ALMA MATER STUDIORUM - Università di Bologna

Dipartimento di Fisica e Astronomia

Dottorato di Ricerca in Fisica
Ciclo XXIX

Study of the associated production of the Higgs boson with a top quark pair in a boosted regime in the ATLAS experiment at LHC.

Candidata: Silvia Biondi

Coordinatore Dottorato:
Chiar.mo Prof. Gastone Castellani

Relatori:
Dr. Roberto Spighi
Chiar.mo Prof. Mauro Villa
Correlatore:
Dr. Matteo Franchini

Settore Concorsuale di afferenza: 02/A1
Settore Scientifico disciplinare: FIS/01

Esame Finale - Anno 2017

Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.

Albert Einstein

Dedicated to the source of my pride and happiness: my family.

Contents

Contents

Abstract

The measurements of the $pp \rightarrow t\bar{t}H + X$ signal strength ($\mu = \sigma_{obs}/\sigma_{t\bar{t}H_{SM}}$) and its upper limit at a center-of-mass energy of pp collision of 13 TeV are presented in this thesis. The pp data collected with the ATLAS detector in 2015 and 2016, corresponding to an integrated luminosity of 36.5 fb^{-1}, have been submitted to a detailed analysis. The $t\bar{t}H$ signal strength and its upper limit have been measured searching for the decays $H \rightarrow b\bar{b}$ and $t\bar{t} \rightarrow l\nu bq\bar{q}'\bar{b}$. Two different analysis approaches are used. In the first (namely "resolved"), the procedure uses standard identification and reconstruction algorithms. In the second (namely "combined"), the events are separated in two exclusive regimes, including events containing the hadronically decaying top quark with a low transverse momentum ($p_{\mathrm{T}} < 250$ GeV) and the boosted ones with the opposite requirement, where the partially overlapped jets coming from high p_{T} tops are suitably identified, reconstructed and tagged. The boosted sample is sensitive to a significantly different kinematic region and provides additional information. The measured signal strength is 1.4±0.5 and 1.2±0.5 respectively for the resolved and the combined analysis. A signal strength larger than 2.3 and 2.0 can be excluded at the 95% confidence level respectively for the resolved and the combined analysis. Both results are compatible with the prediction of the Standard Model ($\mu = 1$).

x

Sommario

In questa tesi è presentata la misura della "*signal strength*" ($\mu = \sigma_{obs}/\sigma_{t\bar{t}H_{SM}}$) e del limite superiore della produzione associata di un bosone Higgs con una coppia di quark top, $pp \rightarrow t\bar{t}H + X$ ad un'energia del centro di massa di 13 TeV. I dati sono stati acquisiti dall'esperimento ATLAS nel 2015 e 2016, corrispondenti ad una luminosità integrata di 36.5 fb^{-1}. La *signal strength* del canale $t\bar{t}H$ e il suo limite sono stati misurati studiando i canali di decadimento $H \rightarrow b\bar{b}$ e $t\bar{t} \rightarrow l\nu bq\bar{q}'\bar{b}$, con due differenti approcci. Nel primo (detto "*resolved*"), sono stati usati gli algoritmi standard di identificazione e ricostruzione come nelle precedenti analisi già pubblicate. Nel secondo (detto "*combined*"), gli eventi sono separati in due regimi esclusivi, includendo quelli contenenti il quark top adronico con un basso momento trasverso ($p_{\mathrm{T}} < 250$ GeV) e quelli boosted con la richiesta opposta, dove i jet parzialmente sovrapposti provenienti da top con alto p_{T} vengono ricostruiti e identificati con metodi innovativi. Il campione di eventi *boosted* è sensibile ad una regione cinematica significativamente differente e, per questo, fornisce informazioni aggiuntive. La *signal strength* misurata è 1.4±0.5 e 1.2±0.5 rispettivamente per l'analisi *resolved* e *combined*. Una *signal strength* più grande di 2.3 e 2.0 può essere esclusa con un livello di confidenza del 95%, rispettivamente per l'analisi *resolved* e *combined*. Entrambi i risultati sono compatibili con la previsione del Modello Standard ($\mu = 1$).

Introduction

During the last 60 years, much effort has been devoted to the search for a theoretical model able to explain the fundamental interactions among fundamental particles. The Standard Model of particle physics represents the most complete theory and, at the same time, it is greatly supported by experimental evidences. This theory predicts the existence of a particle responsible for giving mass to all the fundamental massive particles of the Standard Model: the Higgs boson. The measurement of all the properties of this new particle is essential for a confirmation of the Standard Model or for the discovery of new physics processes.

The Large Hadron Collider, located at CERN laboratories in Geneva, is the largest particle accelerator in the world and it is designed to reach energies of 13 TeV in the center of mass of the colliding protons, allowing the search for processes at the energy frontier such as the Higgs production. On the 4th of July 2012 the CERN laboratories announced the discovery of a 125 GeV mass resonance with properties similar to the SM Higgs boson with a confidence level of more than 5σ. Of particular importance among the main Higgs boson production mechanisms at LHC is the $t\bar{t}H$ channel, in which the Higgs boson is produced in association with a top quark pair. Although the production cross section for the $t\bar{t}H$ channel is the smallest ($\sigma_{t\bar{t}H} \sim 500$ fb, at 13 TeV) compared to the other Higgs production mechanisms, it deserves special attention for many reasons. First of all, the $t\bar{t}H$ cross section is expected to increase faster with respect to the other Higgs production channels, as a function of the centre-of-mass energy of the pp collision. This allows to obtain a better precision on the upper limit of the signal strength ($\mu = \sigma_{obs}/\sigma_{t\bar{t}H_{SM}}$) with respect to the previous data taking carried on 2012 at 8 TeV and, consequently, to have a stronger check on the SM predictions. The increase of energy at LHC gives the opportunity to face a completely new physics regime never explored before, opening an exceptional window on possible new physics phenomena. For this reason it is crucial also to focalize the measurements in a boosted regime selecting the events with the involving particles emitted at high energy.

Another important aspect of this channel is the possibility to have a direct measurement of the quark top-Higgs Yukawa coupling which is expected to be the largest among the other particles because of the large mass (173 GeV) of top quark. The determination of the $t\bar{t}H$ production cross section will also be crucial for a more precise determination of the intrinsic characteristics of the Higgs boson. In particular the most recent simulations at the Next-to-Leading-Order (NLO) approximation show a clear dependence of the $t\bar{t}H$ cross section by the Higgs parity. With the actual statistic, a determination of the Higgs CP state could be reached by comparing the $t\bar{t}H$ cross section obtained in two different kinematic ranges of the Higgs boson, which further emphasizes the importance to perform the analysis in different kinematic regimes. For these reasons, this channel has become of particular interest in the study of the Higgs boson production with the increasing of LHC luminosity and center-of-mass energy. It presents a quite complex signature with high jet multiplicity and a varying number of leptons in the final state. The different final states are determined from the different decay modes of the top quark pair and of the Higgs boson.

The analysis presented in this work concerns the channel in which the Higgs boson decays in a pair of b quarks and the $t\bar{t}$ system decays semileptonically ($t\bar{t} \rightarrow l\nu bqq'\bar{b}$). The analysis has been performed in two complementary ways. The first follows the procedure used in the previous pubblished papers, namely "resolved", through standard identification and reconstruction algorithms. In the second way, the

events are separated in two exclusive regimes, namely the resolved one including events containing the hadronically decaying top quark with a low transverse momentum ($p_T < 250$ GeV) and the boosted one with the opposite requirement (namely "combined").

The topology of the events belonging to the boosted regime ($p_T > 250$ GeV) is different compared to the resolved one because the decay products of the boosted particles are almost totally collimated, bringing to an overlapping of the decay jets. In this situation, the standard jet reconstruction algorithms loose their efficiency and new innovative techniques are needed. The *boosted techniques* are widely used in this analysis because they exploit the characteristic substructure of the objects (the "large-R jets") that include the decay products.

The low $t\bar{t}H$ production with respect to a wide variety of background processes leads the adoption of the most advanced approaches both regarding the signal identification and its extraction. The process identification has been performed with a MultiVariate Analysis technique in order to determine a discriminating variable distribution that allows to increase the significance of the selection with respect to a cut-based procedure. The signal has been extracted by defining many independent sub-samples with different requirements and selecting the ones with the highest significance. At the end all the selected events have been subject to a fit under a theoretical assumption. In this procedure, the systematic uncertainties have a crucial role and the best way to treat them has to be found in order to correctly consider several uncertainty sources: from the theory, from the detector and from object reconstruction techniques.

The analysis has been performed with the full ATLAS recorded data-set in the first two years of Run-2 (2015 and 2016), corresponding to 36.5 fb^{-1} at 13 TeV. The results on the signal strength for this channel have been obtained both using the standard procedure already published and the combination of the resolved and boosted regimes. Both are, at the moment, the most precise with respect the ones obtained by the ATLAS and CMS collaborations with the previous data taking at 8 TeV and with the first part of data at 13 TeV (13.2 fb^{-1}).

This thesis is organized as follows. Chapter 1 focuses on the theoretical basis of the Standard Model, on the Higgs mechanism and on the physics of the $t\bar{t}H$ production. In Chapter 2, a description of the LHC collider, as well as of the ATLAS detector, is provided. Chapter 3 presents a wide overview of the data and Monte Carlo simulation samples used in this analysis. In Chapter 4 the description of the objects, reconstruction and tagging techniques is discussed. Chapter 5 describes in detail the MultiVariate Analysis and the signal extraction technique, as well as their application in the specific case of the analysis. Chapter 6 discusses the sources of uncertainties, their treatment and the results of the resolved and combined analyses, as well as the future perspectives of the analysis presented in this thesis. The conclusion summaries briefly the main results obtained in the analysis.

For this study, performed almost entirely at CERN, Silvia Biondi covered several aspects through all the analysis chain. The first contribution concerned the development of a general code (Chapter 3) usable also by different analysis teams, and finalized to the pre-selection of the events. She has personally developed part of the analysis codes, based on the official ATLAS framework, in order to integrate the analysis procedure into the ATLAS community for the LHC Run-2 data. She followed the application and optimisation of the top-tagging technique (Chapter 4), developing also the possibility to extend it to the Higgs boson for a future analysis development of the boosted channel. She implemented the setup and the fine-tuning of the MVA technique, together with a study on the performance of the b-tagging algorithm in the boosted regime (Chapter 5). She used the fitting tool to extract the final results on the signal strength and its upper limits (Chapter 6). During her ATLAS qualification task, Silvia Biondi has collaborated to the study on the upgrade of Muon detector for the LHC Phase-2; she optimised and simulated the newly proposed trigger performances, to evaluate the muon efficiency. This work is briefly described in appendix A.

Chapter 1

Top quark, Higgs boson and $t\bar{t}H$ production

The top quark and the Higgs boson are the last discovered and the two heaviest particles of the Standard Model and, correspondingly, the study of their properties and their couplings is one of the most interesting item in particle physics field in these years. During the Run-1 and Run-2 data taking periods, the LHC increased its centre-of-mass energy, $\sqrt{s}$, from 8 to 13 TeV and the prediction of the Higgs boson production in association with a top quark pair (the argument of this thesis) shows the biggest relative increase in cross section with respect to all the other Higgs production processes. Moreover, the $t\bar{t}$ process provides access to a direct measurement of the Higgs-Top Yukawa coupling which is expected to be near 1, due to the large mass of the top quark, and might provide an important hint for the searches on beyond the Standard Model theories, such as the Supersymmetry.

In the next paragraphs, a brief description of the Standard Model, the top quark, the Higgs boson and the $t\bar{t}$ production is given in order to better understand the principal characteristics of the $t\bar{t}H$ channel.

1.1   The Standard Model of particle interactions

The discovery of the electron, in 1897 by Joseph J. Thomson, gave origin to the study of elementary particle physics. Many particles have been discovered during the following years, using initially cosmic rays and, subsequently, accelerators of progressively increasing energy. During the years, the experimental techniques and the intensity of the beams improved allowing a long history of discoveries and knowledge that results in the well-known experimental-theoretical framework called Standard Model (SM). The SM describes, from the 70's on, the fundamental particles and their interactions, up to distances of 1 fm (10^{-15} m), in terms of a specific type of relativistic quantum field gauge theory. At the moment the gravitation is not included in the Standard Model, however its contribution in the particle physics is negligible and this does not degrade the level of prediction.

In the SM, there are 12 fundamental particles, with their 12 related anti-particles, three historical fundamental forces (electromagnetic, strong and weak), 3 sets of gauge bosons (the mediators of the forces) and the Higgs Boson, the last discovered particle that confirms the SM theory.

1.1.1   The fundamental particles

The elementary particles described by the Standard Model can be divided into two main categories: fermions and bosons, reported in figure 1.1. Fermions are spin 1/2 particles, consisting of leptons and
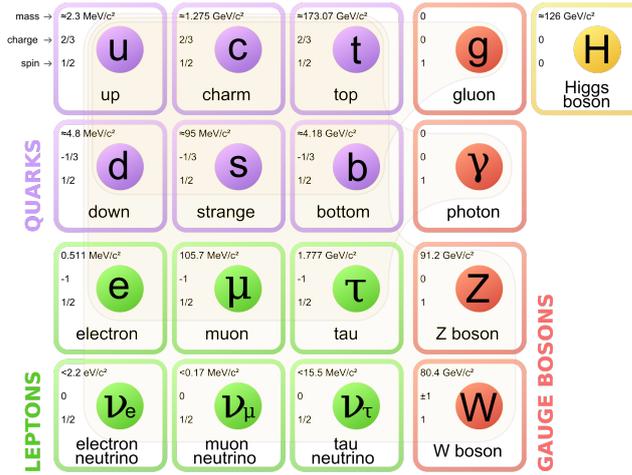
Figure 1.1: The fundamental particles which constitute the SM, divided into fermions (violet and green) and gauge bosons (red). The Higgs boson, responsible for the spontaneous symmetry breaking, is also illustrated. Each particle is reported with its mass value or, in the neutrinos case, with an upper limit on the mass, its spin and electric charge, provided by the Particle Data Group [1].

quarks, which obey to the Fermi-Dirac statistics and to the Pauli exclusion principle.

Leptons are divided into three families by weak interactions:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix} \tag{1.1.1}$$

with a large mass range (from about 0 to almost 2 GeV). The SM regards neutrinos as massless particles, in contrast with the experimental evidence of their oscillation. According to this phenomenon, neutrinos should have mass but, up to now, only limits are provided by direct measurements [1]. Neutrinos are electrically neutral particles and can interact only by means of the weak force. The leptons e, μ and τ have negative charge allowing them to interact both electromagnetically and via the weak force. A leptonic (electronic, muonic and tauonic) number, with value 1 (-1) for leptons (anti-leptons) is associated to each lepton family and it is conserved in all the interactions. Violations of the leptonic number have been observed only in neutrino oscillations.

Quarks are the fundamental fermions forming the hadrons. They can interact via the electromagnetic, weak and strong interactions and differently from the leptons, quarks are not directly observed, because they are always bound in hadrons, due to the so-called *confinement* phenomenon. As for leptons, quarks are divided into three families

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \tag{1.1.2}$$

where u, d, etc., are called flavours. Quarks are the only known particles with fractional electric charge (2/3 for u, c, t and -1/3 for d, s, b). Different quantum numbers can be associated to the quarks: colour, barionic number and flavour:

- the colour (red, blue and green) is the "charge" of the strong force, conserved in all the interactions, and it is the responsible for confinement: quarks cannot be observed in free states but only in a combination which nulls the colour charge;

- the barionic number, which is 1/3 (-1/3) for all quarks (anti-quarks), is additive and conserved by all the three interactions;

- the flavour number is conserved in all interactions except the weak force.

1.1.2   The fundamental interactions

In Nature there are four fundamental forces: the electromagnetic, the weak, the strong and the gravitational, that handle all the interactions between particles and are precisely described by gauge theories. The SM includes all these forces except the gravitational one because at the moment it is described by a quantum theory not renormalizable, nevertheless this does not involve any modification in the SM prediction due to its negligible contribution (about 10^{-36} with respect to the electromagnetic forces at the GeV scale) in the particle physics.

The basis of the gauge theories are laid down by Maxwell in 1864, studying classical electrodynamic; beginning from the Maxwell's equation in absence of sources for the magnetic field $\vec{B}$,

$$\mathrm{div}\vec{B} = \vec{\nabla} \cdot \vec{B} = 0, \tag{1.1.3}$$

with $\vec{\nabla}$ the vector operator *nabla*, $\vec{\nabla} = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$, invites to write the magnetic field as

$$\vec{B} = \mathrm{rot}\vec{A} = \vec{\nabla} \times \vec{A} \tag{1.1.4}$$

where $\vec{A}$ is the vector potential of the $\vec{B}$ field. If a gradient of an arbitrary scalar function Λ is added to the vector potential

$$\vec{A} \rightarrow \vec{A} + \vec{\nabla}\Lambda \tag{1.1.5}$$

the magnetic field results unchanged by this transformation, because

$$\vec{B} = \vec{\nabla} \times (\vec{A} + \vec{\nabla}\Lambda) = \vec{\nabla} \times \vec{A}. \tag{1.1.6}$$

Moving to the equation related to the electric field $\vec{E}$ (in absence of sources)

$$\vec{\nabla} \times \vec{E} = -\partial\vec{B}/\partial t, \tag{1.1.7}$$

which, from eq. (1.1.4), can be rewritten as

$$\vec{\nabla} \times (\vec{E} + \partial\vec{A}/\partial t) = 0. \tag{1.1.8}$$

This suggests the identification

$$\vec{E} + \partial\vec{A}/\partial t = -\vec{\nabla}V, \tag{1.1.9}$$

where V is the scalar potential of the electric field $\vec{E}$. In order to have also $\vec{E}$ invariant under the transformation (1.1.5), it is necessary to require

$$V \rightarrow V - \partial\Lambda/\partial t. \tag{1.1.10}$$

The Maxwell's equations eq. (1.1.3) and (1.1.7) can be compactly expressed in covariant notation:

$$F^{\mu\nu} = \partial^\nu A^\mu - \partial^\mu A^\nu \tag{1.1.11}$$

where μ and ν are four dimensions indices running on the space-time coordinates, $\partial^\mu = (\frac{\partial}{\partial t}, \vec{\nabla})$, $\partial_\mu = (\frac{\partial}{\partial t}, -\vec{\nabla})$, $F^{\mu\nu}$ is the electromagnetic strenght tensor

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} = -F^{\nu\mu} \tag{1.1.12}$$

and A^μ is the four-vector potential

$$A^\mu = (V, \vec{A}). \tag{1.1.13}$$

$F^{\mu\nu}$ is unchanged by the "gauge transformation"

$$A^\mu \to A^\mu - \partial^\mu \Lambda. \tag{1.1.14}$$

Following this covariant notation, the Maxwell's equations (eq. (1.1.3) and (1.1.7) and, consequently (1.1.11)) can be expressed also likewise

$$\partial_\mu {}^*F^{\mu\nu} = 0 \tag{1.1.15}$$

where ${}^*F^{\mu\nu}$ is formally determined from $F^{\mu\nu}$, replacing $\vec{E} \to \vec{B}$ and $\vec{B} \to -\vec{E}$. Obviously, also in this notation, it is unchanged by gauge transformations.

The remaining Maxwell's equations are

$$\vec{\nabla} \cdot \vec{E} = \rho \tag{1.1.16}$$

and

$$\vec{\nabla} \times \vec{B} = \vec{J} + \frac{\partial \vec{E}}{\partial t}, \tag{1.1.17}$$

where $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$ = 1 (μ_0 and ϵ_0 are respectively permeability and permettivity of the vacuum), ρ and $\vec{J}$ are the electric charge density and the current density respectively.

They can be expressed in covariant notation:

$$\partial_\mu F^{\mu\nu} = -J^\nu \tag{1.1.18}$$

with the electromagnetic current given by

$$J^\nu = (\rho, \vec{J}). \tag{1.1.19}$$

Also these Maxwell's equations in covariant notation are unchanged by the gauge transformation (1.1.14) leading to two consequences:

- the first one is the conservation of the electromagnetic current

$$\partial_\nu J^\nu = -\partial_\nu \partial_\mu F^{\mu\nu} = 0; \tag{1.1.20}$$

- the second is that the eq. (1.1.18), in the case of absence of sources and in a Lorenz gauge ($\partial_\mu A^\mu = 0$), satisfy the relation

$$\Box A^\nu = 0 \tag{1.1.21}$$

where $\Box$ is the D'Alembert operator, meaning that each component of the vector potential (identified with the photon field) satisfies the Klein-Gordon equation for a massless particle.

The fact that many different four-vector potentials lead to the same electromagnetic fields and the same physics, without changing the resulting interaction, is a manifestation of the gauge invariance of the classical electrodynamics.

These hidden symmetries at the roots of electromagnetism worth to be described more in detail. When the field describing an interaction changes in each point of the space-time of the same given quantity under certain transformation, its equation presents a global symmetry with respect to that transformation. On the contrary, when the field is invariant under a trasformation which depends on a parameter which is locally defined, its Lagrangian presents a local symmetry. All the fundamental forces of the SM are based on lagrangians which are invariant under local symmetries; according to which particle interactions are described through the exchange of field quanta, called gauge bosons, with integer spin 1. The gauge bosons belonging to the SM [1] are:

- γ (photon) is the massless electromagnetic interaction mediator;

- $W^\pm$ and Z are the weak interaction mediators with a mass, respectively, of 80.34 and 91.19 GeV;

- g (gluons) are the 8 massless strong interaction mediators.

The SM is based on a local symmetry extending the gauge invariance of QED to a number of charges defined in abstract spaces. Since the symmetry breaking involves the electroweak part (the Weinberg-Salam model) of the SM, the stress will go on the electroweak part of the SM Lagrangian. Therefore, to summarize the standard (Weinberg-Salam) model, the Lagrangian [2], taking into account all the ingredients, can be written as

$$
\begin{aligned}
L = &-\frac{1}{4}\boldsymbol{W_{\mu\nu}} \cdot \boldsymbol{W^{\mu\nu}} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \\
&+ \bar{L}\gamma^\mu(i\partial_\mu - g\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{W_\mu} - g^{'}\frac{Y}{2}B_\mu)L + \bar{R}\gamma^\mu(i\partial_\mu - g^{'}\frac{Y}{2}B_\mu)R \\
&+ |(i\partial_\mu - g\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{W_\mu} - g^{'}\frac{Y}{2}B_\mu)\phi|^2 - V(\phi) \\
&- (G_1\bar{L}\phi R + G_2\bar{L}\phi_C R + \text{hermitian conjugate}).
\end{aligned}
\tag{1.1.22}
$$

where $\boldsymbol{W_\mu}$ is the isotriplet vector field coupled to the weak isospin current with coupling factor g, B_μ is the single vector field coupled to the weak hypercharge (Y) current with coupling factor $g^{'}/2$, $\boldsymbol{\tau}$ are the Pauli Matrices, ϕ are four real scalar fields, $V(\phi)$ is the Higgs potential (see paragraph 1.3.1), $G_{1,2}$ are the matrices of Yukawa couplings, L denotes a left-handed fermion (lepton or quark) doublet and R denotes a right-handed fermion singlet.

A brief description of the electroweak part of the Lagrangian of the SM (1.1.22) can be summarized as follow:

- the first line contains the kinetic energy and the self coupling of the $\boldsymbol{W_\mu}$ fields (where $\boldsymbol{W_{\mu\nu}} = \partial_\mu\boldsymbol{W_\nu} - \partial_\nu\boldsymbol{W_\mu} - g\boldsymbol{W_\mu}\times\boldsymbol{W_\nu}$) and the kinetic energy of the B_μ fields (where $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$). The $W^\pm$, Z and γ bosons arise from a linear combination of the $\boldsymbol{W_\mu}$ and B_μ fields;

- the second line contains the kinetic energy of the lepton and quark (the $i\partial_\mu$ term) and their interactions with the $W^\pm$ (the $-g\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{W_\mu}$ term) and with the Z and γ (the $-g'\frac{Y}{2}B_\mu$ term);

- the third line contains the Higgs Mechanism so that the $W^\pm$ and Z become massive and the photon remains massless (see paragraph 1.3.1 for all details);

- the last line finally contains the lepton and quark coupling to the Higgs field to generate their masses.

The Standard Model symmetry group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{1.1.23}$$

is defined by the product of the single interaction ones, where C is the colour charge, L is the left-handed weak-isospin I_3 doublets and Y is the weak hypercharge defined to satisfy the Gell-Mann-Nishijima relation $Q = I_3 + \frac{1}{2}Y$ (Q is the electric charge). In equation (1.1.23):

- $SU(3)_C$ is the non-Abelian group associated to the strong interaction between quarks and gluons in hadrons, governed by the colour charge C; its mediators are eight massless gauge bosons (gluons), which interact with quarks according to quantum chromodynamics (QCD);

- $SU(2)_L \otimes U(1)_Y$ is the symmetry group which describes the electroweak interaction, better known as the Glashow-Weinberg-Salam [3] [4] [5](GWS) theory. The two groups ($SU(2)_L$ and $U(1)_Y$) are associated to the weak ispospin I_3 and the hypercharge Y respectively (the generators of the $SU(2)_L$ and $U(1)_Y$ groups respectively). The vector bosons $W^\pm$, Z and γ arise from a linear combination of a representation of these groups.

A spontaneous symmetry breaking has to be introduced in the SM in order to justify the mass of the weak mediators ($W^\pm$ and Z), which should be massless to preserve the gauge invariance of the model. This spontaneous symmetry breaking, better known as the *Higgs Mechanism*, is shown in more details in section 1.3.1.

1.2 The top quark

The experimental evidence of the top quark arrived only in the 1995, after about 20 years since its prediction, as a consequence of the discovery of the b-quark. Many indirect evidences of the top-quark were obtained from limits on the FCNC (Flavour Changing Neutral Current) and from the determination of the weak isospin $I_3 = -\frac{1}{2}$ of the b-quark, at LEP and SLC (Stanford Linear Collider).

The discovery was announced in 1995 by the CDF [7] and D0 [8] experiments at the proton-antiproton ($p\bar{p}$) collider Tevatron (Fermilab, Chicago). Since its discovery, and using data collected up to 2011 ($\sqrt{s}$ =1.96 TeV), the properties of the top quark and its interactions have been studied in detail and are now being continued at LHC. Differently from $p\bar{p}$ collisions at Tevatron, the LHC collides protons with protons at a higher center-of-mass energy ($\sqrt{s}$ from 7 TeV to 13 TeV). With these features, the $t\bar{t}$ production cross section at LHC is 20 times the one at Tevatron and the statistics is moreover increased by the higher luminosity. The energy and luminosity reached by LHC allow to perform very precise measurements on the top quark especially in kinematic regions never studied before.

Up to now the top quark is one of the most interesting research field in the high energy physics due to its peculiarities: for example, it is an optimal place where to search for new physics phenomena, allowed

by theories but not discovered up to now. One of the most important characteristic of the top-quark is the large mass $m_t = 173.34 \pm 0.27 \pm 0.71$ GeV (figure 1.2 and table 1.1) which leads to various consequences peculiar of this quark. First of all being heavier than a W boson, it is the only quark that decays into a real W boson and a b-quark ($t \rightarrow Wb$), before hadronization can occur. As a consequence, the lifetime of the top quark is very short, $5 \cdot 10^{-25}$ s, to be compared with $\sim 3 \cdot 10^{-24}$ s which is the typical hadronization lifetime; for this reason the top can be studied as a bare quark and not aggregated to another quark inside an hadron.

In addition, it is the only quark whose Yukawa coupling to the Higgs boson, $y_t = m_t/v$ (where $v \simeq 246$ GeV is the vacuum expectation value), is order of unity (see section 3.4).
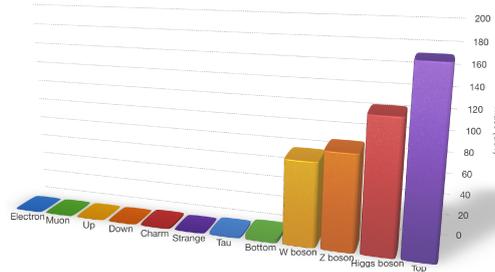
Figure 1.2: Comparison between the masses of the fundamental particles of the Standard Model.

Particles	Mass
electron	$0.510998928 \pm 0.000000011$ MeV
muon	$105.6583715 \pm 0.0000035$ MeV
up	$2.3^{+0.7}_{-0.5}$ MeV
down	$4.8^{+0.7}_{-0.3}$ MeV
strange	95 ± 5 MeV
charm	1.275 ± 0.025 GeV
tau	1776.82 ± 0.16 MeV
bottom	4.18 ± 0.03 GeV
W boson	80.385 ± 0.015 GeV
Z boson	91.1876 ± 0.0021 GeV
Higgs boson	125.6 ± 0.3 GeV
top	$173.34 \pm 0.27 \pm 0.71$ GeV

Table 1.1: Masses of the leptons, quarks and bosons [1], in ascending order. For the Higgs boson, the reported mass is the combination of the ATLAS and CMS results [1]. For the top quark, the reported mass is the combination of Tevatron and LHC (ATLAS and CMS collaborations) results.

Furthermore, the top quark contributes substantially more than other quarks in higher order loop diagrams of the electroweak theory, which implies that m_t is a crucial parameter of this theory. Precise measurements of m_t provide indirect constraints on the mass of the Higgs boson, together with other parameters of the electroweak theory as the mass of the W-boson m_W. For these reasons, the top quark plays a special role in the SM [1]. An accurate knowledge of its properties (mass, couplings, production cross section, branching ratios decay, etc.) can bring key information on fundamental interactions at the electroweak breaking scale and beyond.

The top quark plays also an important role in many scenarios for new physics beyond the SM [1], [9]; this constitutes one of the main motivations for the top quark physics program at the Large Hadron Collider (LHC). Several models predict the existence of new particles decaying predominantly into top quark pairs, making attractive the searches for resonances in the top quark pair invariant mass distribution. New particles may also be produced in top quark decays, for instance a real charged Higgs boson, as in the process $t \rightarrow H^+ b$, since $m_t > m_{H^+} + m_b$, supposing m_{H^+} similar to m_H. In addition, precise measurements of the properties of the top quark and its interactions may reveal effects from new physics. This concerns in particular the study of differential distributions, such as the asymmetry in the rapidity distributions of top quark and anti-quark, but also the search for FCNC in top quark decays (that means considering different decay modes, as $t \rightarrow Zc, u$) and for the production of same-sign top quark pairs.

1.2.1 Top-quark production

Top quarks can be produced in pairs $t\bar{t}$ via the strong interactions or in singlet via the electroweak interaction. Though the first production mode is dominant, and it will be discussed in detail in section 1.4, there is a significant number of top quarks which are produced singly. In this case, the production of the top quark involves mostly a vertex with a quark b and a W boson, since $|V_{tb}| \gg |V_{td}|, |V_{ts}|$, as can be seen by the flavour changing coefficients given by the Cabibbo-Kobaiashi-Maskawa (CKM) matrix (1.2.24) [1]:

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix} \quad (1.2.24)$$

Three different production modes are possible, as shown in figure 1.3:

- the t-channel mode, where a space-like W-boson is scattered off by a b-quark, which is either considered through the b-quark PDF in the proton (flavor excitation, massless scheme) or produced via gluon splitting $g \rightarrow b\bar{b}$ (W-gluon fusion, massive scheme);

- the s-channel mode, where a time-like W-boson is produced from two quarks belonging to an isospin doublet, e.g., $u\bar{d}$, and subsequently decays into $t\bar{b}$;

- the Wt-channel mode, also called associated production, where the top quark is produced in association with a close-to real W-boson.

a   **b**   **c**
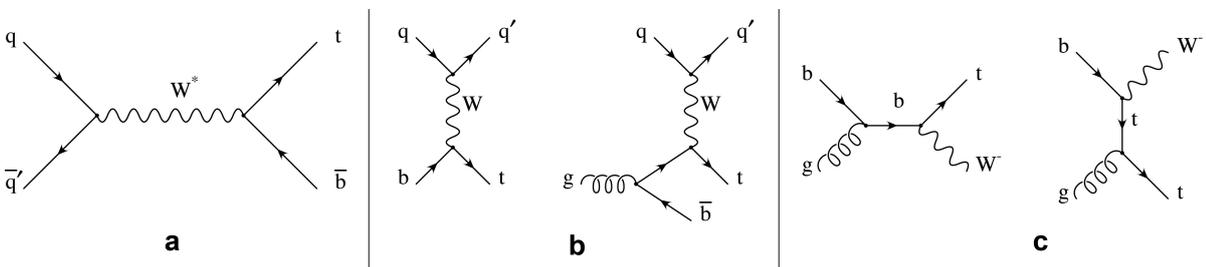
Figure 1.3: Example of Feynman diagrams for single top quark production at LO QCD. From left to right: s-channel, t-channel and Wt-channel production.

Single top quark production is important for various reasons: first of all, its proof of existence provides a relevant test of the Standard Model, because it is the only way to measure V_{tb} directly and, moreover,

the three production modes are all sensitive to the tWb vertex but in different ways. Besides, new physics could appear in single top events because it provides the possibility to investigate the structure of tWb coupling, FCNC, fourth family of the quarks and the existence of W'.

The dominant production mode is the t-channel, observed at both Tevatron and LHC; on the contrary, the s-channel gives the smallest contribution, while the Wt-channel has a very small cross section at Tevatron but it becomes significant at LHC. Fig. 1.4 provides a summary of all single top cross section measurements at the Tevatron and the LHC as a function of the center-of-mass energy. All cross section measurements are very well described by the theory calculation within their uncertainty, providing an important test of the SM. The complete procedure, and the relative formula, to calculate the cross section is explained in detail in section 1.4.2.
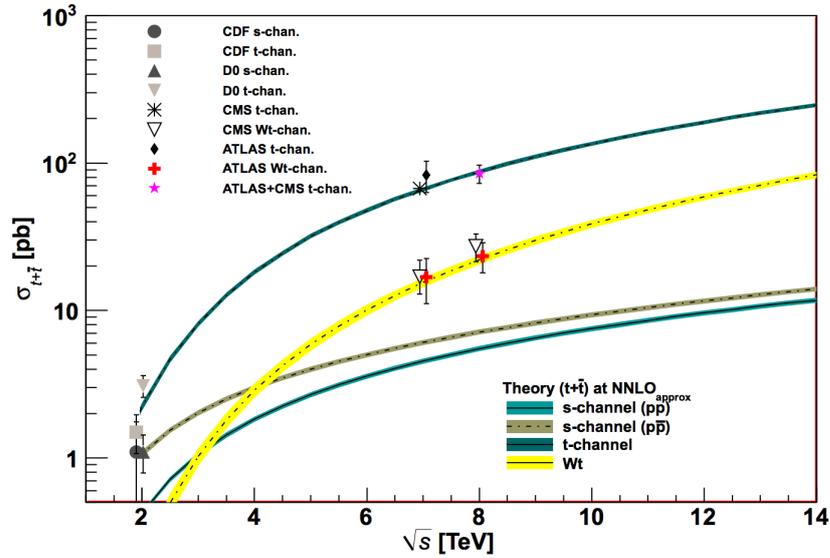
Figure 1.4: Measured and predicted single top production cross sections from Tevatron energies in $p\bar{p}$ collisions $\sqrt{s} = 1.96$ TeV to LHC energies $\sqrt{s} = 7$ TeV in pp collisions.

