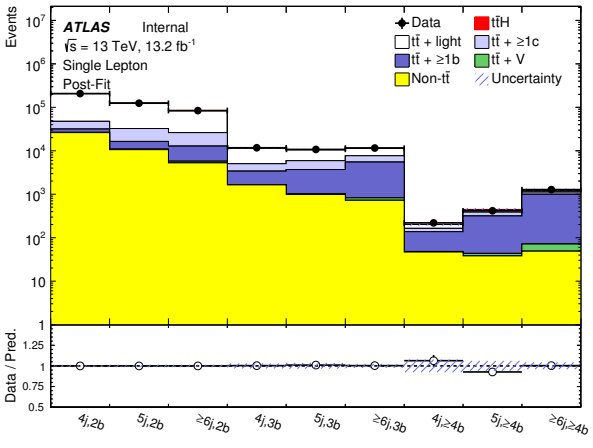


Figure 8.8.: Comparison between data and prediction after the fit to data using the BDT.

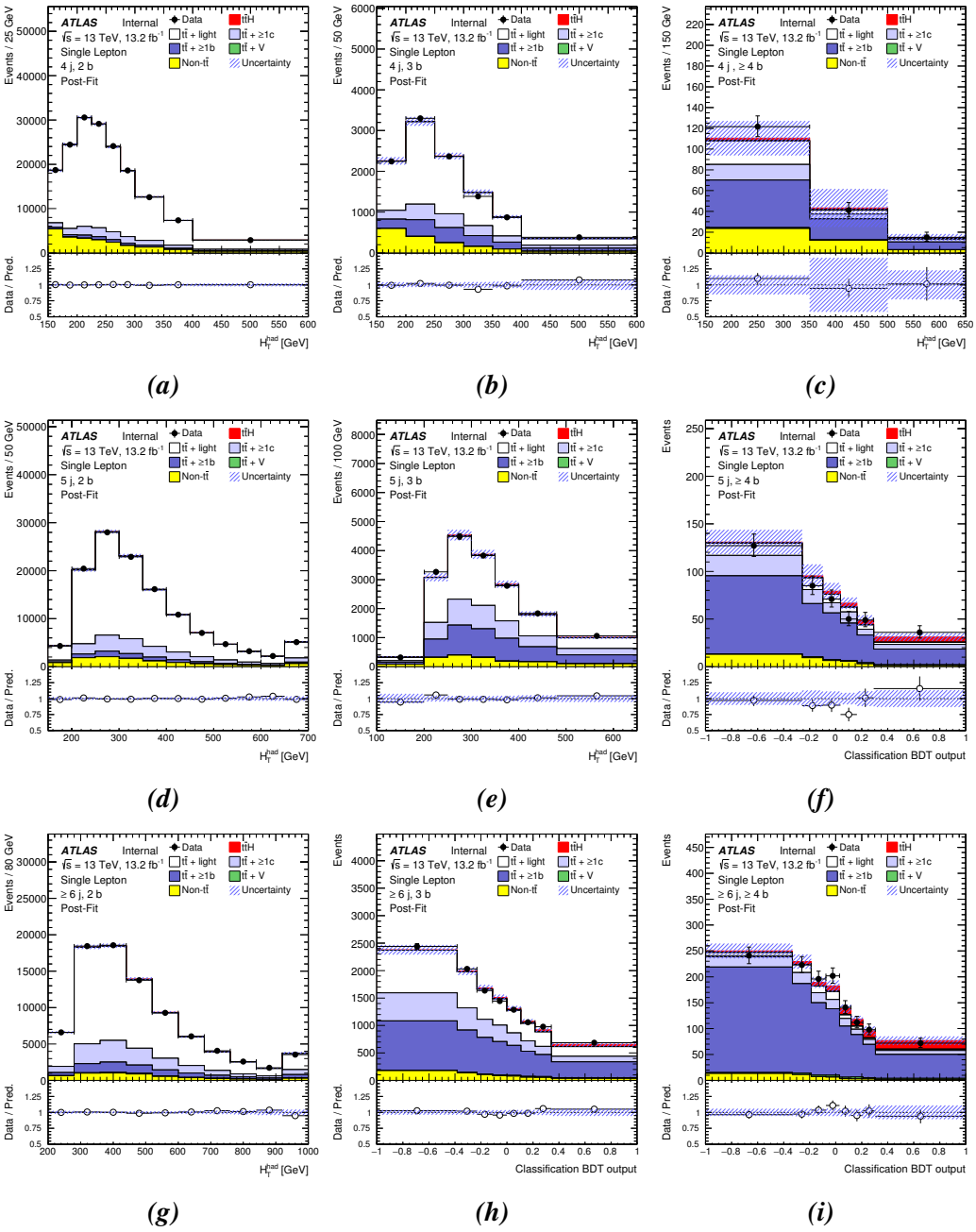


Figure 8.9.: (a) to (i), post-fit distributions of H_T^{had} and BDT output in the regions used for the fit to data using the BDT in the signal regions.

8.2.2. Combination with the Dilepton Analysis

A similar search of the $t\bar{t}H(H \to b\bar{b})$ production was performed in the opposite-charge lepton decay channel [61]. This analysis categorised events in five jet and b-tagged jet multiplicity regions, with an MVA technique implemented in the signal regions. Since the single lepton and dileptonic analyses are designed to be orthogonal and use the same set of systematic uncertainties, a combination of the two is performed through a simultaneous fit of the five regions of the dilepton analysis and the nine of the single lepton analysis. The fitted signal strength for the combined fit is $\mu = 2.1^{+1.0}_{-0.9}$, Fig. 8.10a, and a value of μ greater than 4.0 is excluded at 95% CL, Fig. 8.10b. The result corresponds to an observed significance of 2.4σ, while 1.2σ would be expected in the absence of the SM signal [148]. Yields at pre-fit and post-fit level are compared in Table 8.1, while post-fit yields of signal and total background per bin, ordered by $\log(S/B)$, are shown in Fig. 8.10c.

	4j,2b		4j,3b		4j,≥4b	
	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t}$ + light	160000 ± 30000	158800 ± 4800	5300 ± 1500	6300 ± 440	17 ± 11	36 ± 14
$t\bar{t}$+ ≥ 1c	10800 ± 2400	16800 ± 4100	880 ± 300	1680 ± 350	11.7 ± 5.4	24.4 ± 6.3
$t\bar{t}$+ ≥ 1b	4580 ± 930	5760 ± 980	1570 ± 470	1930 ± 320	76 ± 24	94 ± 13
$t\bar{t}$ +V	212 ± 27	218 ± 24	18.4 ± 3.8	20.4 ± 3.6	1.60 ± 0.42	1.73 ± 0.33
Single top	10300 ± 1700	10400 ± 1300	390 ± 87	476 ± 80	9.6 ± 3.5	12.9 ± 3.2
W/Z+jets	6500 ± 2400	7800 ± 2200	220 ± 100	410 ± 150	2.1 ± 1.2	2.6 ± 1.2
Diboson	420 ± 220	390 ± 190	15 ± 10	19 ± 11	3.9 ± 3.3	3.5 ± 3.0
Non-prompt	9200 ± 4200	7800 ± 1500	770 ± 360	770 ± 240	29 ± 27	23 ± 23
tH	9.3 ± 1.3	9.3 ± 1.2	4.41 ± 0.66	4.55 ± 0.57	0.62 ± 0.13	0.64 ± 0.10
Total background	202000 ± 32000	208000 ± 1900	9200 ± 1900	11610 ± 300	152 ± 44	199 ± 28
$t\bar{t}H$	63.8 ± 6.2	134 ± 42	24.6 ± 4.1	54 ± 21	3.32 ± 0.87	7.7 ± 2.9
Total	202000 ± 32000	208200 ± 1900	9200 ± 1900	11660 ± 300	155 ± 45	207 ± 28
Data	208239		11686		218	

	5j,2b		5j,3b		5j,≥ 4b	
	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t}$ + light	91000 ± 17000	91500 ± 3900	3640 ± 880	4580 ± 450	24 ± 15	45 ± 19
$t\bar{t}$+ ≥ 1c	10800 ± 2100	16600 ± 3800	1170 ± 330	2150 ± 410	30 ± 12	64 ± 11
$t\bar{t}$+ ≥ 1b	4440 ± 530	5760 ± 840	2230 ± 460	2830 ± 370	224 ± 62	278 ± 29
$t\bar{t}$ +V	277 ± 33	287 ± 30	35.3 ± 6.1	39.6 ± 5.9	4.9 ± 1.5	5.4 ± 1.4
Single top	4900 ± 1200	4790 ± 690	305 ± 87	338 ± 67	14.3 ± 5.6	16.1 ± 3.9
W/Z+jets	2700 ± 1100	2720 ± 780	200 ± 100	300 ± 120	3.0 ± 3.0	3.3 ± 2.3
Diboson	200 ± 110	210 ± 110	15.7 ± 9.7	16.0 ± 8.7	0.39 ± 0.28	0.43 ± 0.29
Non-prompt	3300 ± 1500	2800 ± 670	300 ± 150	300 ± 110	20 ± 17	16 ± 16
tH	7.4 ± 1.3	7.5 ± 1.3	3.88 ± 0.72	4.14 ± 0.69	0.82 ± 0.16	0.91 ± 0.14
Total background	117000 ± 20000	124600 ± 1400	7900 ± 1400	10560 ± 280	322 ± 78	429 ± 28
$t\bar{t}H$	96.5 ± 7.7	206 ± 61	49.7 ± 6.9	110 ± 42	11.8 ± 2.6	27 ± 10
Total	118000 ± 20000	124900 ± 1400	7900 ± 1400	10670 ± 280	333 ± 79	457 ± 27
Data	124688		10755		418	

	≥6j,2b		≥6j,3b		≥6j,≥4b	
	Pre-fit	Post-fit	Pre-fit	Post-fit	Pre-fit	Post-fit
$t\bar{t}$ + light	54000 ± 24000	58600 ± 4000	2600 ± 1100	3610 ± 500	34 ± 22	74 ± 32
$t\bar{t}$+ ≥ 1c	11500 ± 3700	12500 ± 5200	1550 ± 560	1960 ± 660	71 ± 37	91 ± 36
$t\bar{t}$+ ≥ 1b	4800 ± 1200	7180 ± 920	3240 ± 800	4830 ± 470	670 ± 190	955 ± 70
$t\bar{t}$ +V	470 ± 61	498 ± 49	86 ± 13	98 ± 10	19.1 ± 4.2	22.3 ± 3.5
Single top	2690 ± 840	2430 ± 400	278 ± 100	286 ± 65	29 ± 14	32 ± 12
W/Z+jets	1610 ± 660	1720 ± 520	121 ± 55	169 ± 65	11.9 ± 6.7	12.9 ± 6.4
Diboson	164 ± 88	166 ± 83	14.4 ± 8.3	15.8 ± 8.4	2.0 ± 1.3	2.1 ± 1.3
Non-prompt	1220 ± 560	1050 ± 310	270 ± 150	270 ± 120	1.2 ± 1.2	1.2 ± 1.2
tH	9.6 ± 2.4	9.9 ± 2.3	5.7 ± 1.5	6.2 ± 1.5	1.86 ± 0.53	2.10 ± 0.50
Total background	77000 ± 26000	84200 ± 1400	8200 ± 1900	11250 ± 240	840 ± 230	1191 ± 55
$t\bar{t}H$	198 ± 18	430 ± 120	119 ± 16	261 ± 99	44.9 ± 9.4	107 ± 39
Total	77000 ± 26000	84600 ± 1400	8300 ± 1900	11520 ± 220	890 ± 230	1298 ± 41
Data	84556		11561		1285	

Table 8.1.: Event yields in the single lepton channel. Post-fit yields refer to the combined fit in dilepton and single-lepton channels to data. The uncertainties are the sum in quadrature of statistical and systematic uncertainties. In the pre-fit case, errors do not consider any uncertainty on the $t\bar{t}+ \geq 1b$ or $t\bar{t}+ \geq 1c$ normalisation. In the post-fit case, these errors are computed accounting correlations among nuisance parameters and processes, including the errors on the determination of the $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ normalisation. Concerning $t\bar{t}H$ signal, the pre-fit yield values consider both theoretical prediction and corresponding uncertainties, while the post-fit yield and errors are obtained from the signal strength measurement.

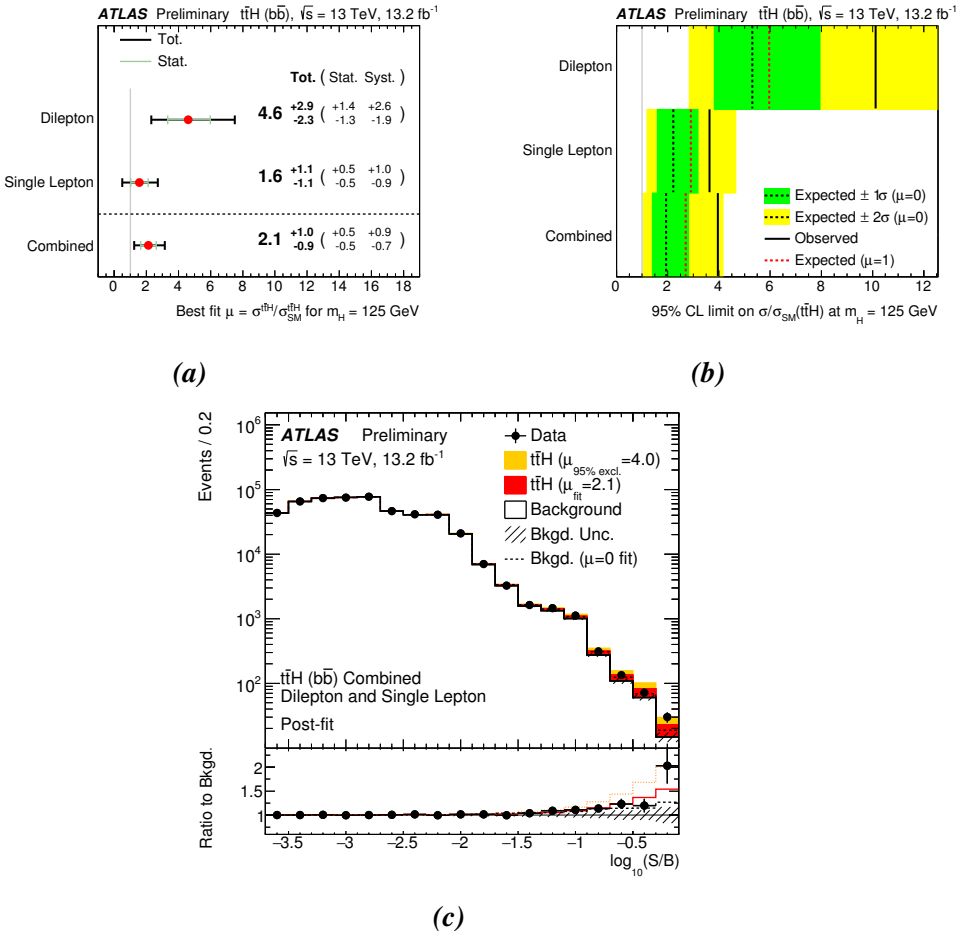


Figure 8.10.: Summary plots of the results of the single lepton, dilepton and combination analyses. (a) Summary of the signal strength measurements in the individual channels and for the combination. (b) Summary of the of 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction in the individual channels and for the combination. (c) Post-fit yields of signal and total background per bin, ordered by $\log(S/B)$, for all bins used in the combined fit of the single lepton and dilepton channels. The signal is shown normalised to the best-fit value and to the excluded value. The background is also shown after the fit to data assuming zero signal contribution [61].

9 Conclusions and Outlook

9.1. Summary and Conclusions

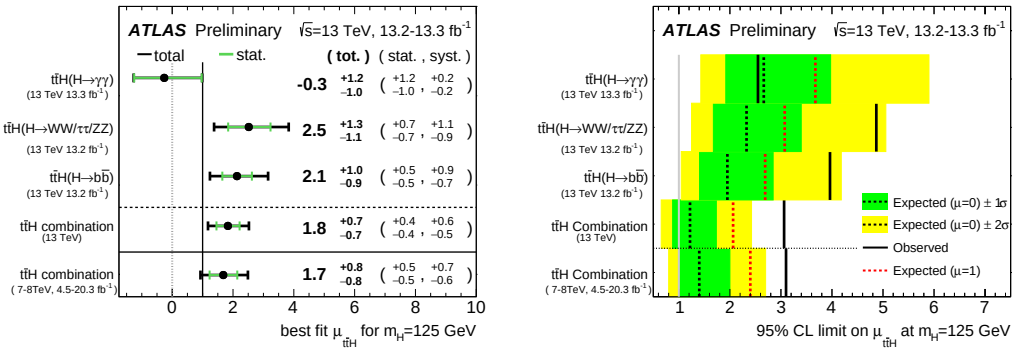
At present, the SM of particle physics is the most successful theoretical framework for understanding the infinitesimal components of the Universe and their interaction. However, it alone cannot express a large fraction of the known phenomena and observations, giving the hint that a more complete theory has still to be discovered. One of the greatest successes of the theory is the prediction of the existence of the Higgs boson, a particle needed within the electroweak symmetry breaking mechanism. A particle compatible with the properties of such particle has been discovered in 2012 at the largest laboratory of the world: CERN. The LHC accelerator at CERN has been employed to perform such discoveries, together with the ATLAS and CMS experiments. The discovery of such a particle paved the way to a large collection of new searches at CERN, since measuring the properties of the Higgs boson can give a direction for the search of new physics with the goal of extending the SM or finding new theories to cover as many unknown phenomena as possible. A great emphasis has been placed to the measurement of the couplings of the Higgs boson with other particles, which are measurable through the production and decay rates of the Higgs boson. The $t\bar{t}H$ production channel with the decay of the Higgs boson into bottom quarks provides a measurement of the coupling of the Higgs boson to heavy quarks both in the production and decay. In particular, the production of the Higgs boson through this channel is of particular interest since it gives access to the direct measurement of the Yukawa coupling to the top quark, which is close to unity in the SM. Possible measurements that show deviations from the predicted couplings will be a hint of new physics.

This thesis presents the search of the $t\bar{t}H(H \to b\bar{b})$ production in the single lepton chan-

nel using the data recorded by the ATLAS experiment at $\sqrt{s} = 13$ TeV in 2015 and up to July 2016, for a total of 13.2 fb^{-1}. The main source of background is the $t\bar{t}$ +jets process, where in particular the $t\bar{t} + b\bar{b}$ component is an irreducible background since it has the same final state as the $t\bar{t}H$ signal. In order to enhance the sensitivity of the analysis and constrain as much as possible the systematic uncertainties, the selected events are categorised in nine regions, defined through the multiplicity of jets and b-tagged jets. Since the signal yields are still very low compared to the background even in the signal rich regions, a sophisticated analysis based on the MVA techniques is developed. In particular, an ANN is developed for separating the signal from the background in the signal richest regions. This ANN is developed in parallel with a BDT discriminant, and the ANN is used as an internal cross check. A likelihood fit is performed simultaneously in the nine regions. The obtained result shows a significant decrease of the systematic uncertainties due to the constraining power of data and by correlations among the different systematic uncertainties introduced by the fit. A combination with the dilepton analysis is performed and the final measured signal strength is $\mu = 2.1^{+1.0}_{-0.9}$. A value greater than 4.0 times the SM prediction is excluded at 95% CL. The result corresponds to an observed significance of 2.4σ, while 1.2σ would be expected in the absence of the SM signal.