At the LHC, the t-channel mode of single top quark production has the cleanest signature with a light quark jet recoiling against the top quark; the latest results obtained at $\sqrt{s} = 8$ TeV by ATLAS with an integrated luminosity $\mathcal{L} = 5.8$ fb^{-1} [10] and by CMS with $\mathcal{L} = 19.7$ fb^{-1} [11] are:

$$\sigma_t^{ATLAS}(t\text{-channel}) = 95 \pm 2(stat) \pm 18(sys) \quad \text{pb}$$
$$\sigma_t^{CMS}(t\text{-channel}) = 83.6 \pm 2.3(stat) \pm 7.4(sys) \quad \text{pb}$$

(1.2.25)

The Wt-channel is important because it interferes at NLO QCD with the top-quark pair production. Different techniques have been developed in the Monte Carlo simulations in order to provide an unambiguous signal definition. During the Run-1, both ATLAS, at $\sqrt{s} = 7$ TeV, and CMS, at $\sqrt{s} = 8$ TeV (for the first time at LHC) have measured the process cross section [12]:

$$\sigma_t^{ATLAS}(Wt\text{-channel, 7 TeV}) = 16.8 \pm 2.9(stat) \pm 4.9(sys) \quad \text{pb}$$
$$\sigma_t^{CMS}(Wt\text{-channel, 8 TeV}) = 23.4 \pm 5.5(stat) \pm 5.4(sys) \quad \text{pb}$$

(1.2.26)

The s-channel has been measured at Tevatron and LHC (only by the ATLAS experiment) [13] [14]:

$$\sigma_t^{CDF,D0}(s\text{-channel, 1.96 TeV}) = 1.29^{+0.26}_{-0.24} \quad \text{pb}$$
$$\sigma_t^{ATLAS}(s\text{-channel, 8 TeV}) = 4.8^{+1.8}_{-1.6} \quad \text{pb}$$

(1.2.27)

Table 1.2 [15] summarizes the approximate NNLO cross section for both s-, t- and Wt channels at Tevatron and LHC (at $\sqrt{s}$ = 7 TeV). At both the Tevatron and the LHC, the t-channel cross section is the largest among the three single top production modes. The s-channel yields the second largest contribution at Tevatron, while it is the smallest at LHC because the $\bar{u}$ and $\bar{d}$ at the initial states come from the quark-gluon sea of the proton.

	Tevatron (pb)	LHC (pb)	
	$t(\bar{t})$	t	$\bar{t}$
t-channel	$2.08^{+0.00}_{-0.04} \pm 0.12$	$41.7^{+1.6}_{-0.2} \pm 0.8$	$22.5 \pm 0.5^{+0.7}_{-0.9}$
s-channel	$1.046^{+0.002+0.060}_{-0.010-0.056}$	$3.17 \pm 0.06^{+0.13}_{-0.10}$	$1.42 \pm 0.01^{+0.06}_{-0.07}$
Wt-channel	$0.28 \pm 0.04^{+0.02}_{-0.04}$	$7.8 \pm 0.2^{+0.5}_{-0.6}$	$7.8 \pm 0.2^{+0.5}_{-0.6}$

Table 1.2: Approximate NNLO cross-section (in pb) of the three single-top production modes at the Tevatron ($\sqrt{s}$ = 1.96 TeV) and at the LHC ($\sqrt{s}$ = 7 TeV). For LHC, cross sections corresponding to exclusive t and $\bar{t}$ productions are shown separately to present the charge asymmetry in the t- and s-channels. The two uncertainties correspond to the scale variation and the PDF uncertainty, respectively.

1.2.2 Top-quark decays

The top quark decays almost exclusively in $t \rightarrow Wb$, as highlighted by the CKM matrix (1.2.24), where $|V_{tb}| \gg |V_{td}|, |V_{ts}|$, strongly suppressing the decays $t \rightarrow W(d,s)$. The Feynman diagrams of the top and antitop decay is shown in figure 1.5. The total width of the top quark Γ_t is a fundamental property
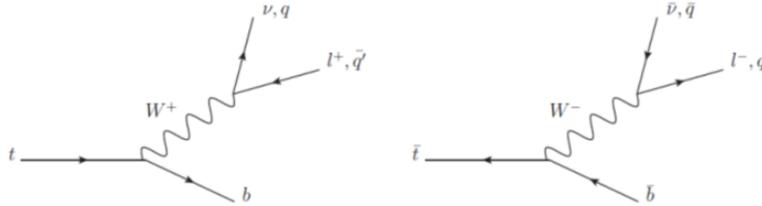
Figure 1.5: Leading decays of top and anti-top quarks.

because it is the only quark that decays before the hadronization, due to its extremely short lifetime (see paragraph 1.2). The theoretical calculation of the total decay width at NLO QCD, neglecting the $t \rightarrow W(d,s)$ decays [17], gives

$$\Gamma_t = \frac{G_F m_t^3}{8\pi\sqrt{2}} |V_{tb}|^2 \left(1 - \frac{m_W^2}{m_t^2}\right)^2 \left(1 + 2\frac{m_W^2}{m_t^2}\right)^2 \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right] \tag{1.2.28}$$

where G_F is the Fermi constant ($G_F/(\hbar c)^3 = 1.166 \cdot 10^{-5}$ GeV^{-2} [1]). The last estimation, obtained as an average of several measurements, is $\Gamma_t = 2.0 \pm 0.5$ GeV [1] leading to a very short lifetime $\tau_t \simeq \frac{1}{\Gamma_t} = 5 \cdot 10^{-25}$ s. Due to this extremely short lifetime, the top quark is the only one that can decay before hadronize. In fact, the typical time of hadron formation is $\sim 3 \cdot 10^{-24}$ s, which is about a factor 6

larger than the top-quark lifetime. Another consequence is that no *toponium* ($t\bar{t}$) bounded state can exist, in a very simple classical simplification (not considering the quantum mechanics), because it could live much less than an orbit revolution time. For this reason the spin information of the top quark is exactly transferred to its decay products.

1.2.3   Top-quark mass

The measurement of the mass of the top-quark has been performed by four experiments, at Tevatron (CDF and D0) and LHC (ATLAS and CMS), and the most recent average value is $m_t = 173.34 \pm 0.27 \pm 0.72$ GeV [1]. The top-quark mass has been directly measured in the production of a $t\bar{t}$ pair exploiting all the possible decay modes. The W decays in a couple of quarks (labelled as jets) or in a lepton and a neutrino (labelled as lepton), so the possible decay channels of a $t\bar{t}$ pair are the lepton+jets, the dilepton, and the all-jets channels (see paragraph 1.4 for a more detailed discussion). The results [1] are summarized in Table 1.3, for both LHC Run-1 and Run-2, and CDF and D0, and illustrated in figure 1.6, for Tevatron and LHC Run-1 only. The lepton+jets channel still yields the most precise measurements because of a very good signal to background ratio and the presence of only a single neutrino in the final state. The momentum of a single neutrino can be reconstructed (up to a quadratic ambiguity) via the missing E_T measurement plus the constraint that the lepton and neutrino momenta derive from the known W boson mass. Since LHC provides very large data samples, measurements in the dilepton channel have the same precision as the other channels, thanks to its very clean signature and the high statistics.

m_t (GeV)	Experiment	Luminosity (fb^{-1})	Channel
$174.94 \pm 1.14 \pm 0.96$	D0 Run-2	3.6	lepton+jets
$172.85 \pm 0.71 \pm 0.85$	CDF Run-2	8.7	lepton+jets
$173.93 \pm 1.64 \pm 0.87$	CDF Run-2	8.7	Missing E_T+jets
$172.5 \pm 1.4 \pm 1.5$	CDF Run-2	5.8	All jets
$172.31 \pm 0.75 \pm 1.35$	ATLAS	4.7	lepton+jets
$173.09 \pm 0.64 \pm 1.50$	ATLAS	4.7	dilepton
$174.9 \pm 2.1 \pm 3.8$	ATLAS	2.04	All jets
$173.49 \pm 0.43 \pm 0.98$	CMS	5.0	lepton+jets
$172.5 \pm 0.4 \pm 1.5$	CMS	5.0	dilepton
$173.49 \pm 0.69 \pm 1.21$	CMS	3.54	All jets
$173.34 \pm 0.27 \pm 0.71$	Tevatron+LHC	$\leq 8.7 + \leq 4.9$	

Table 1.3: Measurements of top-quark mass from Tevatron and LHC experiments. Statistical uncertainties are listed first, followed by the systematic ones.

Indirect constraints on m_t can be obtained from precision measurements of the parameters of the electroweak theory. The mass of the W-boson [2] can be expressed as a function of the electromagnetic coupling $\alpha(M_Z^2)$, the Fermi constant G_F and the electroweak mixing angle θ_W

$$m_W^2 = \frac{\pi \alpha(M_Z^2)/(\sqrt{2}G_F)}{\sin^2 \theta_W (1 - \delta_r)}, \tag{1.2.29}$$

where

$$\sin^2 \theta_W = 1 - \frac{m_W^2}{m_Z^2}. \tag{1.2.30}$$
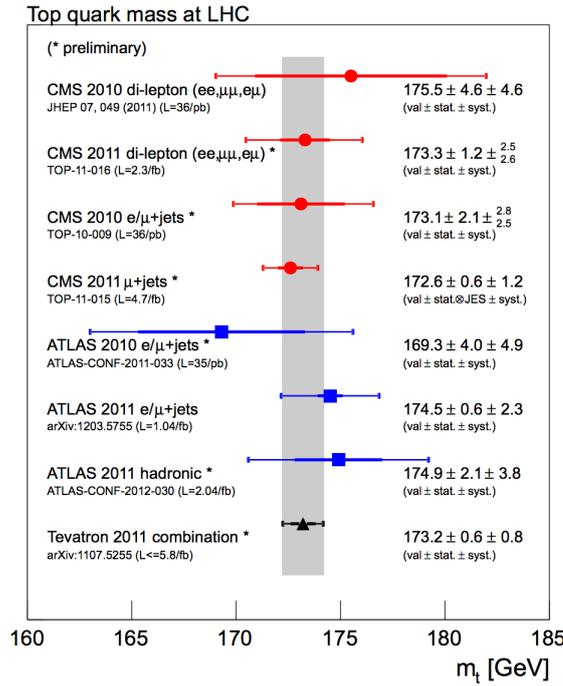
Figure 1.6: Summary of mass measurements of the top-quark performed at LHC Run-1, compared with the Tevatron average.

The term δ_r contains contributions from higher order electroweak loop diagrams involving the top quark which depend quadratically on m_t. The most recent indirect constraint on m_t based on electroweak precision measurements is $m_t = 179.7^{+11.7}_{-8.7}$ GeV [9], in good agreement with the direct measurements.

1.3 The Higgs boson

In the last decades, the hypothesis that the strong, weak, electromagnetic and gravitational interactions are just different aspects of a single universal interaction has grown. This unified interaction, however, would be manifested at some hugely high energy. At the energies reached in laboratories, it is necessary to assume that this symmetry is broken.

The electromagnetic and weak forces had been unified in the '60s, thanks to the works of Glashow, Weinberg and Salam. They showed how it would be possible to treat electromagnetic and weak interactions as different aspects of a single *electroweak interaction*, with two couplings and four mediating bosons (three weak mediators, $W^{\pm}$ and Z and an electromagnetic one, γ). The main problem was the mass of the mediators of the weak field that for a gauge field theory should be zero and with an infinite interaction range. This is in contrast with the experimental observations that mediators have an high mass and consequently a short range for the weak interactions. According to the Glashow-Weinberg-Salam model, the symmetry between electromagnetic and weak interactions would be manifest at very large momentum transfers ($q^2 \gg 10^4$ GeV^2), while at low energies, it would be a *spontaneous symmetry breaking*, that provide a mass to the bosons without spoiling the renormalisability of the theory. In 1964 Higgs [18], Brout and Englert [19] postulated the existence of a scalar field responsible of this spontaneous symmetry breaking, the so-called *Higgs Mechanism*.

1.3.1   The Higgs mechanism

Let's consider a system described by a lagrangian L and its symmetries. If the system has a non-degenerate energy level, the energy eigenstate will be unique and invariant for the symmetries of the Lagrangian. On the contrary, if that level is degenerate, the correspondent eigenstates are not invariant under those transformations.

In each system, the lowest level of energy is the ground state (or vacuum); if it is non-degenerate it will possess the same symmetries of the Lagrangian, on the contrary, if it is degenerate, there will not be only one eigenstate for the representation of the lowest energy level. Any of the degenerate states of the fundamental level might not possess the symmetries of the Lagrangian anymore. The realization of an asymmetric state is known as spontaneous symmetry breaking.

In order to better understand the mechanism, one can consider a scalar field ϕ that interacts with itself. For this field, the Lagrangian [20] is

$$L = \frac{1}{2}(\partial_\mu \phi)^2 - V(\phi) \tag{1.3.31}$$

choosing a potential V in order to respect the invariance of L under the symmetry operation that replaces $\phi \to -\phi$, of the type

$$V(\phi) = \frac{1}{2}\mu^2 \phi^2 + \frac{1}{4}\lambda \phi^4 \tag{1.3.32}$$

where μ is the scalar particle mass and λ is a dimensionless positive constant, representing the coupling of the 4-boson vertex.
The minimum value of V (the ground state) occurs at $\phi = \phi_{min}$, corresponding to $\partial V / \partial \phi = 0$, that is

$$\phi(\mu^2 + \lambda \phi^2) = 0 \tag{1.3.33}$$

If $\mu^2 > 0$, the system describes a scalar field with mass μ and the potential $V(\phi)$ exhibits the ground
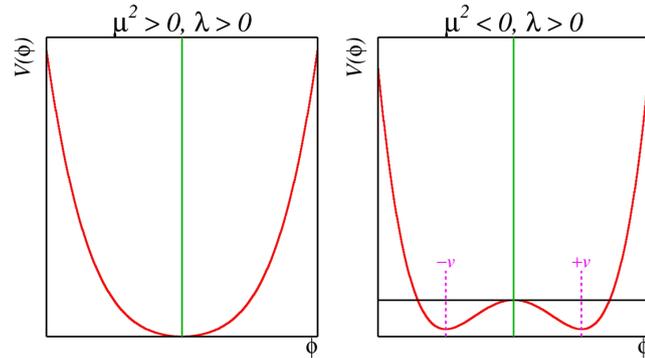
Figure 1.7: $V(\phi)$ potentials as a function of the complex scalar field ϕ in $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right) cases (Higgs potential).

state on the origin at $\phi_{min} = 0$ (figure 1.7, left) and it obeys the reflection symmetry of the Lagrangian. If $\mu^2 < 0$, $\phi = 0$ is a relative maximum (figure 1.7, right) and the potential has minima in

$$\phi_{min} = \pm\sqrt{\frac{-\mu^2}{\lambda}} = \pm v \tag{1.3.34}$$

where v is called the vacuum expectation (the energy of the ground state) of the scalar field ϕ. Eq. (1.3.34) shows two degenerate lowest-energy states, either of which may be chosen to be the vacuum. From the relation

$$v^2 = \frac{1}{\sqrt{2}G_F}, \tag{1.3.35}$$

obtainable from the Weinberg-Salam model, the vacuum expectation value results to be not null and in particular to be equal to $\simeq 246$ GeV. Because of the parity of the Lagrangian ($V(\phi) = V(-\phi)$), the physical consequences must be independent of this choice; whatever it is, the symmetry is spontaneously broken, meaning that the parity transformation is an invariance of the Lagrangian but not of the vacuum state.

Without loss of generality it is possible to choose

$$\phi = +v = \sqrt{\frac{-\mu^2}{\lambda}}. \tag{1.3.36}$$

Applying perturbative calculations around the minimum, ϕ becomes

$$\phi = v + \eta(x) \tag{1.3.37}$$

where $\eta(x)$ represents the quantum fluctuations about this minimum. Substituting (1.3.37) in the Lagrangian (1.3.31) the new Lagrangian is

$$L' = \frac{1}{2}(\partial_\mu \eta)^2 - \lambda v^2 \eta^2 + \text{const.} + \text{cubic and quartic terms on } \eta \tag{1.3.38}$$

where const contains all the constant terms not depending on η. Identifying the first term of the (1.3.38) with the kinematic energy and the second one with the mass term of the field η (that has the correct sign), it gives:

$$m_\eta = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}. \tag{1.3.39}$$

The higher-order terms in η represent the interaction of the η field with itself. The way this mass is generated (or, better, reveled) is the *spontaneous symmetry breaking*. In the L' version of the scalar theory, the reflection symmetry of the Lagrangian has been broken by the choice of the ground state $\phi = +v$, rather than $\phi = -v$.

The mechanism to generate a mass for the gauge boson can be generalized to a complex scalar field

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \tag{1.3.40}$$

described by the Lagrangian

$$L = (\partial_\mu \phi)^*(\partial^\mu \phi) - \frac{1}{2}\mu^2 \phi^* \phi - \frac{1}{4}\lambda(\phi^* \phi)^2 \tag{1.3.41}$$

which is invariant under the U(1) global gauge symmetry $\phi \to e^{i\alpha}\phi$. In this case there is a circle of radius v of degenerate minima of the potential $V(\phi)$ in $\phi_1\phi_2$ plane (see figure 1.8), such that $\phi_1^2 + \phi_2^2 = v^2$ with $v^2 = -\mu^2/\lambda$. As before, expanding the $\phi(x)$ field around a minimum energy position

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\epsilon(x)), \tag{1.3.42}$$
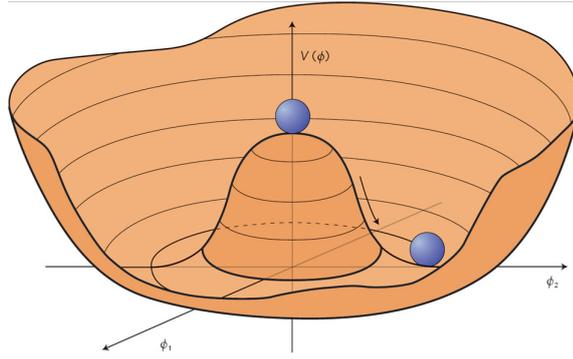
Figure 1.8: $V(\phi)$ potential as a function of the complex scalar field ϕ in $\mu^2 < 0$ case.

where $\eta(x)$ ($\epsilon(x)$) is the shift along the ϕ_1 (ϕ_2) direction. Substituting eq. (1.3.42) in eq. (1.3.41), the Lagrangian becomes

$$L' = \frac{1}{2}(\partial_\mu \epsilon)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu^2 \eta^2 + \text{const} + \text{cubic and quartic terms on } \eta \text{ and } \epsilon, \tag{1.3.43}$$

where the first two terms are the kinematic energy of the ϵ and η fields and the third term is the mass term for the η field

$$m_\eta = \sqrt{-2\mu^2}, \tag{1.3.44}$$

(as the scalar field case). L' does not contain a mass term for the $\epsilon(x)$ field, meaning that the theory has also a massless scalar field (Goldstone boson).

In order to obtain a massive scalar field, it is necessary to require that the Lagrangian is invariant under a U(1) local gauge transformation $\phi(x) \rightarrow e^{i\alpha(x)}\phi(x)$ and substituting the fields $\eta(x)$ and $\epsilon(x)$ of the eq. (1.3.42) with the field $h(x)$ and $\theta(x)$ as

$$\phi(x) = \frac{1}{\sqrt{2}}(v + h(x))e^{\frac{i\theta(x)}{v}} \tag{1.3.45}$$

with $\theta(x)$ chosen so that $h(x)$ is real and requiring ∂_μ to be replaced by the covariant derivative $D_\mu = \partial_\mu - ieA_\mu$.

Neglecting the intermediate steps, the Lagrangian will be independent on θ with the effect that the Goldstone boson disappears and now it contains two interacting massive particles, one relative to the scalar $h(x)$ field (the Higgs particle) and a vector gauge boson for the potential A_μ; this process is the Higgs mechanism for the U(1) gauge symmetry.

The generalization of the Higgs mechanism for a SU(2) gauge symmetry group is obtained by defining a complex doublet scalar field

$$\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \sqrt{\frac{1}{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.3.46}$$

and taking the obvious extension of the Lagrangian (1.3.42) as

$$L = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{1}{4}\lambda(\phi^\dagger \phi)^2. \tag{1.3.47}$$

The Lagrangian (1.3.47) is invariant under the gloabl SU(2) phase transformation

$$\phi = e^{i\alpha_a \frac{\tau_a}{2}}\phi \tag{1.3.48}$$

with a = 1, 2, 3 and τ_a are a set of three traceless 3×3 matrices.

To obtain the invariance of the eq. (1.3.47) for the local SU(2) phase transformation

$$\phi(x) = e^{i\alpha(x)_a \frac{\tau_a}{2}}\phi(x), \tag{1.3.49}$$

it is necessary to replace ∂_μ by the covariant derivative $D_\mu = \partial_\mu + ig\frac{\tau_a}{2}W_\mu^a$, where W_μ^a are the 3 gauge fields.

Considering an infinitesimal transformation of (1.3.49) as

$$\phi(x) = \left[1 + \frac{i\alpha(x)\tau}{2}\right]\phi(x) \tag{1.3.50}$$

the Lagrangian (1.3.47) becomes

$$L = (\partial_\mu\phi + ig\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{W_\mu}\phi)^\dagger(\partial^\mu\phi + ig\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{W^\mu}\phi) - \frac{1}{2}\mu^2\phi^\dagger\phi - \frac{1}{4}\lambda(\phi^\dagger\phi)^2 - \frac{1}{4}\boldsymbol{W_{\mu\nu}}\cdot\boldsymbol{W^{\mu\nu}}, \tag{1.3.51}$$

with $V(\phi) = \frac{1}{2}\mu^2\phi^\dagger\phi + \frac{1}{4}\lambda(\phi^\dagger\phi)^2$ and where there is the presence, in addition, of the kinetic energy terms of the gauge field $\boldsymbol{W_{\mu\nu}}$; all the terms have been already explained in the paragraph 1.1.2.

If $\mu^2 > 0$, the Lagrangian describes a system of four scalar particles (ϕ_i of eq. (1.3.46)), each of mass μ, interacting with three massless gauge bosons (W_μ^a).

If $\mu^2 < 0$ and $\lambda > 0$, the potential $V(\phi)$ of the lagrangian (1.3.51) has its minimum at

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} \tag{1.3.52}$$

Since it is necessary to expand $\phi(x)$ around a particular minimum, it is possible, without loss of generality, to choose

$$\phi_1 = \phi_2 = \phi_4 = 0; \qquad \phi_3^2 = -\frac{\mu^2}{2\lambda} \equiv v^2. \tag{1.3.53}$$

The expansion about this particular vacuum state

$$\phi_0 \equiv \sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.3.54}$$

can be described as below:

$$\phi(x) \equiv \sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.3.55}$$

That is, of the four scalar fields, the only one that remains is $h(x)$, the Higgs field. As in the previous case, the massive scalar h is obtained substituting the expansion (1.3.55) in the $V(\phi)$ potential and keeping only the square term of the $h(x)$ field.

To determine the gauge bosons W_μ^a, it is enough to substitute ϕ_0 of the eq. (1.3.54) in the Lagrangian (1.3.51) and keeping the relevant term

$$\left|ig\frac{1}{2}\boldsymbol{\tau}\cdot\boldsymbol{W_\mu}\phi\right|^2 = \frac{g^2}{8}\left|\begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & W_\mu^3 \end{pmatrix}\begin{pmatrix} 0 \\ v \end{pmatrix}\right|^2 = \frac{g^2v^2}{8}[(W_\mu^1)^2 + (W_\mu^2)^2 + (W_\mu^3)^2]. \tag{1.3.56}$$

where here $|\quad|^2$ has been used as shorthand for $(\quad)^\dagger(\quad)$, that describes the three massive gauge fields, with a mass $M = \frac{1}{2}vg$.

The final extension to a $SU(2)_L \times U(1)_Y$ group symmetry is obtained generalizing the Lagrangian (1.3.51) to

$$L' = |(i\partial_\mu - g\boldsymbol{T} \cdot \boldsymbol{W_\mu} - g'\frac{Y}{2}B_\mu)\phi|^2 - \frac{1}{2}\mu^2\phi^\dagger\phi - \frac{1}{4}\lambda(\phi^\dagger\phi)^2 \qquad (1.3.57)$$

where $|\quad|^2 = (\quad)^\dagger(\quad)$ and $\boldsymbol{T}$ and Y are the generators of the $SU(2)_L \times U(1)_Y$ groups already explained in section 1.1.2 (together with all the other terms).

Neglecting the intermediate steps, the mass terms of the boson mediators are

$$m_\gamma = 0, \qquad m_W = \frac{1}{2}vg, \qquad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}. \qquad (1.3.58)$$

The fermion (lepton and quark) masses are introduced including a $SU(2)_L \times U(1)_Y$ gauge invariant term in the Lagrangian (1.3.57) and substituting the Higgs field with the expansion (1.3.55):

$$L^f = -\frac{g_{ffH}}{\sqrt{2}}\bar{\psi}^f\psi^f - \frac{g_{ffH}}{\sqrt{2}}\bar{\psi}^f\psi^f h(x) \qquad (1.3.59)$$

where g_{ffH} is the Yukawa coupling between the fermion and the Higgs field, the second term in (1.3.59) describes the interaction of any fermion-antifermion couple with the Higgs boson and the mass of the fermion is thus given by

$$M^f = \frac{g_{ffH}}{\sqrt{2}}v \qquad (1.3.60)$$

where v, as already explained, is the vacuum expectation value of the Higgs field ($\simeq 246$ GeV).

1.3.2 The Higgs discovery

As a confirmation of this prediction, the discovery of a new boson was announced by ATLAS [23] and CMS [24] experiments on 4th of July 2012 at CERN. Experiments conducted in last 5 years seem to confirm that this new particle of mass (125.9 ± 0.4) GeV has all the characteristics to be the Higgs boson, but further tests have to be taken into account. ATLAS and CMS performed their researches with similar integrated luminosity, referred to 2011 (4.8 fb^{-1} for ATLAS and 5.1 fb^{-1} for CMS) and 2012 (5.8 fb^{-1} for ATLAS and 5.3 fb^{-1} for CMS) periods. They both observed an excess of events in the invariant mass plots of different Higgs decay channels combining all the measurements of the 2011 and 2012 data taking. They analized data studying different Higgs decay modes: $H \to \gamma\gamma$, $H \to WW^*$, $H \to ZZ^*$, $H \to \tau\tau$ and $H \to b\bar{b}$. In figure 1.9 results are presented: the ATLAS distribution of the four-lepton invariant mass, m_{4l}, deriving from the $H \to ZZ^*$ decay channel, compared to the background expectation in the 80-250 GeV mass region for the combined $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV data [23] and the CMS diphoton invariant mass distribution in the $H \to \gamma\gamma$ decay channel [24].

On 8 October 2013, Peter Higgs and Françoise Englert were awarded with the Nobel Prize. Unfortunately Robert Brout had passed away in 2011 and cannot receive the prestigious award.

1.3.3 Higgs production

The main Higgs production mechanisms are gluon-gluon fusion (ggF), vector boson fusion (VBF), associated production with a gauge boson (VH) and associated production with top quark ($t\bar{t}H$) (figure 1.10) [1].

The study of the Higgs production is very important because it can provide fundamental measurements for the determination of the parameters of the Lagrangian of the SM (1.1.22); in particular, the ggF and the $t\bar{t}H$ processes involve the terms of the fourth line of (1.1.22), giving information on the
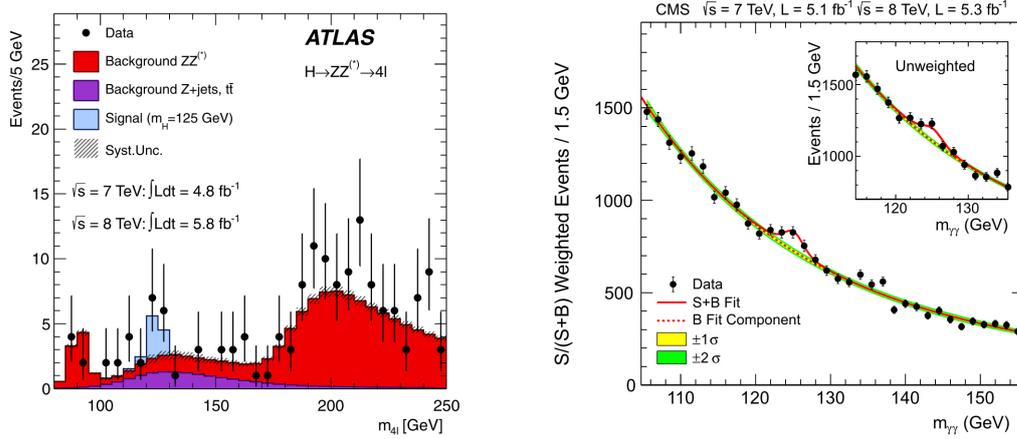
Figure 1.9: The ATLAS and CMS Higgs boson evidence in two different channels. The ATLAS distribution of the four leptons invariant mass, m_{4l} for the combined $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV data for the $H \to ZZ^*$ decay channel (left) and the CMS diphoton invariant mass distribution in the $H \to \gamma\gamma$ decay channel (right).
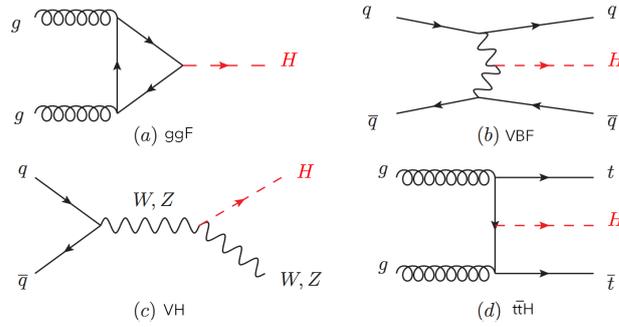
Figure 1.10: Generic LO Feynman diagrams contributing to the Higgs production in (a) gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung (or associated production with a gauge boson) and (d) associated production with top quarks.

Yukawa coupling matrices G_1 and G_2, while VBF and VH involve the third line of (1.1.22), helping the determination of the vector boson couplings with the Higgs boson.

The prediction of the cross sections for the Higgs boson production as a function of $\sqrt{s}$, for pp collisions, including bands indicating the theoretical uncertainties, are summarized in figure 1.11 [25] and in table 1.4, for the different processes.

The $t\bar{t}H$ process (studied in this thesis) has the lower cross section with respect to all the others, but has the larger increase with respect to the others at higher energies.

Gluon-gluon fusion (ggF)

The Higgs boson production mechanism with the largest cross section is the gluon-gluon fusion process, $gg \to H + X$, mediated by the exchange of a virtual, heavy top quark. The ggF accounts for about 87% of Higgs boson production independently by the energy. The theoretical calculation of the cross section for this process is computed at the Next to Next to Leading Order (NNLO) [28]. Among the NLO corrections for the ggF cross section calculation, *virtual* and *real* loops have been considered: while the former does not affect the initial or final states, leading to a process identical to the LO one, the latter
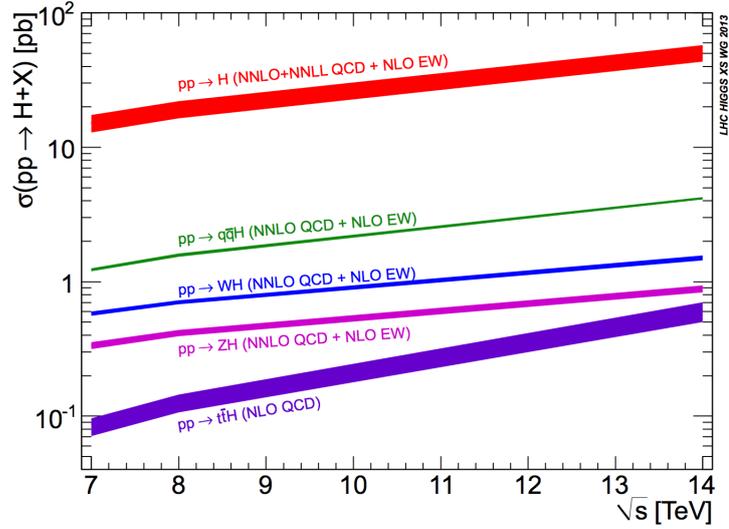
Figure 1.11: The SM Higgs boson production cross sections as a function of the center of mass energy, $\sqrt{s}$, for pp collisions. The theoretical uncertainties [25] are shown as a band.

$\sqrt{s}$ (TeV)	ggF	VBF	WH	ZH	$t\bar{t}H$	total (pb)
7	$15.1 \pm 15\%$	$1.22^{+3\%}_{-2\%}$	$0.58 \pm 4\%$	$0.33 \pm 6\%$	$0.09^{+12\%}_{-18\%}$	17
8	$19.3 \pm 15\%$	$1.58^{+3\%}_{-2\%}$	$0.70^{+4\%}_{-5\%}$	$0.41 \pm 6\%$	$0.13^{+12\%}_{-18\%}$	22
13	$43.9^{+15\%}_{-14\%}$	$3.75 \pm 4\%$	$1.38^{+3\%}_{-4\%}$	$0.87 \pm 6\%$	$0.51^{+15\%}_{-18\%}$	50
14	$49.5^{+15\%}_{-14\%}$	$4.23 \pm 3\%$	$1.52 \pm 3\%$	$0.97 \pm 6\%$	$0.61^{+15\%}_{-18\%}$	57

Table 1.4: The SM prediction for the Higgs boson production cross sections of m_H = 125 GeV in pp collisions, as a function of the center of mass energy, $\sqrt{s}$, [26] and [27].

involves corrections due to addictive partons in the final states ($gg \rightarrow Hg$, $qg \rightarrow Hq$ and $q\bar{q} \rightarrow Hg$). The NLO QCD corrections increase the Higgs production with respect to the leading-order prediction by about 80% and the NNLO corrections further enhance the cross section by approximately 20%.

Vector boson fusion production (VBF)

The second largest Higgs production cross section at LHC is the vector boson fusion. Higgs production via VBF, $qq \rightarrow qqH$, proceeds by the scattering of two quarks, mediated by the exchange of a W or Z boson, with the Higgs boson radiated off the weak-boson propagator. The scattered quarks give rise to two hard jets in the forward and backward regions of the detector. Because of the color-singlet nature of the weak-gauge boson exchange, gluon radiation from the central-rapidity regions is strongly suppressed. For all these reasons, this channel has a distinguishable signature which makes easier the separation between signal and background. Thus, VBF channel provides a particularly clean environment not only for Higgs searches but also for the determination of Higgs boson couplings at LHC.

The production cross section has been calculated with full NLO QCD and EW corrections and ap-

proximate NNLO QCD corrections [29].

WH and ZH associated production (Higgs-Strahlung)

The next most relevant Higgs boson production mechanisms at LHC are the associated production with a W or a Z gauge boson. This mechanism is also known as *Higgs-Strahlung*, where the Higgs boson is irradiated through an off-shell W/Z boson ($pp \rightarrow q\bar{q} \rightarrow W^*(Z^*)H$). The WH and ZH production modes provide a relatively clean environment for studying the decay of the Higgs boson into bottom quarks.

The NNLO QCD prediction on the WH production at LHC contains corrections including the leptonic decays of the W boson and the decay of the Higgs boson into a $b\bar{b}$ pair [30].

Associated production with a pair of quark top ($t\bar{t}H$)

The $t\bar{t}H$ production channel is the channel under analysis in this thesis and it will be presented in detail in section 3.4.

1.3.4 Higgs decays

For the understanding and interpretation of the experimental results, the computation of all the relevant Higgs decay widths is essential, including an estimate of their uncertainties.

The branching ratios for the most relevant decay modes of the Higgs boson as functions of m_H are shown in figure 1.12 [25] and are listed in table 1.5 for a $m_H = 125$ GeV.

The dominant decay modes are $H \rightarrow b\bar{b}$ and $H \rightarrow WW^*$, followed by $H \rightarrow gg$, $H \rightarrow \tau^+\tau^-$, $H \rightarrow c\bar{c}$ and $H \rightarrow ZZ^*$. With much smaller branching ratios, the Higgs decays into $\gamma\gamma$, γZ and $\mu^+\mu^-$.
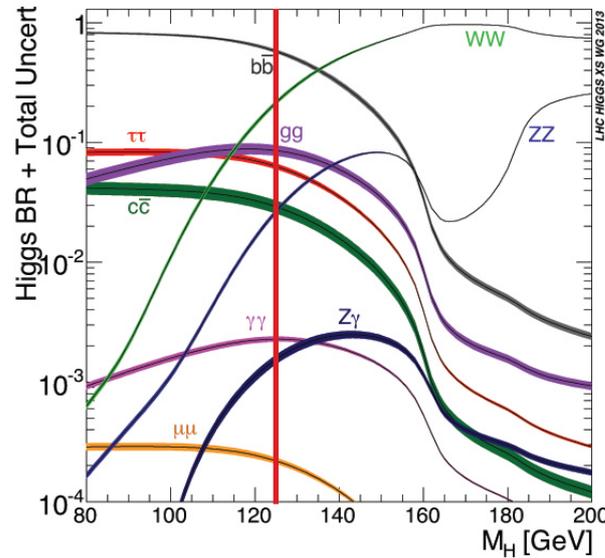
Figure 1.12: The predicted branching ratios for the main decays of the SM Higgs boson as functions of m_H. The theoretical uncertainties are indicated as a band and a vertical line is placed at $m_H = 125$ GeV, close to the experimental value.