9.2. Comparison with Other Searches

Searches of the $t\bar{t}H$ process have also been performed in the ATLAS experiment in the multilepton [149] and diphoton [150] final states. The 95% CL for the signal strength measured by the two analyses has been found to be 4.9 and 2.6, respectively. The most sensitive result is achieved by combining these two analyses with the $t\bar{t}H(H \to b\bar{b})$ one [148]. The combined signal strength is $1.8^{+1.8}_{-0.7}$, which corresponds to an observed significance of 2.8σ. The sensitivity of this combination already exceeds the one obtained from the combination of the Run 1 $t\bar{t}H$ analyses [32]. The signal strength and the limits on μ obtained from the analyses listed above are summarised in Fig. 9.1.
Similar searches for the $t\bar{t}H$ process in several decay channels have been performed by the CMS collaboration using data collected at $\sqrt{s} = 13$ TeV collision energy [151–153]. The observed 95% CL upper limit on the $t\bar{t}H$ cross section is 1.5 times larger than the SM prediction, while the best fitted signal strength for the combined single lepton and dilepton channel is -0.2 ± 0.8. The CMS analysis uses a simplified treatment of the systematic uncertainties compared to ATLAS, and as a result the obtained errors on μ are reduced.



(a) *(b)*

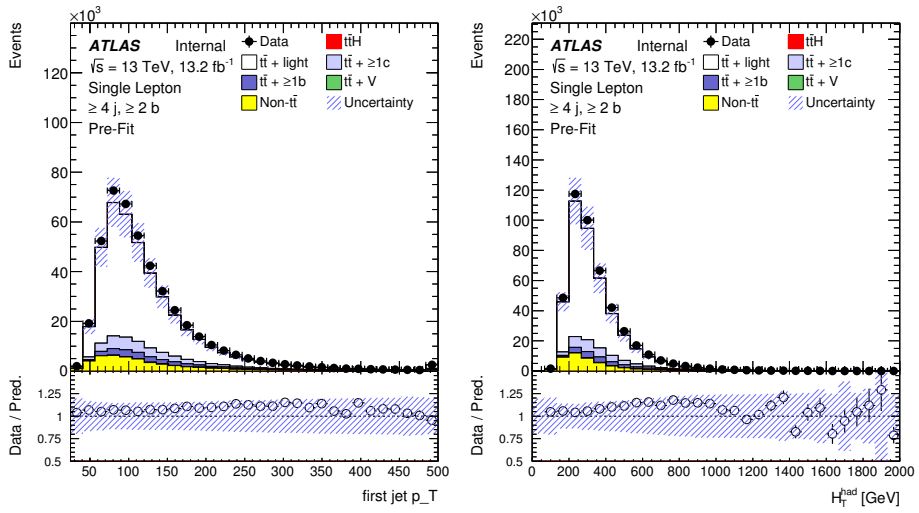
Figure 9.1.: (a) Summary of the observed signal strength measurements from the individual analyses and for their combination. (b) Upper limits on the $t\bar{t}H$ signal strength for the individual analyses as well as their combination at 95% CL.

9.3. Future Improvements

The sensitivity of the analysis presented in this thesis is limited by the amount of data collected and by the systematic uncertainties. It is possible to make an estimation of the needed luminosity for achieving a 3σ measurement of the $t\bar{t}H$ cross section for the single lepton channel at 13 TeV. If systematic uncertainties are kept the same, a data set of at least 100 fb^{-1} is needed. This number for the combined result is considerably smaller. Given that by the end of 2016 the collected dataset is about 36 fb^{-1}, this goal will be achieved in the near future. One challenge is to improve the reconstruction MVA in order to perform the best final state reconstruction, especially considering also boosted regimes, where the combinatorial background might be reduced [154]. Another possible improvement of the MVA techniques come from the usage of the so called *continuous b-tagging* technique, which exploits the shape of the MV2c *b*-tagging discriminant. The finer jet differentiation achievable using continuous *b*-tagging, permits to classify them into five different classes according to the number of calibrated working points with which they are tagged, while in the cumulative *b*-tagging technique just two classes (*b*-tagged or not *b*-tagged) are used. This technique gives the possibility of defining extra pure signal regions and regions enriched in different types of background. Such categorisation of analyses significantly improves their sensitivity compared to the cumulative *b*-tagging cut one. The reconstruction of the final state can be improved developing more sophisticated techniques, such as the matrix-element method which has already been used for the Run 1 analysis [31]. Another necessary improvement requires a better modelling of the $t\bar{t}$ +jets background, and in particular a more precise modelling and reduced systematic uncertainties in the $t\bar{t}$+HF background.

A Plots of Basic Distributions

In this appendix some control plots of basic distributions are shown. The inclusive region (≥4j,≥2b) is considered. Plots for the p_T of the first jet and of the lepton, the η of the lepton, E_T^{miss}, transverse mass of the W boson, the scalar sum of the p_T of all the jets in the final event (H_T^{had}), jet multiplicity and b-tag multiplicity are shown. Some discrepancies between data and prediction can be seen, but those are covered by the statistical and systematic uncertainties which are displayed by the hashed area.



(a) ***(b)***

Figure A.1.: Plots of basic distributions: (a) p_T of the first jet, (b) H_T^{had}

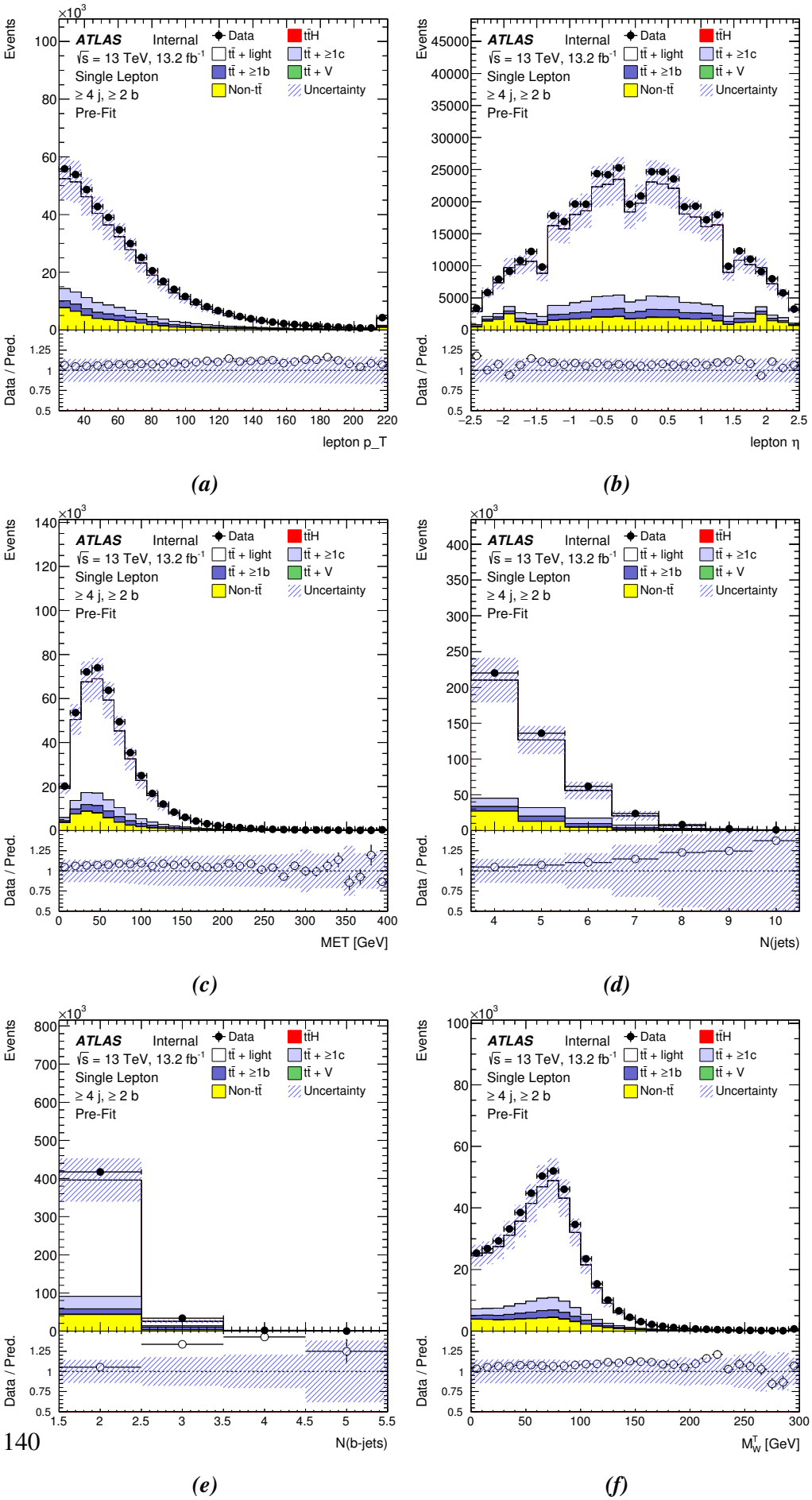


Figure A.2.: Plots of basic distributions: (a) p_T of the lepton, (b) η of the lepton, (c) E_T^{miss}, (d) jets multiplicity, (e) b-tag multiplicity, (f) transverse mass of the W boson

B

Plots of input variables used for the ANN training

In this appendix, the plots of the distributions of the input variables used for the training of the ANN in the three considered signal regions, Figs.B.1 to B.12. As a recap, the used variables are listed in Table B.1. The plots show pre-fit distributions, normalised overlayed plots for signal and background to evaluate their separation, and finally the post-fit distributions. The latter show a good agreement between data and prediction a part of some statistical fluctuations.

(≥6j,≥4b)	(≥6j,3b)	(5j,≥ 4b)
$\Delta R_{bb}^{\text{avg}}$	$p_T^{\text{jet}_5}$	$\Delta R_{bb}^{\text{avg}}$
$\text{M}_{bb}^{\text{min}\Delta R}$	$\text{Centrality}_{\text{all}}$	$\text{M}_{bb}^{\text{min}\Delta R}$
$\text{Centrality}_{\text{all}}$	$H1_{\text{all}}$	$\text{Centrality}_{\text{all}}$
$H1_{\text{all}}$	H_T^{Had}	$H1_{\text{all}}$
$p_T^{\text{jet}_5}$	$\text{Aplan}_{\text{jets}}$	$p_T^{\text{jet}_5}$
N_{30}^{Higgs}	$\Delta R_{bb}^{\text{avg}}$	N_{30}^{Higgs}
H_T	N_{30}^{Higgs}	H_T
$\Delta\eta_{jj}^{\text{max}\Delta\eta}$	$\text{M}_{bj}^{\text{max}p_T}$	$\Delta\eta_{jj}^{\text{max}\Delta\eta}$
$\text{M}_{jj}^{\text{minM}}$	$\text{M}_{bb}^{\text{min}\Delta R}$	$\text{M}_{jj}^{\text{minM}}$
$\Delta R_{l,bb}^{\text{min}\Delta R}$	$\Delta R_{uu}^{\text{min}\Delta R}$	$\Delta R_{l,bb}^{\text{min}\Delta R}$
$\text{Aplan}_{b\text{-jets}}$	$\text{M}_{uu}^{\text{min}\Delta R}$	$\text{Aplan}_{b\text{-jets}}$
$\Delta R_{bb}^{\text{max}p_T}$	$\text{M}_{jj}^{\text{max}p_T}$	$\Delta R_{bb}^{\text{max}p_T}$

Table B.1.: Recap on the input variables used for the ANN in each of the signal regions.

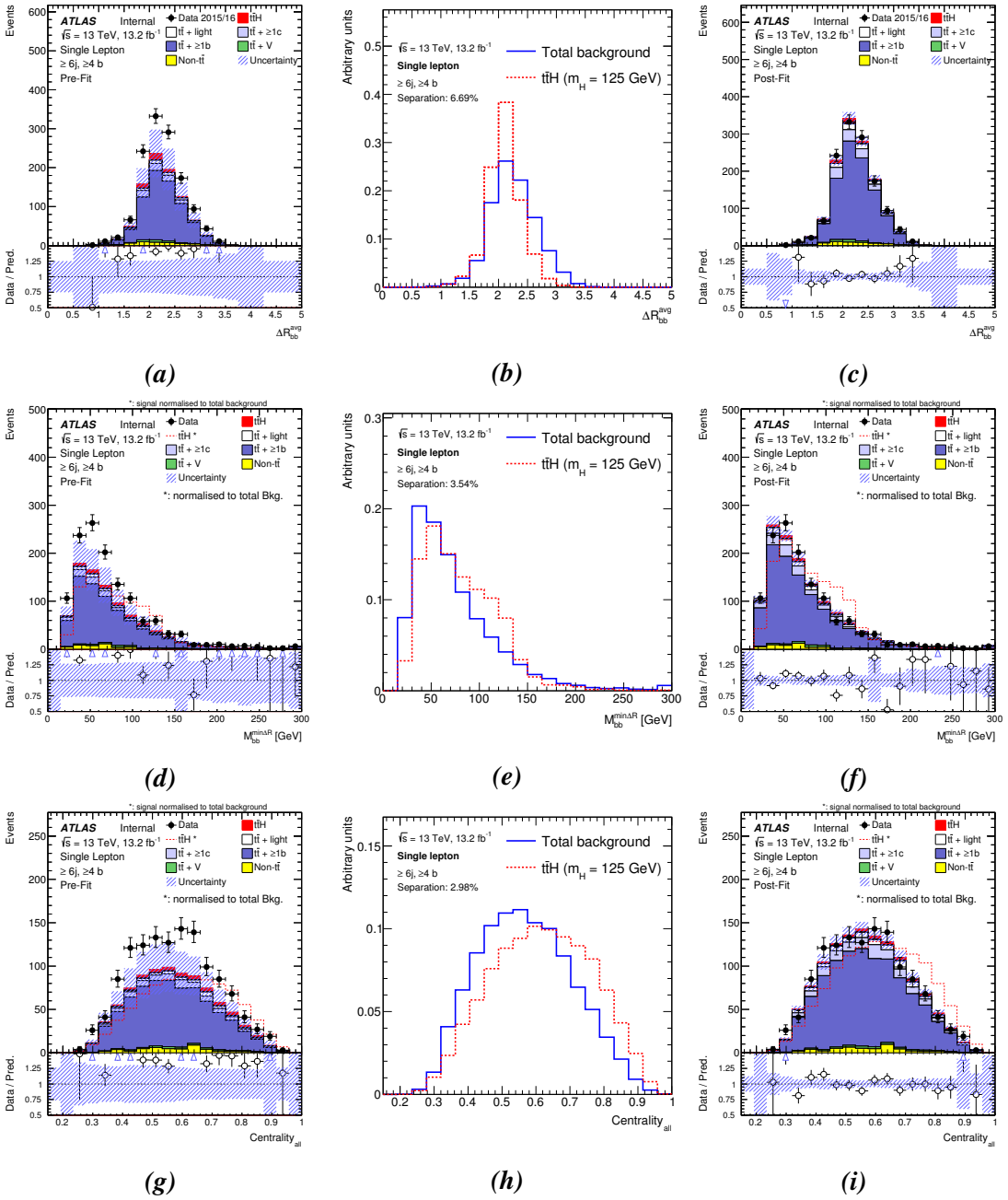


Figure B.1.: Input variables used for training the ANN in the (≥ 6j,≥4b) region. Variables from 1 to 3. The plots show pre-fit distributions, separation plots, post-fit distributions.

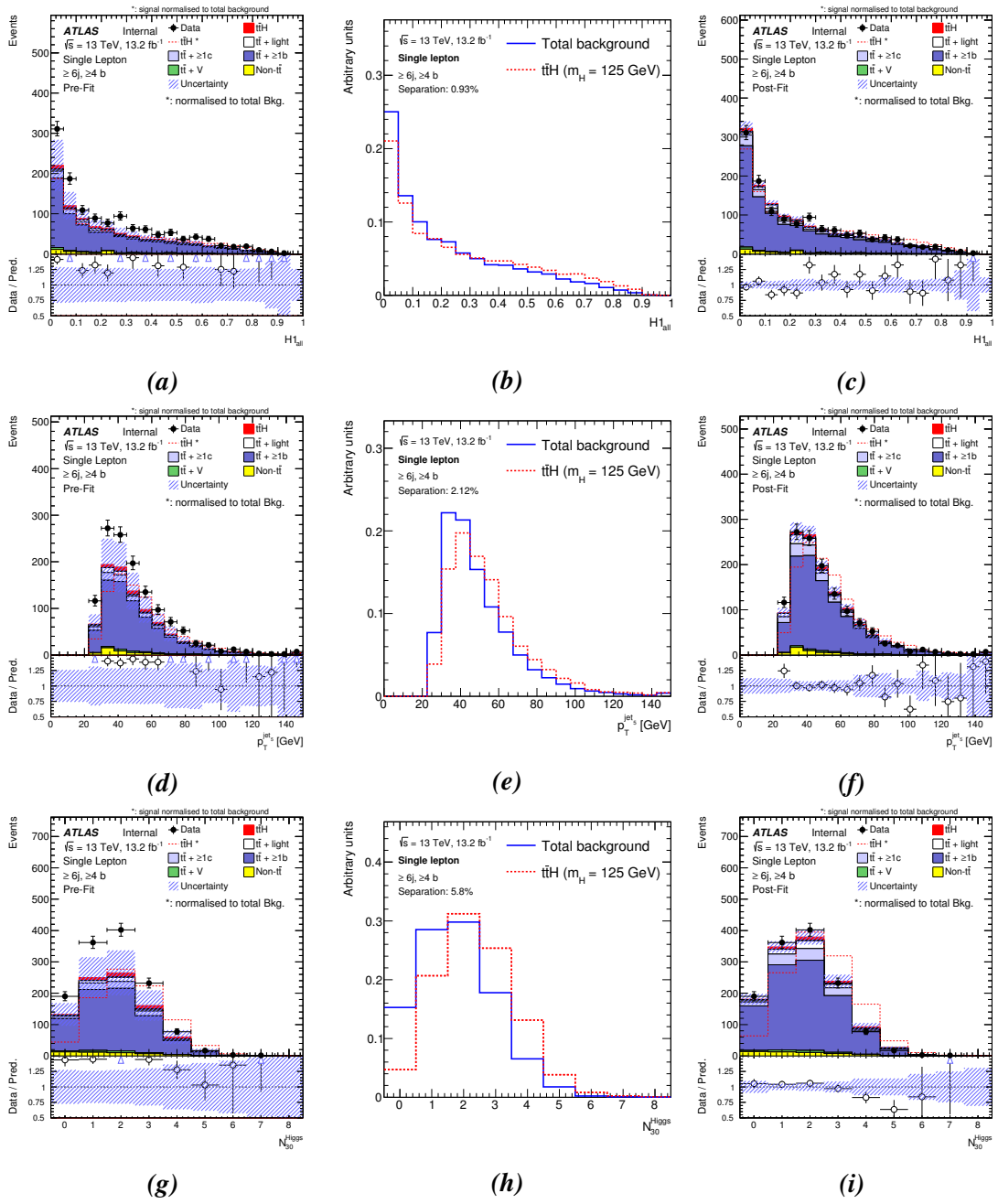


Figure B.2.: Input variables used for training the ANN in the (≥ 6j,≥4b) region. Variables from 4 to 6. The plots show pre-fit distributions, separation plots, post-fit distributions.