Decay channel	Branching Ratio	Rel. uncertainties %
$H \to b\bar{b}$	$5.77 \cdot 10^{-1}$	$^{+3.2}_{-3.3}$
$H \to W^+W^-$	$2.15 \cdot 10^{-1}$	$^{+4.3}_{-4.2}$
$H \to gg$	$8.6 \cdot 10^{-2}$	$^{+10.2}_{-10.0}$
$H \to \tau^+\tau^-$	$6.3 \cdot 10^{-2}$	$^{+5.7}_{-5.7}$
$H \to c\bar{c}$	$2.9 \cdot 10^{-2}$	$^{+12.2}_{-12.2}$
$H \to ZZ$	$2.6 \cdot 10^{-2}$	$^{+4.3}_{-4.1}$
$H \to \gamma\gamma$	$2.3 \cdot 10^{-3}$	$^{+5.0}_{-4.9}$
$H \to Z\gamma$	$1.5 \cdot 10^{-3}$	$^{+9.0}_{-8.9}$
$H \to \mu^+\mu^-$	$2.2 \cdot 10^{-4}$	$^{+6.0}_{-5.9}$

Table 1.5: The predicted branching ratios and the relative uncertainty [31] for a SM Higgs boson with m_H = 125 GeV.

Lepton and quark pair decay channels

In the Born approximation, the width of the Higgs decay into lepton pairs is [32]

$$\Gamma(H \to l^+l^-) = \frac{G_F m_l^2}{4\sqrt{2}\pi} m_H \beta^3 \tag{1.3.61}$$

with $\beta = (1 - 4m_l^2/m_H^2)^{1/2}$, m_l the mass of lepton and m_H the mass of Higgs boson.

For the decay into quark pairs, eq. (1.3.61) has to be corrected by a color factor $N_c = 3$ and by the QCD corrections. The partial decay width becomes

$$\Gamma(H \to q\bar{q}) = \frac{3G_F m_q^2}{4\sqrt{2}\pi} m_H \beta^3 [1 + \frac{4}{3}\frac{\alpha_s}{\pi}\Delta_H^{QCD}] \tag{1.3.62}$$

where m_q is the quark mass, Δ_H^{QCD} is the QCD correction and α_S is the strong force coupling. The QCD corrections in eq. (1.3.62) are not negligible since in the limit $m_H \gg m_q$ the decay width receives many contributions which, in the case of the b quark (the one studied in this thesis) decrease the $H \to b\bar{b}$ decay width by more than 50%.

Among the quark decay channels, the $b\bar{b}$ has the highest BR, but it is not measurable in all the production mechanisms; in the ggF production, for example, it would be totally overwhelmed by background processes, such as $Z \to q\bar{q}$ and $qq \to b\bar{b}$, whose cross sections are many orders of magnitude larger (see figure 1.25).

W, Z and γ decay channels

The partial decay width of the Higgs boson decaying into W and Z is given by [32]

$$\Gamma(H \to VV^*) = \delta_V \frac{\sqrt{2}G_F}{32\pi} m_H^3 (1 - 4x + 12x^2)\beta \tag{1.3.63}$$

where $x = m_V^2/m_H^2$, $\beta = \sqrt{1 - 4x}$ and $\delta_V = 2(1)$ for $V = W(Z)$. Since the mass of the Higgs boson is of 125 GeV, one or both the two bosons are produced *off-shell*. The Higgs boson can also decay into

loop-induced $\gamma\gamma$ decay channels or $Z\gamma$. Although the BR for the $H \rightarrow \gamma\gamma$ is very small, it has a very clean signature and, together with the $H \rightarrow ZZ^*$, it has been the main decay channel used for the initial announcement of the Higgs boson discovery.

1.3.5   Higgs mass measurements

In order to measure the mass of the observed boson, the ATLAS and CMS experiments combine the measurements from the $\gamma\gamma$ and ZZ channels which have both excellent mass resolution and large significance. The mass measured by ATLAS [23] and CMS [24] results 125.5 ± 0.2(sta.)$^{+0.5\%}_{-0.6\%}$(sys.) and 125.7 ± 0.3(sta.) ± 0.3(sys.) GeV respectively. In both experiments the systematic uncertainty has been dominated by the uncertainty of the photon energy and the lepton momentum. All these measurements and the combination of the ATLAS and CMS (125.6 ± 0.3 GeV) results, assuming uncorrelated systematics uncertainties between the two experiments, are summarized in figure 1.13 [1].
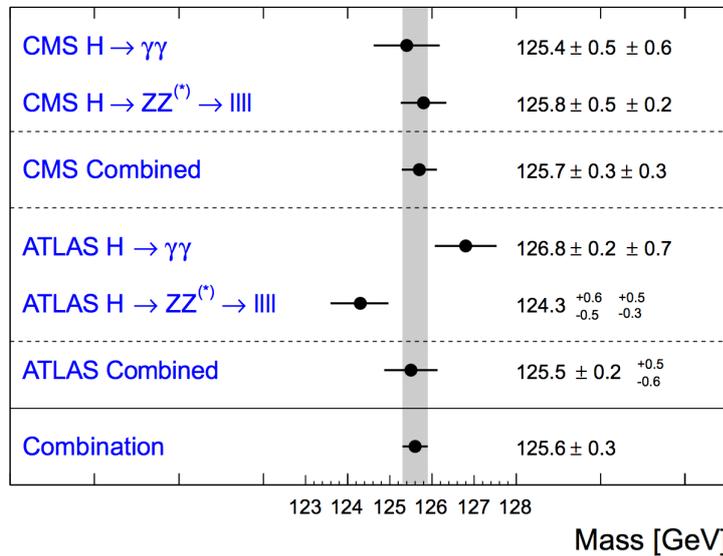
Figure 1.13: Values of all the CMS and ATLAS mass measurements in the $\gamma\gamma$ and ZZ channels and their combination.

1.4   Top anti-top pair ($t\bar{t}$)

In the SM the dominant mechanism for top quark pair ($t\bar{t}$) production is mediated by the strong interaction. At LO, the $t\bar{t}$ production at LHC can be successfully described in terms of quantum chromodynamics (QCD) via the gluon fusion ($gg \rightarrow t\bar{t}$) and quark-antiquark interaction ($q\bar{q} \rightarrow t\bar{t}$), shown in figure 1.14. At the next-to-leading order (NLO) it is also present the contribution of partonic sub-processes with gq ($g\bar{q}$) in the initial state, shown in figure 1.15.

The calculation of the $t\bar{t}$ production is extremely important in this thesis in order to determine both the inclusive cross section of the $t\bar{t}H$ production and its major background contributions (as explained in section 3.4).
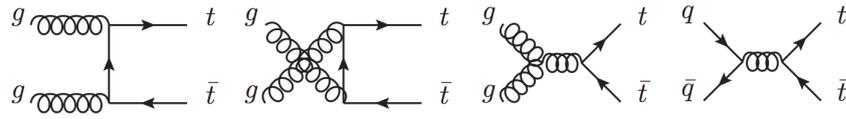
Figure 1.14: Feynman diagrams for $t\bar{t}$ production at leading order.
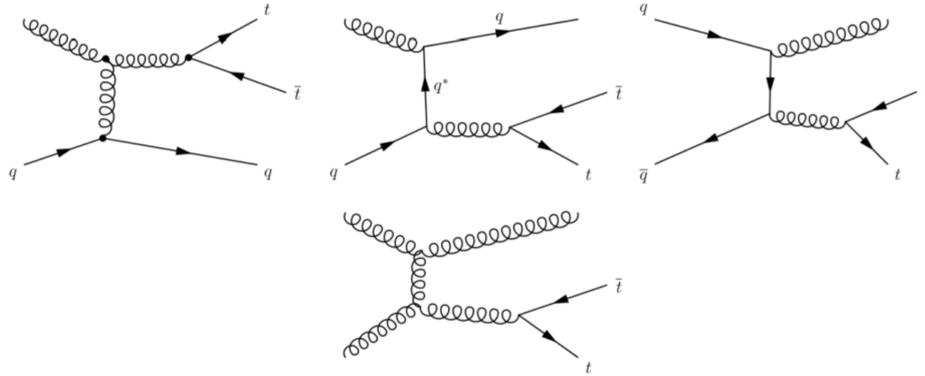
Figure 1.15: Feynman diagrams for $t\bar{t}$ production at next-to-leading order QCD.

1.4.1 $t\bar{t}$ production

At the proton-proton collision energy of $\sqrt{s} = 13\,\mathrm{TeV}$ of LHC, the dominant $t\bar{t}$ production comes from gg processes (85%) and only 15% from $q\bar{q}$ initial state. This situation is exactly the opposite of the Tevatron where, due to the different energy and initial state, the dominant process was $q\bar{q}$ (85%) and only 15% from gg.

The energy of LHC is high enough to consider quarks and gluons as quasi-free particles and consequently pp collisions can be described by the interaction between their constituent partons (quarks and gluons).

In the hard scattering of two colliding protons, top quarks are mostly produced in $t\bar{t}$ pairs through the strong interactions between partons. Using the parton model, this process can be illustrated schematically as figure 1.16.
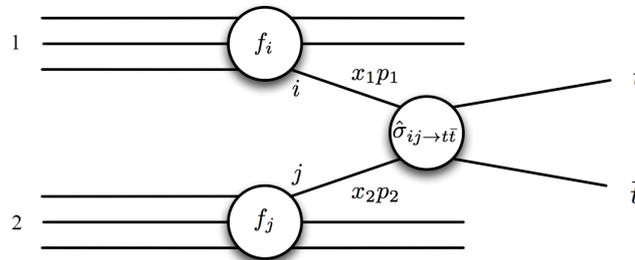
Figure 1.16: The top quark pair production of a hard scattering in the parton model.

The figure shows a collision of two energetic protons 1 and 2 with momentum p_1 and p_2 respectively. The hard scattering of the two protons can be seen as interaction between two partons from the protons.

The probability of finding parton i (j) carrying momentum fraction x_1 (x_2) in proton 1 (2) is described by f_i (f_j), the parton distribution function (PDF, see below).

With the help of the factorization theorem of QCD, which separates the perturbatively calculable part of the interaction from the non perturbative one, the inclusive production cross section of the process $pp \to t\bar{t}$ is expressed as

$$\sigma_{pp \to t\bar{t}}(s, m_t) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \cdot \hat{\sigma}_{ij \to t\bar{t}}(\hat{s}, m_t, \mu_f, \mu_r, \alpha_s) \qquad (1.4.64)$$

where the sum runs over all quarks and gluons contributing, i and j are the two partons interacting, x_i (x_j) is the parton momentum fraction with respect to the proton momentum, $f_i(x_i, \mu_f^2)$ ($f_j(x_j, \mu_f^2)$) is the PDF of the parton i (j) and $\hat{\sigma}_{ij}$ is the elementary cross section of those partons. The partonic cross section $\hat{\sigma}_{ij \to t\bar{t}}$ is evaluated in perturbative QCD and has a dependency on the partonic center-of-mass energy $\hat{s} \sim x_i x_j s$, the top quark mass m_t, the factorization scale μ_f and α_s, the coupling constant in QCD with dependency on the renormalization scale μ_r. The PDFs are not predicted by theory, but have been measured in experiments and can be evolved to the appropriate scale at which the proton is effectively probed. Therefore, the PDFs have an extra dependency on the factorization scale μ_f, which connects the PDFs with the partonic cross section $\hat{\sigma}_{ij \to t\bar{t}}$ that can be calculated in perturbative QCD.

The dependence from the renormalization scale arises from the fact that the partonic cross section is evaluated at a fixed perturbation order, neglecting higher order contributions. Such dependences become weaker and weaker as we add higher order corrections to calculation. The factorization scale, on the other hand, indicates the transition between the perturbative regime, which belongs to the partonic cross section, and the non-perturbative one, included in the PDF definition.

To avoid spoiling the QCD perturbative calculation, a commonly used convention is to set both μ_f and μ_r to the order of the hard-scaling energy characterizing the production process, which is m_t for the top quark production. With $\alpha_s(m_t) < 1$, the partonic cross section can be expanded in a fixed-order series in α_s as

$$\hat{\sigma}_{ij \to t\bar{t}} = \alpha_s^2 \left[\hat{\sigma}_{ij \to t\bar{t}}^{(0)} + \alpha_s \hat{\sigma}_{ij \to t\bar{t}}^{(1)} + \alpha_s^2 \hat{\sigma}_{ij \to t\bar{t}}^{(2)} + ... \right], \qquad (1.4.65)$$

where the first term in square brackets is referred to as leading-order (LO), the second term next-to-leading-order (NLO), the third term next-to-next-to-leading-order (NNLO), and so on.

As a physical observable, the total cross section $\sigma_{pp \to t\bar{t}}$ should not depend on the choice of the scales; but in the QCD perturbative calculation it does. One reason is that the calculation of eq. (1.4.65) is usually truncated to certain fixed order, and the truncated part has dependency on $\alpha_s(\mu_r)$. Possible deviations due to the choice of the scales and the PDFs are usually evaluated as theoretical uncertainties.

The Parton Distribution Functions (PDF) $f_i(x)$ [2] describe the distributions of the fraction of momentum carried by the partons. In the factorization theorem at the bases of eq. (1.3.63) they describe the non perturbatively calculable part of the interaction and they are extracted by data from a plethora of other measurements or studies. A simple interpretation of the PDF is that, at a given factorization scale μ_f, they describe the probability P that a parton i carries a fraction x of the proton momentum p (see figure 1.17) as:

$$f_i(x) = \frac{dP_i}{dx}. \qquad (1.4.66)$$

The available energy for the scattering process is not the nominal energy of LHC, but it depends on the energy carried from each parton by the relation $\hat{s} \sim x_i x_j s$, where x_i and x_j are the proton momentum fractions of the two interacting partons. A deep knowledge of the PDFs is important for the simulation of all the physics processes; different numerical computations are used to extract the PDFs and the main differences between the PDF sets arise from the choice of the included results (expecially from Deep
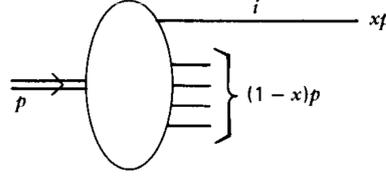
Figure 1.17: Sketch of the evaluation of the momentum of parton i inside a proton of momentum p.

Inelastic Scattering experiments), the treatment of systematic uncertainties, the parametrization at the starting scale, the chosen heavy-quark scheme and the values of the quark masses.

Fig. 1.18 illustrates the MSTW 2008 NLO [16] analytical parameterization of PDFs for all the different quarks, anti-quarks and gluons, for different values of Q^2 (the transfert momentum).
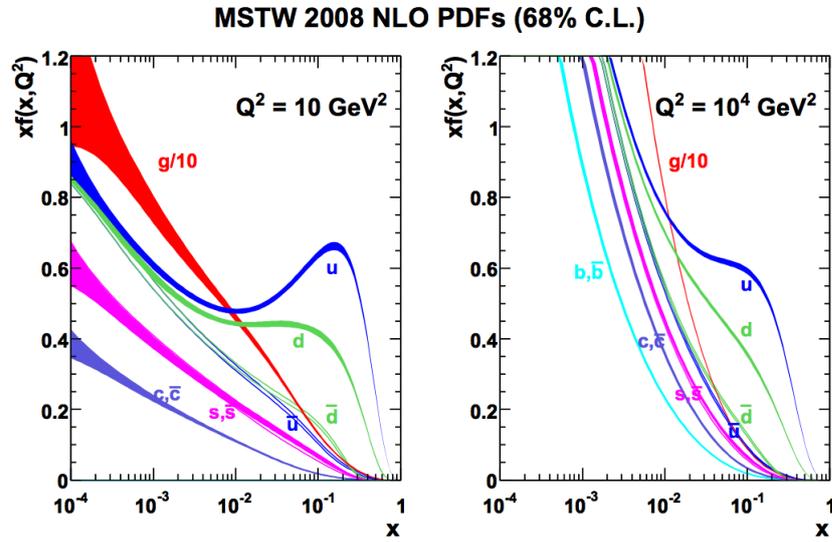
Figure 1.18: Analytical parameterization of the parton distribution functions $xf_a(x, Q^2)$ using the MSTW 2008 NLO [16] estimation for the various quarks, antiquarks, and the gluon, with different values of Q^2. The gluon distribution is scaled by a factor of 0.1.

At LHC, with $\sqrt{s}$ = 13 TeV, the typical value of momentum fraction for the $t\bar{t}$ production is $x = 2m_t/\sqrt{s}$ that corresponds to $x \sim 0.03$, instead at Tevatron (with $\sqrt{s}$ = 1.96 TeV) it was $x \sim 0.2$. This difference reflects in different amount of gluon and quark population inside the hadrons, as shown in figures 1.18. At low values of x the probability to have gluons is much higher than quarks, explaining why at LHC, the dominant production process is the one involving a pair of gluons (gg), instead of a pair of quarks ($q\bar{q}$ or qq).

1.4.2 $t\bar{t}$ cross section

The cross section of a process σ_{proc} is obtained by the following relation

$$\sigma_{proc} = \frac{N_{obs}}{\int Ldt \cdot \epsilon} \tag{1.4.67}$$

where the N_{obs} is the number of observed events of the process (with the background subtracted), L is the instantaneous luminosity and ϵ is the efficiency to detect this process. $\int L dt$ means that the luminosity is integrated overall the period of the data acquisition and it is called *integrated luminosity* $\mathcal{L}$. While, the instantaneous luminosity depends on the features of the collider, and it's obtained by the relation:

$$L = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} n_b \tag{1.4.68}$$

where f is the collider frequency, that is the revolution frequency of the bunches, n_1 and n_2 are the numbers of particles contained in each colliding bunch, n_b is the number of filled bunches, σ_x and σ_y are the transverse profile of the beam. The cross section is related to the probability that a process happens, it has the dimension of an area and it is measured in barn (1 b = 10^{-24} cm^2).

Fig. 1.19 and 1.20 summarize the $t\bar{t}$ production cross section measurements and predictions from both the Tevatron and LHC.
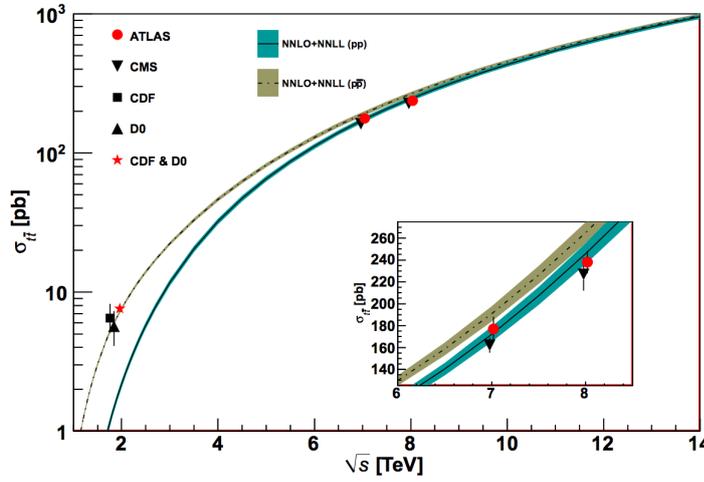
Figure 1.19: Measured and predicted $t\bar{t}$ production cross sections from Tevatron energies ($p\bar{p}$ collisions) to LHC energies (pp collisions) [1].

1.4.3 $t\bar{t}$ decays

The decays of a top quark pair can be classified according to the decay of the produced W-bosons (because $t \to Wb$ is practically the only decay channel):

- di-lepton channel: both W-bosons decay into lepton (electron, muon, tau) and neutrino, $t\bar{t} \to W^+bW^-\bar{b} \to \bar{l}\nu_l bl'\bar{\nu}_{l'}\bar{b}$. The branching ratio (BR), considering all the tree leptons, is $\sim 10\%$. If only electrons and muons are considered, the BR is $\sim 6\%$;

- lepton+jets channel: one W-boson decays into lepton and neutrino, the other one into a quark - anti-quark$'$ pair, $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'bl\bar{\nu}_l\bar{b} + l\bar{\nu}_l bq\bar{q}'\bar{b}$. The BR for this decay mode is $\sim 45\%$;

- hadronic channel: both W-bosons decay into a quark - anti-quark pair, $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}'bq''\bar{q}'''\bar{b}$. The BR for this decay mode is $\sim 45\%$.

The detail of all the different branching ratios are summarized in figure 1.21.
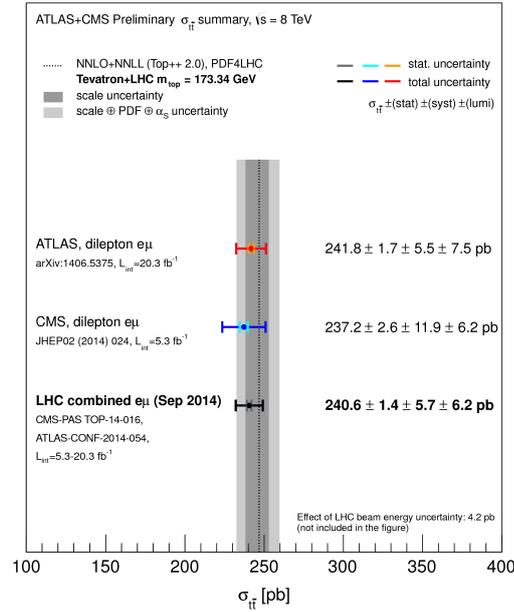
Figure 1.20: $\sigma_{t\bar{t}}$ measurements by the ATLAS and CMS collaborations and the result of the LHC combination, compared with the NNLO+NNLL QCD calculation, for the world average top quark mass of 173.34 GeV [6]. The uncertainty due to the LHC beam energy is 4.2 pb and it is not included in the total uncertainty on the measurements or the combination. This uncertainty has to be added in quadrature to the total uncertainty for the comparison of the measurements with predictions.
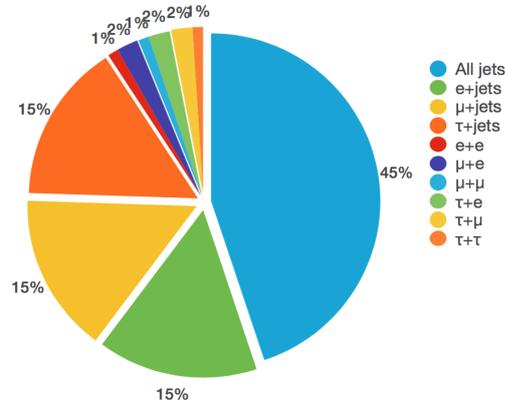
Figure 1.21: All the three decay modes of the $t\bar{t}$ system are illustrated in the pie chart. Different lepton contributions are reported separately.

1.5   Higgs boson associated with a $t\bar{t}$ pair ($t\bar{t}H$)

The $t\bar{t}H$ production mode is fundamental to probe the Yukawa coupling between the quark-top and the Higgs (1.3.59). Among the quarks-Higgs couplings, the most interesting is certainly the coupling between the Higgs boson and the top quark, the two most massive elementary constituents of the SM.

A Higgs mass of about 125 GeV provides an excellent opportunity to explore the Higgs couplings to many SM particles. The strength of the couplings of the Higgs boson to fermions (g_{ffH}) and to the

electroweak gauge bosons (g_{VVH}) is set by the fermion m_f and boson m_V ($V = W, Z$):

$$g_{ffH} = [\sqrt{2}G_F]^{1/2}m_f$$
$$g_{VVH} = 2[\sqrt{2}G_F]^{1/2}m_V^2$$

(1.5.69)

where G_F is the Fermi constant.

According to the Glashow-Weinberg-Salam model, illustrated in section 1.3.1, fermions, as well as gauge bosons, should be massless. Fermions, however, acquire mass by the coupling with the Higgs field, as shown in figure 1.22. In a similar way, the masses of the bosons come up by their direct coupling with the Higgs field.
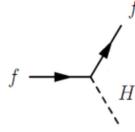
Figure 1.22: Feynman diagram of the direct coupling of the Higgs field with a fermion of the Standard Model.

Due to its large mass, the top quark Yukawa coupling is expected to be near one (~ 0.99), differently by the other quarks that have a coupling of $\sim 10^{-2}$. Since the top quark pair is heavier than the Higgs boson, this latter cannot decay into a top quark pair and thus top-Higgs coupling cannot be measured using this decay channel. However, the Higgs boson coupling to top quarks can be experimentally measured by the gluon fusion production mechanism of the $t\bar{t}H$ channel. In fact, the ggF process proceeds via a fermionic loop in which the top quark provides the dominant contribution, neglecting contributions from theories beyond the Standard Model. In this scenario, the $t\bar{t}H$ production mechanism is extremely important since it provides direct measurement of this coupling.

1.5.1 $t\bar{t}H$ production

The $t\bar{t}H$ production channel is the production of a Higgs boson in association with a top-quark pair. Its Feynman diagram, at LO, is shown in figure 1.10(d).

The first searches for $t\bar{t}H$ production were performed by the CDF and D0 experiments at the Tevatron collider, which only put limits to the SM Higgs boson production. At LHC, during the Run-1 and Run-2, the $t\bar{t}H$ production has been studied in different channels (see section 1.5.2), but also in this case only limits have been put to the SM Higgs production. The ratio between the measured cross section with respect to the predicted one by the SM $\mu = \sigma/\sigma_{SM}$ has been provided by the studies on the multi-leptonic final states [33],[34] (leptonic final states of the $t\bar{t}H$, with the Higgs decaying in a pair of vector bosons WW^* or ZZ^*), di-photon final state [35] and hadronic channels [36],[37] (with the Higgs decaying in a pair of b-quarks). The single measured signal strength μ are shown in figure 1.23, together with the combined value for the Run-1 and Run-2 data taking period. The combination of the different channels in the multi-lepton channel gives $\mu = 2.5^{+1.3}_{-1.1}$ leading to a limit $\mu < 4.9$ at 95%CL, it is $\mu = -0.3^{+1.2}_{-1.0}$ for the di-photon final state case with a limit $\mu < 2.6$ at 95%CL, while it is $\mu = 2.1^{+1.0}_{-0.9}$ for a limit $\mu < 4.0$ at 95%CL in the hadronic channel. Obviously it is necessary to perform more precise measurements to obtain a more significant result.

The peculiar feature of the $t\bar{t}H$ production is the relative increase of its cross section with respect to all the other production processes increasing the center-of-mass energy (figure 1.11 and table 1.4); the $t\bar{t}H$ cross section presents the biggest relative increase from a center-of-mass of $\sqrt{s}$ = 7 TeV to 14 TeV
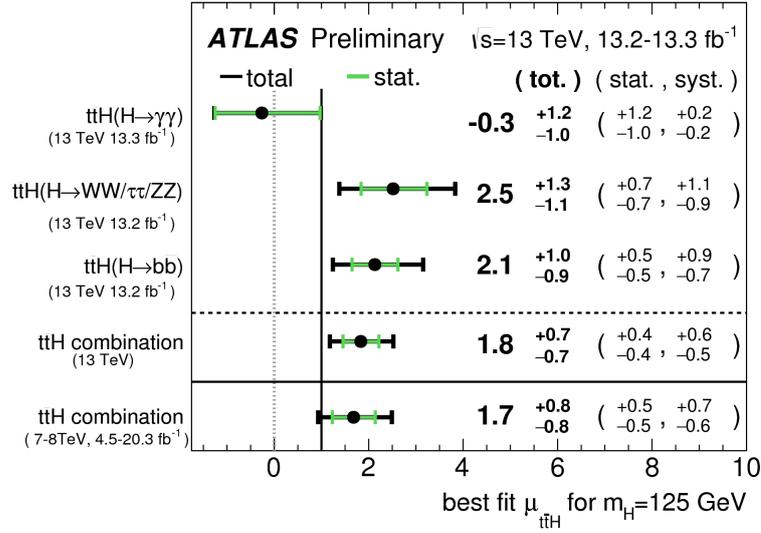
Figure 1.23: The fitted value of the signal strength and its uncertainty for the individual channels and their combination, assuming m_H = 125 GeV. The green line shows the statistical error on the signal strenght.

and with a factor $\sim$ 7 of cross section improvement, the study of the $t\bar{t}H$ production is one the most interesting analyses of Run-2.

A precise prediction for the cross section is needed in order to interpret the evidence for $t\bar{t}H$ production and the measurement of the $t\bar{t}H$ coupling as a verification of the Standard Model or as a signal for new physics. QCD corrections [26] are important and crucial in order to reduce the dependence of the cross section on the arbitrary renormalization and factorization scales. The LO and NLO QCD calculations of the Standard Model process $pp \rightarrow t\bar{t}H$ provide the cross-section values listed in table 1.6.

Centre of mass energy (TeV)	LO cross-section (pb)	NLO cross-section (pb)
8	$0.127^{+0.051}_{-0.033}$	$0.132^{+0.007}_{-0.014}$
13	$0.464^{+0.166}_{-0.115}$	$0.507^{+0.037}_{-0.056}$
14	$0.558^{+0.196}_{-0.136}$	$0.614^{+0.047}_{-0.067}$

Table 1.6: The predicted cross-sections calculated with both LO and NLO QCD corrections [39] for a SM Higgs boson with m_H = 125 GeV. The two uncertainties are respectively due to the factorization and renormalization scales and the PDF uncertainties.

Moreover, many new physics scenarios predict the existence of heavy top quark partners that would decay into a top quark and a Higgs boson: the observation of a significant deviation in the $t\bar{t}$ cross section with respect to the SM prediction would be an indirect indication of unknown phenomena.

1.5.2 $t\bar{t}H$ decays

Since the top quark decays with nearly 100% probability into a W boson and a b quark, the experimental signatures for Higgs production in association with a top quark pair are determined by the decay of the

29

W boson. The different final states of the W pairs combined with the different Higgs decay channels give a wide variety of complex final states.

As seen in section 1.4.3, there are 3 different decay modes for a top-quark pair: *hadronic*, *lepton + jets* and *dilepton* mode. The semileptonic decay mode (*lepton + jets*) is chosen due to its clear signature and its precision. Regarding the Higgs decay channels, there are many decay modes available, but the hadronic decay $H \rightarrow b\bar{b}$ is the chosen one because it contributes almost for the 60% to the total Higgs boson decay width.

In figure 1.24 is shown an example of $t\bar{t}H$ production, with the Higgs boson decaying in $b\bar{b}$ and the $t\bar{t}$ system decaying in a semi-leptonic mode ($t \rightarrow bW^+$, $W^+ \rightarrow q\bar{q}'$ and $\bar{t} \rightarrow bW^-$, $W^- \rightarrow l^-\bar{\nu}_l$), which is the decay mode considered and studied in this thesis.
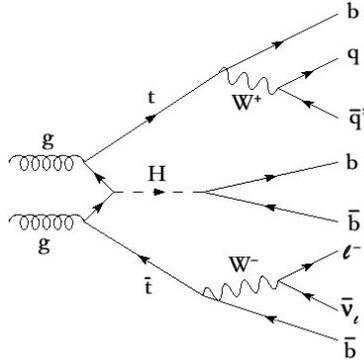
Figure 1.24: Example of leading order Feynman diagram for $t\bar{t}H$ production at pp colliders, followed by Higgs boson decay in $b\bar{b}$ and $t\bar{t}$ decay in a semi-leptonic mode, which is the decay mode studied in this thesis.

However, the inclusive search for the process $H \rightarrow b\bar{b}$ is very challenging, since the production cross section for inclusive $b\bar{b}$ production, which is about 100 μb, is 7 orders of magnitude larger than the Higgs production cross section in the $b\bar{b}$ channel, as shown in figure 1.25.

The associated production of the Higgs boson with a top-quark pair helps to separate the signal from the overwhelming QCD background by reconstructing all the final states of this channel. The analysis described in this thesis exploits this benefit, since the $H \rightarrow b\bar{b}$ in association with a semileptonic decay of the $t\bar{t}$ system is considered in a *boosted regime*. The difference between the *boosted* and the *resolved regime* is in the transverse momentum of the final objects that is considered: the boosted object decays have their decay products collimated in the momentum direction of the boosted mother particle in the rest frame of the detector, with the consequence of a final state with objects that cannot be reconstructed singularly. It leads to a totally different event topology with respect to resolved one, which has all the final products well defined, as shown in figures 1.26a and 1.26b.

The two different regimes will be considered in this analysis, since the final goal is to combine the resolved and the boosted analyses, as discussed in more details in Chapters 4 and 5.
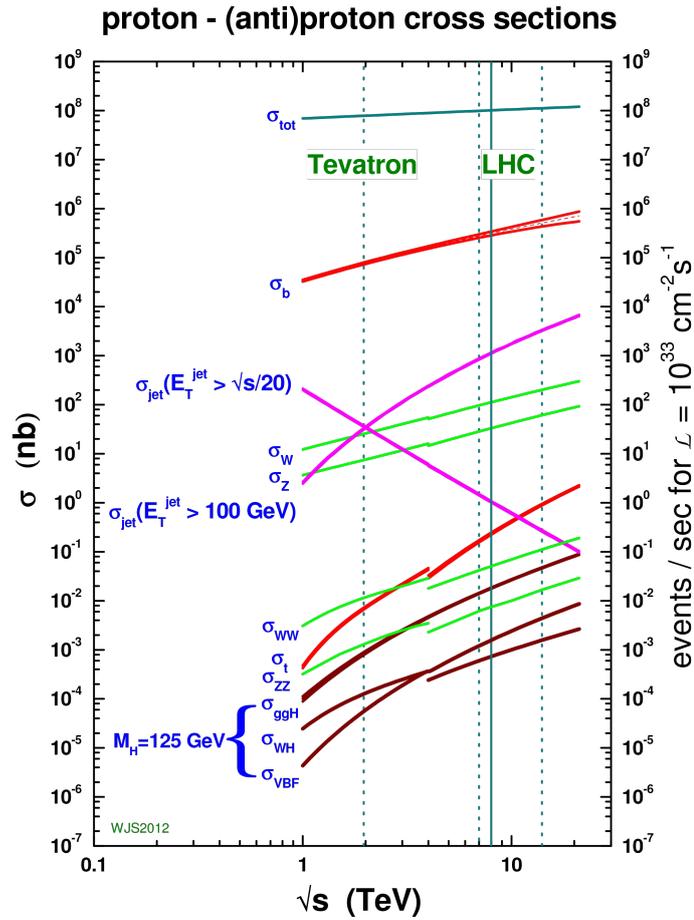
Figure 1.25: Production cross section for different processes as a function of the center of mass energy, $\sqrt{s}$, for Tevatron ($p\bar{p}$ collisions, left) and LHC (pp collisions, right) [40].
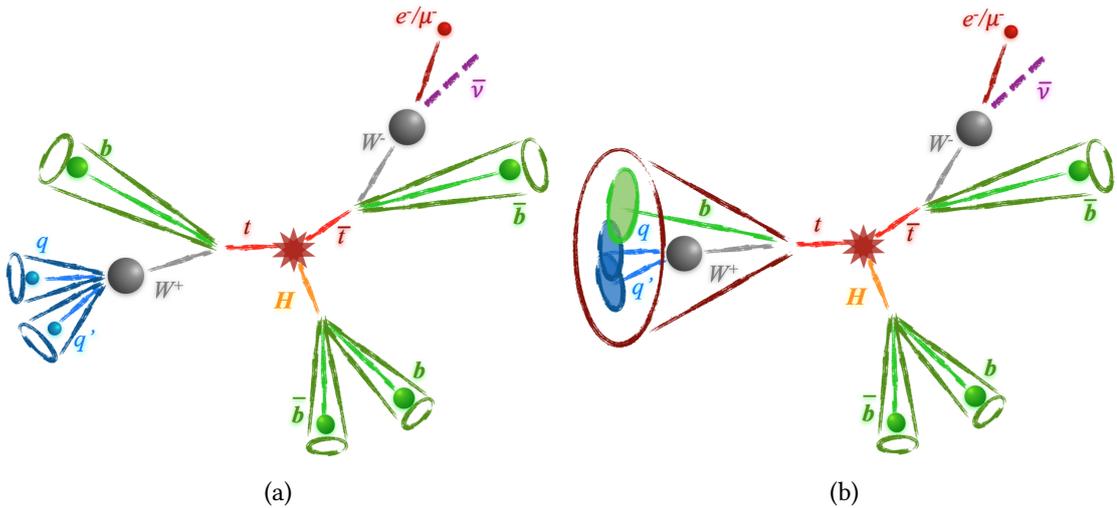
Figure 1.26: (a) A schematic representation of the $t\bar{t}H$ channel in the resolved regime. (b) A schematic representation of the $t\bar{t}H$ channel in the boosted regime, where it is the case of only one large-R jet in the final state, but two large-R jets can be present too.