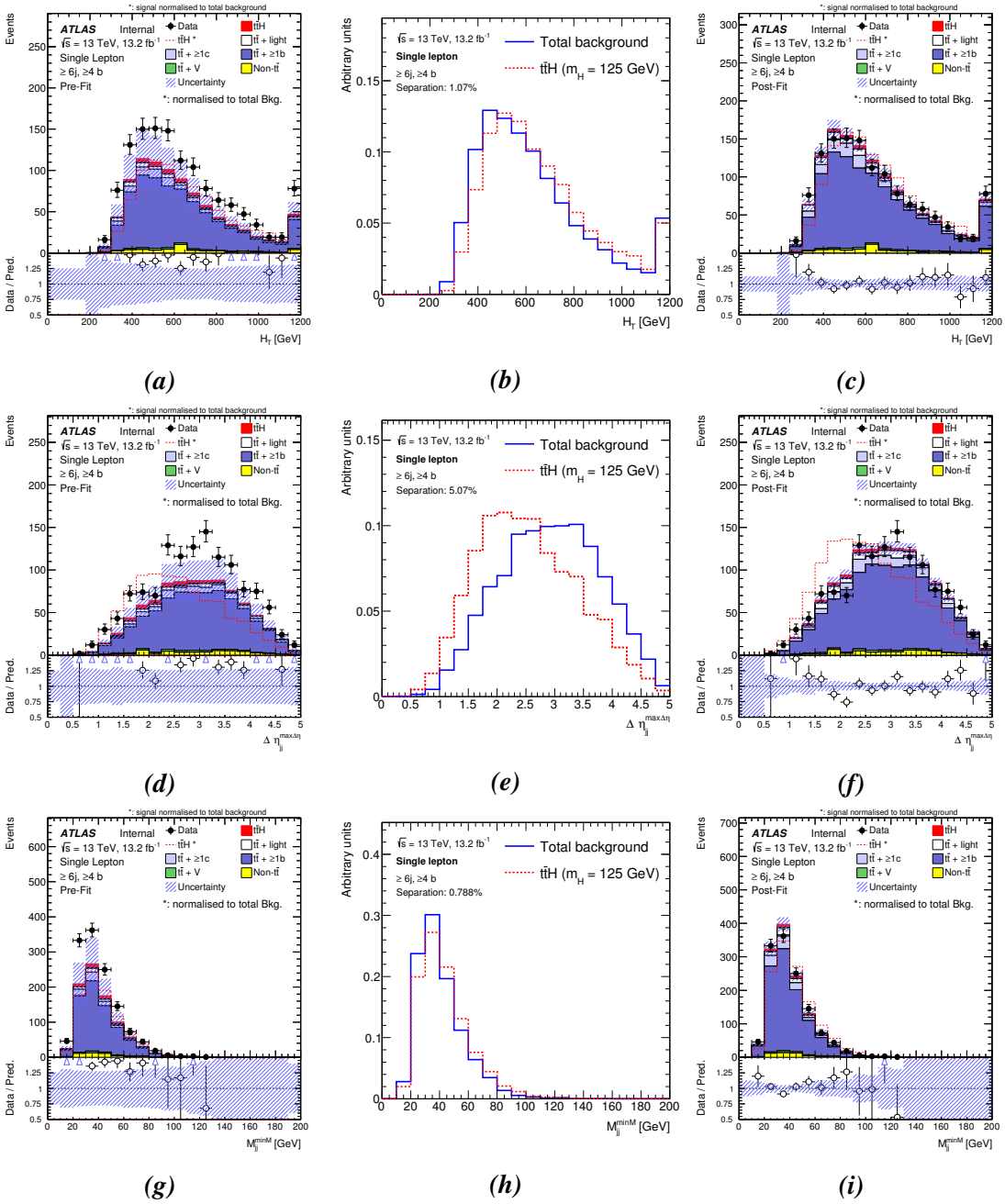


Figure B.3.: Input variables used for training the ANN in the (≥6j,≥4b) region. Variables from 7 to 9. The plots show pre-fit distributions, separation plots, post-fit distributions.

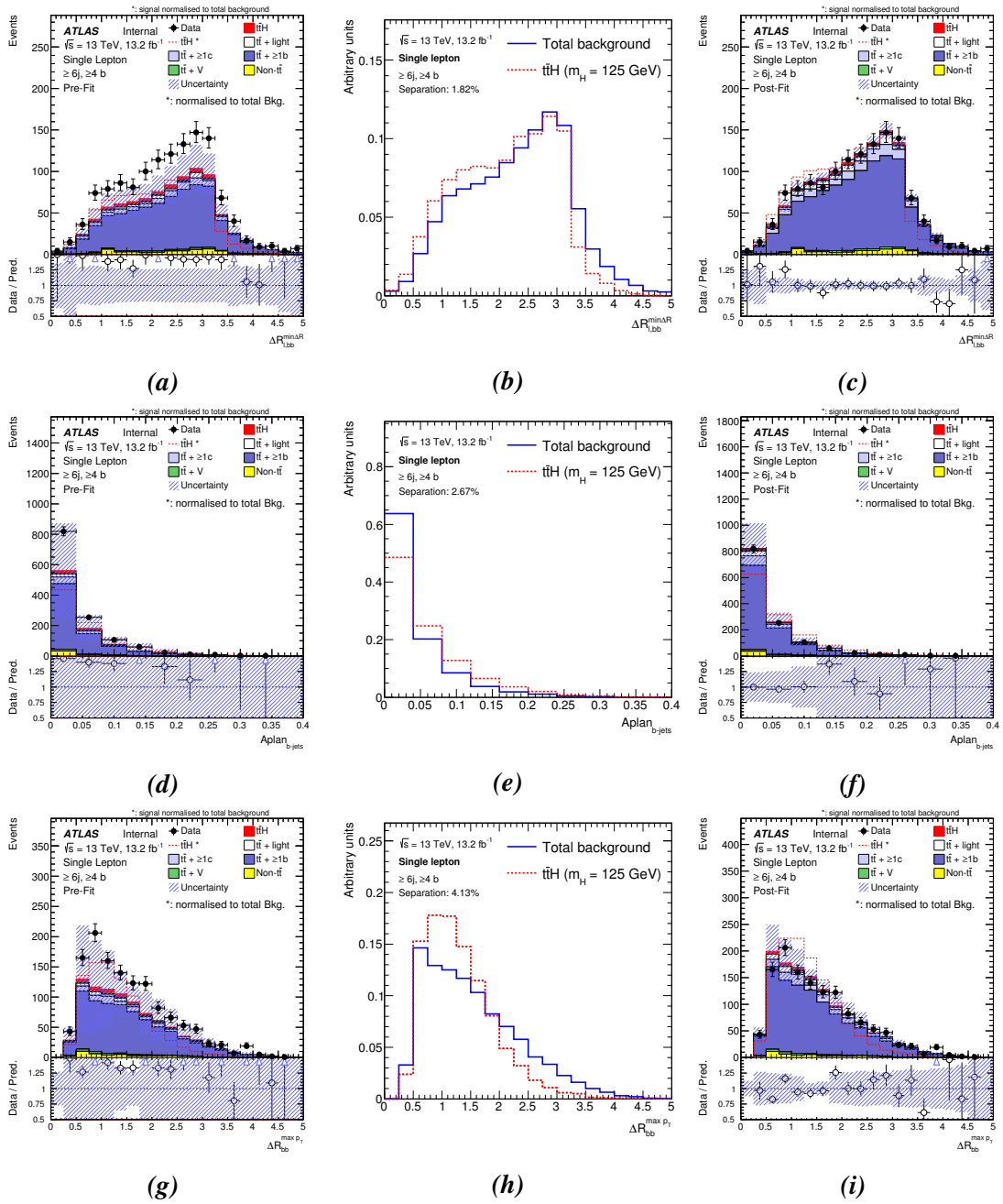


Figure B.4.: Input variables used for training the ANN in the (≥ 6j,≥ 4b) region. Variables from 10 to 12. The plots show pre-fit distributions, separation plots, post-fit distributions.

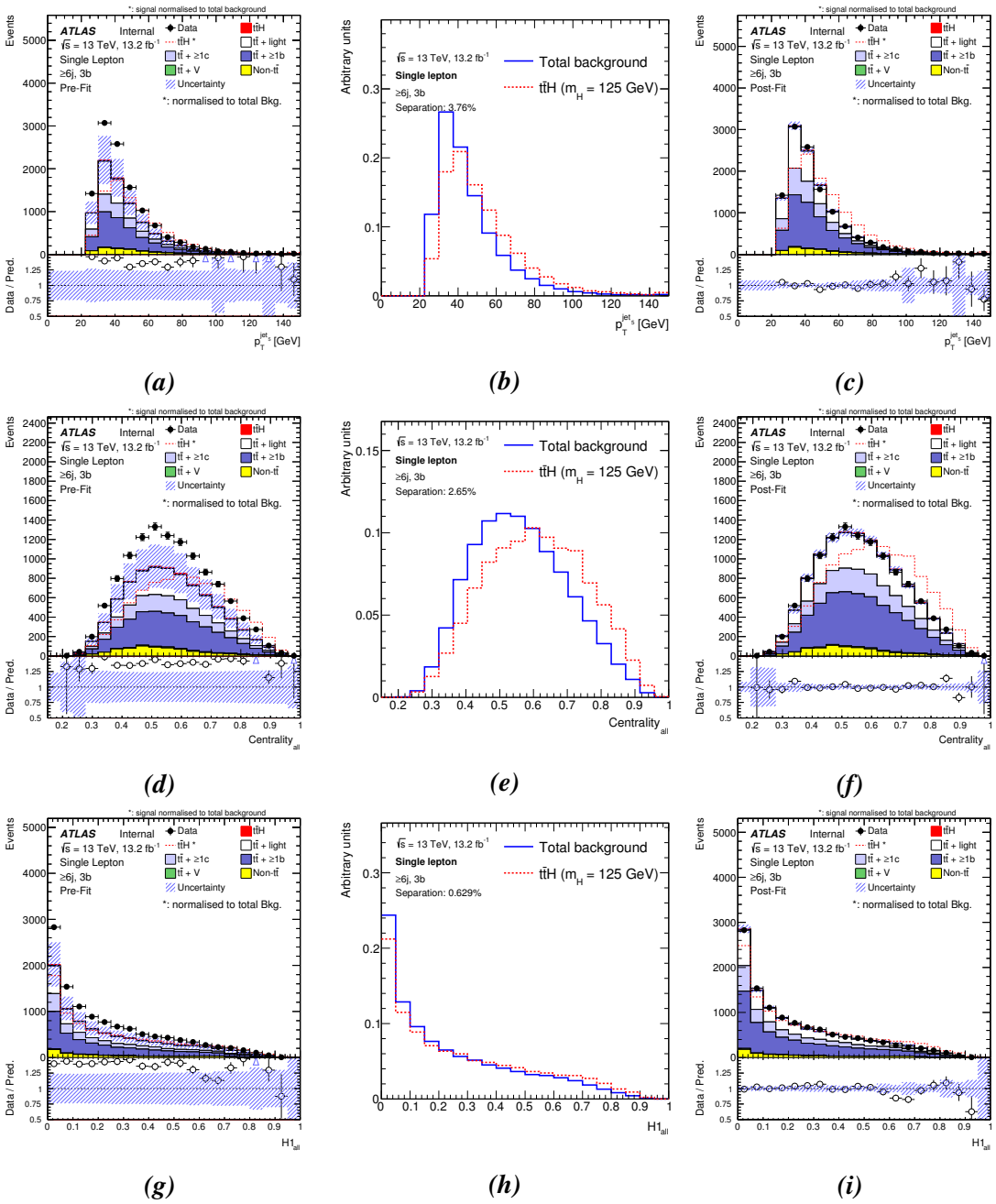


Figure B.5.: Input variables used for training the ANN in the (≥ 6j,3b) region. Variables from 1 to 3. The plots show pre-fit distributions, separation plots, post-fit distributions.

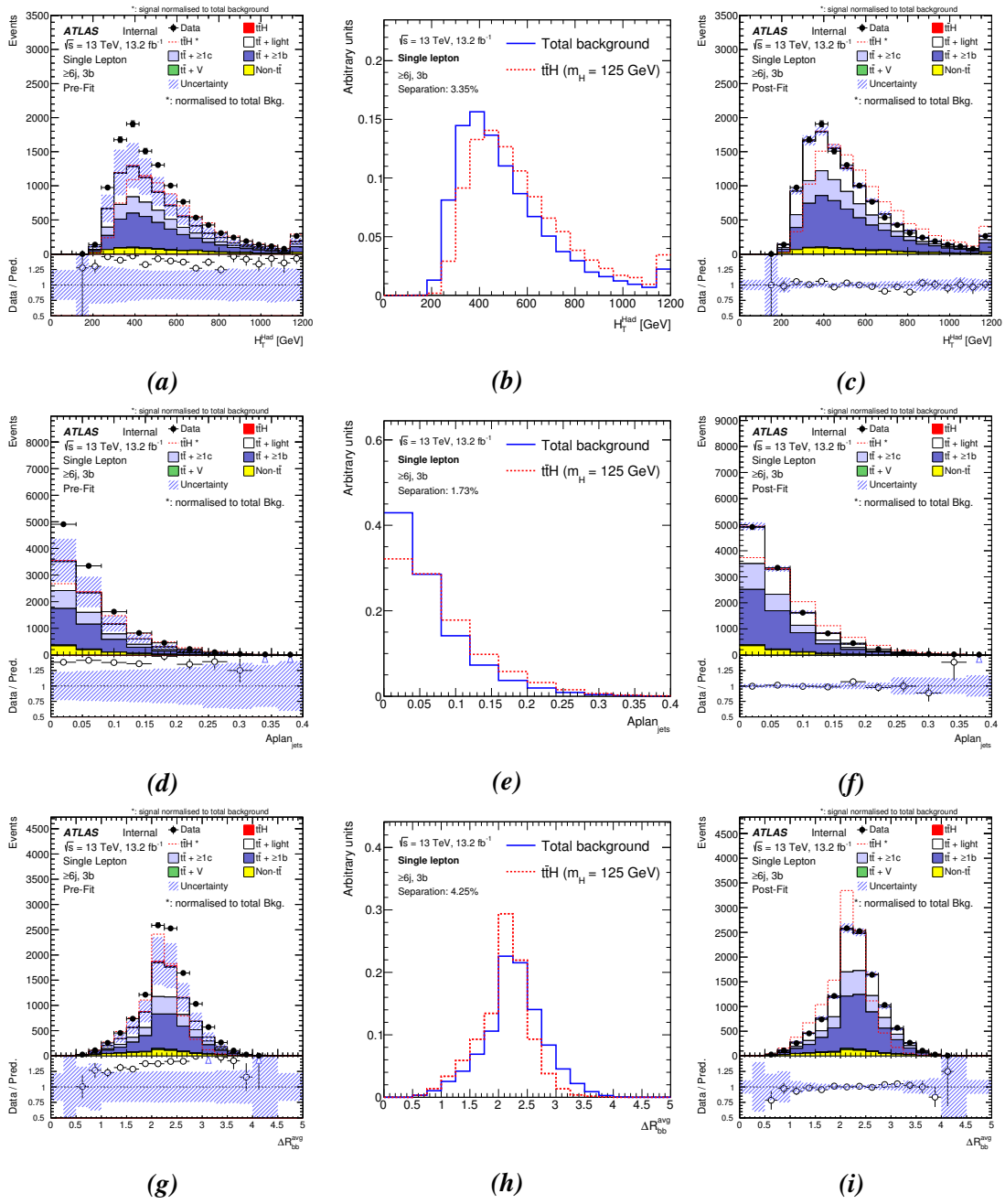


Figure B.6.: Input variables used for training the ANN in the (≥ 6j,3b) region. Variables from 4 to 6. The plots show pre-fit distributions, separation plots, post-fit distributions.

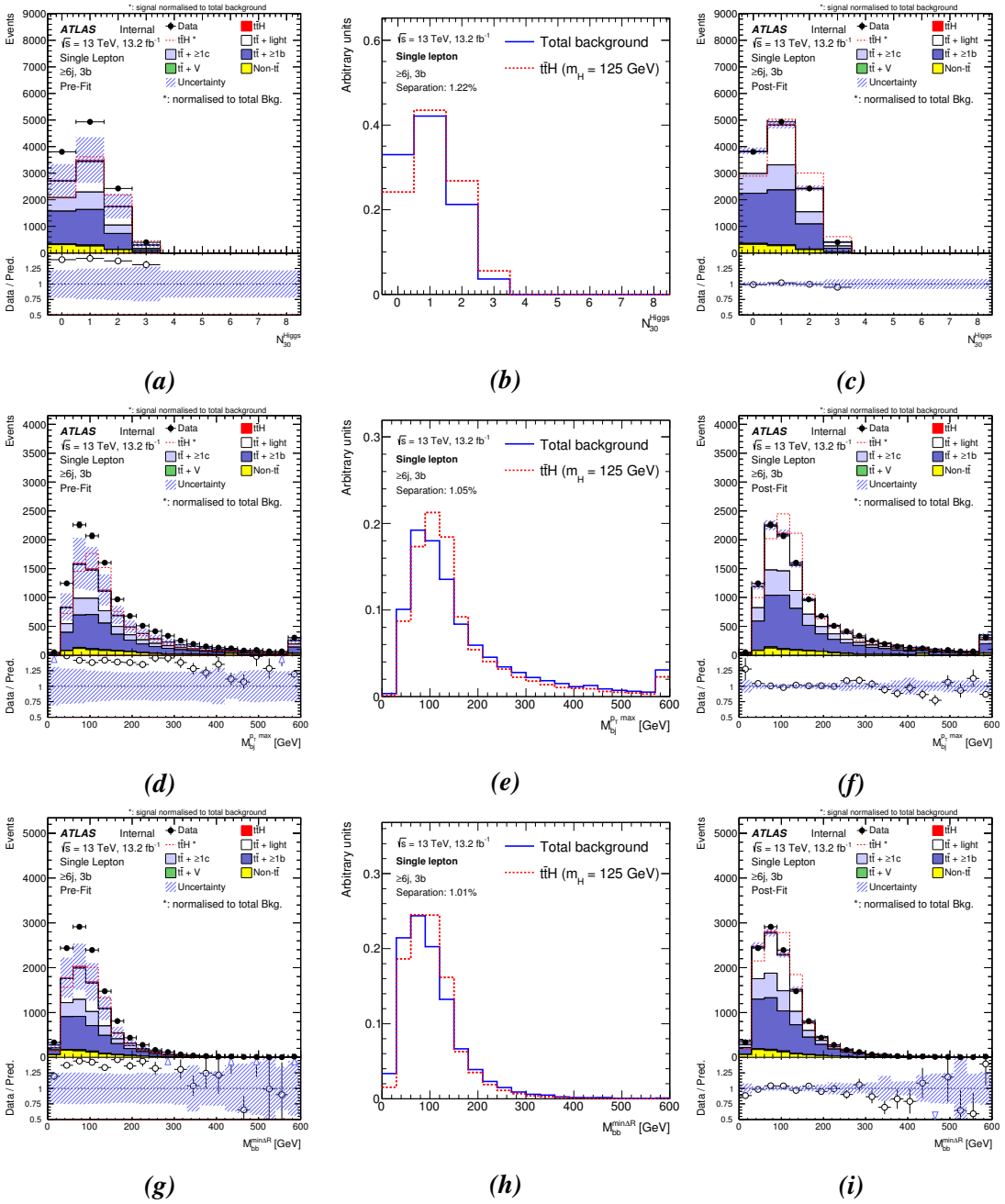


Figure B.7.: Input variables used for training the ANN in the (≥ 6j,3b) region. Variables from 7 to 9. The plots show pre-fit distributions, separation plots, post-fit distributions.

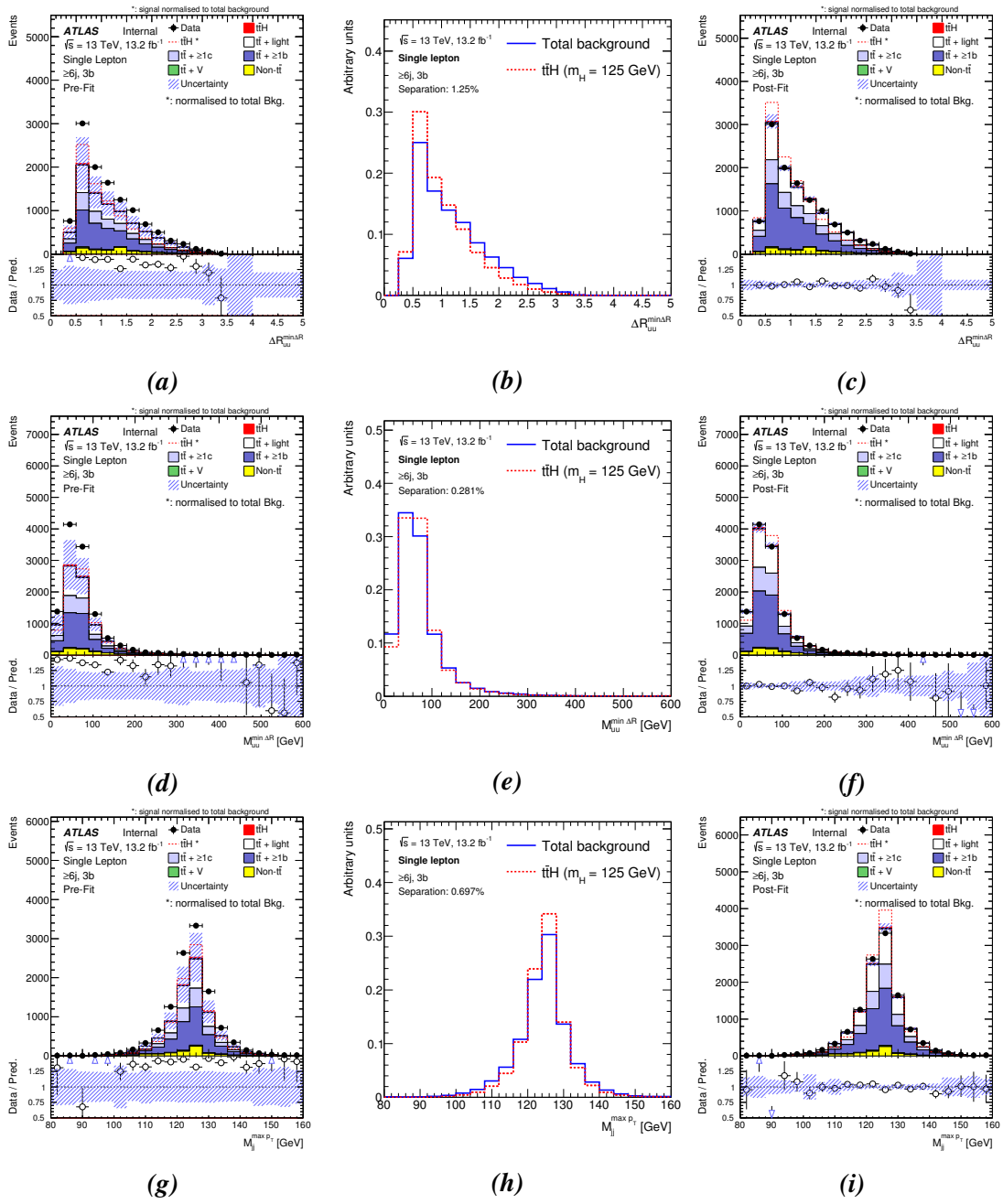


Figure B.8.: Input variables used for training the ANN in the (≥ 6j,3b) region. Variables from 10 to 12. The plots show pre-fit distributions, separation plots, post-fit distributions.

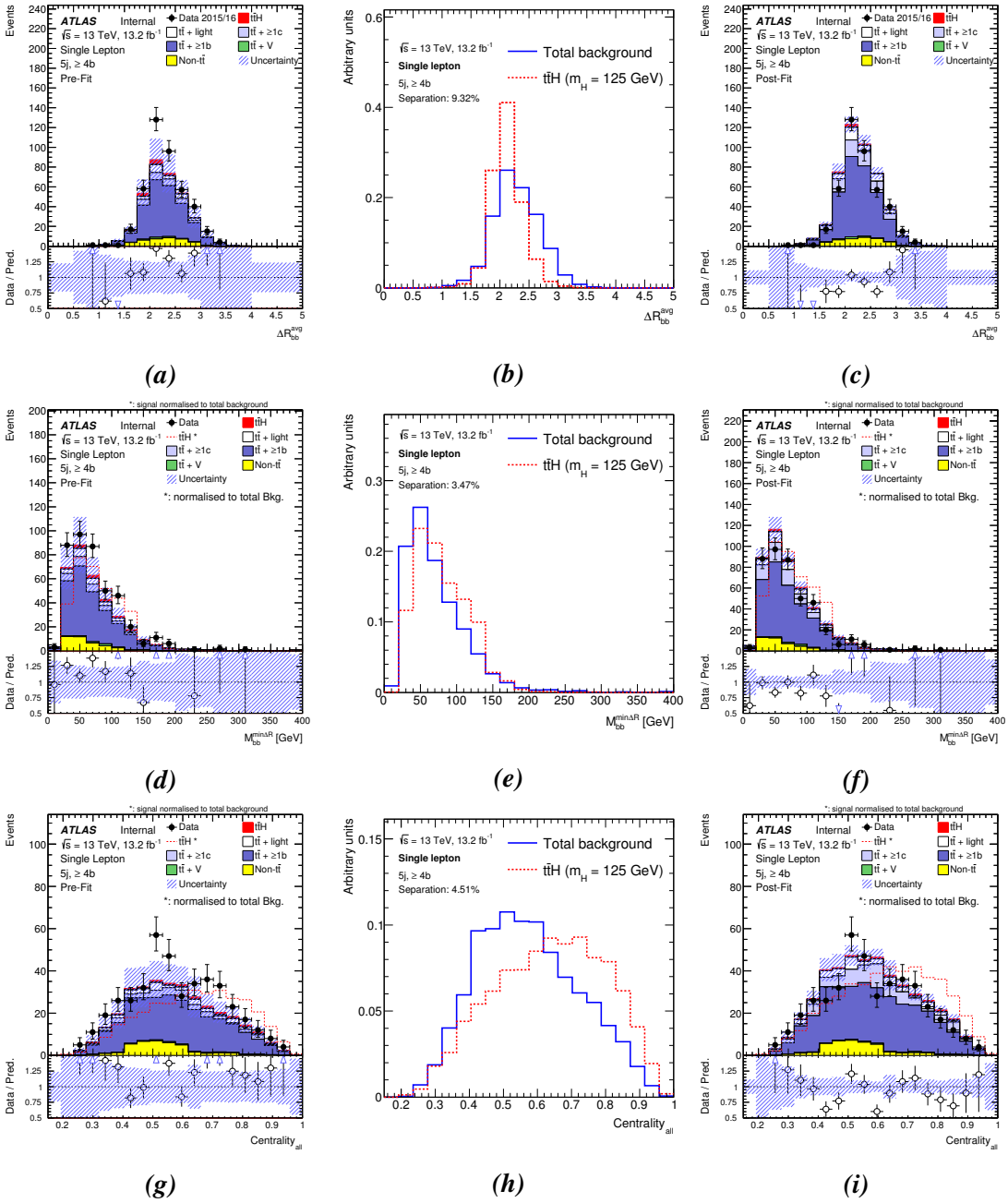


Figure B.9.: Input variables used for training the ANN in the (5j,≥4b) region. Variables from 1 to 3. The plots show pre-fit distributions, separation plots, post-fit distributions.

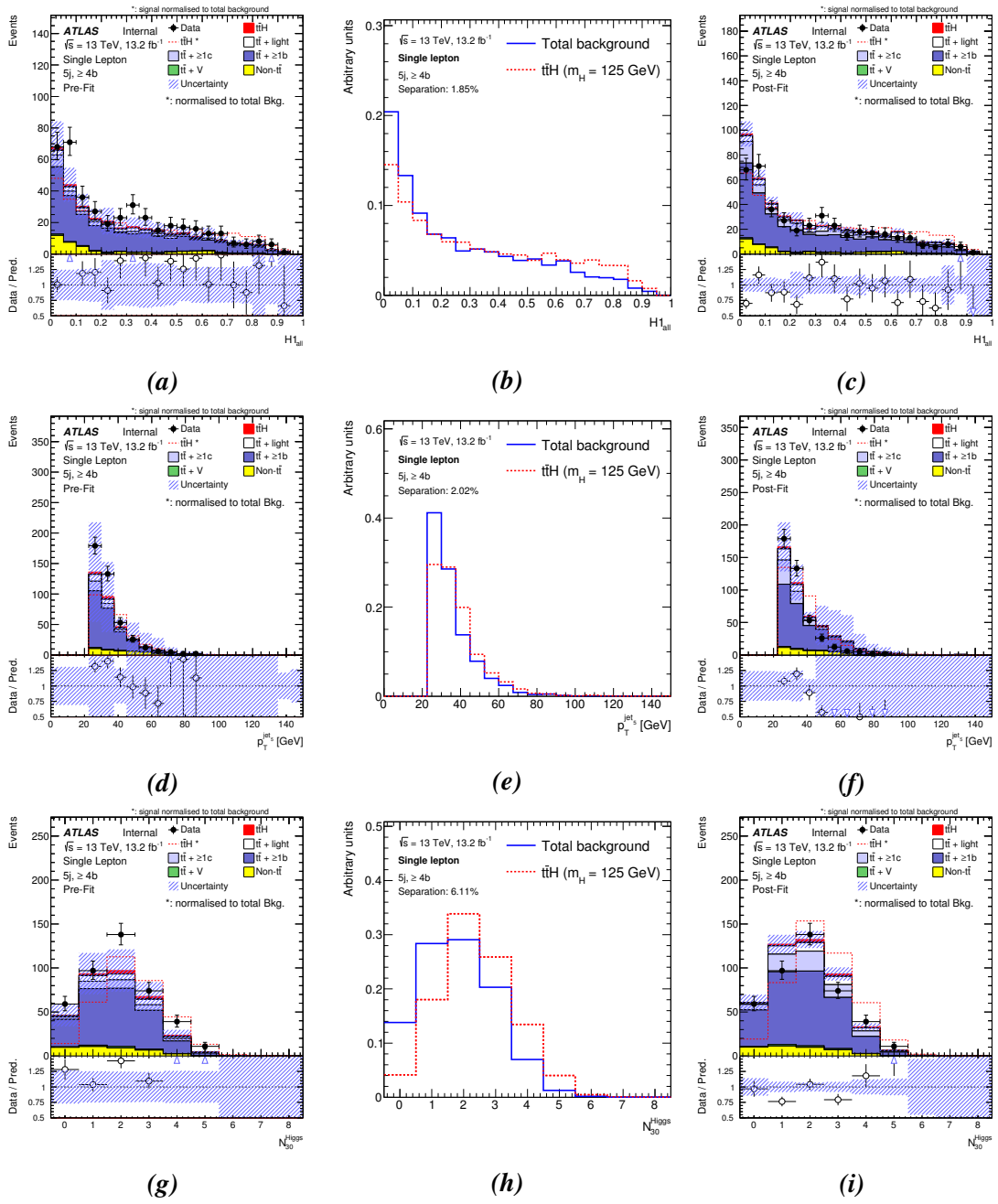


Figure B.10.: Input variables used for training the ANN in the (5j,≥4b) region. Variables from 4 to 6. The plots show pre-fit distributions, separation plots, post-fit distributions.

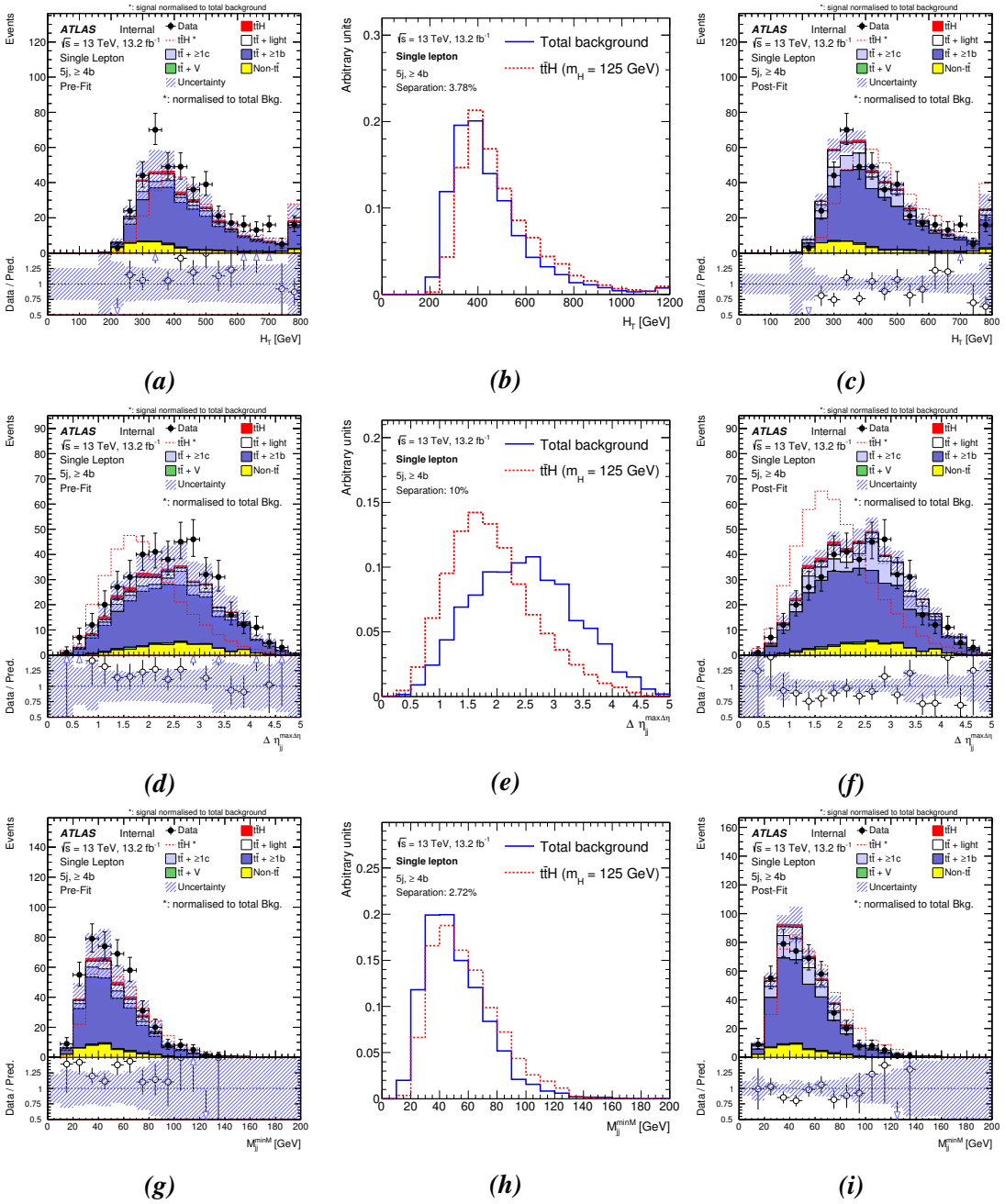


Figure B.11.: Input variables used for training the ANN in the (5j,≥4b) region. Variables from 7 to 9. The plots show pre-fit distributions, separation plots, post-fit distributions.