Chapter 2

LHC and the ATLAS experiment

The CERN, the European Organization for Nuclear Research, is the world largest research center. Founded in 1954, the CERN laboratory sits astride the Franco-Swiss border near Geneva. It was one of Europe's first joint ventures and now has 21 member states that cooperate to probe the fundamental structure of the universe. The instruments used are particle accelerators and detectors; accelerators boost beams of particles to high energy to collide with each other or with stationary targets and detectors observe and record the results of these collisions.

The Large Hadron Collider, LHC, is the newest CERN accelerator complex and it is the world's largest and most powerful particle accelerator. Along the accelerator ring are positioned four particle detectors: ATLAS, CMS, ALICE and LHCb.

2.1 The Large Hadron Collider

LHC is situated from 50 to 175 m under the ground of Geneva, as schematized in figure 2.1, and it consists of a 27-kilometer ring of superconducting magnets with radiofrequency cavities to boost the energy of the particles. Inside the accelerator two high energy particles beams travel in opposite directions in separate beam pipe kept at ultrahigh vacuum.

The beams collide in four points, in correspondence of the four detectors.

The accelerating system consists of 16 radiofrequency cavities with a maximum electric field of 5.5 MV/m. The two beams were structured in a maximum of 3564 bunch slots. At the maximum energy, during the Run-1, the bunches collide every 50 ns, during the Run-2 every 25 ns. Inside the accelerator, the beams are deflected by 1232 electromagnets composed by coils of special electric cables that, operating in superconducting state (temperature of 1.9 K), could endure a circulating current of 11.85 kA to generate a magnetic field of 8.4 T. The focusing system consists of 392 superconducting magnets quadrupoles producing a 6.8 T field each.

The LHC can accelerate protons with a minimum ebergy of 450 GeV. Protons of this energy are produced by a chain of accelerators that take protons from a hydrogen gas bottle and progressively accelerates them with the help of several intermediate accelerators: a linear accelerator *Linac2*, which accelerates the
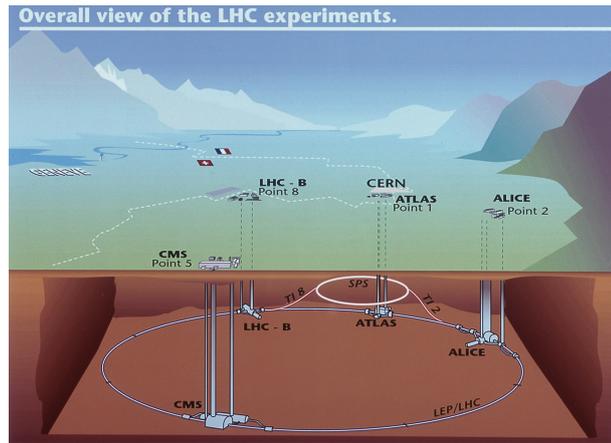
Figure 2.1: The LHC complex in the underground of Geneva.

protons to the energy of 50 MeV, and three synchrotrons, *Proton Synchrotron Booster* (PSB), till 1.4 GeV, *Proton Synchrotron* (PS), till 25 GeV, and *Super Proton Synchrotron* (SPS), till 450 GeV, as illustrated in figure 2.2.
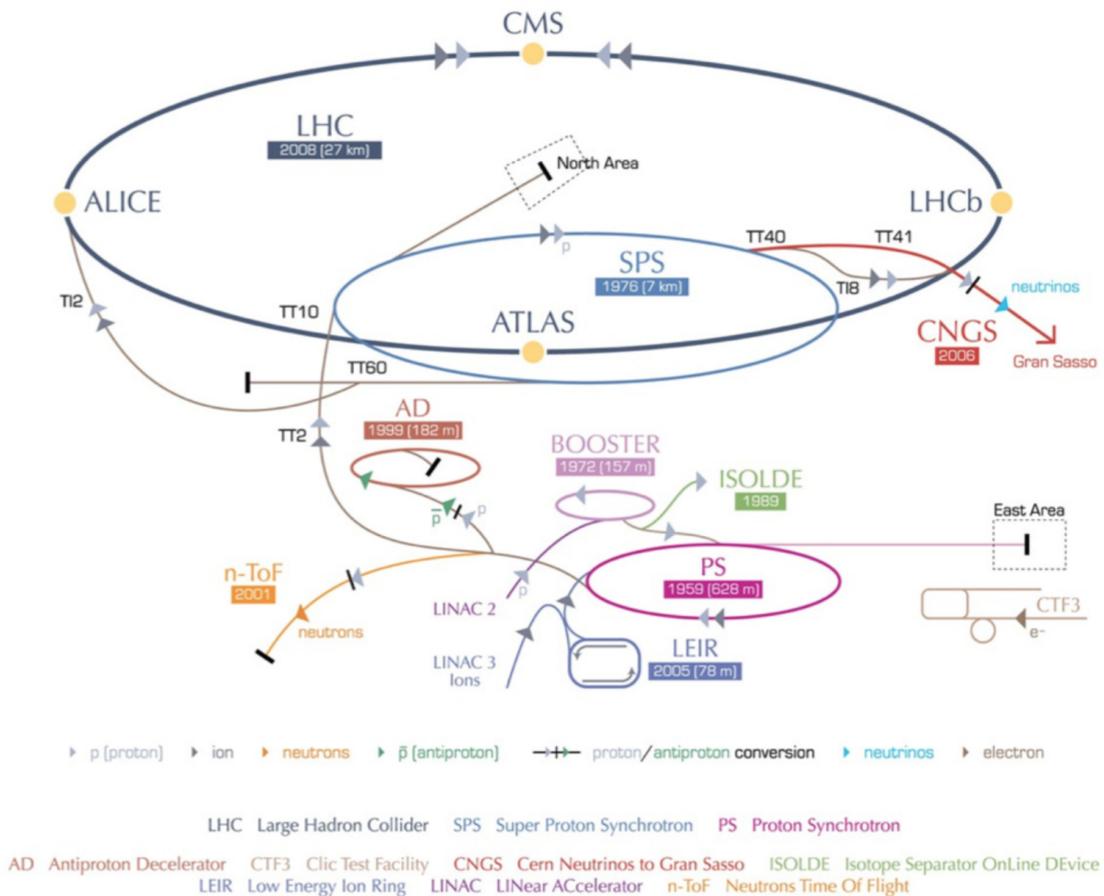
Figure 2.2: Scheme of the CERN accelerator complex.

For the last data taking, the so called Run-2 (see figure 2.3), LHC is designed to accelerate protons (and also heavy ions in dedicated runs) up to an energy of 6.5 TeV producing collisions at a center-of-mass energy of 13 TeV at a maximum instantaneus peak luminosity of $L = 5 \times 10^{33}$ cm^{-2} s^{-1}. Figure 2.3 illustrates all the periods of active work or technical shut down of LHC together with the future plan of upgrades [41], as the Phase-2 (High Luminosity-LHC, from 2023) already under investigation for being approved and applied during the third long shut down (LS3).
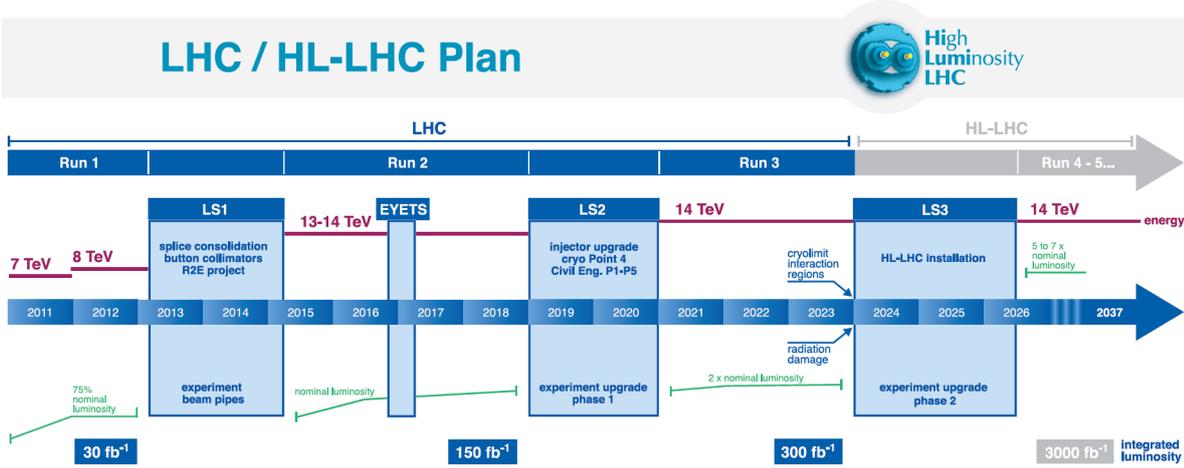
Figure 2.3: Time schedule of LHC, from Run-1 to last upgrade to High Luminosity LHC.

An overview of the LHC luminosity-related parameters during the Run-1 and Run-2 periods (2010-2015) is reported in table 2.1, compared with the design values. In figure 2.4a and 2.4b [44], the luminosity

Parameter	2010	2011	2012	2015	Design value
Beam energy (TeV)	3.5	3.5	4	6.5	7
β^* in IP 1 and 5 (m)	2.0/3.5	1.5/1.0	0.6	0.80	0.55
Filled bunch spacing (ns)	150	75/50	50	25	25
Max. number of filled bunches	368	1380	1380	2244	2808
Protons per bunch	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}	1.15×10^{11}	1.15×10^{11}
Peak luminosity (cm^{-2}s^{-1})	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}	5×10^{33}	1×10^{34}
Pile-up interactions	4	17	37	15	19

Table 2.1: An overview of performance-related parameters during LHC operations in 2010-2015.

delivered to (green) and recorded by (yellow) the ATLAS detector, in 2015 and 2016, during stable pp beam collisions is shown as a function of the time. The delivered luminosity accounts from the start of stable beams to the detector in a safe standby mode to allow a beam dump or beam studies. The recorded luminosity reflects the DAQ inefficiency, as well as the inefficiency of the so-called "warm start": when the stable beam flag is raised, the tracking detectors undergo a ramp of the high-voltage and, for the pixel system, turning on the preamplifiers. The ratio between the recorded and the delivered luminosity is $\sim 93\%$.

The mean number of interactions per filled bunch crossing μ, the so-called *pile-up*, corresponding to the mean of the poisson distribution on the number of interactions per crossing, is shown in figures 2.5a

and 2.5b [44] as a function of time, during pp collisions in 2015 and 2016 (the distributions for the pp collisions in 2010, 2011 and 2012 are reported in [43] and shown in figure 2.6).

(a)   (b)
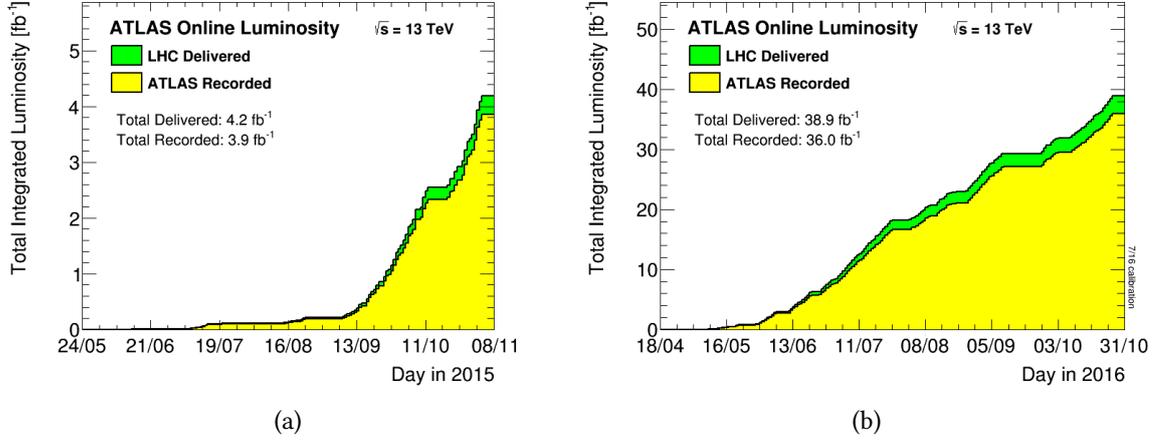
Figure 2.4: Cumulative delivered (green) and recorded (yellow) luminosity versus time during stable beams for pp collisions at 13 TeV centre-of-mass energy in (a) 2015 and (b) 2016.

(a)   (b)
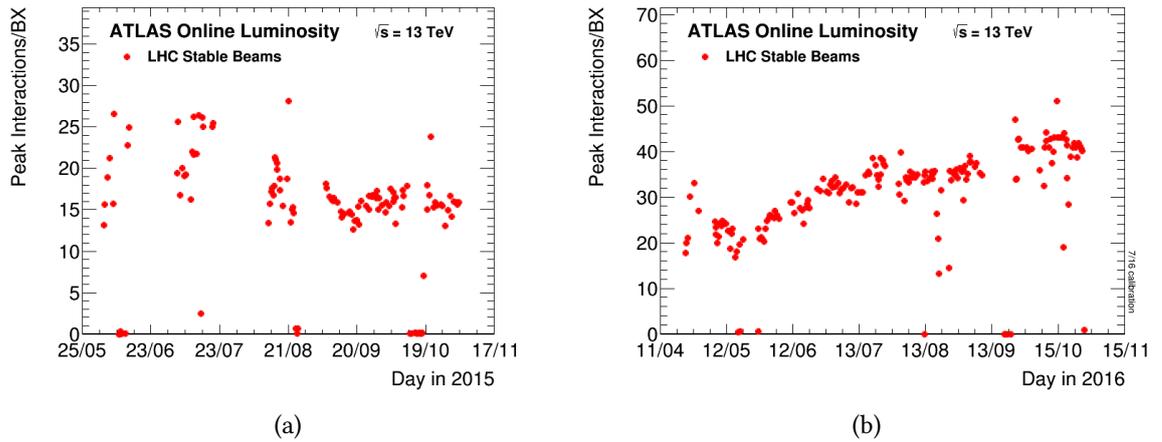
Figure 2.5: The mean number of interactions per filled bunch crossing per lumi-block versus day during the pp runs of (a) 2015 and (b) 2016. The online luminosity measurement is used for this calculation. Only the maximum value during stable beam periods is shown.

| | Integrated Luminosity | | | | |
	2010 (pb^{-1})	2011 (fb^{-1})	2012 (fb^{-1})	2015 (fb^{-1})	2016 (fb^{-1})
Delivered	48.1	5.46	22.8	4.2	38.9
Recorded	45.0	5.08	21.3	3.9	36.0

Table 2.2: Delivered and recorded integrated luminosity $\mathcal{L}$ in 2010 and 2011 ($\sqrt{s}$ = 7 TeV), 2012 ($\sqrt{s}$ = 8 TeV), 2015 and 2016 ($\sqrt{s}$ = 13 TeV), by the ATLAS detector.
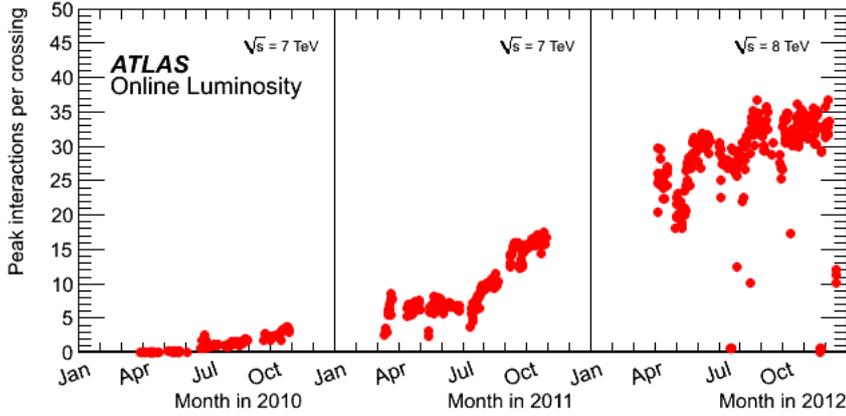
Figure 2.6: The mean number of interactions per filled bunch crossing per lumi-block versus day during the pp runs of 2010, 2011 and 2012.

2.2 The coordinate system and nomenclature

The nominal interaction point is defined as the origin of a right-handed coordinate system, while the beam direction defines the z-axis and, consequently, the x-y plane is transverse to the beam direction. The positive x-axis is defined as pointing from the interaction point to the centre of the LHC ring and the positive y-axis is defined as pointing upwards. The side-A of the detector is defined as that with positive z and side-C is the one with negative z. The azimuthal angle ϕ is measured in a plane transverse to the beam axis, and the polar angle θ in a plane containing the beam axis. The rapidity y is defined as

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z}, \tag{2.2.1}$$

that, in the case of objects with a negligible mass with respect to the energy, corresponds to the pseudorapidity η:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.2.2}$$

The transverse momentum p_T, the transverse energy E_T and the missing transverse energy $E_\mathrm{T}^\mathrm{miss}$ are defined in the x-y plane unless stated otherwise. ΔR is the distance in the η-ϕ space defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

2.3 Physics requirements

Requirements for the ATLAS detector system [45] have been defined, at the end of the '90s, using a set of processes covering much of the new phenomena hopefully to observe at the TeV scale. The high luminosity and increased cross-sections at the LHC enable further high precision tests of QCD, electroweak interactions and flavour physics; in particular the top quark will be produced at a rate of a few tens of Hz, providing the opportunity to test its couplings and spin.

The search for the Higgs boson has been used as a benchmark to make important choices related to the sub-systems of ATLAS, since the small production cross section and the natural width of few MeV of the Higgs boson.

The nature of proton-proton collisions imposes another difficulty; QCD jet production cross-sections dominate over the rare processes, such as the decays of supersymmetric particles or the graviton from the postulated extra-dimensions theory, requiring the identification of experimental signatures characteristic of those processes in question, such as E_T^{miss} or secondary vertices. Identifying such final states for these rare processes imposes further demands on the integrated luminosity needed and on the particle-identification capabilities of the detector.

These benchmark physics goals can be turned into a set of general requirements for the LHC detectors:

- due to the experimental conditions at the LHC, the detectors require fast, radiation-hard electronics and sensor elements. In addition, high detector granularity is needed to handle the high particle fluxes and to reduce the influence of overlapping events;

- large acceptance in pseudorapidity and almost full azimuthal angle coverage;

- good charged-particle momentum resolution and reconstruction efficiency in the inner tracker;

- good resolution on the secondary vertices necessary for tagging of τ-leptons and b-jets;

- good electromagnetic (EM) calorimetry for electron and photon identification and their energy measurements, complemented by full-coverage hadronic calorimetry for accurate jet and missing transverse energy measurements;

- good muon identification and momentum resolution over a wide range of momenta together with the unambiguously determination of the charge of high p_T muons;

- highly efficient triggering on low transverse-momentum objects with sufficient background rejection in order to achieve an acceptable trigger rate for most physics processes of interest.

Both ATLAS and CMS detectors are designed to satisfy these requirements in order to perform precise measurements of SM processes and to discover new physics ones. The four experiments have different detector structure that correspond to different physical purpose:

- **A Toroidal LHC ApparatuS (ATLAS)** is a multipurpose experiment to discover signatures of new physics and to perform precise measurements of Standard Model;

- **Compact Muon Solenoid (CMS)** is the other multipurpose experiment that pursues the same physics goals as ATLAS using different and complementary technologies;

- **LHCb** has a completely different structure with respect to the other experiments and investigates the flavour physics of B meson and the CP violation;

- **A Large Ion Collider Experiment (ALICE)** is dedicated to the study of quark-gluon plasma produced in heavy nucleons collisions, accelerated in LHC instead of the protons in some dedicated runs.

37

2.4   The ATLAS detector

The overall ATLAS detector layout is shown in figure 2.7 [45] and its main performance goals are listed in table 2.3.
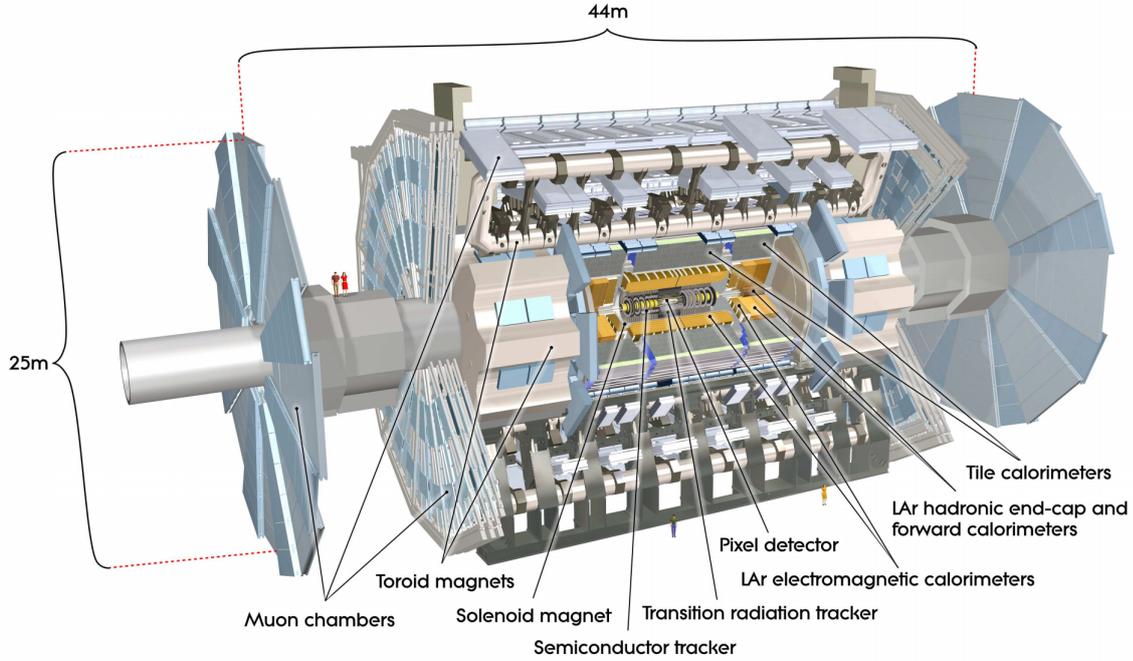
Figure 2.7: Cut-away view of the ATLAS detector. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is approximately 7000 tonnes.

Detector component	Required resolution	Obtained resolution (2015)	η coverage	
			Measurement	Trigger
Tracking	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\% \, p_\mathrm{T} \oplus 1\%$	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.038\% \, p_\mathrm{T} \oplus 1.5\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.2\%$	± 3.2	± 2.5
Hadronic calorimetry (jets)				
barrel (Tile) and end-caps (LAr)	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward (LAr)	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	-	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 10\%$ at $p_\mathrm{T} = 1$ TeV	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 10\%$ at $p_\mathrm{T} = 1$ TeV	± 2.7	± 2.4
combined with tracker	-	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 7\%$ at $p_\mathrm{T} = 1$ TeV	± 2.7	± 2.4

Table 2.3: General performance goals [45] and obtained resolutions in 2015 [46] of the ATLAS detector. If not indicated, the units for E and p_T are in GeV.

The ATLAS detector is nominally forward-backward symmetric with respect to the interaction point. The magnet configuration comprises a thin superconducting solenoid surrounding the inner-detector cavity, and three large superconducting toroids (one barrel and two end-caps) arranged with an eight-

fold azimuthal symmetry around the calorimeters; this choice has driven the design of the rest of the detector.

The *inner detector* is immersed in a 2 T solenoidal field. Momentum, trajectory, vertex measurements and pattern recognition (in particulare the electron identification) are achieved with a combination of high-resolution semiconductor pixel and strip detectors in the inner part of the tracking volume and straw-tube tracking detectors with the capability to generate and detect transition radiation in its outer part.

High granularity *liquid-argon* (*LAr*) electromagnetic sampling calorimeters provide excellent performance in terms of energy and position measurements. The hadronic calorimetry is provided by a scintillator-tile calorimeter, separated into a large barrel and two smaller extended barrel cylinders, one on either side of the central barrel.

The *muon spectrometer* surrounds the calorimeter and defines the overall dimensions of the ATLAS detector. The air-core *toroid system*, with a long barrel and two inserted end-cap magnets, generates strong bending power in a large volume within a light and open structure. For high-p_{T} muons, the muon-spectrometer performance (see table 2.3) is independent of the inner-detector system.

The proton-proton interaction rate at the design luminosity of 10^{34} cm^{-2}s^{-1} is approximately 1 GHz (considering about 30 interactions per bunch crossing), while the event data recording, based on technology and resource limitations, is limited to about 200 Hz. The *Level-1 trigger* (L1) system uses a subset of the total detector information to make a decision on whether or not to continue processing an event, reducing the data rate to approximately 75 kHz (limited by the bandwidth of the readout system). The subsequent two levels, collectively known as the *high-level trigger*, are the *Level-2 trigger* (L2) and the *event filter*. They provide the reduction to a final data-taking rate of approximately 200 Hz.

2.4.1   Tracking

All the detectors involved in the ATLAS tracking system are schematized in figure 2.8.
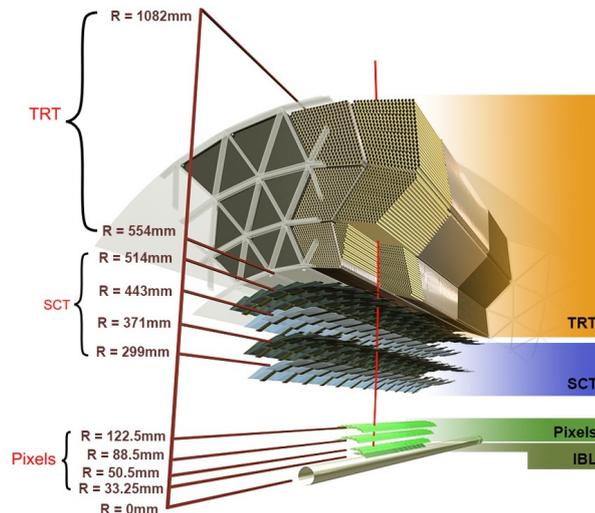
Figure 2.8: Sketch of the ATLAS inner detector showing all its components, including the new insertable B-layer (IBL). The distances to the interaction point are also shown.

Approximately 1000 particles will emerge from the collision point every 25 ns within $|\eta| < 2.5$, creating a very large track density in the detector.

The *Insertable B-Layer* (the 2015 upgrade of the ATLAS detector), the *Pixel* and the silicon microstrip of the *Semi Conductor Trackers* (SCT), used in conjunction with the straw tubes of the *Transition Radiation Tracker* (TRT), allow to reach the momentum and vertex resolution required by the benchmark physics process (see figure 2.9). The resolution of some kinematic quantities are listed in table 2.4.

Parameter	p_T	η	Value
Momentum resolution	100 GeV	~ 0	3.8%
Momentum resolution	100 GeV	~ 2.5	11%
Transverse impact parameter resolution	1000 GeV	~ 0	11 μm
Transverse impact parameter resolution	1000 GeV	~ 2.5	11 μm
Identification efficiency for pions	1 GeV	-	84.0%
Identification efficiency for electrons	5 GeV	-	90.0%

Table 2.4: Main performance of the ATLAS tracking detector.

The **Inner Detector** (ID), composed by the IBL, pixel and SCT, covers the region $|\eta| < 2.5$ (see figure 2.8 and 2.9) and globally provides a transverse impact parameter resolution of $\simeq 35$ ($\simeq 10$) μm for pions with p_T= 5 (100) GeV and a transverse momentum resolution of about 4% for 100 GeV muons.
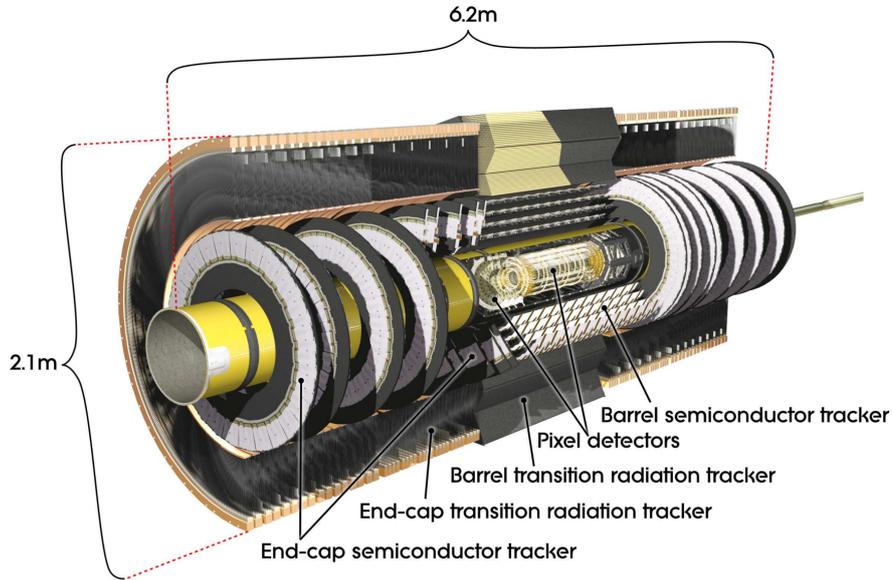
Figure 2.9: Cut-away view of the ATLAS inner detector [45]. Note that the IBL is still missing in this figure.

The ID is immersed in a 2 T magnetic field generated by the central solenoid, which extends over a length of 5.3 m with a diameter of 2.5 m. In the barrel region, the ID is arranged on concentric cylinders around the beam axis while in the end-cap regions it is located on disks perpendicular to the beam axis (the IBL is not present in this region). The highest granularity is achieved around the vertex region using

silicon pixel detectors.

The inner detector system provides tracking, vertex and impact parameter measurements in a range matched by the precision measurements of the electromagnetic calorimeter. The secondary vertex measurement performance is enhanced by the innermost layer of pixels and by the new IBL.

The combination of precision trackers at small radii with the TRT at a larger radius gives very robust pattern recognition and high precision in both R-ϕ and z coordinates.

The Insertable B-Layer

The IBL is the sub-detector of charged tracks nearest to the collision point, made of a single cilindric layer of silicon pixel and contributes to measure the decay vertices of short living particles and the impact parameter. It is positioned between a new thinner Beryllium beam and the inner Pixel layer (the B-Layer, not to be confused with the Insertable B-Layer). Its baseline layout is a barrel layer consisting of 14 staves. The average radial distance of the sensitive area from the beam is 33 mm whereas the total envelope of the IBL in radius is between 31 and 40 mm. The radiation length at $\eta = 0$ of the IBL is 1.54% of X_0.

The IBL is radiation harder than the Pixel detector and the pixel size is smaller ($50 \times 250~\mu\mathrm{m}^2$) in order to lower the occupancy at high luminosity. With its layout, the IBL provides a full ϕ coverage with an intrinsic precision in the coordinate measurements of 23 μm.
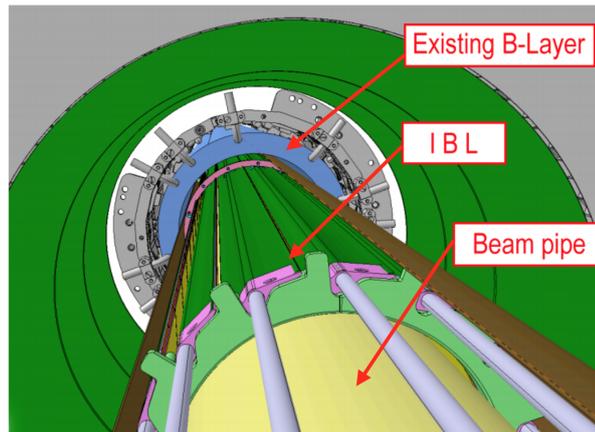
Figure 2.10: Scheme of the insertion of the IBL with the smaller beam pipe. [47].

Because of the low mass and close proximity to the interaction point, the IBL improves the quality of the impact parameter reconstruction and thereby the vertexing and b-tagging performance. For example, in the case of the b-tagging (discussed in detail in Chapter 4) the IBL reduces the probability of tagging the wrong particle type by a factor $\sim$2 [48].

Pixel detector

The pixel detector, figure 2.11, consists of 1744 pixel modules organized in three barrel layers, containing approximately 67 millions of pixels, complemented by three end-cap disks on each side, containing 13 millions of pixels (about 80 milion readout channels). The system covers a total active area of about 1.7 m^2 with an approximate radiation length of 0.1 X_0 at $\eta = 0$. To counterbalance the effect of the Lorentz deviation the pixel modules in the barrel region are tilted 20° with respect to the cylinder's tangent.

The pixel layers are segmented in R-ϕ and z with typically three pixel layers crossed by each track. All pixel sensors are identical and have a minimum pixel size in R-$\phi \times z$ of $50 \times 400\mu\mathrm{m}^2$. The intrinsic
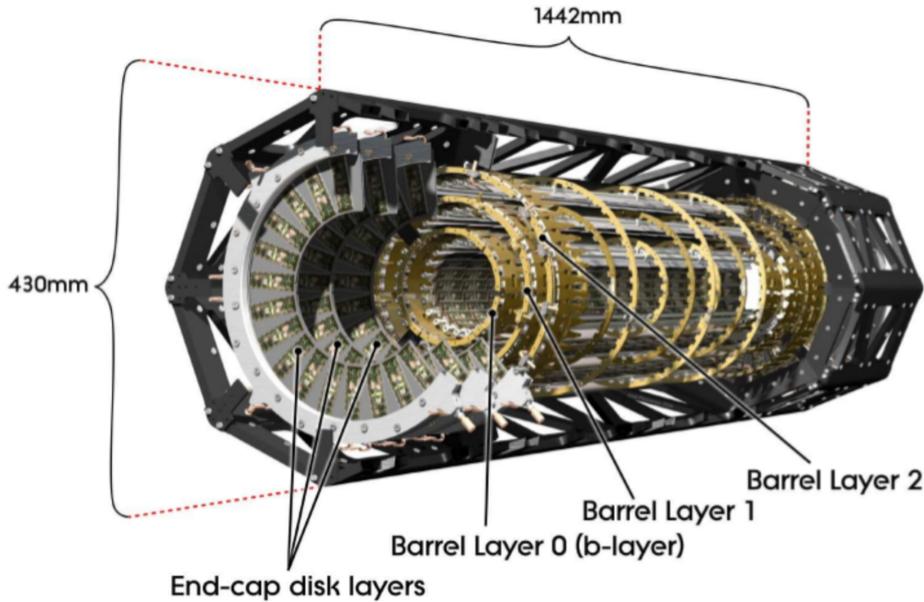
Figure 2.11: Cut-away view of the ATLAS pixel detector [45].

precision in the coordinate measurements in the barrel is 10 μm (R-ϕ) and 115 μm (z) and in the end-caps is 10 μm (R-ϕ) and 115 μm (R).

The micro-strip detector

The layout of the detector is given in figure 2.9.

The SCT system is designed to provide track precision measurements and to contribute to the measurement of momentum, impact parameter and vertex position in the intermediate radial range.

Eight strip layers (four space points) are crossed by each track. In the barrel region, this detector uses small-angle (40 mrad) stereo strips to measure both coordinates, with one set of strips in each layer parallel to the beam direction, measuring R-ϕ. They consist of two 6.4 cm long daisy-chained sensors with a strip pitch of 80 μm. The total number of readout channels in the SCT is approximately 6.3 million. The radiation length at $\eta = 0$ is approximately 0.1 X_0.

In the end-cap region, the detectors have a set of strips running radially and a set of stereo strips at an angle of 40 mrad. The mean pitch of the strips is also approximately 80 μm. The intrinsic measurement accuracies per module in the barrel are 17 μm (R-ϕ) and 580 μm (z) and in the end-caps are 17 μm (R-ϕ) and 580 μm (R).

The straw-tube tracking detector

The TRT is a combination of a tracker (based on the straw tubes) and a Transition Radiation detector for a pattern recognition. A Transition Radiation detector allows to discriminate between a lighter particle and a heavier one through several layers of material of different refraction indices, that produces transition radiation depending on the speed of the incident particle. The high relativistic particles (typically the electrons) can be recognised by the wider emitted radiation with respect to the other incident particles.