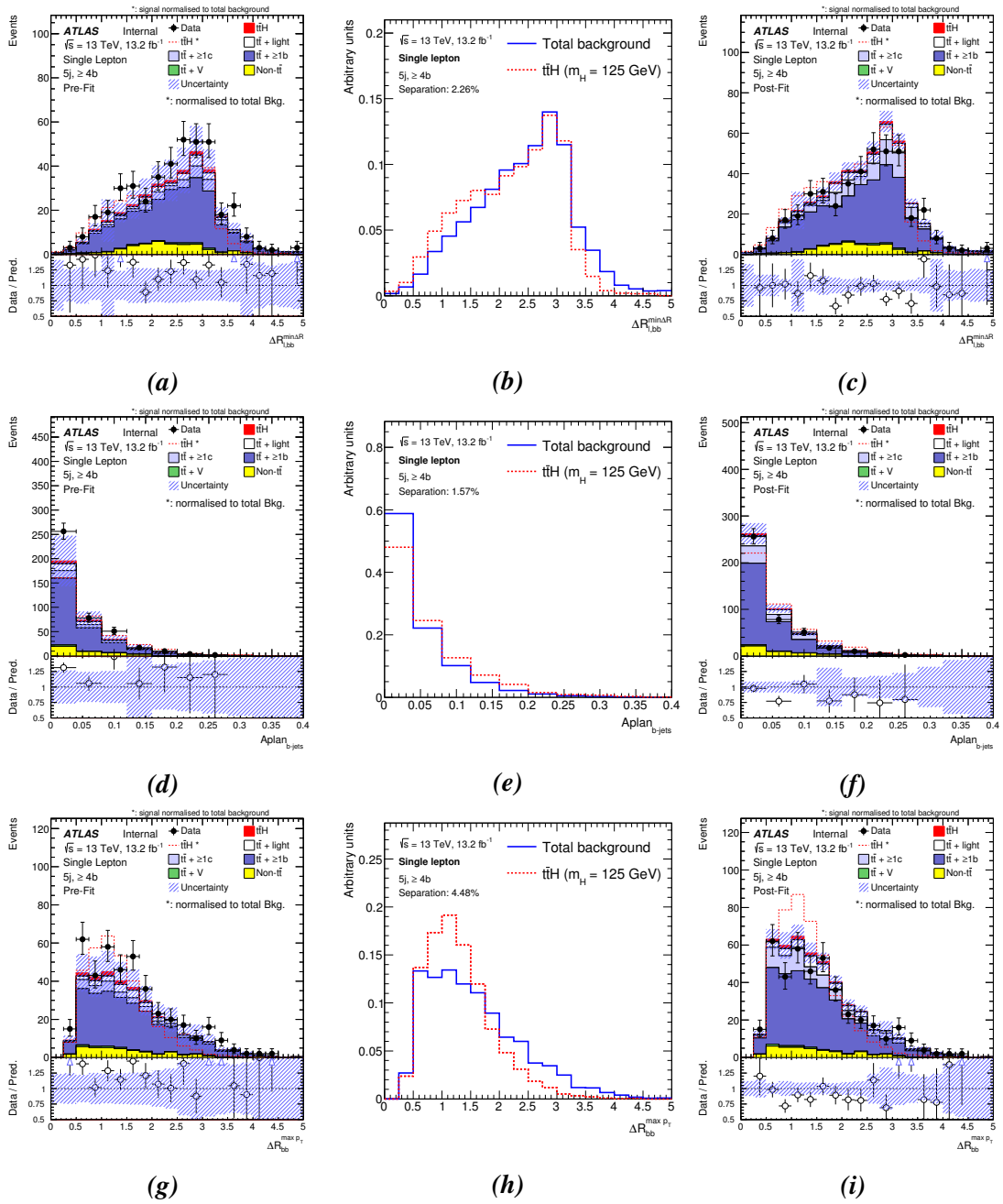


Figure B.12.: Input variables used for training the ANN in the (5j,≥4b) region. Variables from 10 to 12. The plots show pre-fit distributions, separation plots, post-fit distributions.

C

Plots of 1D correlations of the input variables used for the ANN training

This appendix collects the 1D correlation plots for the input variables used for the ANN training which are correlated more than 40% in the MC prediction. The plots show both pre-fit and post-fit distributions. The plots are organised as follows:
for the (5j,≥4b) region:

1. Centrality$_{\text{all}}$ vs: Fig. C.1) $H1_{\text{all}}$, H_T
2. $\text{M}_{bb}^{\min\Delta R}$ vs: Fig. C.2) $\Delta\eta_{jj}^{\max\Delta\eta}$, N_{30}^{Higgs}
3. $\Delta R_{bb}^{\text{avg}}$ vs: Fig. C.3) Aplan$_{\text{b-jets}}$, Centrality$_{\text{all}}$; Fig. C.4) H_T
4. $p_T^{\text{jet}_5}$ vs: Fig. C.4) $\Delta\eta_{jj}^{\max\Delta\eta}$; Fig. C.5) N_{30}^{Higgs}
5. N_{30}^{Higgs} vs: Fig. C.5) $\Delta\eta_{jj}^{\max\Delta\eta}$

for the (≥6j, ≥4b) region:

1. Centrality$_{\text{all}}$ vs: Fig. C.6) H_T , $H1_{\text{all}}$
2. $\Delta R_{bb}^{\text{avg}}$ vs: Fig. C.7) Aplan$_{\text{b-jets}}$, H_T
3. $p_T^{\text{jet}_5}$ vs: Fig. C.8) N_{30}^{Higgs}

for the (≥6j,3b) region:

1. Centrality$_{\text{all}}$ vs: Fig. C.9) Aplan$_{\text{b-jets}}$, $H1_{\text{all}}$
2. H_T^{Had} vs: Fig. C.10) N_{30}^{Higgs}, $p_T^{\text{jet}_5}$
3. Fig. C.11) $\Delta R_{bb}^{\text{avg}}$ vs $\Delta R_{uu}^{\min\Delta R}$, $\text{M}_{bj}^{\max p_T}$ vs $\text{M}_{bb}^{\min\Delta R}$

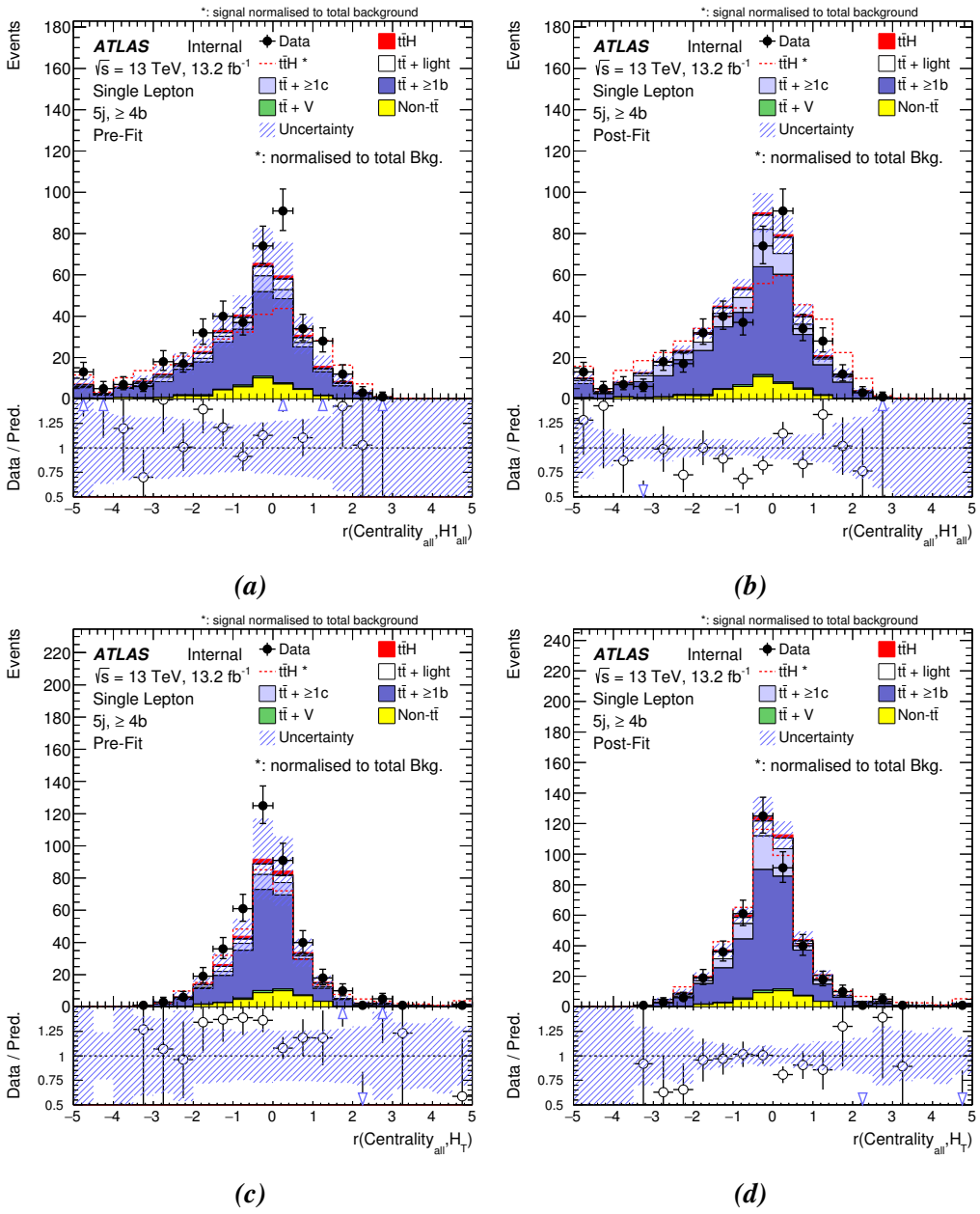


Figure C.1.: 1D correlation for input variables with significant correlation in the (5j,≥4b) region. Part 1.

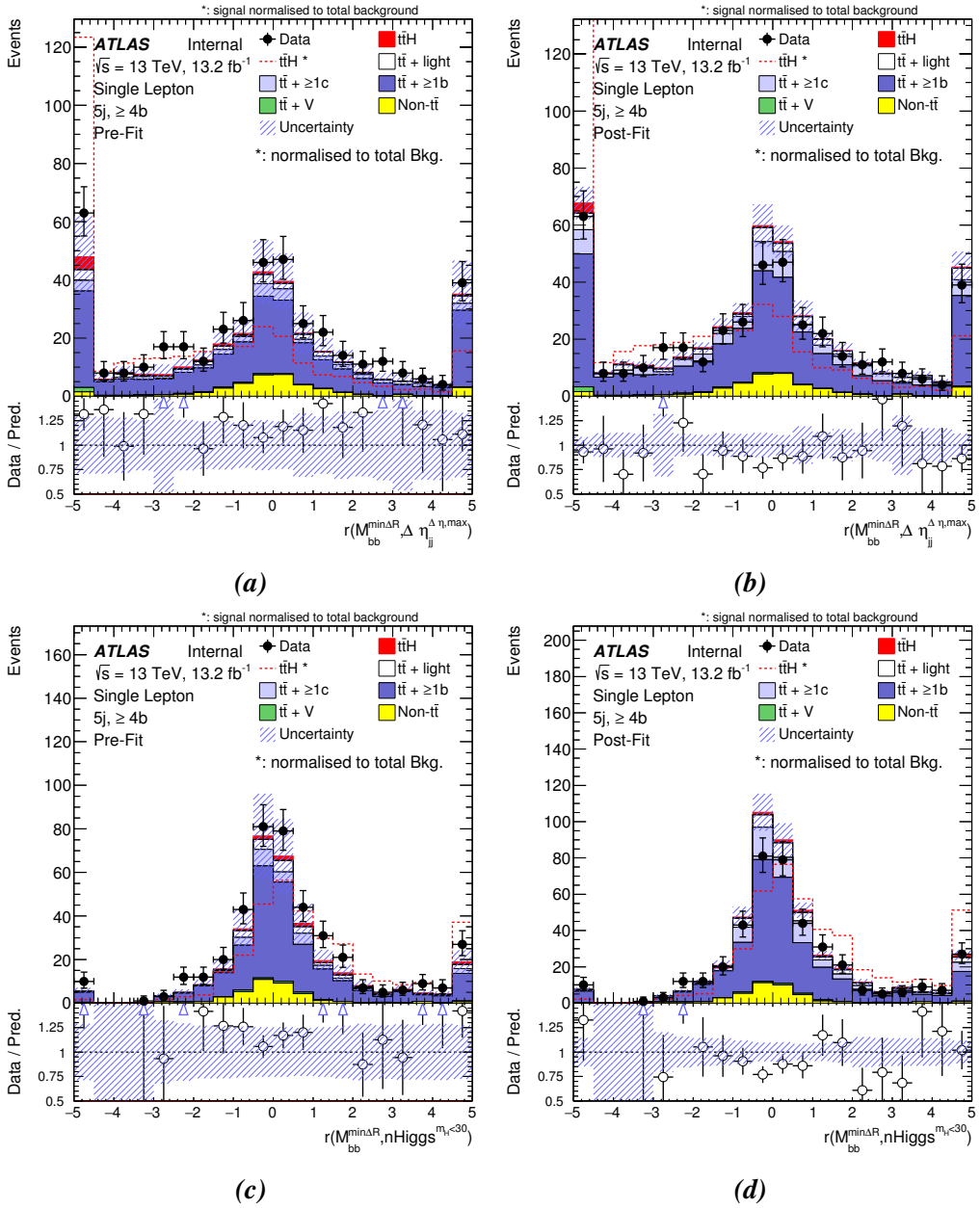


Figure C.2.: 1D correlation for input variables with significant correlation in the (5j,≥4b) region. Part 2.

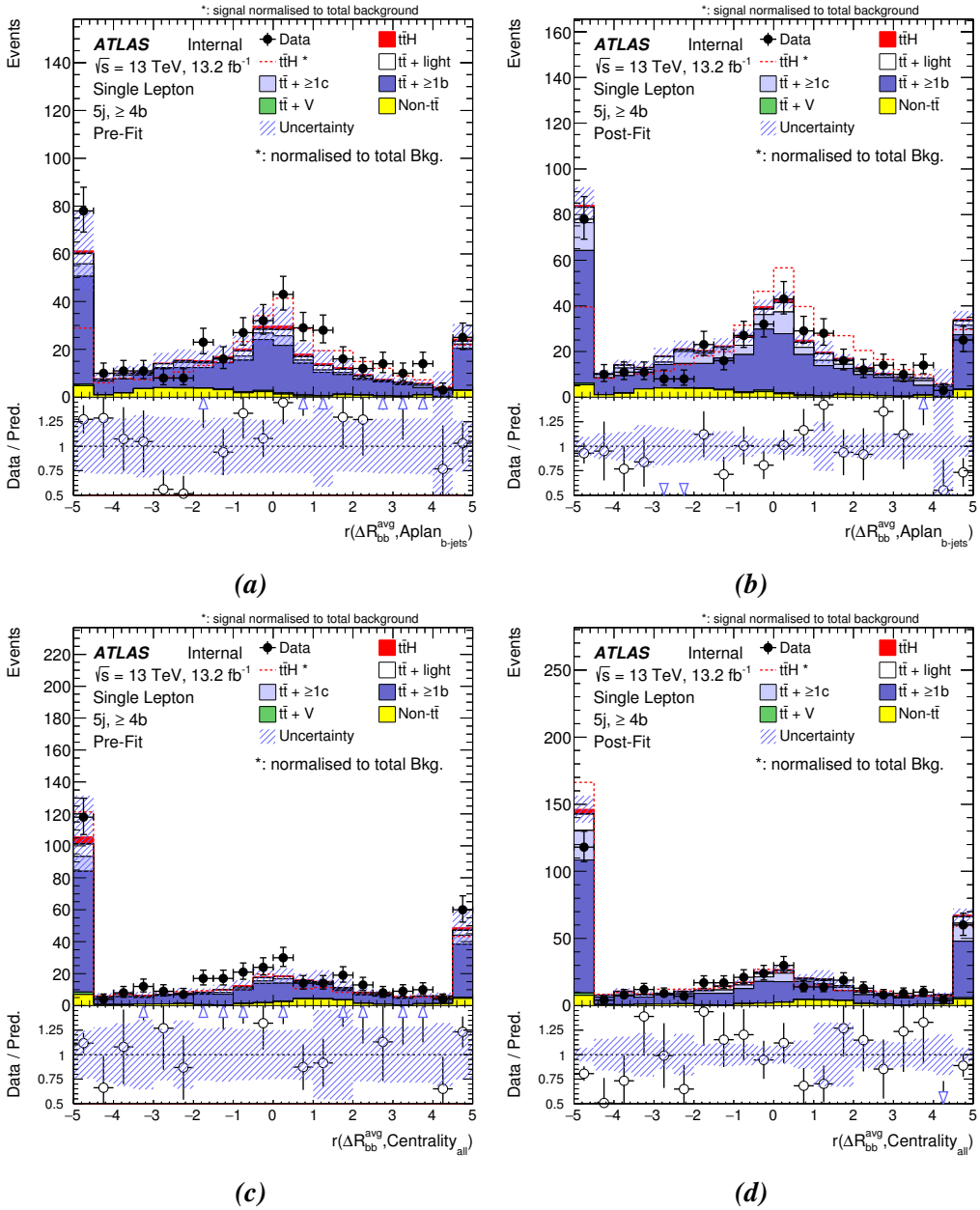


Figure C.3.: 1D correlation for input variables with significant correlation in the (5j,≥4b) region. Part 3.

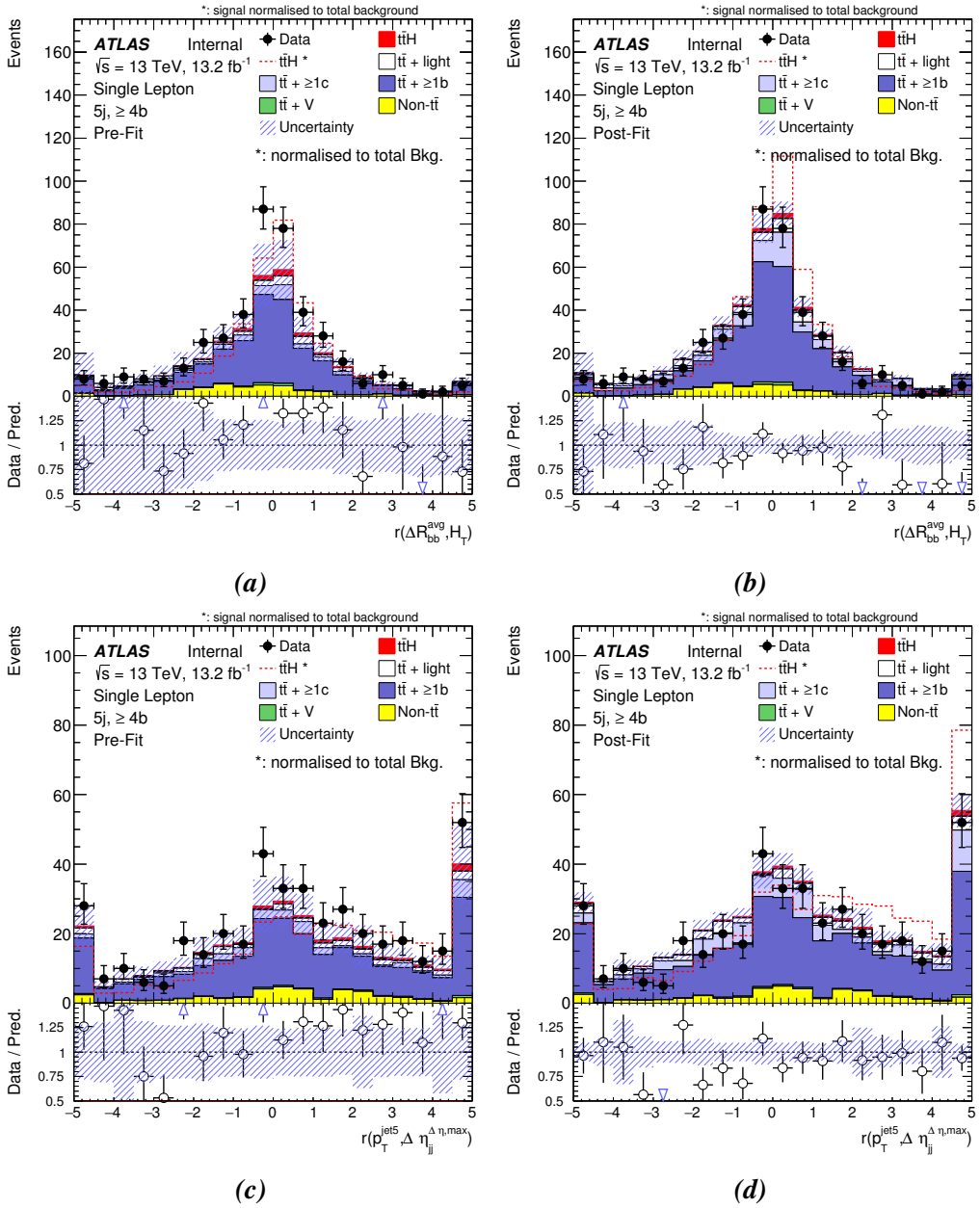


Figure C.4.: 1D correlation for input variables with significant correlation in the (5j,≥4b) region. Part 4.

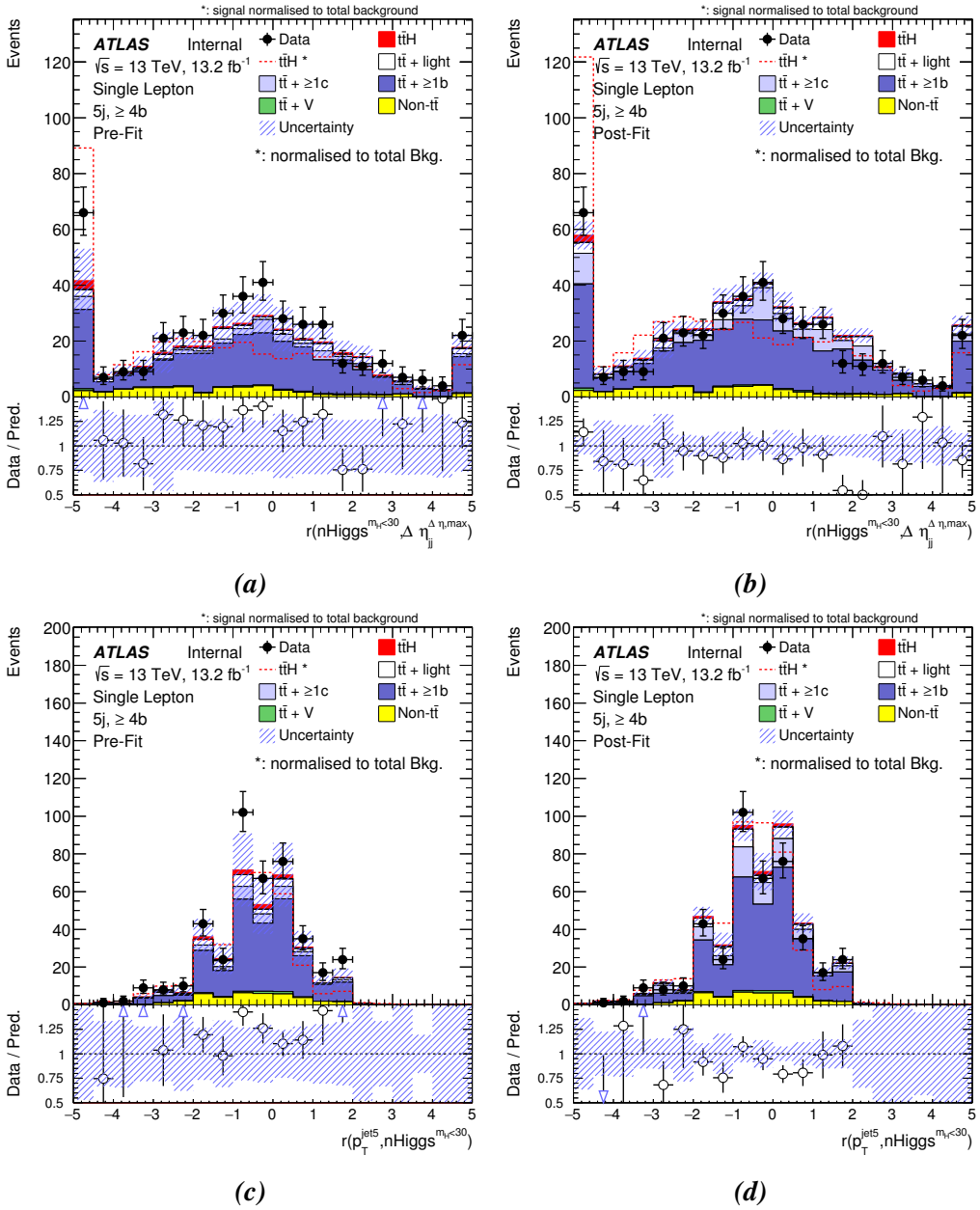


Figure C.5.: 1D correlation for input variables with significant correlation in the (5j,≥4b) region. Part 5.

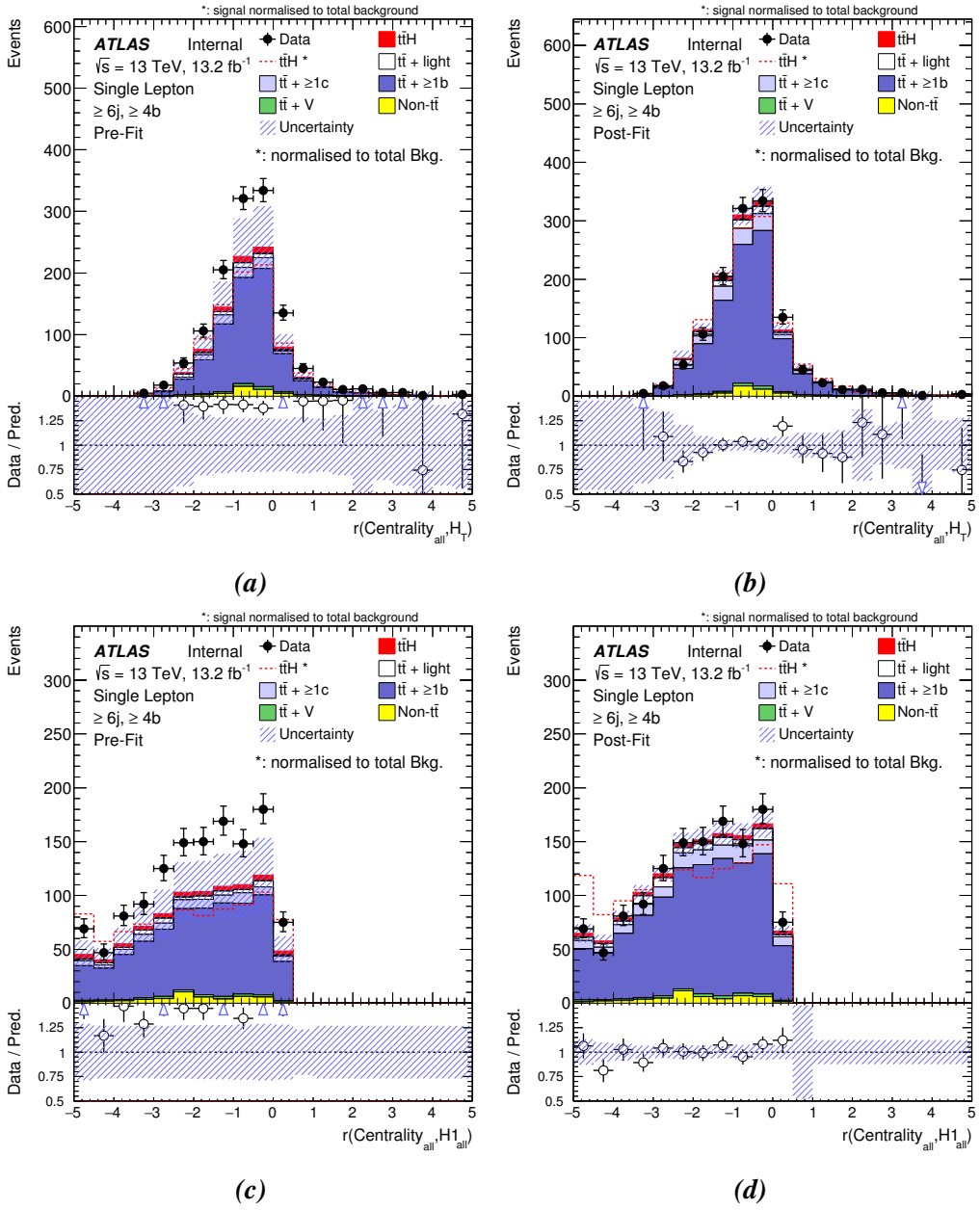


Figure C.6.: 1D correlation for input variables with significant correlation in the (≥6j,≥4b) region. Part 1.

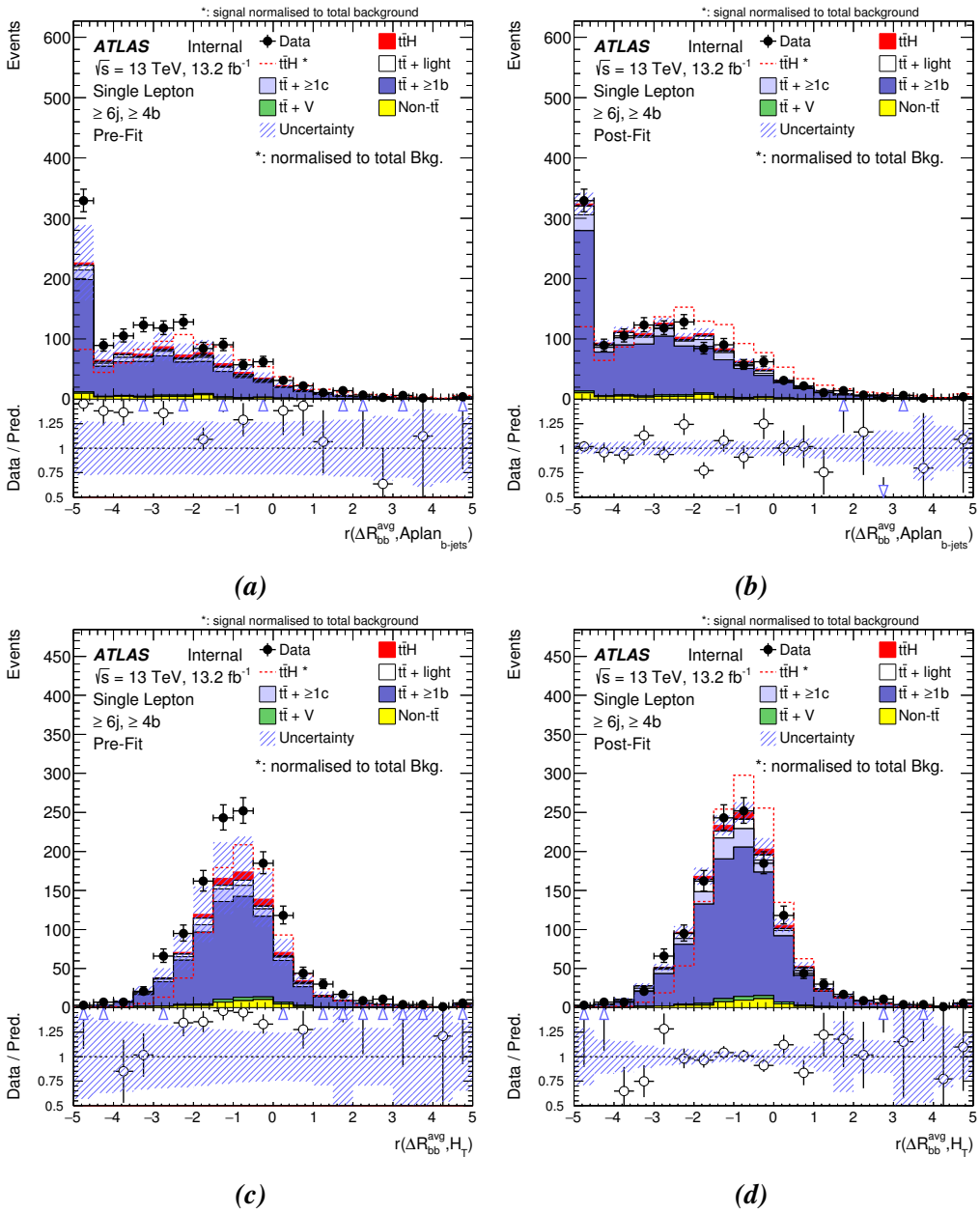
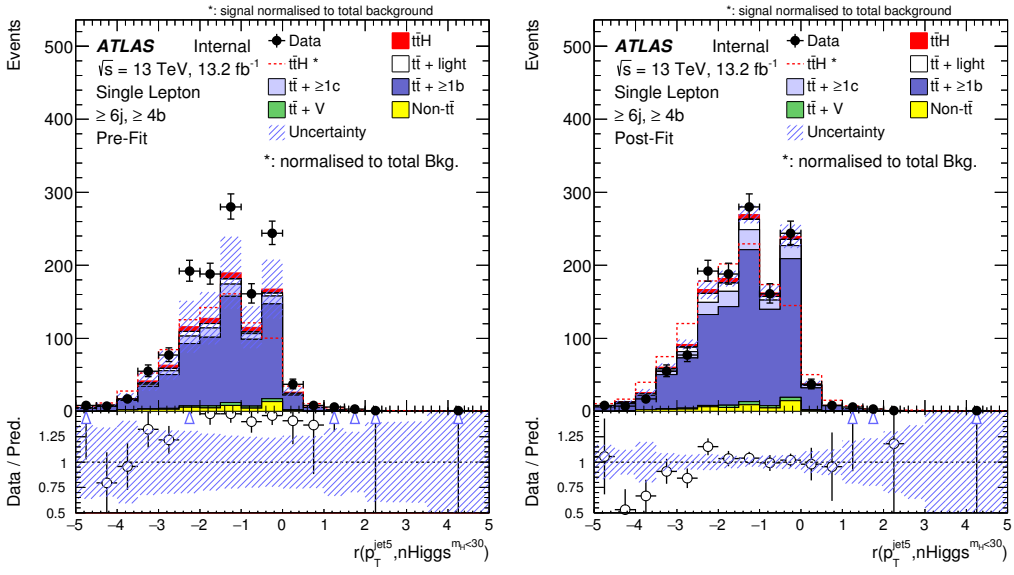


Figure C.7.: 1D correlation for input variables with significant correlation in the (≥6j,≥4b) region. Part 2.



(a) ***(b)***

Figure C.8.: 1D correlation for input variables with significant correlation in the (≥6j,≥4b) region. Part 3.

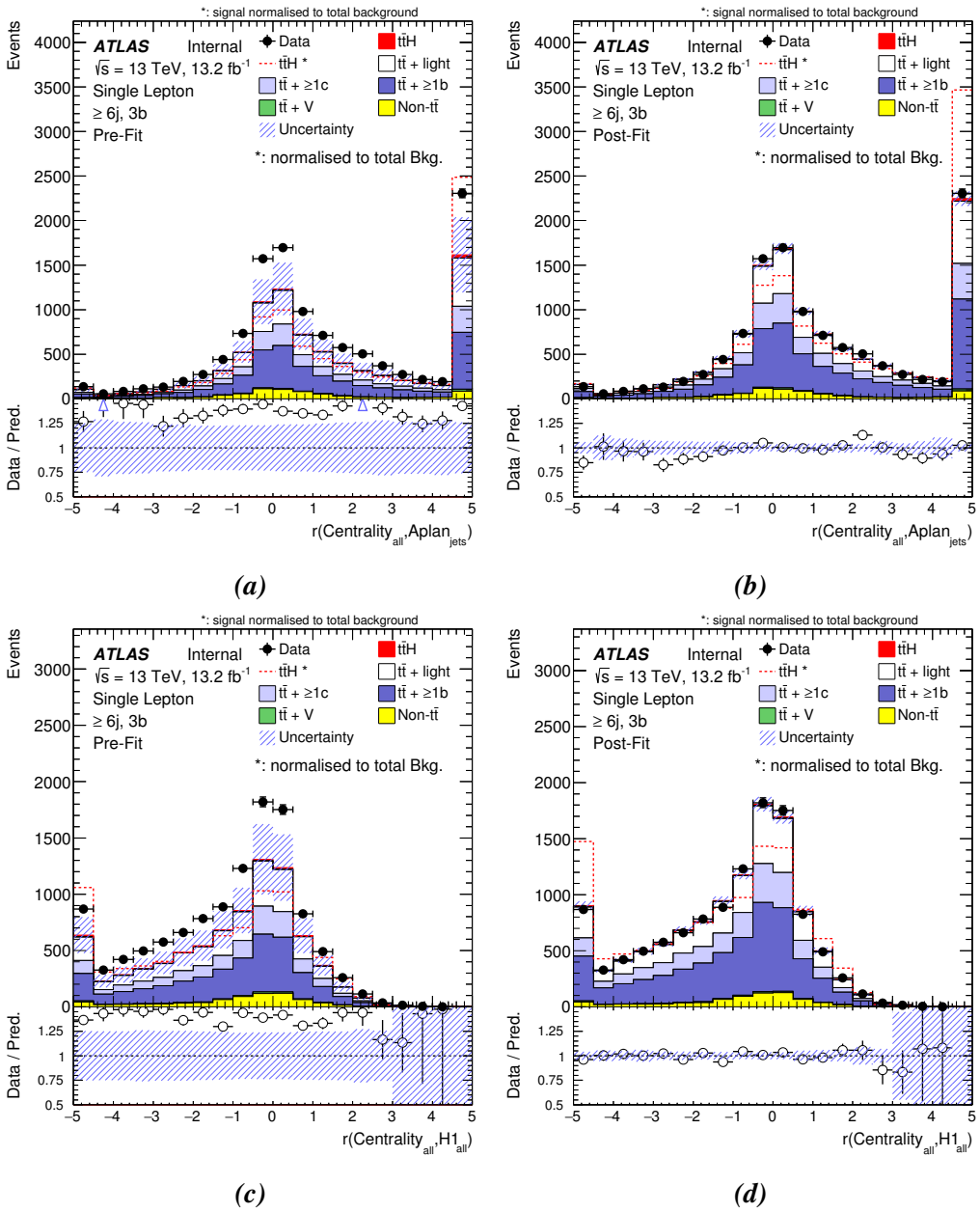


Figure C.9.: 1D correlation for input variables with significant correlation in the (≥6j,3b) region. Part 1.