The electron identification capabilities are enhanced by the detection of transition-radiation photons in the xenon-based gas mixture of the straw tubes.

A single TRT component is composed by Polyimide drift (straw) tubes of 4 mm diameter that contains the anodes: tungsten wires plated gold, directly connected to the front-end electronics and kept at ground potential. The gap between the straw and the wire is filled by a mixture of gases (70% Xe, 27% CO_2 and 3% O_2). The passage of ionizing particle induce a low amplitude signal on the anodes. At the same time, some particles crossing polypropylene fibers cause transition radiation emission (in the X-ray spectrum) which is absorbed by the Xe present in the gas mixture; this last process leads to an high amplitude signal in the TRT electronic that can be distinguished from ionization signal by the voltage intensity.

A large number of hits (typically 36 per track) is provided by the straw tubes of the TRT, which enables track-following up to $|\eta| = 2.0$. The TRT only provides R-ϕ information, for which it has an intrinsic measurement accuracy of 130 μm per straw. In the barrel region, the straws are parallel to the beam axis and are 144 cm long, with their wires divided into two halves, approximately at $\eta = 0$. In the end-cap region, the straws are arranged radially in wheels. The radiation length at $\eta = 0$ is approximately 0.2 X_0.

2.4.2   Magnet system

ATLAS features a unique hybrid system of four large superconducting magnets. This magnetic system is 22 m in diameter and 26 m in length, with a stored energy of 1.6 GJ. Figure 2.12 shows the general layout, the four main layers of detectors and the four superconducting magnets which provide the magnetic field over a volume of approximately 12000 m^3.
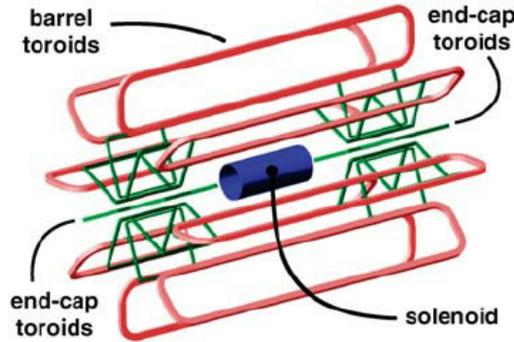
Figure 2.12: Geometry of magnet windings and tile calorimeter steel [45]. The eight barrel toroid coils, with the end-cap coils interleaved are visible. The solenoid winding lies inside the calorimeter volume.

The system consists of

- a **solenoid**, 5.3 m long and 2.5 m of diameter, which is aligned on the beam axis and provides a 2 T axial magnetic field for the inner detector, while minimising the radiative thickness in front of the barrel electromagnetic calorimeter. The layout was carefully optimized to keep the material thickness in front of the calorimeter as low as possible, resulting in the solenoid assembly contributing a total of $\simeq$ 0.66 radiation lengths at normal incidence;

- a **toroid** system, which provides a $\simeq$ 4 T field mostly orthogonal to the muon trajectories. It is constituted by eight Barrel Toroids (BT) 25 m long, with an inner core of 9.4 m and an outer diameter of 20.1 m, and two End-Cap Toroids (ECT) 5 m long (inner core 1.64 m, outer diameter 10.7 m). The end-cap toroid coil system is rotated by 22.5° with respect to the barrel toroid coil system in order to provide radial overlap and to optimise the bending power at the interface between the two coil systems. The toroid structure is open to minimize the uncertainty on the momentum measurements due to multiple scattering. The BT provides the particle bending in the region $\eta < 1$ while in

$1.4 < \eta < 2.7$ charged tracks are bent by the ECTs. In the transition region, $1 < \eta < 1.4$, magnetic deflection is provided by a combination of barrel and end-cap.

2.4.3   Calorimetry

In the difficult environment created by the high luminosity of the LHC machine, the calorimeter is designed to trigger and to provide precision energy and position measurements of electrons, photons, jets, neutrons and missing E_{T}.

A view of the sampling calorimeters and their main characteristics are presented in figure 2.13 and in table 2.5 respectively: the calorimeters cover the range $|\eta| < 4.9$, using different techniques suited to the widely varying requirements of the physics processes of interest and of the radiation environment over this large η-range. Over the η region matched to the inner detector, the fine granularity of the EM calorimeter is ideally suited for precision measurements of electrons and photons. The coarser granularity of the rest of the calorimeter is sufficient to satisfy the physics requirements for jet reconstruction and $E_{\mathrm{T}}^{\mathrm{miss}}$ measurements.
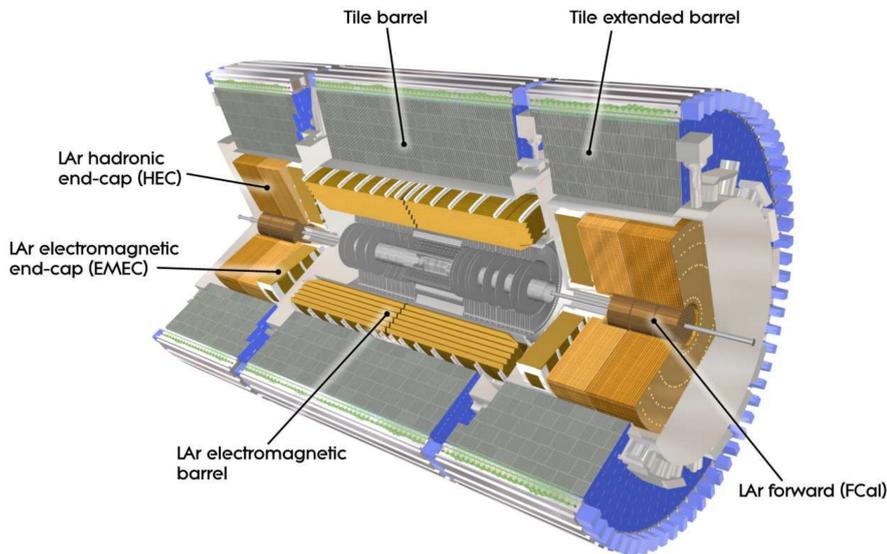
Figure 2.13: Cut-away view of the ATLAS calorimeter system [45].

The calorimeters closest to the beam-line are housed in three cryostats, one barrel and two end-caps. The barrel cryostat contains the electromagnetic barrel calorimeter, whereas the two end-cap cryostats each contain an electromagnetic end-cap calorimeter (EMEC), a hadronic end-cap calorimeter (HEC), located behind the EMEC, and a forward calorimeter (FCal) to cover the region closest to the beam. The Liquid Argon (LAr) forward calorimeters provide both electromagnetic and hadronic energy measurements.

All these calorimeters use liquid argon as the active detector medium, but the hadronic barrel one that uses scillintating tiles; liquid argon has been chosen for its intrinsic linear behaviour, its stability of response over time and its intrinsic radiation-hardness.

Calorimeters must provide good containment for electromagnetic and hadronic showers, and must also limit punch-through into the muon system. The total thickness of the EM calorimeter is more than 22 radiation lengths (X_0) in the barrel and more than 24 X_0 in the end-caps. The 9.7 interaction lengths (λ)

Calorimeter Type	Active material	Passive material	X_0	λ
EM LAr (barrel/end-cap)	LAr	Pb	22/24	2.2
Hadronic Tile	Scintillating tiles	Fe		9.7
Hadronic LAr end-cap	LAr	Cu		10
LAr Forward (EM/Hadronic)	LAr	Cu/W	27/91	2.7/3.7

Table 2.5: Main features of the sampling calorimeters of the ATLAS detector. The interaction (λ) and radiation (X_0) lengths values refer to $|\eta|$ = 0 and $|\eta|$ = 3.2 for the barrel and the end-cap regions respectively.

of the Hadronic calorimeter in the barrel (10 λ in the end-caps) are adequate to provide good resolution for high energy jets. The total thickness, including 1.3 λ from the outer support, is 11 λ at η = 0 and has been shown both by measurements and simulations to be sufficient to reduce punch-through well below the irreducible level of prompt or decay muons. Together with the large η-coverage, this thickness will also ensure a good E_T^{miss} measurement, which is important for many physics signatures and in particular for SUSY particle searches.

The required resolutions of both EM and Hadronic Calorimeter are reported in table 2.3.

LAr electromagnetic calorimeter

The EM calorimeter is a sampling calorimeter, with LAr as active medium and lead as passive medium. It is divided into a barrel part ($|\eta| < 1.475$) and two end-cap components ($1.375 < |\eta| < 3.2$). The position of the central solenoid in front of the EM calorimeter demands optimisation of the material in order to achieve the desired calorimeter performance.

The accordion geometry, which is one of a kind, provides complete ϕ symmetry without azimuthal cracks and is shown in figure 2.14. Over the region devoted to precision physics ($|\eta| < 2.5$), the EM calorimeter is segmented in three sections in depth. For the end-cap inner wheel, the calorimeter is segmented in two sections in depth and has a coarser lateral granularity than for the rest of the acceptance.

The readout system is composed by preamplifiers outside the cryostats, shapers, analog pipelines and an ADC system. The readout electrodes are located in the gaps between the lead absorbers and consist of three conductive copper layers separated by insulating polyimide sheets. The two outer layers are at the high-voltage potential and the inner one is used for reading out the signal via capacitive coupling. Each barrel and end-cap gap between two absorbers is equipped with two electrodes.

This calorimeter allows to achieve an energy resolution, independent on η for photons of p_{T} of 100 GeV better than 1.5% ($\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.2\%$).

Tile calorimeter

The tile calorimeter is a hadronic calorimeter placed directly outside the EM calorimeter. It is a sampling calorimeter using iron as the absorber and scintillating tiles as the active material. The barrel and extended barrels (in the η region between the barrel and the end-caps) are divided azimuthally into 64 modules. It is segmented in depth in three layers, approximately 1.5, 4.1 and 1.8 interaction lengths (λ) thick for the barrel and 1.5, 2.6, and 3.3 λ for the extended barrel. The total detector thickness at the outer edge of the tile-instrumented region is 9.7 λ at η = 0. Two sides of the scintillating tiles are read out by wavelength
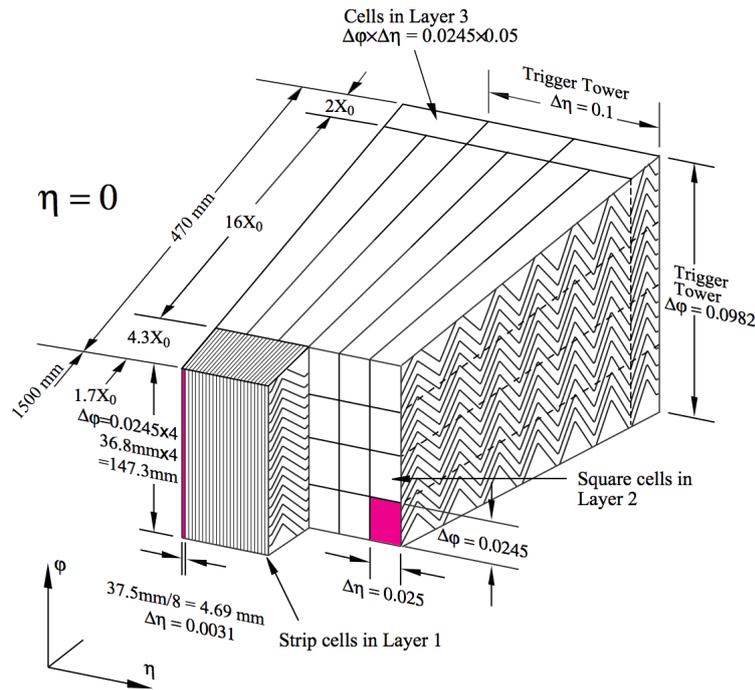
Figure 2.14: Sketch of a barrel module with the accordion geometry. The granularity in η and ϕ cells of each of the three layers and of the trigger towers is also shown.

shifting fibres into two separate photomultiplier tubes. In η, the readout cells built by grouping fibres into the photomultipliers are pseudo-projective towards the interaction region.

The tile calorimeter measures jet energies with a resolution $\sigma(E)/E = 50\%\sqrt{E} \oplus 2.5\% \oplus 5\%/E$, where the energy E is in GeV.

LAr hadronic end-cap calorimeter

The Hadronic End-cap Calorimeter consists of two independent wheels per end-cap, located directly behind the end-cap electromagnetic calorimeter. Each wheel is built from 32 identical wedge-shaped modules, assembled with fixtures at the periphery and at the central bore. Each wheel is divided into two segments in depth, for a total of four layers per end-cap. The wheels closest to the interaction point are built from 25 mm parallel copper plates, as passive material, while those further away use 50 mm copper plates. The copper plates are interleaved with 8.5 mm LAr gaps, providing the active medium for this sampling calorimeter.

LAr forward calorimeter

The Forward Calorimeter (FCal) is an electromagnetic and hadronic calorimeter. The FCal is approximately 10 interaction lengths deep, and consists of three modules (see figure 2.15) in each end-cap: the first, with copper as passive material, is optimised for electromagnetic measurements, while the other two, with tungsten as passive material, measure predominantly the energy of hadronic interactions. Each module consists of a metal matrix, with regularly spaced longitudinal channels filled with the electrode structure consisting of concentric rods and tubes parallel to the beam axis. The LAr in the gap between

the rod and the tube is the sensitive medium. This geometry allows for excellent control of the gaps, which are as small as 0.25 mm in the first section, in order to avoid problems due to ion buildup.
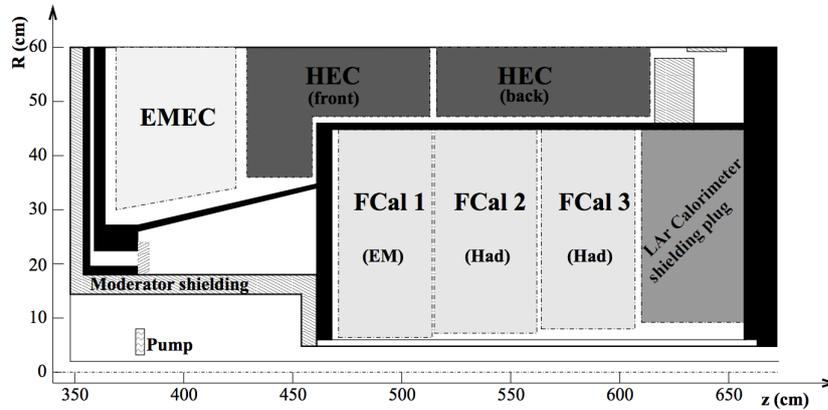
Figure 2.15: Schematic diagram showing the three FCal modules located in the end-cap cryostat. The material in front of the FCal and the shielding plug behind it are also shown. The black regions are structural parts of the cryostat. The diagram has a larger vertical scale for clarity.

2.4.4   Muon system

The muon system will be discussed in more details because my service task concerns the study of its performances for the Phase-2 upgrade (see Appendix 7).

The conceptual layout of the muon spectrometer is shown in figure 2.16; it is composed by separate trigger and high-precision tracking chambers and it measures tracks and momentum of muons through their magnetic deflection in the large superconducting air-core toroid magnets. The anticipated high level of particle flux has had a major impact on the choice and design of the spectrometer instrumentation, affecting performance parameters such as rate capability, granularity, ageing properties, and radiation hardness.

In the barrel region, tracks are measured in chambers arranged in three cylindrical layers around the beam axis; in the transition and end-cap regions, the chambers are installed in planes perpendicular to the beam, also in three layers.

Over most of the η-range, a precision measurement of the track coordinates in the principal bending direction of the magnetic field is provided by *Monitored Drift Tubes* (MDT's). At large pseudorapidities ($2 < \eta < 2.7$), the same measurements are performed by the *Cathode Strip Chambers* (CSC's), the multi-wire proportional chambers with cathodes segmented into strips. The stringent requirements on the relative alignment of the muon chamber layers are met by the combination of precision mechanical-assembly techniques and optical alignment systems both within and between muon chambers.

The trigger system covers the pseudorapidity range $|\eta| < 2.4$ by the use of *Resistive Plate Chambers* (RPC's) in the barrel and *Thin Gap Chambers* (TGC's) in the end-cap regions (the "big wheels" in the ATLAS jargon). The trigger chambers for the muon spectrometer serve a threefold purpose: provide bunch-crossing identification, provide signals above well-defined p_T thresholds, and measure the muon coordinate in the direction orthogonal to the one determined by the precision-tracking chambers.
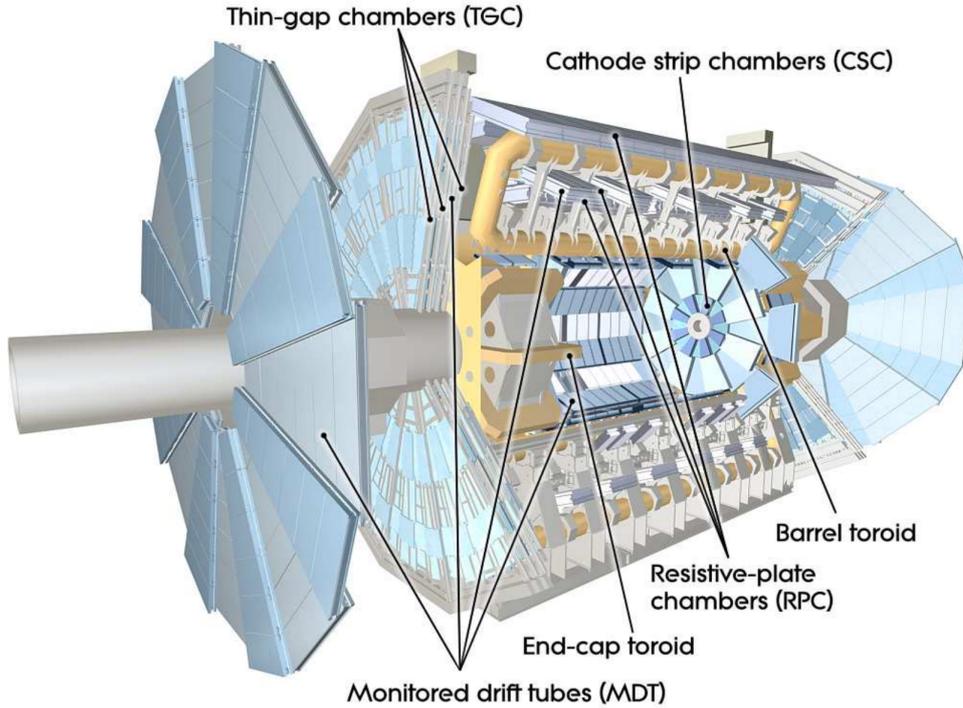
Figure 2.16: Cut-away view of the ATLAS muon system [45].

Table 2.6 gives the main characteristic of the muon detector and the intrinsic time and position resolution of the detectors, to which contributions from signal propagation and electronics have to be added. The design goal was to keep these contributions low enough in order to obtain $\geq 99\%$ probability of beam-crossing identification. Both chamber types (trigger and tracking) deliver signals with a spread of 15-25 ns (including the signal propagation and the electronics contribution), thus providing the ability to tag the beam-crossing. The trigger and CSC chambers measure both coordinates of the track, one (η) in the bending plane and one (ϕ) in the non-bending plane.

Type	Function	Chamber resolution in			Hit/track		Number of	
		z/R	ϕ	time	barrel	end-cap	chambers	channels
MDT	tracking	35 μm (z)	-	-	20	20	1088	339k
CSC	tracking	40 μm (R)	5 mm	7 ns	-	4	32	30.7k
RPC	trigger	10 mm (z)	10 mm	1.5 ns	6	-	544	359k
TGC	trigger	2-6 mm (R)	3-7 mm	4 ns	-	9	3588	318k

Table 2.6: Parameters of the four sub-systems of the muon detector [45]. The quoted spatial resolution (columns 3, 4) does not include chamber-alignment uncertainties. Column 5 lists the intrinsic time resolution of each chamber type, to which contributions from signal-propagation and electronics contributions need to be added.

Monitored Drift Tube Chambers

The basic element of the monitored drift tube chambers is a pressurised drift tube with a diameter of 29.970 mm, operating with Ar/CO_2 gas (93/7) at 3 bar. The electrons resulting from ionisation are collected at

the central tungsten-rhenium wire with a diameter of 50 μm, at a potential of 3080 V. The wire is held in position at the tube ends by a cylindrical end-plug which guarantees the concentricity of the wire with respect to the tube with an accuracy of $\sigma < 10$ μm. The central conductor holding the wire also serves for the gas transfer in and out of the tube.

The choice of the tubes in the building of the precision-tracking chambers offers several advantages:

- the stiffness of the tube assembly allows to combine high mechanical precision with robustness of the chambers;

- a high level of operational reliability can be expected because the failure of a single tube does not affect the operation of most of the others;

- the cylindrical geometry results in a radial electric field: the measurement accuracy, therefore, depends only weakly on the angle of incidence of the particle onto the chamber plane, so the coordinate of the track is determined by the radius of the circle around the wire to which the track is tangential. This is important because the angle of incidence of infinite momentum tracks onto the chamber plane extends up to $45°$.

With these features, the MDT chambers reach a resolution of 80 μm per tube layer (different with respect to the value quoted in table 2.6 because it is not referred to a single plane, like z/R or z).

The operating gas was selected because of the good ageing properties. Deposits on the wires have never been observed in clean samples of this gas mixture, the formation of polymers not being possible in the absence of hydrogen. A disadvantage of this gas mixture is the non-linear space-drift time relation and the drift time of about 700 ns, which is about 50% longer than the typical gases such as Ar/CH$_4$. The non-linearity of the Ar/CO$_2$ gas leads to a reduction of spatial resolution at high counting rates due to the distortion of the electric field created by the positive ions.

Cathode-strip Chambers

The CSC's are multiwire proportional chambers with the wires oriented in the radial direction with respect to the beam. Both cathodes are segmented, one with the strips perpendicular to the wires (providing the longitudinal coordinate) and the other parallel to the wires providing the transverse coordinate. The position of the track is obtained by interpolation between the charges induced on neighbouring cathode strips. The CSC wire signals are not read out. The resolution achieved with this procedure depends on the signal-to-noise ratio and the readout pitch, the latter being the main cost-driving factor for the readout electronics. With a readout pitch of 5.31 mm and 5.56 mm for the large and small chambers respectively in the bending direction, the CSC reaches a resolution of 60 μm per CSC plane, to be compared with the 80 μm resolution of a MDT tube layer. In the non-bending direction the cathode segmentation is coarser leading to a resolution of 5 mm.

As in the case of the MDT's, the CSC's are segmented into large and small chambers in ϕ. The whole CSC system consists of two disks with eight chambers each (eight small and eight large). Each chamber contains four CSC planes resulting in four independent measurements in η and ϕ along each track.

The limit for safe operation of the MDT's is at counting rates of about 150 Hz/cm^2, which will be exceeded in the region $|\eta| > 2$ in the first layer of the end-cap. In this η region of the first layer, the MDT's are replaced by CSC's, which combine high spatial, time and double track resolution with high-rate capability and low neutron sensitivity. Operation is considered safe up to counting rates of about 1000 Hz/cm^2, which is sufficient up to the forward boundary of the muon system at $|\eta| = 2.7$.

Resistive Plate Chambers

The trigger system in the barrel consists of three concentric cylindrical layers around the beam axis. The large lever arm between inner and outer RPC's permits the trigger to select high momentum tracks in the range 9-35 GeV (high-p_T trigger), while the two inner chambers provide the low-p_T trigger in the range 6-9 GeV, as shown in figure 2.17.
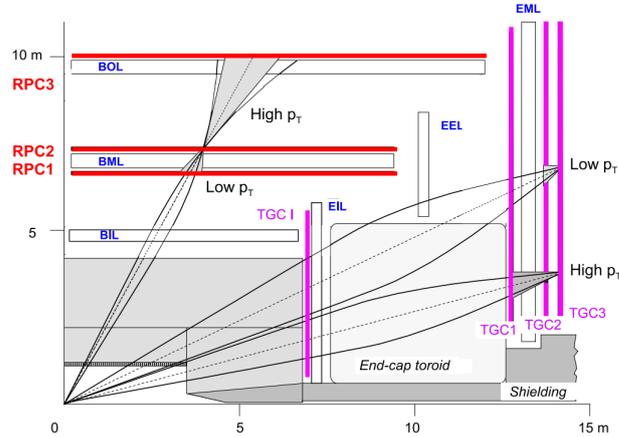
Figure 2.17: Schematics of the muon trigger system [45]. RPC2 and TGC3 are the reference (pivot) planes for barrel and end-cap, respectively.

Each station consists of two independent detector layers, each measuring η and ϕ. A track going through all three stations thus delivers six measurements in η and ϕ. This redundancy in the track measurement allows the use of a 3-out-of-4 coincidence in both projections for the low-p_T trigger (RPC1 and RPC2 stations) and a 1-out-of-2 OR for the high-p_T trigger (RPC3 station). This coincidence scheme rejects fake tracks from noise hits and greatly improves the trigger efficiency in the presence of small chamber inefficiencies.

The RPC is a gaseous parallel electrode-plate detector. Two resistive plates are kept parallel to each other at a distance of 2 mm by insulating spacers. The electric field between the plates of about 4.9 kV/mm allows avalanches to form along the ionising tracks towards the anode. The signal is read out via capacitive coupling to metallic strips, which are mounted on the outer faces of the resistive plates. The gas used is a mixture of $C_2H_2F_4$/Iso-C_4H_{10}/SF_6 (94.7/5/0.3) which combines relatively low operating voltage, non-flammability and low cost, while providing a comfortable plateau for safe avalanche operation.

RPC's can operate both in avalanche and in streamer mode. In the high background environment encountered at the LHC, the avalanche mode offers the benefit of higher rate capability and rate-independent time resolution and has therefore been selected as the operation mode. At the nominal operating voltage of 9.8 kV, a signal with a width of about 5 ns is generated by the track with a streamer probability less than 1%.

During my qualification period, I was involved in the upgrade of the RPC system of the muon spectrometer. The upgrade consists in the addition of an inner layer, the so-called *RPC0*, that improves both the η coverage and the decision rate at higher luminosity. The study is explained in detail in Appendix 7.

Thin Gap Chambers

TGC's are multi-wire proportional chambers with the characteristic that the wire-to-cathode distance of 1.4 mm is smaller than the wire-to-wire distance of 1.8 mm. With a highly quenching gas mixture of CO_2

and n-C$_5$H$_{12}$ (n-pentane), this cell geometry allows for operation in a quasi-saturated mode. The highly quenching gas prevents the occurrence of streamers in all operating conditions.

TGC's provide two functions in the end-cap muon spectrometer: the muon trigger capability and the determination of the second azimuthal coordinate to complement the measurement of the MDT's in the bending (radial) direction. The TGC's need good time resolution to tag the beam-crossing with high efficiency ($\geq 99\%$) and fine granularity to provide a sufficiently sharp cut-off in the momentum of the triggering muon. To match the granularity to the required momentum resolution, the size of the wire groups varies from 6 to 31 mm as a function of η, corresponding to a variation in width from 10.8 mm to 55.8 mm.

The high electric field around the TGC wires and the small wire-to-wire distance lead to very good time resolution (4 ns) for the large majority of the tracks.

2.4.5   Forward detectors

Three smaller detector systems [49] cover the ATLAS forward region: *LUCID* (LUminosity measurement using Cerenkov Integrating Detector), *ALFA* (Absolute Luminosity For ATLAS) and *ZDC* (Zero-Degree Calorimeter). The main functions of the first two detectors is to determine the luminosity measurements delivered to ATLAS and the third to determine the centrality of the heavy ion collisions. At ± 17 m from the interaction point lies *LUCID*; it detects inelastic pp scattering in the forward direction and it is the main online relative-luminosity monitor for ATLAS. *ALFA* is located at ± 240 m and it consists of scintillating fiber trackers located inside Roman pots designed to approach as close as 1 mm to the beam. *ZDC* plays a key role in determining the centrality of heavy-ion collisions. It is located at ± 140 m from the interaction point, just beyond the point where the common straight-section vacuum-pipe divides back into two independent beam-pipes. The ZDC modules consist of layers of alternating quartz rods and tungsten plates which will measure neutral particles at pseudorapidities $|\eta| \geq 8.2$.

LUCID

LUCID is a Cherenkov detector specifically designed for measuring the luminosity. Two detectors are placed around the beam-pipe on both forward ends of the ATLAS detector. Each detector consists of 16 photomultipliers and four quartz fiber bundles. The photomultipliers (PMTs) detect charged particles that traverse their quartz windows, where Cherenkov light is produced. Cherenkov light is produced in the fiber bundles as well and carried to PMTs that are protected by shielding about 2 m away. To increase the detector lifetime, only a subset of the PMTs is used at a given time, the others being available as spares. In addition, 4 PMTs have a reduced window opening to decrease their acceptance and thus avoid saturation of some luminosity algorithms.

The advantages of the LUCID design are that the detector is composed of light material, which is intrinsically radiation hard, and the detector response is fast, within few ns from the moment a particle traverses the detector. The bandwidth of the electronics allow to keep the FWHM (Full Width at Half Maximum) of the PMT signals at a level of 10 ns (fast baseline restoration). These characteristics ensure the capability to separate collisions coming from different bunch crossings (separated by 25 ns) and make LUCID suitable for online monitoring of the bunch structure.

2.4.6   Trigger, readout, data acquisition and control systems

The huge amount of data collected by LHC in each collision cannot be completely acquired and analyzed. ATLAS, through a complex Trigger and Data Acquisition (collectively TDAQ [50]) system, manages the

selections of events, needed to reduce the information rate from $\sim$ 40 MHz to approximately 200 Hz.

The TDAQ systems, the timing- and trigger-control logic, and the Detector Control System (DCS) are partitioned into sub-systems, typically associated with each sub-detector, with the same logical components and building blocks.

Trigger system

The trigger system has two distinct levels: L1 and the High-Level Trigger (HLT). Each trigger level refines the decisions made at the previous level and, where necessary, applies additional selection criteria.

- The **L1 trigger** searches for high transverse-momentum muons, electrons, photons, jets, and τ-leptons decaying into hadrons, as well as large missing and total transverse energy. Its selection is based on information from a subset of detectors. High transverse-momentum muons are identified using trigger chambers in the barrel (RPC) and end-cap regions (TGC) of the spectrometer. Calorimeter selections are based on reduced-granularity information from all the calorimeters. Results from the L1 muon and calorimeter triggers are processed by the central trigger processor, which combines the different trigger selections. Events passing the L1 trigger selection are transferred to the next stages of the detector-specific electronics and subsequently to the data acquisition via point-to-point links. In each event, the L1 trigger also defines one or more Regions-of-Interest (RoI's), i.e. the geographical coordinates in η and ϕ, of those regions within the detector where its selection process has identified interesting features. The RoI data include information on the type of feature identified and the criteria passed. This information is subsequently used by the high-level trigger. The diagram of the L1 operation is reported in figure 2.18.
  The maximum L1 accept rate which the detector readout systems can handle is 100 kHz, reduced from the initial 40 MHz.

- The **HLT** selection is seeded by the RoI information provided by the L1 trigger over a dedicated data path. The selections use, at full granularity and precision, all the available detector data within the RoI's and offline analysis procedures. The system merges together the Run-1 two stage system (L2 and Event Filer) to a single farm in order to reduce the complexity and the dusplication of CPU usage and network transfer. The configurations and thresholds are set to reduce the trigger rate to approximately 1 kHz.

Readout architecture and data acquisition

The *Readout Drivers* (RODs) are detector-specific functional elements of the front-end systems, which achieve a higher level of data concentration and multiplexing by gathering information from several front-end data streams. The front-end electronics sub-system includes different functional components:

- the front-end analogue or analogue-to-digital processing;

- the L1 buffer in which the (analogue or digital) information is retained for a time long enough to accommodate the L1 trigger latency;

- the derandomising buffer in which the data corresponding to a L1 trigger accept are stored before being sent to the following level. This element is necessary to accommodate the maximum instantaneous L1 rate without introducing significant deadtime (maximum 1%);

- the dedicated links or buses which are used to transmit the front-end data stream to the next stage.
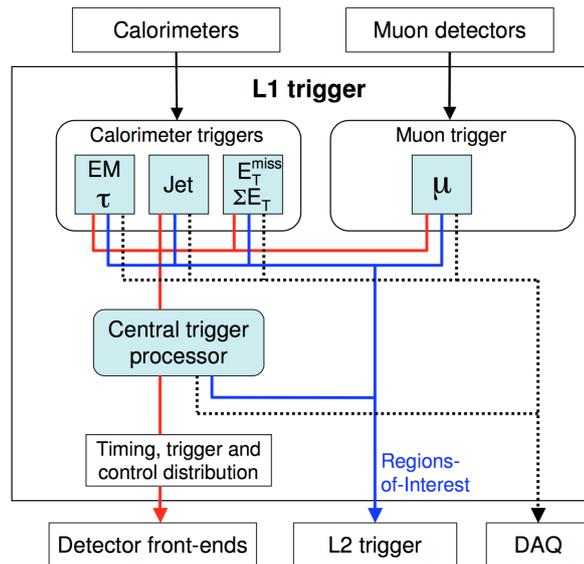
Figure 2.18: Block diagram of the L1 trigger. The overall L1 decision is made by the central trigger processor, taking input from calorimeter and muon trigger results. The paths to the detector front-ends, L2 trigger, and data acquisition system are shown from left to right in red, blue and black, respectively.

After an event is accepted by the L1 trigger, the data from the pipe-lines are transferred off the detector to the RODs. Digitised signals are formatted as raw data prior to being transferred to the DAQ system. The RODs follow some general ATLAS rules, including the definition of the data format of the event (changed significantly from Run-1 to Run-2), the error detection/recovery mechanisms to be implemented, and the physical interface for the data transmission to the DAQ system.

The first stage of the DAQ, the readout system, receives and temporarily stores the data in local buffers. It is subsequently solicited by the HLT that performs the final selection. Selected events are moved to permanent storage at the CERN computer centre. In addition to the movement of data, the data acquisition system also provides for the configuration, control and monitoring of the hardware and software components which together provide the data-taking functionality.

Events stored by the TDAQ are made available to all the ATLAS analysis teams so it is necessary to apply further selection cuts during the offline analysis to reject events selected by triggers not specific for particular analysis on which we are interested for. The event selection applied in this thesis will be presented in Chapter 3.

The Detector Control System

The Detector Control System (DCS) permits the coherent and safe operation of the ATLAS detector hardware and serves as a homogeneous interface to all sub-detectors and to the technical infrastructure of the experiment. It controls, continuously monitoring and archiving the operational parameters, signals indicating any abnormal behaviour to the operator, and allows automatic or manual corrective actions to be taken. The DCS also enables bi-directional communication with the data acquisition system in order to synchronise the state of the detector with data-taking. It also handles the communication between the sub-detectors and other systems which are controlled independently, such as the LHC accelerator, the CERN technical services, the ATLAS magnets, and the detector safety system.

Chapter 3

Data and Monte Carlo samples

The analysis uses the data collected by the ATLAS detector in 2015 and 2016, corresponding to an integrated luminosity of 36.5 fb^{-1}. Several Monte Carlo samples have been used to simulate the physics processes both for signal and background.

In this section the data samples and the simulations will be discussed in details.

3.1 Data samples

The analysis described in this thesis has been performed using the full set of data collected by the ATLAS detector in 2015 and 2016, at $\sqrt{s}$ = 13 TeV, corresponding to $\int \mathcal{L}dt$ = 36.5 fb^{-1}. In figures 2.4a and 2.4b, the luminosity collected in the full period is shown as a function of the time. The difference between the recorded and delivered luminosity accounts for stable conditions, of the beam and all the sub-detector systems, that must work properly during the acquisition time. For the overall data taking, the ratio between the recorded and the delivered luminosity is $\sim 93\%$.

The corresponding integrated luminosity used in this analysis refers to the 2015 data set (periods D-J5), corresponding to 3.2 fb^{-1} and the full 2016 data set (periods A-L) corresponding to 33.3 fb^{-1}.

As described in section 3.4, one of the most interesting $t\bar{t}H$ decay channels contains a charged lepton coming from a leptonic decay of a W. Since this lepton has usually a large (> 20 GeV) transverse momentum with respect to other leptons produced in pp collisions, it is well suited for the very first event selection, happening already during data taking. For the rest of the thesis, the events of interest are those acquired with at least one lepton (electron or muon) identified at the first level trigger (L1), confirmed and well reconstructed at the next step, called High Level Trigger (HLT). The trigger sequence applied on a given lepton at L1 and HLT is called in the following "trigger chain". The exact selection on the lepton candidates performed at L1 and at HLT is differentiated for the 2015 and 2016 data sets to take into account differences in the instantaneous luminosity. The main differences are in the p_{T} threshold, "e24"

for electrons or "mu50" for muons; in the isolation requirement, "lhmedium" for electrons or "iloose" for muons (described in details in Chapter 4); in some further cuts indicated by acronyms like "VH" (for those triggers seeded by L1 items with η-dependent thresholds and a hadronic leakage requirement [51] [52]) and "nod0", used for the 2016 data.

When using the 2015 data set, the following trigger chains are used:

- electrons: `HLT_e24_lhmedium_L1EM20VH` OR `HLT_e60_lhmedium` OR `HLT_e120_lhloose`,

- muons: `HLT_mu50` OR `HLT_mu20_iloose_L1MU15`.

When using the 2016 data set, the following trigger chains are used:

- electrons: `HLT_e24_lhtight_nod0_ivarloose` OR `HLT_e60_lhmedium_nod0` OR `HLT_e140_lhloose_nod0`,

- muons: `HLT_mu50` OR `HLT_mu24_ivarmedium`.

All the data taking information are summarised in table 3.1.

Periods	Luminosity (fb^{-1})	Triggers (electrons)	Trigger (muons)
2015 (D-J5)	3.2	`HLT_e24_lhmedium_L1EM20VH` OR `HLT_e60_lhmedium` OR `HLT_e120_lhloose`	`HLT_mu20_iloose_L1MU15` OR `HLT_mu50`
2016 (A-L)	33.3	`HLT_e24_lhtight_nod0_ivarloose` OR `HLT_e60_lhmedium_nod0` OR `HLT_e140_lhloose_nod0`	`HLT_mu24_ivarmedium` OR `HLT_mu50`

Table 3.1: A summary of run periods and triggers used in this analysis.

3.1.1   Derivation Framework

The physics analysts need to be able to run over the data sample frequently, for the purposes of adding new variables and cuts, fixing bugs, etc. To speed up this process, it is customary to have a careful selection of interesting events where all the analysis (on signal and minimal needed background) can be performed, reducing at the same time the amount of data on storage to be processed. The analysis model used during Run-1 foresaw that this job was usually performed by single users, duplicating similar selections and loosing the possibility to an overall control of the procedure.

In Run-2, the reduction of the acquired data (referred to as "derivations" from now on) has been made centrally, and the purpose of the Derivation Framework is to provide the same offline software tools for all the analyzers in a transparent way. In this contest, I had the fundamental role to coordinate the needs of the analysis research group in order to build a suitable derivation that fulfils the requirements of both the analysis and the central framework. For 2015 and 2016 dataset the derivations have been performed starting from the output of the general reconstruction framework ATHENA, in a format called xAOD (see
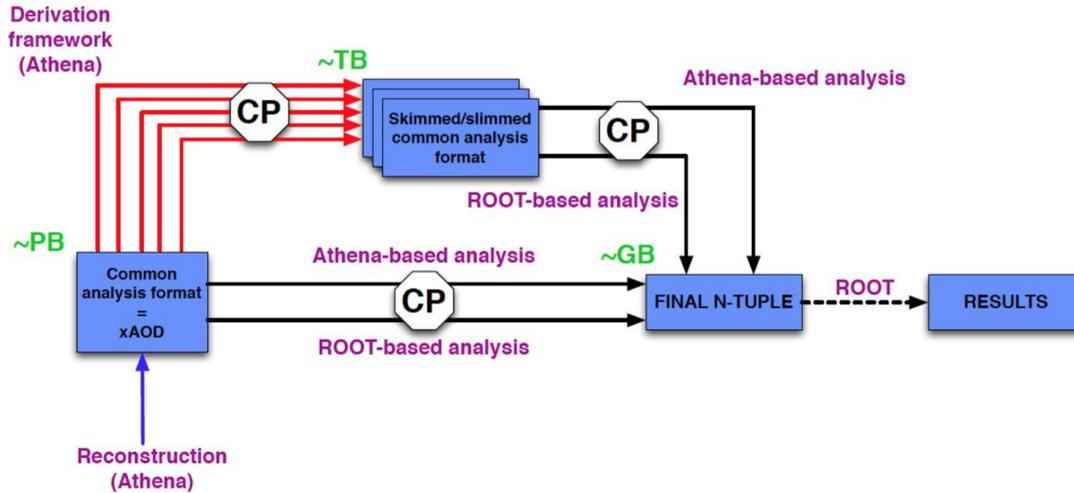
Figure 3.1: Scheme of the Derivation Framework adopted by the ATLAS collaboration for the Run-2. The data are reconstructed through Athena and derived by the Derivation Framework, under the Athena system; the data samples size decreases from the order of PB to few GB. The Combined Performances (CP) help the derivation giving informations about the objects to use and the calibrations to apply on them.

figure 3.1), and providing an output in the same general format, as xAOD, but containing less data (so called Derived-xAOD, DxAOD).

Derivations are built from the full reconstructed data via four operations:

- skimming: removing whole events;

- thinning: removing whole reconstructed objects (tracks, calorimeter clusters, jets, vertices, etc) from within an event, but keeping the rest of the event;

- slimming: removing not necessary information from objects and keeping the rest;

- augmentation: adding information not found in the input data.

Derivations have been defined by individual physics analysis teams or subgroups according to the specific analyses needs with a target size of a few TB for the input data. The total derivation chain decreases the size by a factor 10^3. The framework is generally run on the production system and through the concept of train production can produce a number of independent output formats from a single input file.

3.1.2 Derivation used for the $t\bar{t}H$ analysis

The analysis described in this thesis uses a derivation called TOPQ1, which is one of the four derivations of the Top Group. It has been defined by

- skimming: it requires at least one leptons (muon OR electron) with $p_\mathrm{T} > 20$ GeV inside $|\eta| < 2.5$;

- thinning: it thins tracks, calo clusters and the truth record in the case of Monte Carlo simulations according to the needs of the analysis. The object is removed if it does not pass some quality and isolation cuts, e.g. the number and the p_T of the tracks associated to a vertex;

- slimming: in this step, if a variable or collection is not explicitly asked to be retained, it will not be written out to the DxAOD. The slimming is performed on variable lists for the following analysis objects: calibrated jets and b-jets, tracks, primary vertices, electrons, muons, jets originated from taus and $E_{\mathrm{T}}^{\mathrm{miss}}$;

- augmentation: on the MC samples, it adds a flag (called "decoration") for the $t\bar{t}$ classification ($t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$, $t\bar{t}$+light) recovered from the truth record and used in the background identification (section 3.5.1).

This pre-selection is very loose in order to have all the events needed for several analyses. The objects "muon", "electron", "jets" and "b-jets" are used in according to the Combined Performances recommendations and their definitions will be discussed in more details in Chapter 4.

3.2   Signal and backgrounds in boosted $t\bar{t}H$ analysis

The analysis described in this thesis studies the production of the $t\bar{t}H$ channel with the Higgs boson decaying in $b\bar{b}$ pair and the $t\bar{t}$ system decaying semileptonically (*lepton + jets*, as in figure 1.24, in Chapter 1) and it will be discussed in details in Chapters 4 and 5. The final state is required to be in the *boosted* regime.

The background sources affecting the $t\bar{t}H$ channel (see figure 3.2) are all those processes presenting a final signature similar to the $t\bar{t}H$ one. It could depends on the similarity of the decay products or on the not negligible probability of objects misidentification. The main background contributions are given by the following production processes:

- $t\bar{t}$+jets (classified by quarks flavour in $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$ and $t\bar{t}$+light quarks, called $t\bar{t}$+light);

- single top;

- W/Z+jets;

- $t\bar{t}+V$ (V stays for a Vector Boson);

- diboson;

- multi-jets.

All simulated event samples (signal and background) are passed through the full ATLAS detector simulation using GEANT4 (see section 3.3) and are processed through the same reconstruction software of the data. Simulated MC events are corrected so that the object identification efficiencies, energy scales and energy resolutions match those determined from data.

Table 3.2 provides a summary of the basic characteristics of the MC samples of each process.

3.3   Monte Carlo simulations

The aim of using the MC simulations is to better comprehend the experimental apparatus behaviour by the comparison with the experimental data, to determine the efficiency of a certain analysis and to study for future experiments.
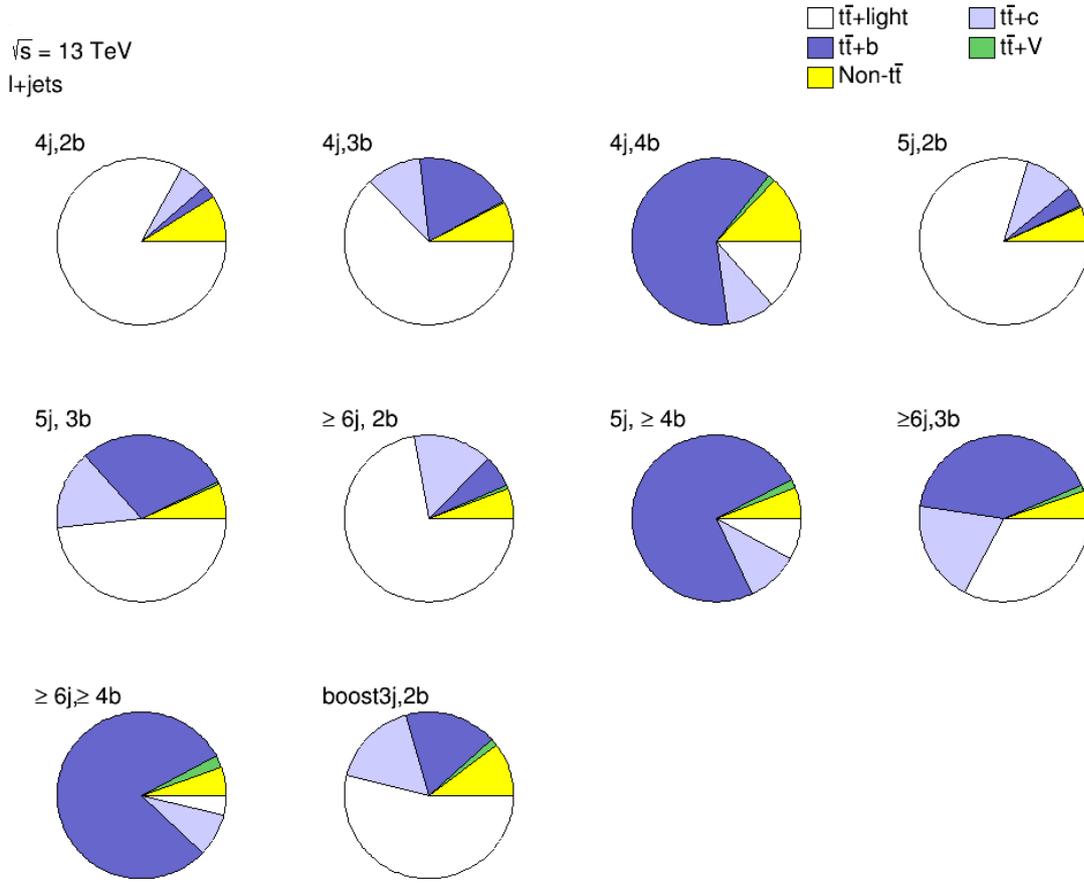
Figure 3.2: Background composition for the resolved and boosted analysis regions (described in Chapter 5) relative to an integrated luminosity $\mathcal{L}$ = 36.5 fb^{-1}. The Heavy Flavour classification is shown (see section 3.5.1).

Sample	Generator	PDF	Parton Shower	Normalisation
$t\bar{t}H$	aMC@NLO	NNPDF3.0NLO	Pythia 8.2	(N)NLO
$t\bar{t}$+jets	Powheg	CTEQ6L1	Pythia 6.428	NNLO+NNLL
W/Z + jets	Sherpa	CT10	Sherpa 2.1.1	NNLO
Single top (s-, Wt-channels)	Powheg	CT10	Pythia 6.428	aNNLO
Single top (t-channel)	Powheg	CT10f4	Pythia 6.428	aNNLO
$t\bar{t} + V$	Madgraph	CTEQ6L1	Pythia 6.425	NLO
Diboson	Sherpa	CT10	Sherpa 2.1.1	NLO

Table 3.2: A summary of basic generator parameters used to simulate all the signal and background processes.

The physics processes and the interactions of the final state particles with the detector are simulated by general-purpose Monte Carlo (GPMC) softwares, that produce the samples according to both theoretical and phenomenological models. They are built from several components, that describe the physics starting from very short distance scales, up to the typical scale of hadron formation and decay. Since QCD is weakly interacting at short distances (below a femtometer), the components of the GPMC dealing with short-distance physics are based upon perturbation theory. At larger distances, all soft hadronic phenomena, like hadronization and the formation of the underlying event, cannot be computed from first principles and, consequently, follow QCD-inspired models.

3.3.1   The event generation

A pp collision is a complicated process because of the composite internal structure of hadrons, an illustrative example of a $t\bar{t}H$ production event is shown in figure 3.3. The final state is described as the superposition of several contributions: the initial state radiation, the partonic hard scattering (depicted as a red blob) that generates the highest transverse momentum partons, followed by the final state radiation, the hadronization (the dark green objects) and the eventual decay. Other final state partons are produced in additional multiple parton interactions and remnants, the so-called *underlying events* (the purple blob), that did not participate in the hard scattering.

The event generation is divided in several steps. In the first step, the scattering probability of the hard process is calculated through the evaluation of the Matrix Elements (MEs) at a fixed perturbative order in the strong coupling constant α_S. The momenta of the initial partons (continuous green lines in figure 3.3) are randomly sampled on the proton PDFs (see Chapter 1) and the outgoing partons (particles going out of the red blob) are randomly distributed in the available phase space. The cross section of any process can be calculated as a convolution of the hadrons PDFs with the partonic cross sections (factorization theorem [1], see eq. 1.4.64). However, the partonic cross sections show collinear divergences connected to long-distance soft interactions. In order to make this calculation viable, the divergences are factored out and absorbed into a redefinition of the PDFs introducing a *factorization scale*, that separates long-distance soft physics (non perturbative part of the interaction) from the short-distance hard process (calculable perturbatively). The factorization scale is often set to the same value at which α_S is evaluated, i.e. the *renormalization scale*, usually set to the mass or the transverse momentum of the final state system. The ME is calculated differently depending on the MC generators, as explained in the following.

This step is followed by the *parton shower* (PS), the QCD cascade generation that describes the final state radiation, sketched with blue curly and straight lines in figure 3.3, as higher order QCD effects. Parton radiation is described as successive parton emissions from the hard interaction scale to the hadronization scale ~ 1 GeV. The development is characterized by an evolution variable, whose typical choices are virtual squared mass of the showering parton (referred to as *virtuality*), the transverse squared momentum of the emitted and remained partons, and $E^2(1 - \cos\theta)$, where E is the energy of the parent parton and θ the angle with the emitted partons. In the gluon splitting processes ($g \rightarrow q\bar{q}$, $g \rightarrow gg$) in the collinear approximation, the distribution of the split pair is not uniform in azimuth. The split correlations are handled in different ways depending on the MC generator: Pythia [56] averages the azimuthal correlation between two successive splitting over the polarizations; instead, Herwig [57] (and Herwig++ as well) fully includes the spin correlation effects [1].

Several other corrections, as the soft emissions, have to be taken into account. In QCD, there are two sources of large logarithms of infrared origin; the first comes from collinear singularities, which arise when two final state particles or a final state and an initial state particle become collinear and the second is the emission of soft gluons at arbitrary angles. Within the conventional parton-shower formalism, based on collinear factorization, it was shown that the last effect can be correctly described by using the
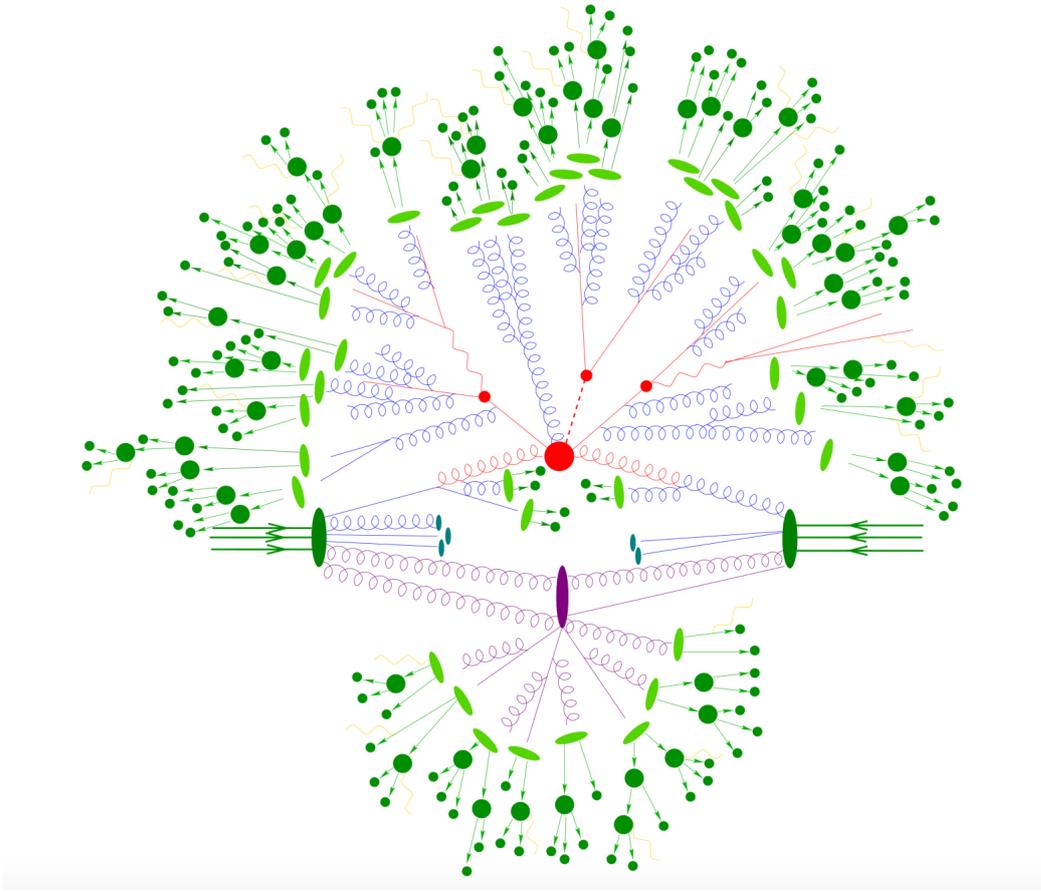
Figure 3.3: Sketch of a $t\bar{t}H$ event [55]. The partonic hard scattering is shown as a red blob, whereas the purple blob represents the additional multi-parton interactions. The initial and final state radiation are depicted as curly and straight lines. The hadrons generated during the hadronization step are shown in light green, whereas the final stable hadrons are shown in dark green.

angle of the emissions as the ordering variable, rather than the virtuality, and that the argument of α_S at the splitting vertex should be the relative parton transverse momentum after the splitting. Without this effect, the particle multiplicity would grow too rapidly with energy. For this reason, angular ordering is used as the evolution variable in both the Herwig and Herwig++ [57] programs. A radical alternative formulation of QCD cascades focuses upon soft emission, rather than collinear emission, as the basic splitting mechanism. It then becomes natural to consider a branching process where it is a parton pair (i.e. a dipole) rather than a single parton, that emits a soft parton, Sherpa and Pythia 6 and 8 [56] use this approach.

Another important correction is related to the quarks masses, since they act as a cut-off on collinear singularities. If the mass of a quark is below, or of the order of the mass of the top quark, its effect in the shower is small. For larger quark masses, like in c, b, or t, it is the mass, rather than the typical hadronic scale, that cuts off collinear radiation. This feature can be implemented with different levels of sophistication: in Pythia (both 6 and 8) [56] mass effects are included using a kind of ME correction method, while in Herwig++ [57] and Sherpa [58] a generalization of the Altarelli-Parisi [59] splitting kernel is used for massive quarks.

The PS algorithms are based on a combination of the collinear (low angle) and soft (low energy)

approximations and are thus inaccurate for hard, large-angle emissions. For this case, it is necessary to use the full ME amplitudes. The use of ME generators together with PS simulations can lead to a double counting of the final state configurations that can be generated by both tools. In order to avoid this double counting, a scale has been introduced below which the additional radiation is modeled by the parton shower and above which it is simulated using the full ME calculations. This scale, called *matching scale*, is arbitrary because of the final results should not be affected by its choice.

In order to improve the PS description of hard collisions, two main methods are adopted: the so-called Matrix Elements and Parton Shower matching (ME+PS) and the matching of the NLO (Next-to-Leading-Order) calculations and Parton Showers (NLO+PS). The former allows to use tree-level matrix elements for hard, large-angle emissions. This method starts by generating exact matrix elements for the production of the basic process plus a number $\geq n$ of other partons. A minimum separation is imposed on the produced partons, requiring, for example, that the relative transverse momentum in any pair of partons is above a given cut Q_{cut} (the *matching scale*). The Q_{cut} must be chosen to be large enough for fixed-order perturbation theory, but small enough so that the shower is accurate for emissions below it. The NLO+PS method, instead, extend the accuracy of the generation of the basic process at the NLO level in QCD. They must thus include the radiation of an extra parton with tree-level accuracy, since this radiation constitutes a NLO correction to the basic process. They must also include NLO virtual corrections.

Several ME+PS implementations use existing LO generators, like MadGraph [60] for the calculation of the matrix elements and feed the partonic events to a GPMC like Pythia or Herwig. Sherpa and Herwig++ also include their own matrix-element generators.

Several NLO+PS processes are implemented in the MC@NLO program [61], together with the new aMC@NLO development [62] and in the Powheg framework [63], [64]. Herwig++ also includes its own Powheg implementation, suitably adapted with the inclusion of vetoed and truncated showers, for several processes. Sherpa instead implements a variant of the MC@NLO method. All the characteristics of the generators used in this analysis are summarized in table 3.3.

Sample	Generator	Matching method
$t\bar{t}H$	aMC@NLO	NLO+PS
$t\bar{t}$+jets	Powheg	NLO+PS
W/Z + jets	Sherpa	ME+PS (LO)
Single top (s-, Wt-channels)	Powheg	NLO+PS
Single top (t-channel)	Powheg	NLO+PS
$t\bar{t} + V$	Madgraph	ME+PS (LO)
Diboson	Sherpa	ME+PS (LO)

Table 3.3: A summary of different matching methods between ME and PS simulators used in each signal and background process.

Two examples of the comparison between different ME generators and PS setups for the two main background processes ($t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$) are shown in figures 3.4a and 3.4b [65] [66], respectively. The default generator (plus PS) for the background simulation, used in the analysis described in this thesis, is the Powheg+Pythia6, as mentioned in section 3.2 and in table 3.2. For the $t\bar{t}+b\bar{b}$ case, the effect

of parton showering and hadronization is estimated by comparing the Powheg+Pythia6 sample with Powheg+Herwig++. In addition, the samples are compared to Powheg+Pythia8. Disagreements of the order of 10-20% are present in the $p_{\mathrm{T}}{}^{b\bar{b}}$ distribution in figure 3.4a. The $t\bar{t}+c\bar{c}$ process is studied considering, in addition to the samples used in the first case, different normalization and factorization scale in the matching method. The $H_T/4$ scale, in figure 3.4b is defined as

$$H_T = \frac{1}{4} \sum_{i \in FS} E_{T,i}, \tag{3.3.1}$$

where $E_{T,i} = \sqrt{m_i^2 + p_{\mathrm{T}i}^2}$ is the trasverse energy of any final state (FS) parton.

(a)   (b)
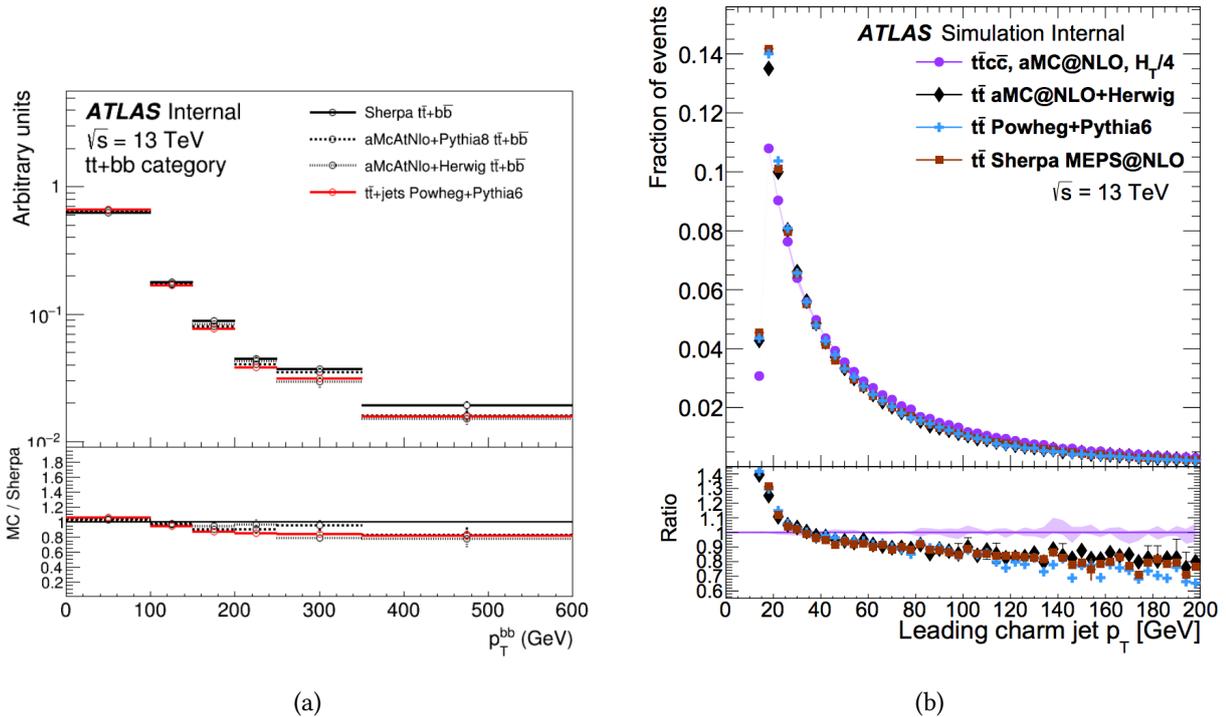
Figure 3.4: (a) Comparison of different generator setups produced at $\sqrt{s}=$ 13 TeV for the $b\bar{b}$ pair p_{T} in $t\bar{t}$ *lepton+jets* events. [65] (b) Normalized distributions of the p_{T} of the leading c-jet, for different predictions. The purple band around the $t\bar{t}+c\bar{c}$ $H_T/4$ sample and the bars on the MG5_aMC $t\bar{t}$ sample show the effect of varying the renormalization and factorization scales up or down by a factor of 2. Other errors are statistical. [66]

    The scale variations and their consequences in the distributions and background modelling, using different ME and PS matching, will be taken into account in the study of the systematics uncertainties in this analysis (see Chapter 6).

    After the PS, the hadronization process, meaning the aggregation of partons into hadrons (light green ovals in figure 3.3), takes place. Perturbation theory becomes invalid and the dynamics enter in a non-perturbative phase. Two phenomenological iterative models are predominantly used to describe the *hadronization*: the Lund string model [67] and the cluster fragmentation model [68], [69]. The former considers gluons between $q\bar{q}$ pairs as color field lines, which are attracted to each other owing to gluon self-interaction. As the q and $\bar{q}$ move apart, the color strings are stretched until it is more energetically favorable for them to fragment and create an additional $q\bar{q}$ pair. The procedure is iterated until all the strings are too light for further fragmentation. As for the cluster fragmentation model, gluons are split

into $q\bar{q}$ pairs which are clustered with the other quarks and di-quarks to form color-singlet hadrons. Clusters are further fragmented until stable hadrons are formed. When a cluster is too light to fragment into a hadron, a light hadron replaces the cluster. Multiple parton interactions, characterized by small momentum transfers, are also simulated at this stage, being dependent on the description of hadronization.

Of the so-called primary hadrons, originating directly from string breaks and/or cluster decays, many are unstable and so decay further, until a set of particles is obtained that can be considered stable ($c\tau \geq$ 10 mm) on time scales relevant to the given measurement. The decay modeling can therefore have a significant impact on final particle yields and spectra, especially for the lowest-lying hadronic states, which receive the largest relative contributions from decays.

In an MC decay package, on the other hand, all information must be quantified and consistent, with all branching ratios summing to unity. When adapting particle summary information for use in a decay package, a number of choices must therefore be made. The amount of ambiguity increases as more excited hadron multiplets are added to the simulation, about which less and less is known from experiment, with each GPMC making its own choices. A related choice is how to distribute the decay products differentially in phase space, in particular which matrix elements to use. Historically, MC generators contained matrix elements only for selected (generator-specific) classes of hadron and τ decays, coupled with a Breit-Wigner smearing of the masses, truncated at the edges of the physical decay phase space. More recently, Herwig++ and Sherpa include helicity-dependence in τ decays, with a more limited treatment available in Pythia8. The Herwig++ and Sherpa generators have also included significantly improved internal simulations of hadronic decays, which include spin correlations between those decays for which matrix elements are used. Moreover, Herwig++ and Pythia include the probability for B mesons to oscillate into $\bar{B}$ ones before decay. Sherpa also includes CP-violating effects and, for common decay modes of the neutral meson and its antiparticle, the interference between the direct decay and oscillation followed by decay.

3.3.2   The detector simulation

After the parton production, parton shower, hadronization and the decay, the interactions of the final state particles with the material of the experiment have to be simulated to allow a direct comparison between simulated and real events. The ATLAS collaboration uses two types of detector simulation. The first is the GEometry ANd Tracking (GEANT4) software [70] that provides a model for the particle interaction through matter. For this reason a detailed description of the ATLAS detector geometry and of the trigger system is necessary. This is the so called Full Simulation (FS). Due to computer elaboration time necessities, some Monte Carlo samples have not been processed using the full detector simulation but with an approximate and faster simulator, namely ATLAS FAST II (AF-II) [71]. It has been developed to simulate and reconstruct events including detector effects, but saving computing resources. The AF-II uses FastCaloSim in the calorimeter [72]; the energy of single particle showers is deposited directly using parameterizations of their longitudinal and lateral energy profile. Because the standard reconstruction is used, it is possible to work with a combination of events obtained from GEANT4 and AF-II without modifying the analysis code. The approach taken by FastCaloSim is intrinsically less accurate, but the parametrisations can be tuned on data. It has been used since 2011 for the production of large MC samples needed for new physics searches as well as precision measurements. AF-II has been validated against the GEANT4 based full simulation for electrons, jets and missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$).

The nominal generator for the $t\bar{t}$ process in this analysis is Powheg+Pythia6 (FS), shown in figure 3.5a, while different alternative generators have been studied for two reasons: to check if the AF-II simulation is suitable to be use in the analysis as the nominal one (that would decrease the time needed to run the analysis) and to study the systematic variations due to the choice of the nominal generator and to the mis-

modeling of the $t\bar{t}$ (see Chapter 6). Figure 3.5b shows the data and MC comparison using all the nominal generators and $t\bar{t}$ Powheg+Pythia6 generators and the AF-II simulation, which is shown to be overestimating the data, instead of the agreement shown in figure 3.5a. In order to understand the behaviour
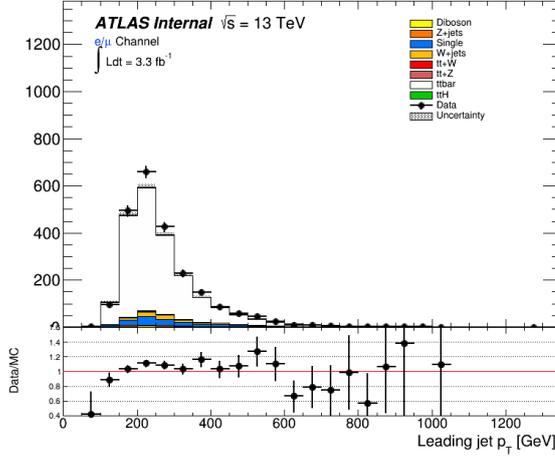
(a)
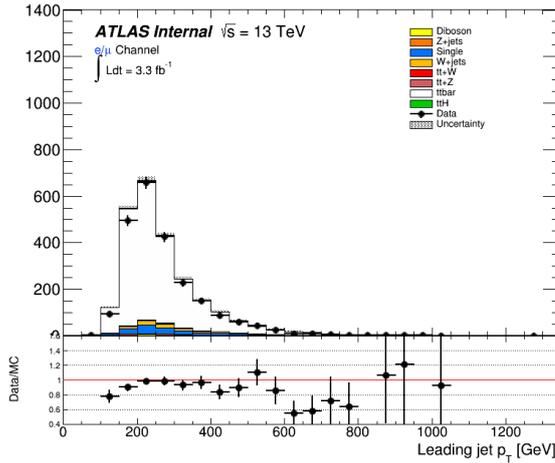
(b)

Figure 3.5: Distribution of the leading jet p_{T}, in the events passing the baseline selection of the $t\bar{t}H$ analysis, using the nominal FS $t\bar{t}$ generator Powheg+Pythia6 (a) and using the alternative AFII $t\bar{t}$ generator Powheg+Pythia6 (b).

of different AFII generators, a comparison of three MC samples used in this analysis, Powheg+Pythia6, aMC@NLO+Herwig++ and Powheg+Herwig++, is shown in figures 3.6a and 3.6b, for the $t\bar{t}$ process as a function of two different observables, that are very interesting for the final results of the analysis.

Nevertheless all the efforts and expertises on the simulation, if a region of the phase space, that is not well known, has to be simulated, differences between data and MC simulations could happen. These differences are not reducible to detector effects, but they are due to the approximated description of the process to be studied. If it is the case of common discrepancies between different experiments, the simulation can be corrected empirically through some suitable weights that take into account this irreducible variation and reduce (or cancel) the differences on a series of test processes (like the $t\bar{t}$ for example). This process is called *modelling*. In a second moment, with a dedicated extrapolation, the obtained weights
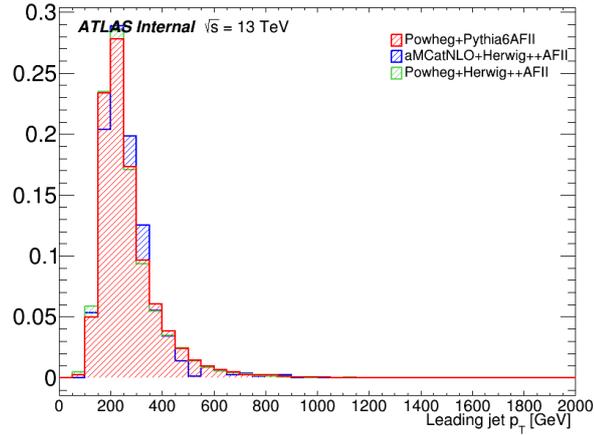
(a)
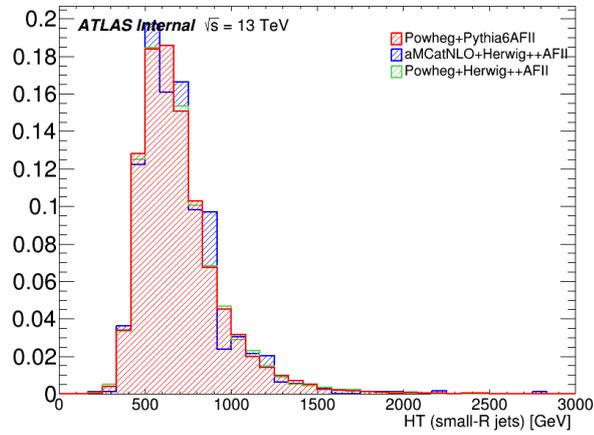
(b)

Figure 3.6: Distribution of the leading jet p_T(a) and the H_T (the scalar sum of the p_T of all the small-R jets in the event) (b), using the three different AFII $t\bar{t}$ generators: Powheg+Pythia6, aMC@NLO+Herwig++ and Powheg+Herwig++.

can be applied to similar processes, in this case, as the $t\bar{t}H$.