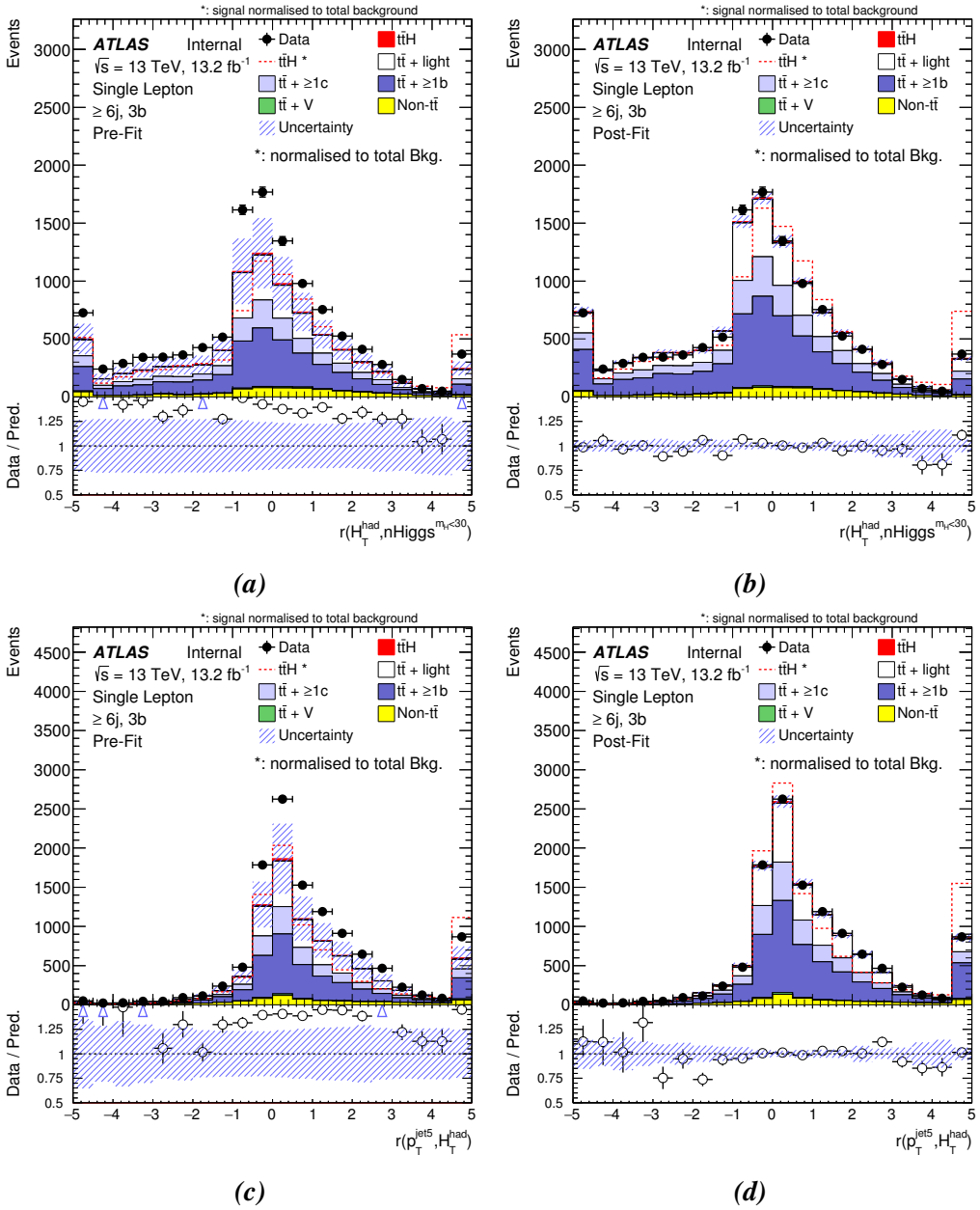


Figure C.10.: 1D correlation for input variables with significant correlation in the (≥6j,3b) region. Part 2.

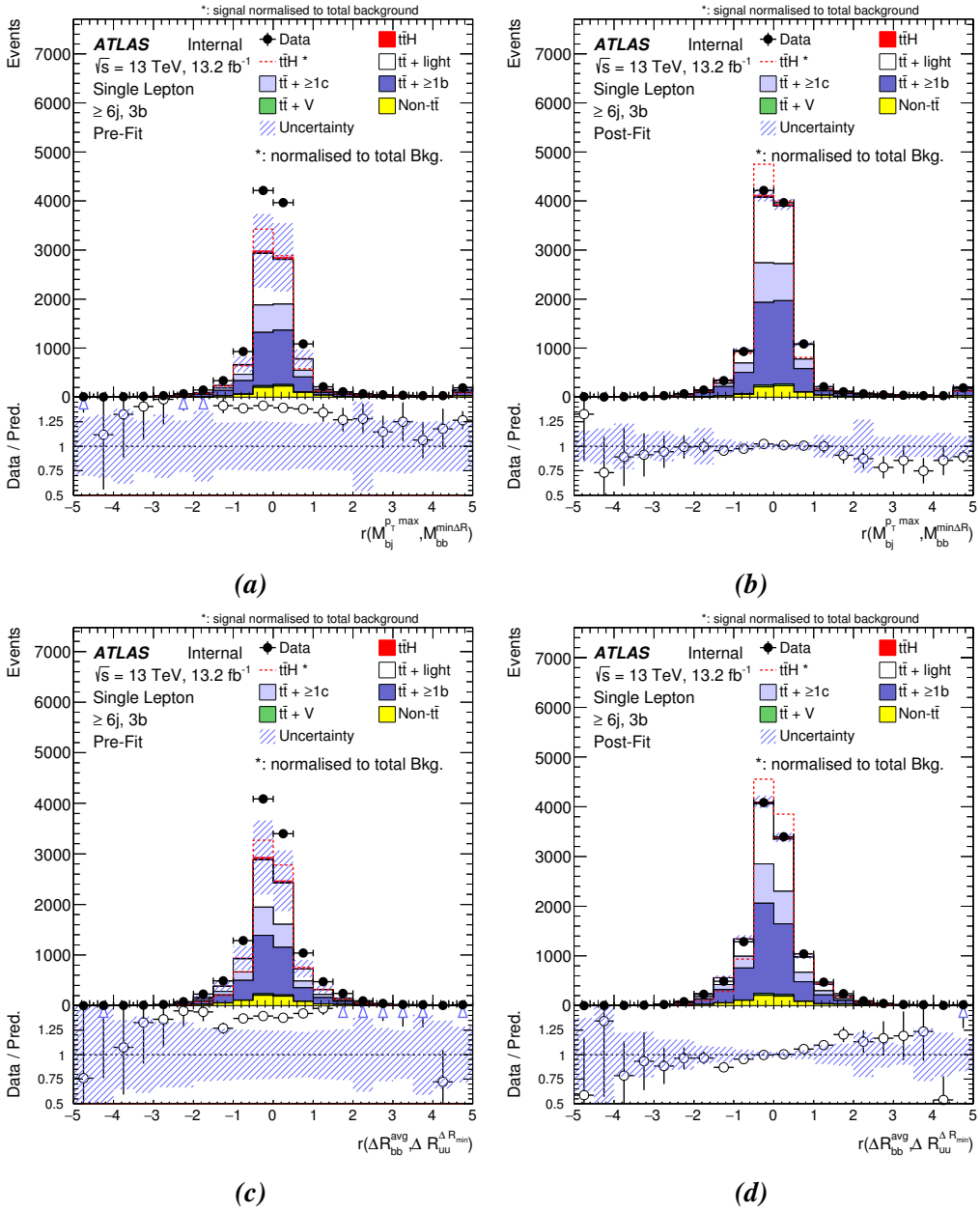


Figure C.11.: 1D correlation for input variables with significant correlation in the (≥6j,3b) region. Part 3.

D

Comparison of ANN and BDT Most Important Systematic Variations

This appendix shows the most relevant systematic variations in the signal regions with respect to the different discriminants considered in this thesis. Only the most important systematic variation is considered, with respect to the ranking plots shown in Chapter 8. In the case of ANN the $t\bar{t}+ \geq 1b$ radiation residual, Fig. D.1, and in the case of the BDT $t\bar{t}+ \geq 1b$ generator residual, Fig. D.2. The shape differences are also enhanced by the smoothing procedure.

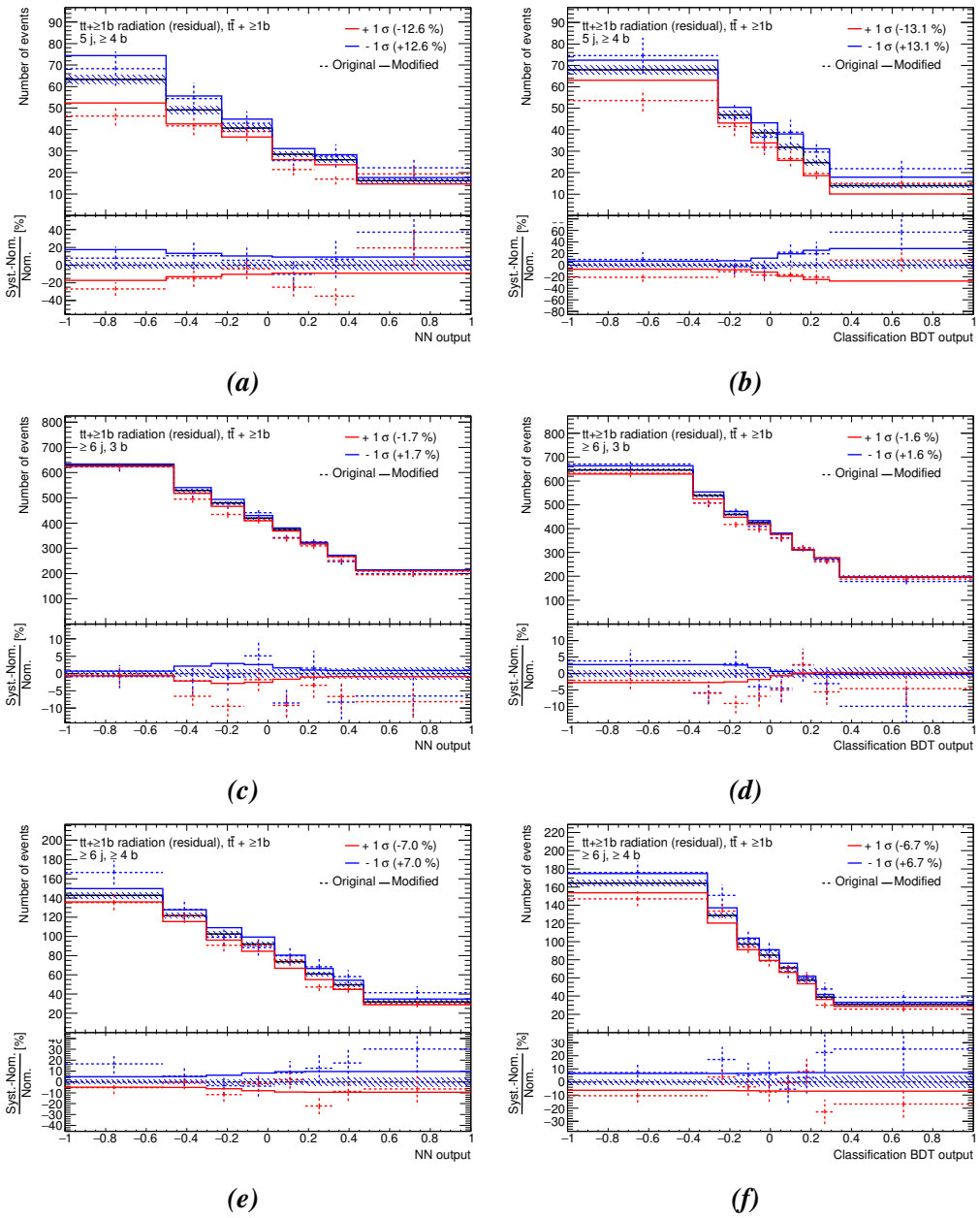


Figure D.1.: Comparison of the systematic uncertainty associated to $t\bar{t}+ \geq 1b$ radiation residual variation for ANN and BDT

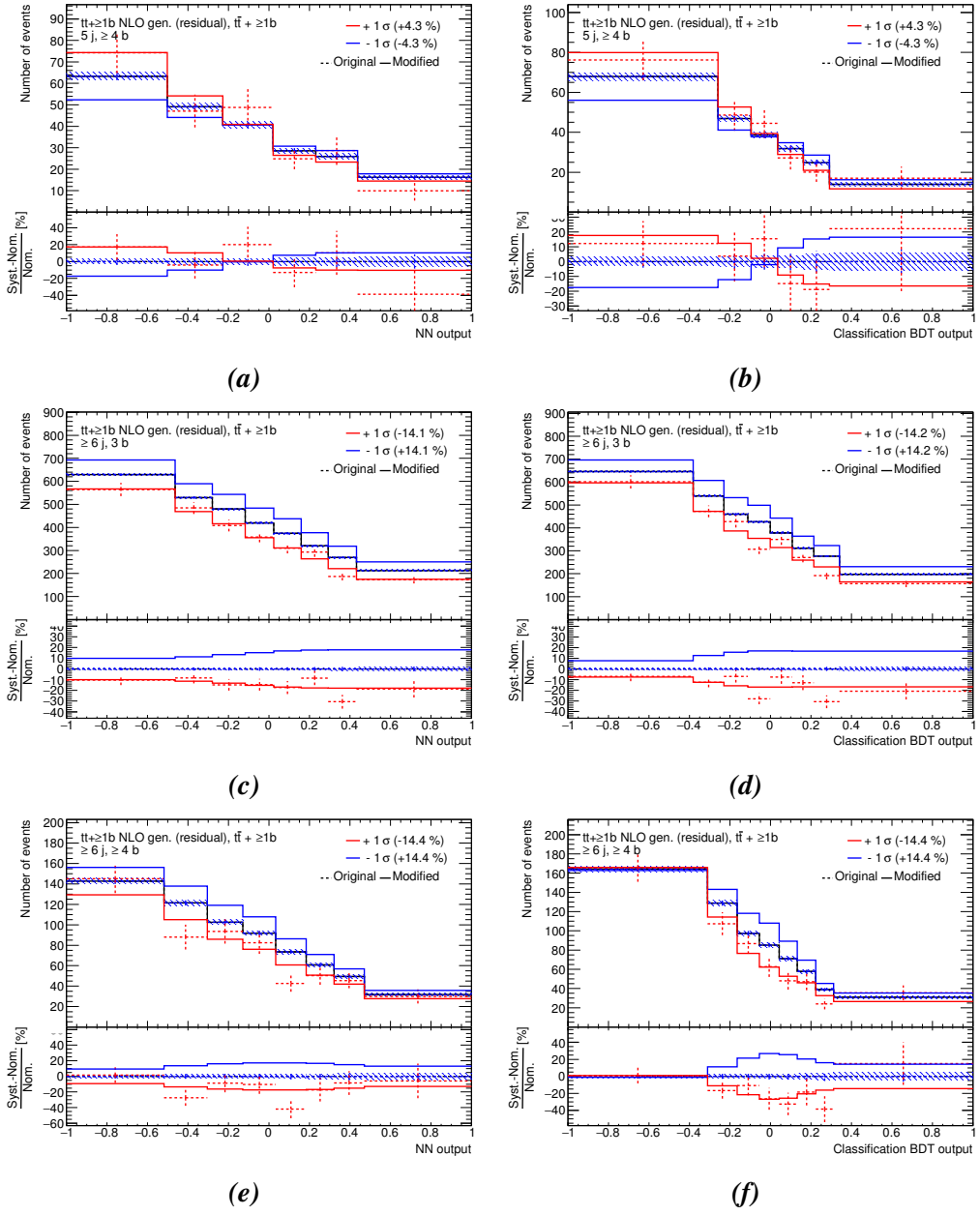


Figure D.2.: Comparison of the systematic uncertainty associated to $t\bar{t}+ \geq 1b$ generator residual variation for ANN and BDT

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1.

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) 30.

[3] F. Abe et al., *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74** (1995) 2626.

[4] S. Abachi et al., *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632.

[5] Y. Ne'eman and Y. Kirsh, *The Particle Hunters*, Cambridge Univ. Pr. (1986).

[6] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579.

[7] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264.

[8] A. Salam, *Weak and Electromagnetic Interactions*, Proceedings of the 8th Nobel symposium, Ed. N. Svartholm, Almqvist & Wiskell, 1968, Conf. Proc. **C680519** (1968) 367.

[9] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl. Phys. **B44** (1972) 189.

[10] C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. **C40** (2016) 100001.

[11] T. Morii, C. S. Lim, and S. N. Mukherjee, *The physics of the standard model and beyond*, River Edge, World Scientific (2004) 298 p (2004).

[12] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I*, Phys. Rev. **122** (1961) 345.

[13] Y. Nambu and G. Jona-Lasinio, *Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. II*, Phys. Rev. **124** (1961) 246.

[14] R. Adam et al., *Planck 2015 results. I. Overview of products and scientific results*, Astron. Astrophys. **594** (2016) A1.

[15] M. S. Turner and D. Huterer, *Cosmic Acceleration, Dark Energy and Fundamental Physics*, J. Phys. Soc. Jap. **76** (2007) 111015.

[16] J. Wess and B. Zumino, *A Lagrangian Model Invariant Under Supergauge Transformations*, Phys. Lett. **49B** (1974) 52.

[17] L. Canetti, M. Drewes, and M. Shaposhnikov, *Matter and Antimatter in the Universe*, New J. Phys. **14** (2012) 095012.

[18] B. Pontecorvo, *Mesonium and anti-mesonium*, Sov. Phys. JETP **6** (1957) 429.

[19] Nobelprize.org, *The Nobel Prize in Physics 2015*, URL: `http://www.nobelprize.org/nobel_prizes/physics/laureates/2015/` (visited on 04/25/2017).

[20] S. F. King, *Neutrino mass models*, Rept. Prog. Phys. **67** (2004) 107.

[21] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612.

[22] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001.

[23] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **1212** (2012) 054.

[24] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **1301** (2013) 080.

[25] 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. **110** (2013) 252004.

[26] M. Aaboud et al., *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, (2016), arXiv: `1609.03920 [hep-ex]`.

[27] J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, (2013), ed. by S. Heinemeyer et al., arXiv: `1307.1347 [hep-ph]`.

[28] A. Djouadi, M. Spira, and P. Zerwas, *Production of Higgs bosons in proton colliders: QCD corrections*, Phys. Lett. **B264** (1991) 440.

[29] CMS Collaboration, *Search for the associated production of the Higgs boson with a top-quark pair*, JHEP **09** (2014) 087, [Erratum: JHEP10,106(2014)].

[30] CMS Collaboration, *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method*, Eur. Phys. J. **C75** (2015) 251.

[31] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, Eur. Phys. J. **C75** (2015) 349.

[32] ATLAS Collaboration, *Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s}$=8 TeV with the ATLAS detector*, JHEP **05** (2016) 160.

[33] ATLAS and CMS Collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$=7 and 8 TeV*, JHEP **08** (2016) 045.

[34] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001, ed. by L. Evans.

[35] O. S. Brüning et al., *LHC Design Report*, CERN Yellow Reports: Monographs, Geneva: CERN, 2004.

[36] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, Geneva, July 2015.

[37] S. Ballestrero, W. Vandelli, and G. Avolio, *ATLAS TDAQ system: current status and performance*, Phys. Procedia **37** (2012) 1819.

[38] ATLAS Collaboration, *Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System*, (Sept. 2013), Final version presented to December 2013 LHCC.

[39] M. Backes, *The ATLAS Trigger System: Ready for Run-2*, PoS **LeptonPhoton2015** (2016) 045.

[40] T. Hryn'ova and K. Nagano, *Trigger Menu Strategy for Run 2*, ATL-COM-DAQ-2014-054, Geneva, May 2014.

[41] E. Simioni et al., *Upgrade of the ATLAS Level-1 Trigger with event topology information*, J. Phys. Conf. Ser. **664** (2015) 082052.