3.4 $t\bar{t}H$ Monte Carlo signal

The simulation of the $t\bar{t}H$ signal production for the full data taking (2015 and 2016) has been done using: aMC@NLO [61] and Pythia8 [56], with the same Higgs boson mass, set to m_H = 125 GeV. aMC@NLO is a generator framework that allows the implementation of any production process, computing the cross section at NLO accuracy. Pythia8 is a coherent set of physics models for the evolution from few-body hard process to a complex multihadronic final state and, for this reason, it is used to simulate the parton shower for the $t\bar{t}H$ process.

For this process, both the generators use the parton distribution function (PDF) parametrization

NNPDF [73] (based on LO, NLO and NNLO QCD theory, including also electroweak corrections) that includes a global dataset of results from HERA-II deep-inelastic inclusive cross-sections, the combined HERA charm data, jet production from ATLAS and CMS, vector boson rapidity and transverse momentum distributions from ATLAS, CMS and LHCb, $W + c$ data from CMS and top quark pair production total cross sections from ATLAS and CMS.

The Monte Carlo simulation includes three samples of events, depending on the $t\bar{t}$ systems decay: the *dilepton* events (where $t\bar{t}$ decays leptonically), the *l+jets* events (where $t\bar{t}$ decays semi-leptonically) and the *all-hadronic* (where there are no leptons in the final state). The signal channel, as explained in Chapter 1, is considered to include only the Higgs boson decay $H \to b\bar{b}$ and the semi-leptonic $t\bar{t}$ decay.

3.5 Background sources and their evaluation

The background sources depend on the event selection criteria and cut threshold, that will be discussed extensively in Chapter 4. Nonetheless it is possible to distinguish in general two categories of backgrounds whose topologies mimic the $t\bar{t}H$ signature:

- exclusive channels whose cross sections are known, and are due to hard scattering perturbative processes;

- other channels due to soft QCD processes, with unknown cross sections or with wrong particle identifications or wrong reconstructions.

Background channels in the first category are evaluated through extensive Monte Carlo simulations, provided that their yields are known. Channels in the second category cannot be estimated with MC simulations. This might happen for a variety of reasons: the overall yield is not known, the production process cannot be evaluated perturbatively (soft QCD) or it will require CPU power beyond current availability, the efficiencies on wrong identification or reconstruction have too large uncertainties. In this case it is customary to use *data driven* techniques, where the background contribution is evaluated directly from acquired data, taking into account only appropriate regions (control regions in the following), in which the background processes are dominant.

The background contribution to the signal $t\bar{t}H$ in a semi-leptonic final state is given by:

- $t\bar{t}$**+jets** is the dominant background of the $t\bar{t}H$ signal (see figure 3.2), where the $t\bar{t}$ pair decays leptonically and semi-leptonically. This background channel is generated using the Powheg [63][64] at NLO generator, with CTEQ6L1 PDFs set, interfaced with Pythia v6.42 as the parton shower generator. The decays of heavy-flavour hadrons are modelled using the EvtGen [75] package, that provides a framework for the implementation of physics processes relevant to decays of B mesons and other resonances. The h_{damp} parameter in Powheg, which controls the p_{T} threshold above which the first additional gluon is emitted beyond the Born approximation (tree-level, that is the very first interaction, particles from the red blob in figure 3.3) and thus regulates the p_{T} of the recoil emission against the $t\bar{t}$ system, is set to 1.5 the mass of the top quark (m_{top} = 172.5 GeV);

- ***single top*** **production** is the second largest contribution to the background, after the $t\bar{t}$+jets one (see figure 3.2). It considers all the three production channels (s-, t-, Wt-channel); its contribution (in the case of the t-channel, which has the higher production cross-section) is about a factor two smaller than the $t\bar{t}$ cross-section (see Chapter 1). These events are generated by Powheg [63][64] generator interfaced to Pythia v6.42, as for the $t\bar{t}$+jets events;

- *V +**jets**: production of Z or W boson with additional jets ($W + jets$ and $Z + jets$). $Z + jets$ events can contribute in both the electron and muon Z decays, where one lepton is not detected, and in the tau decay case, where one τ lepton decays leptonically and the other hadronically. The main contribution from the $W + jets$ events comes from the leptonic W boson decays, that give the needed lepton in the final state. These contributions have been evaluated using Sherpa v2.1.1 [58] as generator with CT10 [74] as PDF sets. Sherpa contains a flexible tree-level ME generator for the calculation of hard scattering processes within the SM and beyond. The emission of additional QCD partons off the initial and final states is described through a parton-shower model;

- ***diboson***: production of boson pairs (WW, ZZ or WZ). These events are generated and showered using Sherpa [58] v2.1.1 with CT10 [74] PDF sets, as the V+jets events;

- *$t\bar{t}$+V*: production of $t\bar{t}$ pairs, decaying leptonically and semi-leptonically, with an additional boson ($t\bar{t}$+W and $t\bar{t}$+Z). This type of events can be mismatched for $t\bar{t}H$ processes both when the Z/W boson decays hadronically ($Z \rightarrow q\bar{q}$, $W^{\pm} \rightarrow qq'$) and the $t\bar{t}$ system decays semi-leptonically and when the Z/W boson decays leptonically ($Z \rightarrow l^+l^-$, $W^{\pm} \rightarrow l^{\pm}\nu_{l\pm}$) and the $t\bar{t}$ system decays hadronically. The events have been modelled by samples generated using MadGraph [60] interfaced to Pythia v8.1 for the parton shower, with NNPDF set of PDF. MadGraph is a tool for automatically generating ME for High Energy Physics processes at NLO accuracy and it has been developed in order to merge to showering/hadronization codes for complete event simulation in the optimal way both at LO and NLO;

- ***multi-jets***: events from the QCD processes, evaluated entirely via a data driven technique called *Matrix Method*, see section 3.5.3.

3.5.1   Heavy Flavour classification for $t\bar{t}$+jets

Since the $t\bar{t}$+jets contribution is the largest (see figure 3.2) and affects significantly the analysis due to the extreme similarity with respect to the signal final state, its accurate study is crucial.

To categorize the $t\bar{t}$+jets events depending on the flavour parton originated from the jets not belonging to the $t\bar{t}$ system, it has been defined the "particle jet" as the jet reconstructed from the all stable truth particles (not counting muons and neutrinos) with the anti-k_t algorithm (described in detail in Chapter 4) with a radius parameter R = 0.4 requiring $p_{\mathrm{T}} > 15$ GeV and $|\eta| < 2.5$. Events where at least one such particle jet is matched within $\Delta R < 0.4$ to a truth b-hadron with $p_{\mathrm{T}} > 5$ GeV not originating from a top quark decay are labelled as $t\bar{t} + b\bar{b}$ events. Similarly, events which are not already categorised as $t\bar{t} + b\bar{b}$, and where at least one particle jet is matched to a charm quark not originating from a W boson decay, are labelled as $t\bar{t} + c\bar{c}$ events. Events labelled as either $t\bar{t} + b\bar{b}$ or $t\bar{t} + c\bar{c}$ are generically referred to as $t\bar{t}$+HF events (Heavy Flavour). The remaining events are labelled as $t\bar{t}$+light-jet events, including those with no additional jets. A finer categorisation of different topologies in $t\bar{t}$+HF is, for example, two particle jets matched to an extra b-hadron or c-hadron each (referred to as $t\bar{t} + b\bar{b}$ or $t\bar{t} + c\bar{c}$), a single particle jet matched to a single b-hadron or c-hadron (referred to as $t\bar{t} + b$ or $t\bar{t} + c$), or a single particle jet matched to a b-hadron or c-hadron pair (referred to as $t\bar{t} + B$ or $t\bar{t} + C$); this categorization is made for the purpose of comparisons to other $t\bar{t}$+jets event generators and the propagation of systematic uncertainties related to the modelling of $t\bar{t}$+HF. This $t\bar{t}$+HF categorisation is identical to that used in the 8 TeV resolved analysis [76].

The modelling of the $t\bar{t}+b\bar{b}$ background is improved by reweighting the Powheg+Pythia6 prediction to an NLO prediction of $t\bar{t}+b\bar{b}$ including parton showering [79], based on Sherpa+OL [80][81] using the CT10

PDF set [77]. This reweighting is performed for different topologies of $t\bar{t}+b\bar{b}$ in such a way that the inter-normalisation of each of the categories ($t\bar{t}+b\bar{b}$, $t\bar{t}+b$, $t\bar{t}+B$, etc) and the relevant kinematic distributions are at NLO accuracy.

Figure 3.7 shows the contributions of the different $t\bar{t}+b\bar{b}$ event categories to the total $t\bar{t}+b\bar{b}$ cross section at generator level for the Powheg+Pythia 6 and SherpaOL samples. Unfortunately some of the $t\bar{t}+b\bar{b}$ contributions are not included in the NLO prediction. In particular, two topologies are identified: $b\bar{b}$ pairs arising from Multiple Parton Interaction (MPI) overlaying a $t\bar{t}$+jets event, gluon to $b\bar{b}$ splitting where the gluon is radiated from the top decay products. This second contribution is labelled as Final State Radiation (FSR). The MPI and FSR contributions are excluded from the comparisons since they are not contained in the NLO prediction.
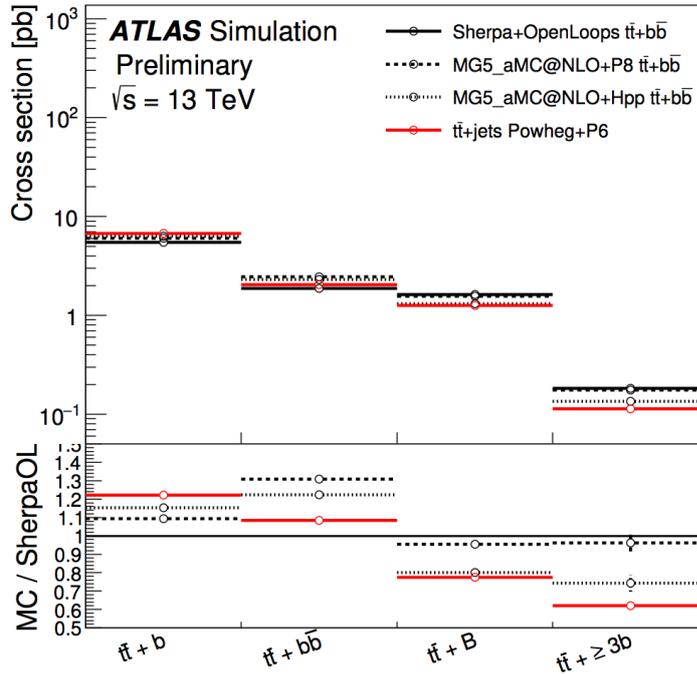
Figure 3.7: The predicted cross-sections for the $t\bar{t}+\geqslant 1b$ sub-categories. The inclusive Powheg+Pythia6 prediction is compared to four-flavour $t\bar{t}+b\bar{b}$ calculations from SherpaOL and MG5_aMC with different parton showers. [78]

The relative distribution across categories is such that SherpaOL predicts a higher contribution in the $t\bar{t}+B$ category, as well as in every category where the production of a second $b\bar{b}$ pair is required. The modelling of the relevant kinematic variables in each category is in reasonable agreement between Powheg+Pythia and SherpaOL. Some differences are observed in the very low regions of the mass and p_{T} of the $b\bar{b}$ pair, and in the p_{T} of the top quark and $t\bar{t}$ systems.

The prediction from SherpaOL is expected to model the $t\bar{t}+b\bar{b}$ contribution more accurately than Powheg+Pythia. Thus, $t\bar{t}+b\bar{b}$ events are reweighted from Powheg+Pythia to reproduce the NLO $t\bar{t}+b\bar{b}$ prediction from SherpaOL for the relative contributions of different categories, as well as their kinematics.

3.5.2 Alternative Monte Carlo simulations

To evaluate the systematic uncertainties in the simulations, both for the main channel and for the background channels specific MC simulations have been performed with different generators. Hadronization

and parton showering uncertainties for the background have been estimated using a sample generated with Powheg and showered by Herwig++ v2.7.1. The treatment of the uncertainties will be faced in Chapter 6.

3.5.3   QCD multi-jets background estimation

The selection of events with top quarks is often based on the identification of one or more charged isolated leptons (well identified electrons or muons) from the W decay, referred to as "real" leptons in the following. Quality and isolation requirements are applied to select these leptons: the isolation requires selections on the longitudinal and transverse impact parameters, $|z_0 \sin \theta| < 0.5$ mm (where z_0 is the z coordinate of the point along the extrapolated lepton track closest to the primary vertex) and $d_0^{sig} = |\frac{d_0}{\sigma_{d_0}}| <$ 5 (where d_0 and σ_{d_0} are respectively the distance of the closest approach and its uncertainty of the lepton track to the primary vertex), following in this paragraph. Well identified leptons not coming from primary vertex (called "non-prompt leptons") and hadrons incorrectly identified as leptons (called "non leptonic particles") may satisfy these selection criteria, giving rise to so called "fake" lepton background.

In the case of electrons, these include contributions from semileptonic decays of b- and c-quarks, photon conversions and jets with large electromagnetic energy (from the π^0 contributions in the calorimeter). Muon labelled as "fake" leptons, as defined before, can originate from semileptonic decays of b- and c-quarks, from charged hadron decays in the tracking volume or in hadronic showers, or from punch-through particles emerging from high-energy hadronic showers.

The QCD multi-jet events, characterised by a cross-section several orders of magnitude larger than for W boson or top events, can present fake leptons arised from jet misidentification and consequently enter in the background contribution of this analysis. The QCD background is highly detector dependent, hence the better way for its estimation is via data driven methods; the one used for the 2015-2016 data taking is the *Matrix Method* (MM), already used by the CDF and D0 experiments at Tevatron [82], and it is based on the efficiency determination of real and fake events selected with different lepton requirements. The first step of the Matrix Method consists in the selection of two different event samples, called "tight" and "loose", that differ only in the lepton cut definition. The tight selection is exactly the one applied in standard analysis while the loose is a selection with a looser requirement in the leptonic cut; in the specific case of this analysis, the lepton isolation cut has been removed in the loose selection.

The number of events surviving to the tight and loose (N^{tight} and N^{loose}) selection can be expressed as the number of events containing a real lepton ($N_{real}^{tight(loose)}$) plus the number of events containing a fake lepton ($N_{fake}^{tight(loose)}$):

$$N^{tight} = N_{real}^{tight} + N_{fake}^{tight} \qquad\qquad N^{loose} = N_{real}^{loose} + N_{fake}^{loose}. \tag{3.5.2}$$

The equation system (3.5.2) can be more conveniently rewritten as

$$N^{tight} = \epsilon_{real} N_{real}^{loose} + \epsilon_{fake} N_{fake}^{loose}, \tag{3.5.3}$$

defining the real and fake efficiencies as

$$\epsilon_{real} = \frac{N_{real}^{tight}}{N_{real}^{loose}} \qquad\qquad \epsilon_{fake} = \frac{N_{fake}^{tight}}{N_{fake}^{loose}}. \tag{3.5.4}$$

Consequently, the number of events containing fake lepton in the analysis and corresponding to the background contribution from the QCD is

$$N_{fake}^{tight} = \frac{\epsilon_{fake}}{\epsilon_{real} - \epsilon_{fake}}(\epsilon_{real}N^{loose} - N^{tight}). \tag{3.5.5}$$

The real and fake efficiencies ϵ_{real} and ϵ_{fake} are measured in control regions dominated by real and fake leptons respectively and do not be necessarily belong to the analysis sample. The efficiencies are determined as the ratio between the number of tight and loose events and they are expected to be dependent on the kinematic variables that characterise the events in these regions. For this reason, the efficiencies are parametrised as a function of different set of variables, as explained in the following.

The ϵ_{real} is estimated via a tag-and-probe technique from an enriched sample of $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ events that, with high approximation, are free from fake lepton contaminations. For each event, this method selects an unbiased loose lepton (*probe*) from the Z decay and checks for a tight selection on the other object produced from the same particle's decay (*tag*). The efficiency is determined by selecting the tight component on the probe lepton.

The ϵ_{fake} is measured in data samples dominated by events containing fake leptons. Each event of these control regions, denoted by CR$_f$, contains one loose lepton, at least one jet and has the following requirements:

- only for e+jets events:

$$E_T^{miss} + m_T^W < 60 \,\text{GeV} \tag{3.5.6}$$

  where E_T^{miss} is the event missing transverse energy, defined as

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2} \tag{3.5.7}$$

  and m_T^W is the transverse mass of the W defined as

$$m_T^W = \sqrt{2p_T^{lepton}E_T^{\text{miss}}(1 - \cos \Delta\phi)} < 20 \,\text{GeV} \,, \tag{3.5.8}$$

  where $\Delta\phi$ is the angle between the E_T^{miss} and the lepton;

- only for μ+jets events:

$$d_0^{sig} = |\frac{d_0}{\sigma_{d_0}}| \tag{3.5.9}$$

  where d_0^{sig} is the muon impact parameter significance.

These requirements have been introduced in order to enhance a sample with a large background contribution arising from fake leptons. Table 3.4 summarises the definition of the different control regions for electron and muon channel, respectively.

Channel	n_{jet} cut	Other cuts
e+jets	≥ 1	m_T^W <20 GeV, $E_T^{miss} + m_T^W$ <60 GeV
μ+jets	≥ 1	$|d_0^{sig}|$ >5

Table 3.4: Summary of the requirements for the different electron/muon control regions. [83]

In the background evaluation, an essential step is to note that, as clear from formula 3.5.5, its accuracy is related to the uncertainty of the difference between the two efficiencies (ϵ_{real}, ϵ_{fake}). Since these

might depend on selection or classification criteria on identified tracks, for a more precise background evaluation they are evaluated based on the trigger, on the isolation requirement, on the lepton p_T and on other kinematic or topological quantities (see table 3.5). For practical reasons, the value of ϵ_{real} and ϵ_{fake} are measured as a function of discrete variables (called x variables), like the number of jets on an event, and of continuous variables (called y variables), like lepton p_T or η. The discrete variables will have no more than three values, while the continuous one will be binned, possibly with variable width bins in order to optimize the uncertainties on the resulting efficiencies and on the final background estimation. For each efficiency type (electron or muon, real or fake) only a sub-set of x and y variables are used, as summarized in table 3.5.

	x variables			y variables				
	Trigger	n_{jet}	n_{b-jet}	$\|\eta^l\|$	$p_T{}^l$	$p_T{}^{lead-jet}$	$\Delta R(l, jet)$	$\Delta\phi(l, E_T^{miss})$
$\epsilon_{real}(e)$	✓	✓		✓	✓		✓	
$\epsilon_{real}(\mu)$	✓	✓		✓	✓		✓	
$\epsilon_{fake}(e)$	✓		✓	✓		✓		✓
$\epsilon_{fake}(\mu)$	✓		✓	✓	✓		✓	

Table 3.5: Summary of variables used to parametrise the real and fake lepton efficiencies in the Matrix Method (all these variables will be completely defined and described in Chapter 4). The column "Trigger" refers to the specific lepton trigger, n_{jets} (n_{b-jets}) is the jets (b-jets) multiplicity, $p_T{}^{lead-jet}$ ($p_T{}^l$) stays for the p_T of the leading jet (lepton), $\Delta R(l, jet)$ is the angular distance between the lepton and the jet, $\Delta\phi(l, E_T^{miss})$ is the azimuthal difference between the lepton and the missing energy in the event.

These efficiencies are used to compute the weights

$$w_i = \frac{\epsilon_{fake}}{\epsilon_{real} - \epsilon_{fake}}(\epsilon_{real} - \delta_i), \tag{3.5.10}$$

where δ_i equals unity if the loose event i passes the tight event selection and 0 otherwise.

To correctly account these variables correlations and dependencies, the weights are computed as a function of the different combinations of the variables, listed in table 3.5, through:

$$\epsilon_k(x_1, ..., x_N; y_1, ..., y_N) = \frac{1}{\epsilon_k(x_1, ..., x_N)^{M-1}} \cdot \Pi_{j=1}^M \epsilon_k(x_1, ..., x_N; y_i). \tag{3.5.11}$$

Here the expression $\epsilon_k(x_1, ..., x_N)$ represents the efficiency measured as a function of all the x variables. The expression $\epsilon_k(x_1, ..., x_N; y_i)$ represents instead the efficiency measured as a function of all the x variables and of the variable y_j. The M in the denominator represents the matrix of the method, built from the combination of the real and fake leptons:

$$\begin{pmatrix} N_t \\ N_l \end{pmatrix} = M \begin{pmatrix} N_r \\ N_f \end{pmatrix}. \tag{3.5.12}$$

The equation 3.5.11 implies that the full correlation between the variables x (discrete variables) and each of the variables y (continuous variables) is taken into account, while the correlation between the y variables is neglected. In particular, for each of the efficiencies, the assumption of no correlation between the

variables y is checked by comparing the observed dependency on the variable y_j, i.e. $\epsilon_k(x_1, ..., x_N; y_i)$, and the efficiency $\epsilon_k(x_1, ..., x_N; y_1, ..., y_N)$ averaged over all the other $y_{j'}$ variables, with $j' \neq j$.

The background estimation in a given bin of the final observable is given by the sum of w_i over all the events in that bin.

Since the method described above is complex, an analysis needs some prescriptions from the experts in order to evaluate the specific efficiencies and weights. For the Run-2 data taking, only a preliminary evaluation has been provided for the resolved analysis, while the boosted channel needs more studies to determine the multi-jet contribution in its signal region. For these reasons, the multi-jet background (and consequently its systematic uncertainties) has not been included in the analyses described in this thesis.

3.6 Data and Monte Carlo comparison plots

In this section the agreement between the data and MC distributions is verified and shown for the boosted signal region (described in Chapter 4), including all the signal and background contributions. The distributions of the observables of interest for this analysis have been chosen: the number of small-R jets, the number of small-R jets tagged as b-jets, the p_T, η and ϕ of the lepton, the small-R jets (as well as the large-R jets, explained in more details in Chapter 4) features, like the p_T, η, mass.

The statistical (of MC and data separately) and systematic uncertainties are also included in the distributions and they will be discussed in detail in Chapter 6. It has been chosen to not include MC modelling related systematics in these distributions, to stress the luminosity and detector systematic uncertainties only. While the MC modelling systematics will be studied in detail during the fit procedure, as explained in Chapter 6.

(a)

(b)

(c)
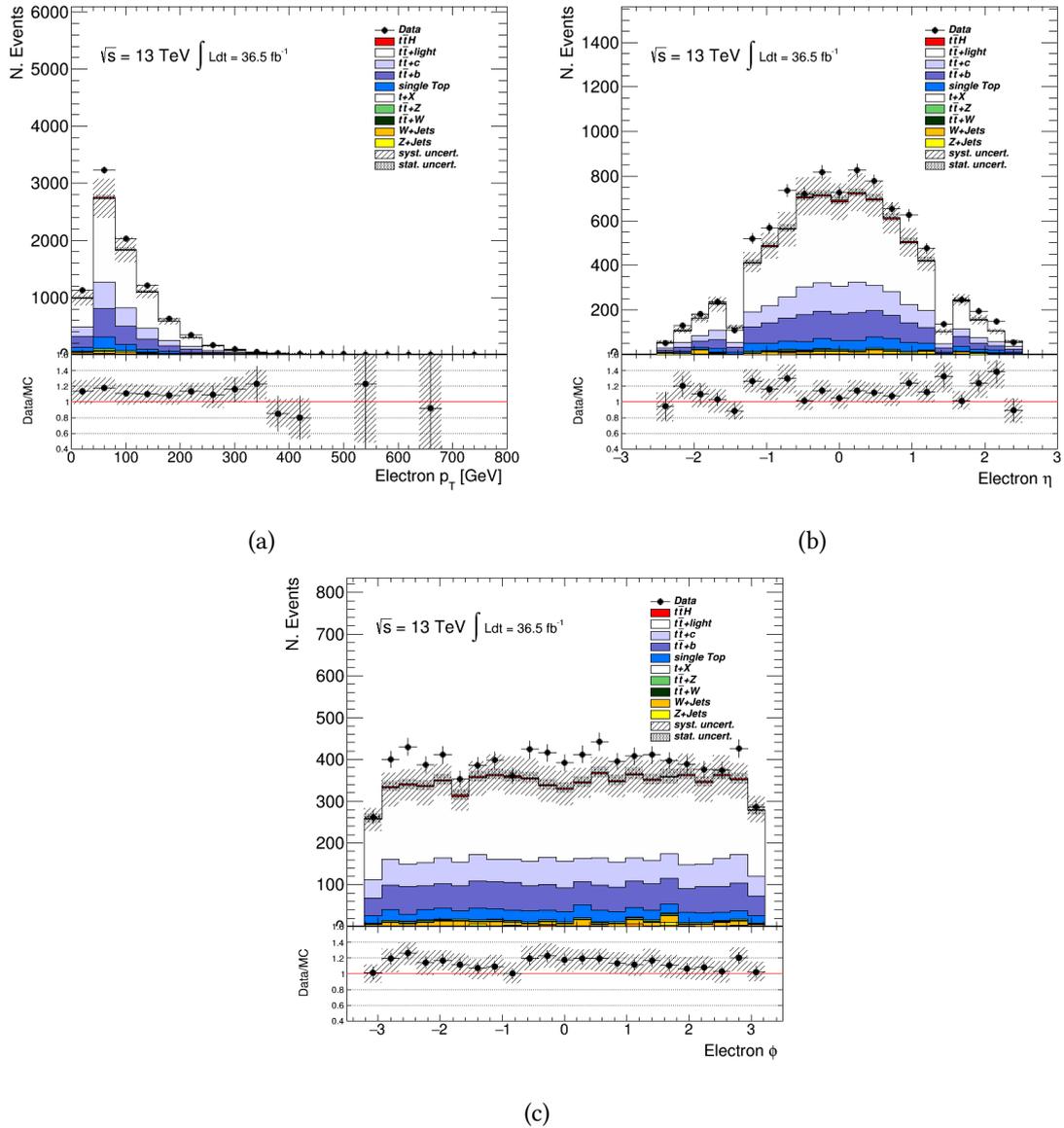
Figure 3.8: Distributions of the p_T (a), η (b) and ϕ (c) of the electron in the event, for MC and data.
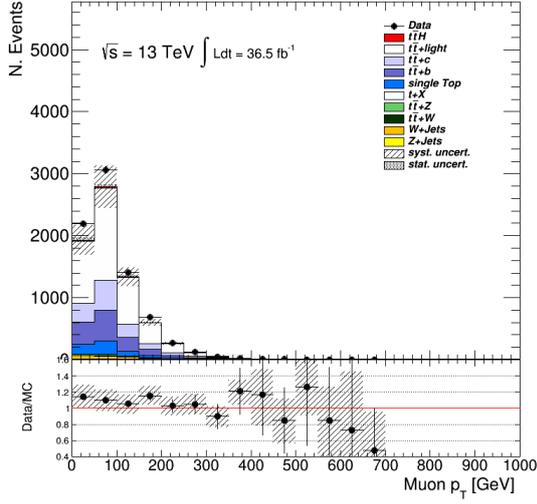
(a)
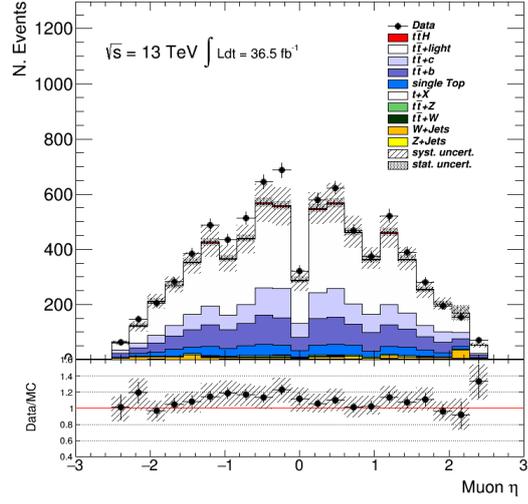
(b)
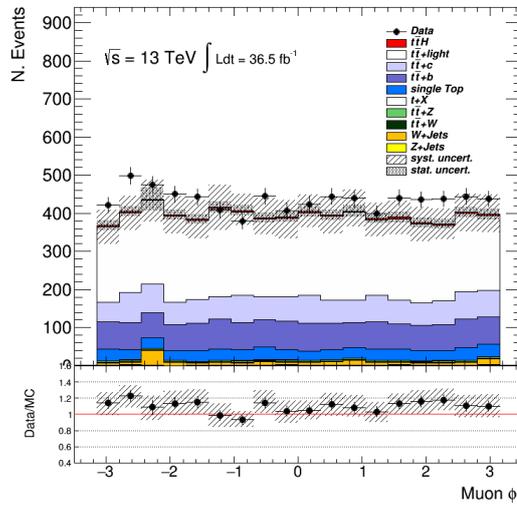
(c)

Figure 3.9: Distributions of the p_T (a), η (b) and ϕ (c) of the muon in the event, for MC and data.

(a)   (b)
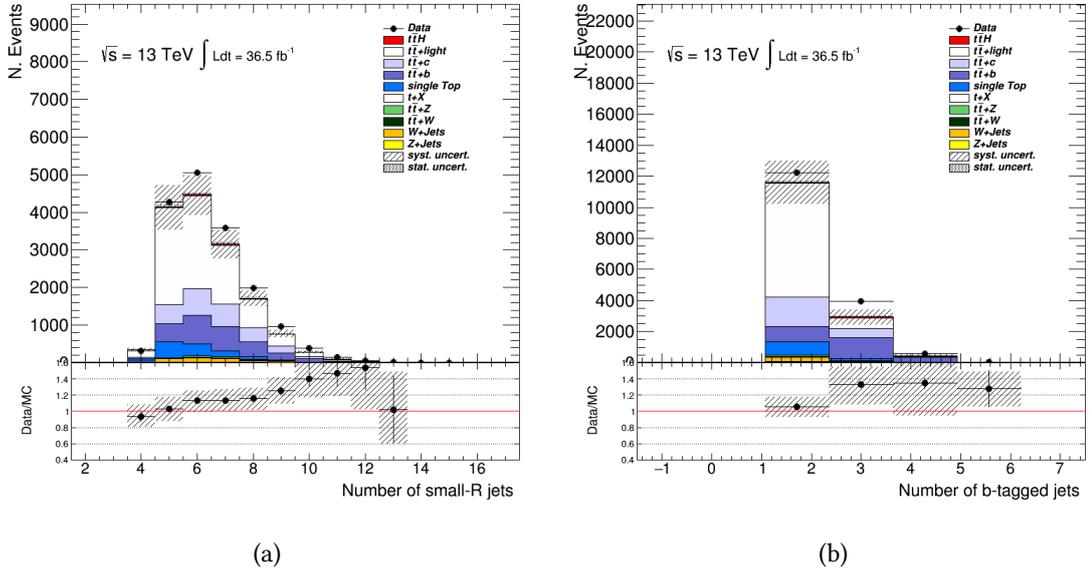
Figure 3.10: Distributions of the number of small-R jets (a) and b-tagged small-R jets (b) in the event, for MC and data.

(a)   (b)
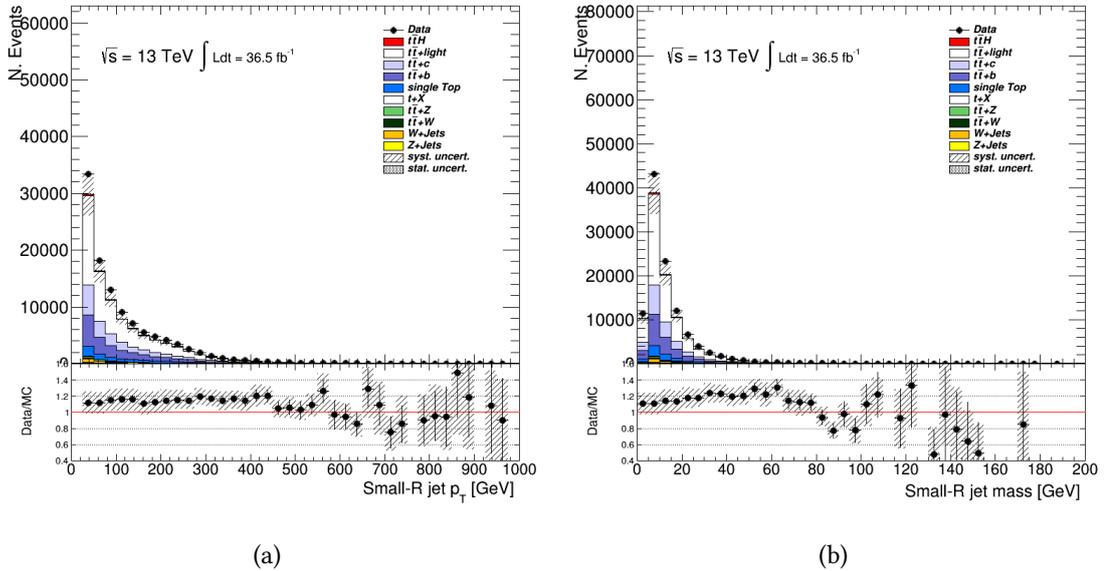
Figure 3.11: Distributions of the p_{T} (a) and mass (b) of small-R jets in the event, for MC and data.

(a)   (b)
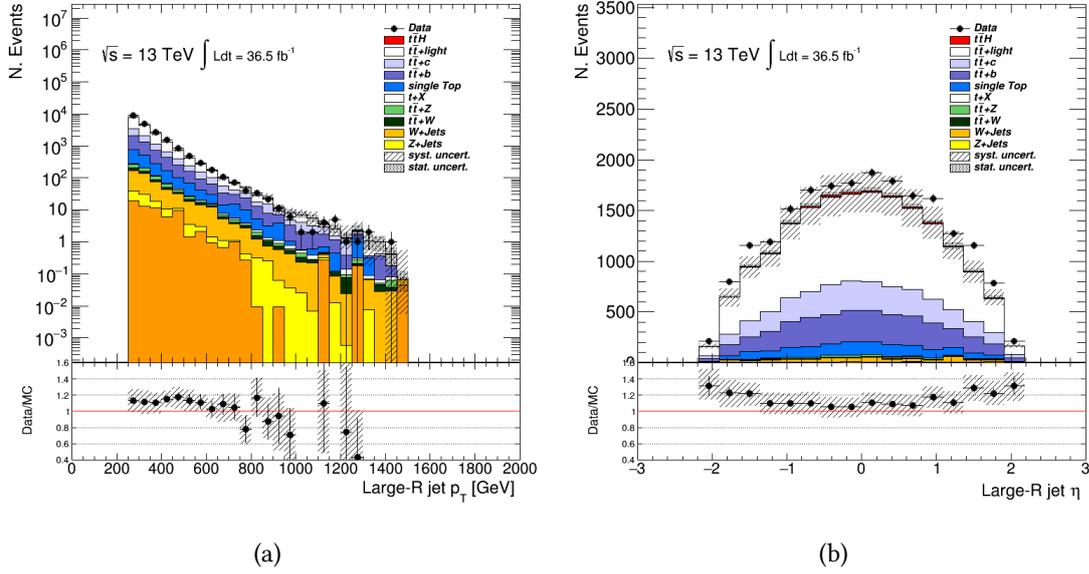
Figure 3.12: Distributions of the p_T (a) and η (b) of large-R jets in the event, for MC and data.