[42] J. Glatzer, *Operation of the Upgraded ATLAS Level-1 Central Trigger System*, J. Phys. Conf. Ser. **664** (2015) 082013.

[43] G. Pásztor, "The Upgrade of the ATLAS Electron and Photon Triggers towards LHC Run 2 and their Performance", *Proceedings, Meeting of the APS Division of Particles and Fields (DPF 2015): Ann Arbor, Michigan, USA, 4-8 Aug 2015*, 2015, arXiv: 1511.00334 [hep-ex].

[44] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823.

[45] J. C. Collins, D. E. Soper, and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1.

[46] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004.

[47] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015).

[48] J. Pumplin, H. L. Lai, and W. K. Tung, *The Charm Parton Content of the Nucleon*, Phys. Rev. **D75** (2007) 054029.

[49] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) 189.

[50] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31.

[51] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. **B238** (1984) 492.

[52] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026.

[53] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*, Cambridge monographs on particle physics, nuclear physics, and cosmology, Photography by S. Vascotto, Cambridge: Cambridge Univ. Press, 2003.

[54] D. Amati and G. Veneziano, *Preconfinement as a Property of Perturbative QCD*, Phys. Lett. **83B** (1979) 87.

[55] G. Corcella et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010.

[56] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009).

[57] J. Alwall et al., *MadGraph 5 : Going Beyond*, JHEP **1106** (2011) 128.

[58] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411** (2004) 040.

[59] M. Asai, *Geant4-a simulation toolkit*, Trans. Amer. Nucl. Soc. **95** (2006) 757.

[60] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[61] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, ATLAS-CONF-2016-080, Geneva, Aug. 2016.

[62] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079.

[63] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852.

[64] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021 (2014), `https://cds.cern.ch/record/1966419`.

[65] P. Artoisenet et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **03** (2013) 015.

[66] R. Raitio and W. W. Wada, *Higgs Boson Production at Large Transverse Momentum in QCD*, Phys. Rev. **D19** (1979) 941.

[67] W. Beenakker et al., *NLO QCD corrections to $t\bar{t}H$ production in hadron collisions*, Nucl. Phys. **B653** (2003) 151.

[68] S. Dawson et al., *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, Phys. Rev. **D68** (2003) 034022.

[69] Y. Zhang et al., *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider*, Phys. Lett. **B738** (2014) 1.

[70] S. Frixione et al., *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, JHEP **06** (2015) 184.

[71] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A Program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) 56.

[72] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **0711** (2007) 070.

[73] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **1006** (2010) 043.

[74] J. M. Campbell et al., *Top-pair production and decay at NLO matched with parton showers*, JHEP **04** (2015) 114.

[75] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D **82** (2010) 074018.

[76] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930.

[77] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612.

[78] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t}$*, Phys. Rev. Lett. **109** (2012) 132001.

[79] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **1212** (2012) 054.

[80] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the quark-gluon reaction*, JHEP **1301** (2013) 080.

[81] M. Czakon, P. Fiedler, and A. Mitov, *The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004.

[82] M. Czakon, D. Heymes, and A. Mitov, *High-precision differential predictions for top-quark pairs at the LHC*, Phys. Rev. Lett. **116** (2016) 082003.

[83] M. Czakon, D. Heymes, and A. Mitov, *Dynamical scales for multi-TeV top-pair production at the LHC*, (2016), arXiv: 1606.03350 [hep-ph].

[84] ATLAS Collaboration, *Search for the Standard Model Higgs boson decay to $\mu^+\mu^-$ with the ATLAS detector*, Phys. Lett. B **738** (2014) 68.

[85] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601.

[86] F. Cascioli et al., *NLO matching for $t\bar{t}b\bar{b}$ production with massive b-quarks*, Phys. Lett. B **734** (2014) 210.

[87] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. **C71** (2011) 1547.

[88] S. Alioli et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **09** (2009) 111, [Erratum: JHEP02,011(2010)].

[89] S. Frixione et al., *Single-top hadroproduction in association with a W boson*, JHEP **0807** (2008).

[90] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-*, Phys. Rev. D **82** (2010).

[91] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010).

[92] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011).

[93] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **0812** (2008) 039.

[94] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **0803** (2008) 038.

[95] S. Höche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013).

[96] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s}$ = 13 TeV with the ATLAS Detector*, ATLAS-CONF-2015-039 (2015).

[97] ATLAS Collaboration, *Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2015-018, 2015.

[98] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, *Adaptive vertex fitting*, J. Phys. **G34** (2007) N343.

[99] T. G. Cornelissen et al., *Updates of the ATLAS Tracking Event Data Model (Release 13)*, ATL-SOFT-PUB-2007-003, Geneva, June 2007.

[100] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$ =13 TeV*, Eur. Phys. J. **C76** (2016) 292.

[101] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data*, Eur. Phys. J. C **74** (2014) 2941.

[102] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, Geneva, Apr. 2008.

[103] J. Alison et al., *Description and Performance of the Electron Likelihood Tool at ATLAS using 2012 LHC Data*, ATL-COM-PHYS-2013-378, Geneva, Apr. 2013.

[104] Anastopoulos et al., *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-COM-CONF-2016-028, Geneva, Mar. 2016.

[105] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 063.

[106] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-015, Geneva, July 2015.

[107] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, ATLAS-CONF-2015-037, Geneva, Aug. 2015.

[108] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, Geneva, Aug. 2013.

[109] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, Geneva, May 2014.

[110] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, Geneva, June 2016.

[111] D0 Collaboration: V. M. Abazov et al., *Measurement of the* $t\bar{t}$ *production cross section in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV using secondary vertex b-tagging*, Phys. Rev. D **74** (2006) 112004.

[112] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at* $\sqrt{s}$*= 13 TeV*, ATL-PHYS-PUB-2015-027, Geneva, July 2015.

[113] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at* $\sqrt{s}$*=8 TeV with the ATLAS detector*, ATLAS-CONF-2014-058, Geneva, Oct. 2014.

[114] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, ATL-DAQ-PUB-2016-001, Geneva, Mar. 2016.

[115] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. **C74** (2014) 2941.

[116] ATLAS Collaboration, *Lepton isolation recommendations.* URL: `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/IsolationSelectionTool` (visited on 04/25/2017).

[117] G. Sciolla et al., *Muon Combined Performance in Run 2 (25 ns runs)*, ATL-COM-MUON-2015-093, Geneva, Nov. 2015.

[118] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, Geneva, July 2015.

[119] B. Müller, J. Reinhardt, and M. Strickland, *Neural Networks, an Introduction*, Physics of Neural Networks, Berlin Heidelberg: Springer-Verlag, 1995.

[120] W. S. McCulloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity*, The bulletin of mathematical biophysics **5** (1943) 115.

[121] C. Van Der Malsburg, "Frank Rosenblatt: Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms", *Brain Theory: Proceedings of the First Trieste Meeting on Brain Theory, October 1–4, 1984*, ed. by G. Palm and A. Aertsen, Berlin, Heidelberg: Springer Berlin Heidelberg, 1986 245.

[122] P. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Harvard Univ., 1974.

[123] R. H. Byrd, J. Nocedal, and R. B. Schnabel, *Representations of quasi-Newton matrices and their use in limited memory methods*, Mathematical Programming **63** (1994) 129.

[124] A. Krogh and J. A. Hertz, *A Simple Weight Decay Can Improve Generalization*, ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 4, 1992 950.

[125] M. Feindt, *A Neural Bayesian Estimator for Conditional Probability Densities*, (2004), arXiv: `physics/0402093 [physics.data-an]`.

[126] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. **A559** (2006) 190.

[127] V. Blobel and E. Lohrmann, *Statistische und numerische Methoden der Datenanalyse*, Springer Fachmedien Wiesbaden (1998).

[128] D. J. C. MacKay, *A Practical Bayesian Framework for Backpropagation Networks*, Neural Comput. **4** (May 1992) 448.

[129] L. Breiman et al., *Classification and regression trees*, Chapman and Hall/CRC, 1984.

[130] P. C. Bhat, *Advanced analysis methods in high-energy physics*, AIP Conf. Proc. **583** (2001) 22.

[131] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm", 1996.

[132] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81.

[133] J. Therhaag, *TMVA: Toolkit for multivariate data analysis*, AIP Conf. Proc. **1504** (2009) 1013.

[134] C. Bernaciak et al., *Fox-Wolfram Moments in Higgs Physics*, Phys. Rev. **D87** (2013).

[135] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2518.

[136] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015).

[137] G. Aad et al., *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11** (2016) P04008.

[138] *Simulation of top quark production for the ATLAS experiment at $\sqrt{s}$ = 13 TeV*, ATL-PHYS-PUB-2016-004, Geneva, Jan. 2016.

[139] *Studies of tt+cc production with MadGraph5_aMC@NLO and Herwig++ for the ATLAS experiment*, ATL-PHYS-PUB-2016-011, Geneva, May 2016.

[140] ATLAS Collaboration, *Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s}$ =8 TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) 11.

[141] ATLAS Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, ATL-PHYS-PUB-2016-002 (2016).

[142] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^{+-}$ production and decay at NLO*, JHEP **1207** (2012).

[143] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) 1554, [Erratum: Eur. Phys. J.C73,2501(2013)].

[144] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016, New York, Jan. 2012.

[145] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007.

[146] A. Wald, *Tests of statistical hypotheses concerning several parameters when the number of observations is large*, Transactions of the American Mathematical Society **54** (1943) 426.

[147] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instrum. Meth. **A434** (1999) 435.

[148] ATLAS Collaboration, *Combination of the searches for Higgs boson production in association with top quarks in the $\gamma\gamma$, multilepton, and $b\bar{b}$ decay channels at $\sqrt{s}$=13 TeV with the ATLAS Detector*, ATLAS-CONF-2016-068, Geneva, Aug. 2016.

[149] ATLAS Collaboration, *Search for the Associated Production of a Higgs Boson and a Top Quark Pair in Multilepton Final States with the ATLAS Detector*, ATLAS-CONF-2016-058, Geneva, Aug. 2016.

[150] ATLAS Collaboration, *Measurement of fiducial, differential and production cross sections in the $H \to \gamma\gamma$ decay channel with 13.3 fb^{-1} of 13 TeV proton-proton collision data with the ATLAS detector*, ATLAS-CONF-2016-067, Geneva, Aug. 2016.

[151] CMS Collaboration, *Search for associated production of Higgs bosons and top quarks in multilepton final states at $\sqrt{s}$=13 TeV*, CMS-PAS-HIG-16-022, Geneva, 2016.

[152] CMS Collaboration, *Search for* $\mathrm{t\bar{t}H}$ *production in the* $\mathrm{H} \to \mathrm{b\bar{b}}$ *decay channel with 2016 pp collision data at $\sqrt{s}$=13 TeV*, CMS-PAS-HIG-16-038, Geneva, 2016.

[153] CMS Collaboration, *Updated measurements of Higgs boson production in the diphoton decay channel at $\sqrt{s}$=13 TeV in pp collisions at CMS.*, CMS-PAS-HIG-16-020, Geneva, 2016.

[154] N. Moretti et al., *Measuring the signal strength in $t\bar{t}H$ with $H \to b\bar{b}$*, Phys. Rev. **D93** (2016) 014019.

List of Abbreviations

ALICE A Large Ion Collider Experiment

ANN Artificial Neural Network

ATLAS A Toroidal LHC ApparatuS

BDT Boosted Decision Tree

BFGS Broyden-Fletcher-Goldfarb-Shanno

BT Barrel Toroid

CERN Conseil Européen pour la Recherche Nucléaire

CKM Cabibbo-Kobayashi-Maskawa

CL Confidence Level

CMS Compact Muon Solenoid

CMB Cosmic Microwave Background

CP Charge Parity

CS Central Solenoid

CSC Cathode Strip Chamber

CTP Central Trigger Processors

DGLAP Dokshitzer-Gribov-Lipatov-Altarelli-Parisi

DT Decision Tree

ECT End-Cap Toroids

EM Electromagnetic

EW Unified ElectroWeak

FCAL Forward CALorimeter

FSR Final State Radiation

GUT Grand Unified Theory

HEC Hadronic End-Cap

HLT High-level trigger

IBL Insertable B-Layer

ID Inner Detector

ISR Initial State Radiation

JES Jet Energy Scale

JER Jet Energy Resolution

JVF Jet Vertex Fraction

JVT Jet Vertex Tagger

L1 Level-1

L1Calo Level-1 calorimeter trigger

L1Muon Level-1 muon trigger

L1Topo Level-1 topological trigger

LAr Liquid-Argon

LEP Large Electron-Positron Collider

LHC Large Hadron Collider

LINAC LINear ACcelerator

LO Leading Order

LS1 Long Shutdown 1

MC Monte Carlo

MDT Monitored Drift Tubes

MPI Multi Parton Interaction

MS Muon Spectrometer

MVA Multi Variate Analysis

NLO Next-to-Leading-Order

NNLL Next-to-Next-to-Leading-Logarithmic

NNLO Next-to-Next-to-Leading-Order

NPs Nuisance Parameters

PDFs Parton Distribution Functions

PS Proton-Syncrotron

QED Quantum-Electro-Dynamics

QCD Quantum-Chromo-Dynamics

QFT Quantum Field Theory

RPC Resistive Plate Chambers

ROC Receiver Operating Characteristic

ROS read-out system

SCT Semi-Conductor Tracker

SM Standard Model

SPS Super-Proton-Synchrotron

SUSY Supersymmetry

TDAQ Trigger and Data Acquisition

TGC Thin Gap Chambers

TMVA Toolkit for Multivariate Analysis

TOE Theory of Everything

TRF Tag Rate Function

TRT Transition Radiation Tracker

UE Underlying Event

VEV Vacuum Expectation Values

WIMP Weakly Interacting Massive Particles

List of Figures

List of Tables

Acknowledgements

Thank you. [1]

It was May 2013, when I decided to leave my house for a new adventure. Following the path of the father of my father, who worked in the foundries of Westfalen before being pushed away because Italian (due to WWI), I started a new chapter of my life in Germany, not as a manual laborer, but as member of one of the most famous universities of the world. The list of people to whom I must express my gratitude is very large, and since I am speaking of a journey, I will list them in the same order I met them. I will often refer to *difficulties*, since this journey has not been a fairy tale, so there's no shame in saying that it has not been an easy one. My adventure started with a smile since, arriving at the train station of Göttingen, I was welcomed by two friends, met in Trieste some months before, Enrico and Timon. In the same day I also met Trudi, who gave me a room for one year. I will never forget our dinners and our conversations, she helped me feeling at home again. I say a special thanks to Lucie, who was driving me at the workplace at the very first day and being always helpful and kind in the following years, firstly as secretary of the institute, and then freely giving me support after some terrible moments. I thank also the other secretaries I met, Bernadette, Heidi and Christina for having been kind and patient with an italoanarcobureaucratist.

I cannot find the right words to express my gratitude and feelings to Prof. Quadt for these years. He gave me the possibility to make a dream of my childhood to become reality: to be a scientist. I would like also to express my gratitude to the professors that kindly

[1] Whole transcription of Alfred Hitchcock's Academy Award acceptance speech in 1967.

agreed to be part of my examination committee.
While living and studying at Göttingen, I was surprised to meet many Italians who independently left for my same journey. I thank all of them for their friendship and for the relaxing moments, Piero, Simone and Paolo among the others. At the institute I met many good and interesting people, the space in this page is not enough for all of them, I just give a special thanks to Mohammad and to the elder ones, Olaf, Boris, Jason, Michael, Gerhard, Gen, Joern, Jens and others. I would like also to say a special thanks to the collegues that helped me in the preparation of my thesis defense, Tomas, Royer, Clara, Piero and Paolo.
From spring 2015 to spring 2016 I moved to CERN as part of my research program. I met there Lisa for the first time. I will always be grateful to her for the help and support in all my work, and also for being a reference in the bad moments. I continued to work also with Maria, to whome I must express my gratitude for the constant support. At CERN I could give a face to many voices I knew from my working meetings. I had the pleasure and the luck to meet special and smart people, Mirko among the others, as well my landlord Antoni. I also had an occasion to meet again Michele, who was supporting me in Trieste, during my master thesis preparation. I am grateful to him, as I was in the past, for all the times he replied to my questions and for our informal meetings.
Being myself at CERN was the fulfilling of one of the greatest dreams of my childhood. Staying there just not gave me the opportunity to grow as a scientist, but also gave me the occasion to learn one of the most important lessons of my life, which is the following. I met people from many countries: (Italy), Germany, Spain, Russia, Mexico, Japan, Iran, Iraq, Georgia, France, United States, Netherlands, United Kingdom, Sweden, Portugal, and many more. We were all there with the same goal. Making research. What I learnt about this is that there are not differences between people from distinct cultures, we all share the same human nature and curiosity, since we are just one race. There's no meaning in country borders, in cultural walls, in past war memories. A new era for humankind can happen, if we only have the courage to admit that we are all brothers.
Dulcis in fundo, I want to say my greatest thank you to the only person that was really plenty sharing with me the good and the bad from these years, my girlfriend Elisa. I couldn't go trough all the difficulties without her, her support has been the greatest source of strength I could get. I hope I can use my lifetime to pay back this large debt. This chapter of my life, is also her chapter, and if this can be counted as a success of mine, this is indeed also a success of hers. I also say a great thank you to my parents for always believing in me and for giving me courage when needed. I would like to express my

apologies to my uncle Ernesto, since for the research I made in these years, I will surely not get a Nobel Prize. I must be also grateful to all the old friends left behind, who I didn't lose in space. Gabro, Dario, Giacomo, Potter, Ingegnere, Spion, Chemical, Sara, Barbara and few others that I can barely count on the fingers of my hands, as always happens when counting true friends.

I would like to express my gratitude also to the city of Göttingen, one of "my cities", from now on. It has been an inspiration to live in the same city as some of the greatest genii of the humankind, I was honoured to behold the ashes of their past glory. Göttingen is such as Rome, Venice, whole Italy perhaps: it is great especially thanks to its past. I hope this city will have a good memory of me, and if the words written on Rathaus' walls remind me that "Extra Gottingam non est vita, si est vita, non est ita" (There is no life outside Göttingen, even if it is life, it is no life like here), I promise that I will do my best to live plenty my life and enjoy my time. Forgive me if I leave, but as Marco Polo said: "credo che fosse piacere di Dio nostra tornata, acciò che si potessoro sapere le cose che sono per lo mondo" (I think that it was God will that we returned back, so we could tell to others what we found in our journey).

I arrived here as a young man, I leave as a wiser one. I met humans and Humans. I learnt many things, things that changed me. I learnt what I want to be, and what I don't want to. I promise, I will always remember the academical style of living. I will freely give knowledge to others.

To all the people who tried to push me down I say in my vernacular marilenghe: "Ancje Dio al è furlan; sa nol pae vuei, al pae doman" (also God is friulan, if he doesn't pay back today, he will pay back tomorrow). To the good people I met I say: "You are the Göttingen I will remember for my lifetime". Emperor Augustus' last words, are good now to say farewell: "Acta est fabula, plaudite!" So I say myself to everybody: "farewell, from a friulan scientist". I am leaving as I came here: with a large smile.