(a)   (b)
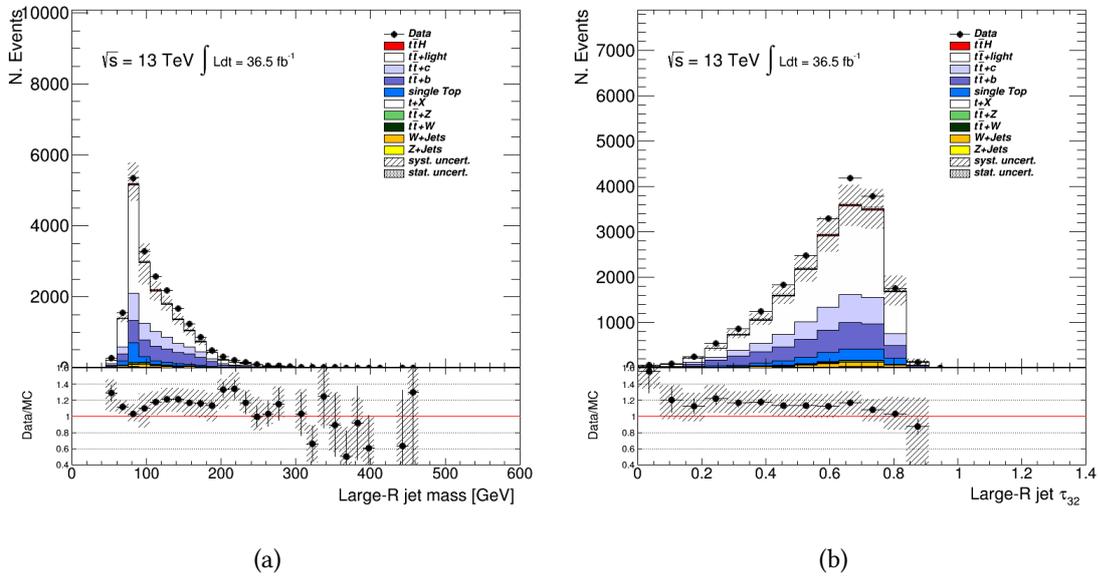
Figure 3.13: Distributions of the mass (a) and substructure variable τ_{32} (b) (discribed in detail in Chapter 4) of large-R jets in the event, for MC and data.

Chapter 4

Object reconstruction

The foundation of any ATLAS physics analysis is based on reconstructed and identified objects that represent the observed characteristics of the particles produced by the pp interactions and travelling through the detector volume. This chapter provides a description of the reconstructed objects used in the analysis presented in this thesis. The description comprehends electrons, muons and jets of different radius.

4.1   Electrons

The electrons reconstruction performed in ATLAS is based on the matching between Inner Detector (ID) tracks and ElectroMagnetic (EM) calorimeter clusters. The information from the EM calorimeter defines the energy of the electron while the ID tracks give the angular direction at the production point.

The ATLAS reconstruction algorithm is based on information coming from the electromagnetic calorimeter, the energy leakage in the hadronic one, the track quality criteria from ID objects and the cluster-track matching. The ATLAS recipe ensures a good discrimination from background objects by mainly requiring electron isolation (see further for isolation details). The Particle IDentification (PID) algorithms use quantities related to the electron cluster and track measurements including calorimeter shower shapes, information from the transition radiation tracker, track-cluster matching related quantities, track properties, and variables measuring bremsstrahlung effects for distinguishing signal from background. The baseline PID algorithm used for Run-2 data analyses is the likelihood-based (LH) method. It is a multivariate analysis (MVA) technique that simultaneously evaluates several properties of the electron candidates when making a selection decision. The LH method uses the signal and background probability density functions (PDFs) of the discriminating variables to calculate an overall probability for the object to be signal or background. The signal and background likelihoods ($\mathcal{L}_S$ and $\mathcal{L}_B$ respectively) for a given

electron are then combined into a discriminant $d_{\mathcal{L}}$ (strongly related to a better known statistical test in the physics field: the likelihood ratio test, used in this analysis and explained in details in Chapter 5) on which a requirement is applied:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \qquad \mathcal{L}_{S(B)}(\vec{x}) = \Pi_{i=1}^{n} P_{s(b),i}(x_i) \qquad (4.1.1)$$

where $\vec{x}$ is the vector of discriminating variable values and $P_{s,i}(x_i)$ is the value of the signal probability density function of the ith variable evaluated at x_i. In the same way, $P_{b,i}(x_i)$ refers to the background probability function. This allows a better background rejection for a given signal efficiency than a "cut-based" algorithm that would use selection criteria sequentially on each variable. In addition to the variables used as input to the LH discriminant, simple selection criteria are used for the variables counting the number of hits on the track.

Three levels of identification operating points are provided for electron PID, referred to, in order of increasing background rejection, as Loose, Medium, and Tight; electrons selected by Medium are all selected by Loose, and Tight electrons are all selected by Medium. The distributions of electron shower shapes depend on the amount of material the electrons pass through, and therefore vary with the pseudorapidity of the electron candidates. In addition, significant changes to the shower shapes and track properties are expected with increasing energy. The PID operating points were consequently optimised in several bins in $|\eta|$ and E_{T}. The efficiency of the LH identification algorithm is illustrated in figure 4.1.
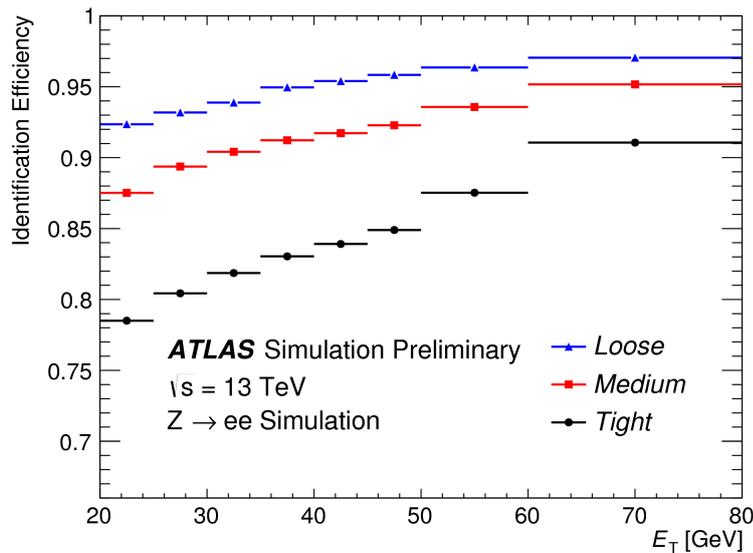
Figure 4.1: The efficiency to identify electrons from $Z \to ee$ decays estimated using simulated di-lepton samples. The efficiencies are obtained using Monte Carlo simulations, and are measured with respect to reconstructed transverse energy of the electrons. The candidates are matched to true electron candidates for $Z \to ee$ events. [84]

After the identification criteria, further selection cuts are applied to the electrons in order to select only objects that are necessary for the specific analyses. The electrons selection chosen in this analysis require a Loose identification working point and further cuts as in the following:

- no error occured in the Liquid Argon electromagnetic calorimeter during the data taking;

- the longitudinal impact parameter, $|z_0 \sin \theta| < 0.5$ mm and the significance of the transverse impact parameter $|\frac{d_0}{\sigma_{d_0}}| < 5$;

- candidates in the calorimeter transition region $1.37 < |\eta| < 1.52$ are excluded;

- $p_T > 25$ GeV;

- candidates must fulfil the isolation requirements on $E_T^{\text{cone0.2}}$ and $p_T^{\text{varcone0.2}}$ variables. The $E_T^{\text{cone0.2}}$ is the calorimetric isolation energy, defined as the sum of transverse energies of topological clusters, calibrated at the electromagnetic scale, within a cone of $\Delta R = 0.2$ around the candidate electron cluster. $p_T^{\text{varcone0.2}}$ is the track isolation, defined as the sum of transverse momenta of all tracks, satisfying quality requirements [84], within a cone of $\Delta R = \min(0.2, 10 \text{ GeV}/E_T)$ around the candidate electron track and originating from the reconstructed primary vertex of the hard collision, excluding the electron associated tracks. The working points for both variable cuts are collected in a η-p_T matrix, characterized by an efficiency higher than 90%.

4.2   Muons

Muon reconstruction is first performed independently in the Inner Detector (ID) and Muon Spectrometer (MS). The information from individual subdetectors is then combined to form the muon tracks that are used in physics analyses. Muon reconstruction in the MS starts searching for hit patterns inside each muon chamber to form segments. In each MDT chamber and nearby trigger chamber, a particular pattern recognition algorithm, the Hough transform [85], is used to search for hits aligned on a trajectory in the not-bending plane of the detector. The MDT segments are reconstructed by performing a straight-line fit to the hits found in each layer. Segments in the CSC detectors are built using a separate combinatorial search in the η and ϕ detector planes. Muon track candidates are then built by fitting together hits from segments in different layers.

The combined ID-MS muon reconstruction [86] is performed according to various algorithms based on the information provided by the ID, MS, and calorimeters. Four muon types, shown in figure 4.2, are defined depending on which sub-detectors are used in reconstruction:

- **Combined (CB) muon:** track reconstruction is performed independently in the ID and MS, and a combined track is formed with a global refit that uses the hits from both the ID and MS subdetectors;

- **Extrapolated (ME) muons, known also as *Standalone Muons*:** the muon trajectory is reconstructed based only on the MS track and a loose requirement on compatibility with originating from the IP;

- **Segment-tagged (ST) muons:** a track in the ID is classified as a muon if, once extrapolated to the MS, it is associated with at least one local track segment in the MDT or CSC chambers;

- **Calorimeter-tagged (CT) muons:** a track in the ID is identified as a muon if it is matched to an energy deposit in the calorimeter compatible with a minimum-ionizing particle.

Four muon identification selections (Loose, Medium, Tight, and High-p_T) are provided to address the specific needs of different physics analyses. Loose, Medium, and Tight are inclusive categories in that muons identified with tighter requirements are also included in the looser categories.

- **Loose muons** The Loose identification criteria are designed to maximise the reconstruction efficiency while providing good-quality muon tracks. They are specifically optimised for reconstructing Higgs boson candidates in the four-lepton final state. All muon types are used. All CB and ME
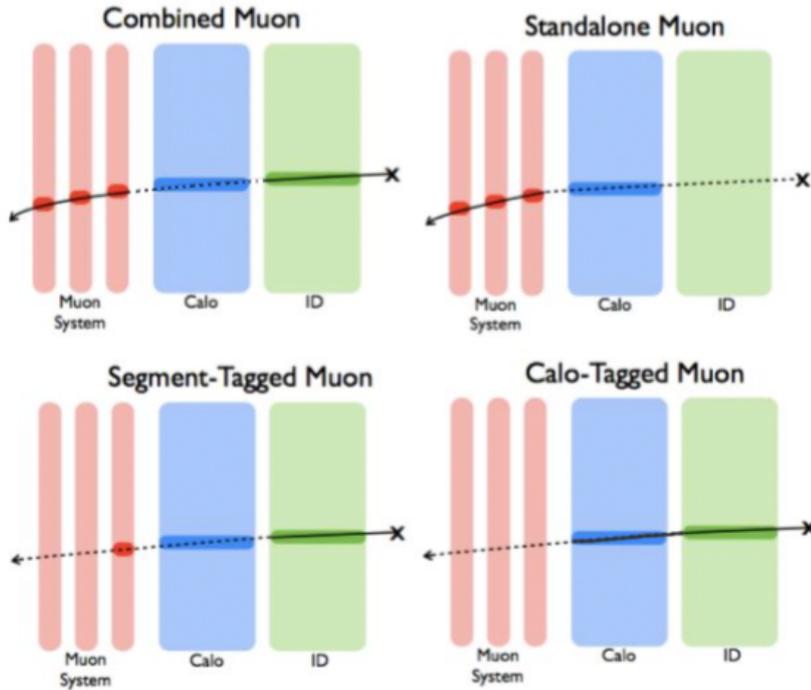
Figure 4.2: The four schematised types of reconstructed muon candidates in ATLAS.

muons satisfying the Medium requirements are included in the Loose selection. CT and ST muons are restricted to the $|\eta| < 0.1$ region. In the region $|\eta| < 2.5$, about 97.5% of the Loose muons are combined muons, approximately 1.5% are CT and the remaining 1% are reconstructed as ST muons.

- **Medium muons** The Medium identification criteria provide the default selection for muons in ATLAS. This selection minimises the systematic uncertainties associated with muon reconstruction and calibration. Only CB and ME tracks are used. The former are required to have ≥ 3 hits in at least two MDT layers, except for tracks in the $|\eta| < 0.1$ region, where tracks with at least one MDT layer but no more than one MDT hole layer are allowed. The latter are required to have at least three MDT/CSC layers, and are employed only in the $2.5 < |\eta| < 2.7$ region to extend the acceptance outside the ID geometrical coverage. A loose selection on the compatibility between ID and MS momentum measurements is applied to suppress the contamination due to hadrons misidentified as muons. In the pseudorapidity region $|\eta| < 2.5$, about 0.5% of the muons classified as Medium originate from the inside-out combined reconstruction strategy.

- **Tight muons** Tight muons are selected to maximise the purity of muons at the cost of some efficiency. Only CB muons with hits in at least two stations of the MS and satisfying the Medium selection criteria are considered. The normalised χ^2 of the combined track fit is required to be < 8 to remove pathological tracks and to ensure stronger background rejection for momenta below 20 GeV where the misidentification probability is higher.

- **High-p_{T} muons** The High-p_{T} selection aims to maximise the momentum resolution for tracks with transverse momentum above 100 GeV. The selection is optimised for searches for high-mass Z' and W' resonances. CB muons passing the Medium selection and having at least three hits in three MS stations are selected. Specific regions of the MS where the alignment is suboptimal are

vetoed as a precaution. Requiring three MS stations, while reducing the reconstruction efficiency by about 20%, improves the p_{T} resolution of muons above 1.5 TeV by approximately 30%.

In this analysis the muon candidates have to pass the following criteria:

- to be identified as a Medium muon, that imply to be reconstructed through CB or ME tracks with at least three MDT+CSC hits;

- the selections on the transverse and longitudinal impact parameters: $|\frac{d_0}{\sigma_{d_0}}| < 3$ and $|z_0 \sin\theta| < 0.5$ mm;

- $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$;

- to fulfil the isolation requirements on $E_{\mathrm{T}}^{\mathrm{cone}0.2}$ and $p_{\mathrm{T}}^{\mathrm{varcone}0.2}$ variables (see section 4.1). The working point choosen for this analysis provides a 99% efficiency, constant in η and p_{T};

- a separation $\Delta R > 0.4$ from the nearest selected jet.

The performance of the Medium muon reconstruction as a function of the p_{T} of the muon, with the 2015 data, is shown in figure 4.2 [86].
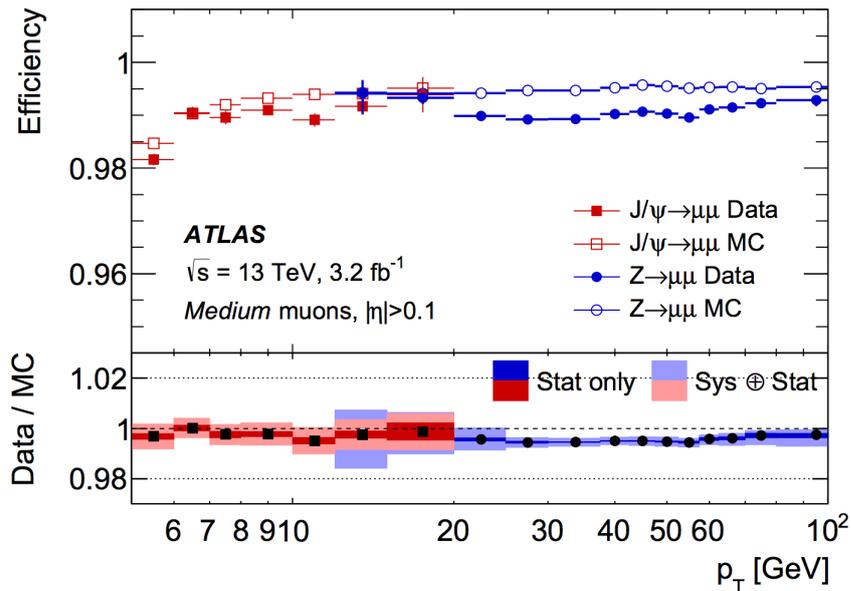
Figure 4.3: Reconstruction efficiency for the Medium muon selection as a function of the p_{T} of the muon, in the region $0.1 < |\eta| < 2.5$ as obtained with $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events. The muons considered in this performance study are required to fulfil only the first requirement of the selection of the analysis described in this thesis. The error bars on the efficiencies indicate the statistical uncertainty. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties. [86]

4.3   Missing transverse momentum

The missing transverse momentum in the event is defined as the magnitude of the negative vector sum p_{T} of all selected and calibrated physics objects in the event, with an extra term added to account for soft

energy in the event that is not associated to any of the selected objects. This soft term is calculated from inner detector tracks matched to the primary vertex to make it more resilient to pile-up contaminations. The missing transverse momentum is not used for event selection but it is used in the event reconstruction.

4.4   Jets

The basic structures of the jet reconstruction process in ATLAS are locally calibrated, three-dimensional topological clusters (*topo-clusters*), built from cells of the calorimeter [87]. Topo-clustering reconstruction starts with the identification of seed cells that should have energy significance at least 4σ above the noise level, where the noise is defined as the sum in quadrature of electronic and pile-up signals. Neighbor cells with energy significance higher than 2σ with respect to the noise are then iteratively added to form the clusters. An extra ring of direct neighbor cells is finally added to the clusters. After topo-clusters reconstruction, a splitting algorithm divides clusters in energy categories using a local energy maxima criterion. Individual clusters are calibrated using local properties such as energy density, calorimeter depth and isolation with respect to nearby clusters. This local cluster weighting (LCW) calibration classifies topological clusters along a continuous scale as being electromagnetic or hadronic, using shower shapes and energy densities. Energy corrections are applied to hadronic clusters based on this classification scheme, derived from single-pion MC simulations and verified with appropriate test beams.

In this analysis jets are reconstructed using the anti-k_t algorithm [88], a method based on a sequential cluster recombination algorithm with the following iterative procedure:

- for each cluster or "pseudo-jet" (the intermediate reconstruction object) i evaluate the distance d_{ij} with each other cluster or pseudo-jet j

$$d_{ij} = \min(p_{T,i}^{2k}, p_{T,j}^{2k}) \frac{\Delta R_{ij}^2}{R^2} \qquad (4.4.2)$$

  where ΔR_{ij}^2 is the angular distance between i and j, defined as

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2, \qquad (4.4.3)$$

  $p_{T,i(j)}$, $\eta_{i(j)}$ and $\phi_{i(j)}$ are respectively the transverse momentum, the pseudorapidity and the azimuthal angle of the i (j) object; R is an input parameter that limits the cone radius of the jet and k is a parameter of the anti-k_t algorithm fixed equal to -1;

- for each cluster or pseudo-jet i evaluate the distance

$$d_{iB} = p_{T,i}^{2k} \qquad (4.4.4)$$

  with the Beam (B);

- find the minimum distance among d_{ij} and d_{iB};

- if the minimum value is d_{ij} then combine i and j into a single *pseudo-jet* and repeat from the first step. Otherwise consider i as a final state and do not consider it in further iterations.

The anti-k_t algorithm favours the clusterization around hard particles rather than soft ones, as it instead happens in the case of the k_t algorithm ($k = 1$); the Cambridge/Achen (C/A) algorithm ($k = 0$) has energy-independent clustering. The anti-k_t algorithm is an infrared and collinear safe algorithm (IRC) for its

distance definition. IRC safety indicates that the set of hard jets remains unchanged even in case of a collinear splitting or the addition of a soft emitted gluon.

The reconstructed jets are calibrated through the ATLAS LCW+JES (where JES is the jet energy resolution) scheme that applies corrections as a function of the jet energy and pseudo-rapidity to the jets reconstructed at the electromagnetic scale.

The jets used in this analysis are required to be calibrated through the LCW+JES chain and reconstructed with the anti-k_t algorithm, with $R = 0.4$.

4.4.1 Pile-up corrections

Due to multiple pp collisions within the same bunch crossing, a variety of particles not belonging to the primary interaction vertex are produced (*pile-up*). The *pile-up* products can interfere or by overlapping with the objects of physics interest, requiring an additional calibration correction, or generating new jets. The *pile-up* contribution to jet calibration is accomplished by subtracting the average additional energy due to *pile-up* interactions from the energy measured by the calorimeters. The correction constants used for that purpose are obtained by *in situ* measurements coming from minimum bias data and depend on the number of reconstructed primary vertices (N_{PV}), the jet pseudo-rapidity (η) and the bunch spacing.

Some pile-up jets remain even after pile-up subtraction mainly due to localised fluctuations in pile-up activity. The matching between the tracks and each jet is used to further reject any jets not originating from the hard-scatter interaction. ATLAS has developed three different track-based tagging approaches for the identification of pile-up jets:

- Jet Vertex Fraction (JVF) algorithm, used in almost all physics analyses in Run-1;

- improved variables (corrJVF and R_{p_T}) for pile-up vertex identification;

- Jet Vertex Tagger (JVT), a new combined discriminant, for optimal performance.

The analysis described in this thesis uses the JVT method, but since it uses the JVF and the improved variables, it is important to consider them more in details.

Jet Vertex Fraction

The Jet Vertex Fraction (JVF) is a variable used to identify the primary vertex (PV) from which the jet originated. Once the hard-scatter PV is identified, the JVF variable can be used to select jets having a high probability of originating from that vertex. Tracks are assigned to calorimeter jets following the ghost-association procedure, which consists of assigning tracks to jets by adding tracks with infinitesimal p_T to the jet clustering process. The JVF is the ratio of the scalar sum of the p_T of matched tracks that originate from a given PV to the scalar sum of p_T of all matched tracks in the jet, independently of their origin.

JVF is defined for each jet with respect to each PV. Tracks are assigned to vertices by requiring $|\Delta z \cdot \sin \theta| < 1$ mm. In case more than one vertex satisfies this criterion, ambiguity is resolved by choosing the vertex with the largest summed $p_T{}^2$ of tracks. For a given jet$_i$, its JVF with respect to the primary vertex PV$_j$ is given by:

$$\mathrm{JVF}_{\mathrm{jet}_i,\mathrm{PV}_j} = \frac{\sum_m p_T(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_j)}{\sum_n \sum_l p_T(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_n)}, \tag{4.4.5}$$

where m runs over all tracks originating from PV$_j$ matched to jet$_i$, n over all primary vertices in the event and l over all tracks originating from PV$_n$ matched to jet$_i$. Only tracks with $p_T > 500$ MeV are

considered in the JVF calculation. JVF is bounded by 0 and 1, but a value of -1 is assigned to jets with no associated tracks.

The principle of the JVF variable is shown schematically in figure 4.4a and the figure 4.4b shows the discriminating power of JVF variable between hard-scatter and pile-up jets.

(a)   (b)
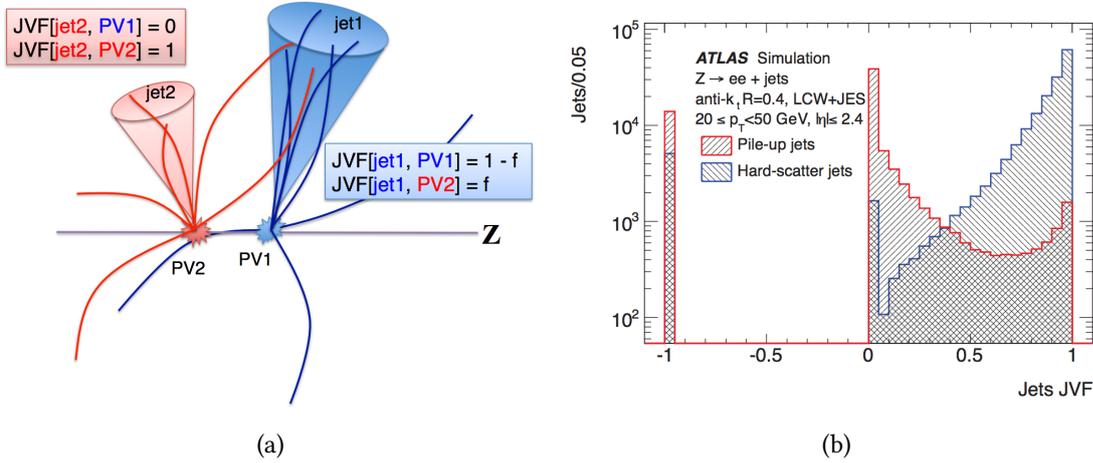
Figure 4.4: (a) A schematic representation of the JVF principle where f denotes a fraction. (b) JVF distribution for hard-scatter (blue), after pile-up subtraction, and pile-up (red) jets with $20 < p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$. Jets are calibrated with the jet energy scale correction in simulated Z+jets events. [89]

Improved variables (corrJVF and $R_{p_{\mathrm{T}}}$) for pile-up vertex identification

While a JVF selection is very effective in rejecting pile-up jets, it has limitations in higher luminosity conditions: as the denominator of JVF (eq. 4.4.5) increases with the number of reconstructed primary vertices in the event, the mean JVF for signal jets (which is 1 by definition) is shifted to smaller values. This pile-up sensitivity is addressed in two different ways: first, by correcting JVF for the explicit pile-up dependence in its denominator (corrJVF) and second, by introducing a new variable defined entirely from hard-scatter observables ($R_{p_{\mathrm{T}}}$). The corrJVF is similar to JVF, but keeping into account the pile-up interactions as

$$\mathrm{corrJVF}_{\mathrm{jet}_i,\mathrm{PV}_j} = \frac{\sum_m p_{\mathrm{T}}(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_j)}{\sum_l p_{\mathrm{T}}(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_j) + \frac{\sum_{n \neq j} \sum_l p_{\mathrm{T}}(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_n)}{w \cdot n_{track}^{\mathrm{PU}}}}, \quad (4.4.6)$$

where $\sum_m p_{\mathrm{T}}(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_j)$ is the scalar sum of the p_{T} of the tracks that are associated with the jet and originate from the hard-scatter vertex PV_j; the term $\sum_{n \neq j} \sum_l p_{\mathrm{T}}(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_n) = p_{\mathrm{T}}^{PU}$ is the scalar sum of the p_{T} of the associated tracks originated from any of the pile-up interactions; n_{track}^{PU} is the total number of pile-up tracks per event and w is a scaling factor chosen equal to 0.01, used to correct for the linear increase of $< p_{\mathrm{T}}^{PU} >$.

The corrJVF variable uses a modified track-to-vertex association method different from the one of the JVF and that consists in two steps. In the first step, the vertex reconstruction is used to assign tracks to vertices. If a track is attached to more than one vertex, priority is given to the vertex with higher $\sum_{tracks} p_{\mathrm{T}}^2$. In the second step, if a track is not associated with any primary vertex after the first step but

it satisfies $|\Delta z| < 3$ mm with respect to the hard-scatter primary vertex, it is assigned to the hard-scatter primary vertex the same. The second step flags tracks from decays in flight of hadrons originated from the hard-scatter but not likely attached to any vertex.

The variable R_{p_T} is defined as the scalar sum of the p_T of the tracks associated to the jet and originate from the hard-scatter vertex divided by the fully calibrated jet p_T, which includes pile-up subtraction:

$$R_{p_\mathrm{T}(\mathrm{jet}_i, \mathrm{PV}_j)} = \frac{\sum_m p_\mathrm{T}(track_m^{\mathrm{jet}_i}, \mathrm{PV}_j)}{p_\mathrm{T}^{jet_i}}. \tag{4.4.7}$$

For pile-up jets, R_{p_T} is peaked at 0 and is steeply falling, since tracks from the hard-scatter vertex rarely contribute. For hard-scatter jets, however, R_{p_T} has the meaning of a charged p_T fraction and its mean value and spread are larger than for pile-up jets.

The corrJVF and R_{p_T} distributions for pile-up and hard-scatter jets are shown in figures 4.5a and 4.5b, using a Monte Carlo simulation sample of dijet events with the Pythia8 generator.

Jet Vertex Tagger

The jet-vertex-tagger (JVT) is a new discriminant derived by R_{p_T} and corrJVF, using simulated dijet events. The reconstruction algorithm is a multivariate (MVA) technique called k-Nearest Neighbourhood method (k-NN) [90]. A multivariate technique is any "classification" method that allows to recognize an object (test sample) belonging to a certain category, starting from N possible categories, through a certain knowledge previously built (training sample). A brief introduction of the MVA is given to understand the JVT method, but the complete description is given in Chapter 5, since the MVA is a crucial part of the analysis. "Classification" concerns the separation of variables of different categories from each other. If variables are grouped in two categories, this is a problem of binary classification, while, if the number of categories exceeds two, this is a multiclass problem. In binary classification, the first category is described as "signal" and the second as "background". In order to classify as correctly as possible, a MVA technique separates the studied sample in a *training* and a *test* samples. The former is the one the model is built on, needed to "learn" the method about the features of the signal, through a set of different discriminant variables. The latter tests if the method is performing a correct classification (without bias or too large statistical errors), separating the signal from the background.

The k-NN method compares an observed (test) event to reference events from a training data set and has best performance when the boundary that separates signal and background events has irregular features that cannot be easily approximated by parametric learning methods. The k-NN method uses the concept of "local neighbourhood", measured through a metric function. The simplest metric choice is the Euclidean distance

$$R = \left(\sum_{i=1}^{n_{var}} |x_i - y_i|^2 \right)^{\frac{1}{2}} \tag{4.4.8}$$

where n_{var} is the number of input variables used for the classification, x_i are coordinates of an event from a training sample and y_i are variables of an observed test event. The k events with the smallest values of R are the k-nearest neighbours. The value of k determines the size of the neighbourhood for which a probability density function is evaluated.

In the JVT method, for each point in the two-dimensional corrJVF-R_{p_T} plane, the relative probability for a jet to be of "signal type" (meaning to be arised from a hard-scatter vertex) is computed as the ratio of the number of hard-scatter jets to the number of hard-scatter plus pile-up jets found in a local neighbourhood around the point. Figure 4.6a shows the fake rate versus efficiency curves comparing the
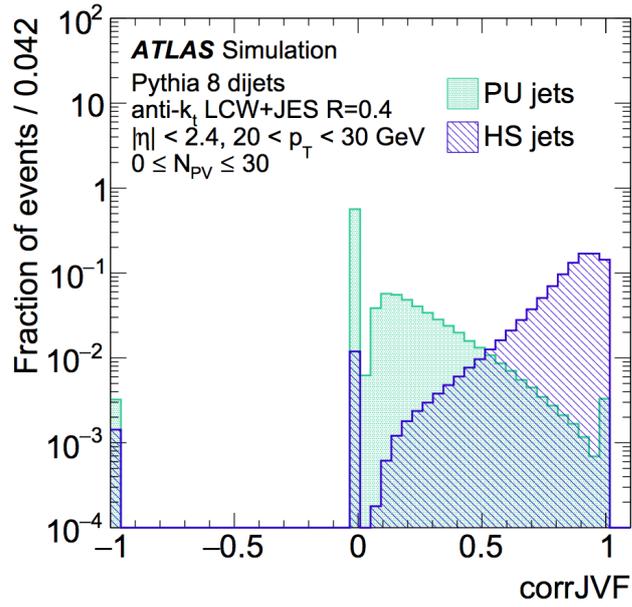
(a)
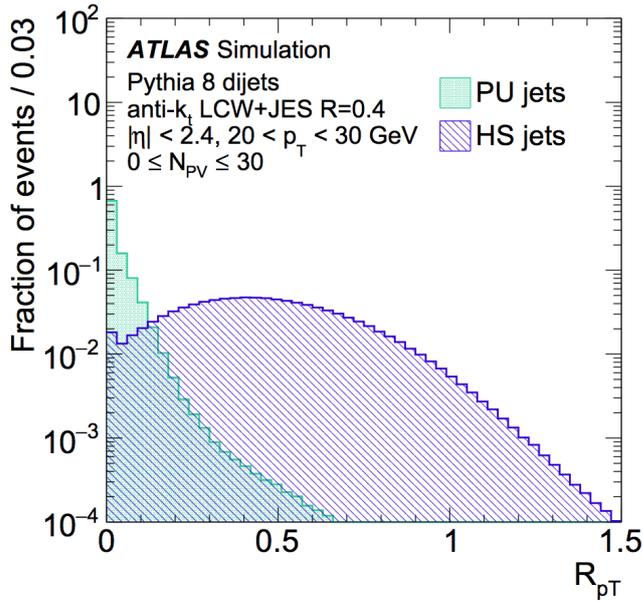
(b)

Figure 4.5: (a) Distribution of corrJVF for pile-up (PU) and hard-scatter (HS) jets with $20 < p_T < 30$ GeV. (b) Distribution of R_{p_T} for pile-up (PU) and hard-scatter (HS) jets with $20 < p_T < 30$ GeV. [89].

performance of the four variables JVF, corrJVF, R_{p_T}, and JVT when selecting a sample of jets with $20 < p_T < 50$ GeV, $|\eta| < 2.4$ in simulated dijet events. The stability of the hard-scatter jet efficiencies on N_{PV}, with a selection based on JVT, is shown in figure 4.6b.

On low p_T ($p_T < 60$ GeV) jets in the central ($|\eta| < 2.4$) region of the detector a requirement of JVT > 0.59 has been applied both to data and simulation.
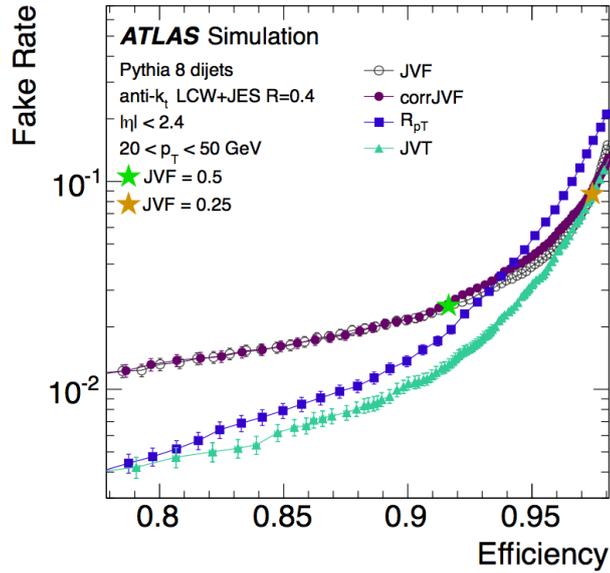
(a)
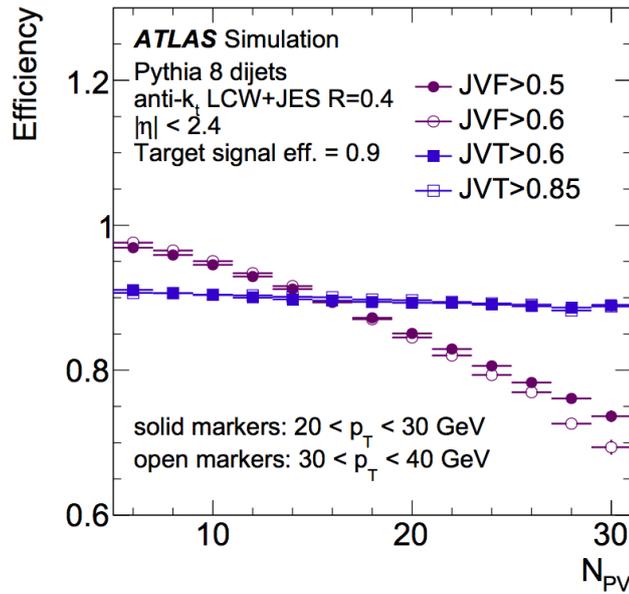
(b)

Figure 4.6: (a) Fake rate from pile-up jets versus hard-scatter jet efficiency curves for JVF, corrJVF, R_{p_T}, and JVT. The JVF working points are indicated with gold and green stars. (b) Hard-scattering jet efficiency dependence on the number of primary vertices for $20 < p_T < 30$ GeV (solid markers) and $30 < p_T < 40$ GeV (open markers) jets for fixed cuts of JVT (blue square) and JVF (violet circle) such that the inclusive efficiency is 90%. The stability of the JVT method is within 1%. [89].

4.4.2   Overlap removal

During jet reconstruction, no distinction is made between identified electrons and jet energy deposits. Therefore, if any of the jets lie within ΔR of 0.2 of a selected electron, the single closest jet is discarded

in order to avoid double-counting of electrons as jets. After this, electrons which are within ΔR of 0.4 of a remaining jet are removed (see section 4.1).

4.4.3 b-tagging algorithm

The lifetime-based tagging algorithms take advantage of the relatively long lifetime of hadrons containing a b quark, of the order of 1.5 ps ($c\tau \approx 450~\mu$m). A b hadron with p_{T}= 30 GeV will have a significant mean flight path length $<l> = \beta\gamma c\tau$, travelling on average about 3 mm in the transverse direction before decaying and therefore leading to topologies with at least one vertex displaced from the point where the hard-scatter collision occurred.

Two classes (see table 4.1) of algorithms aim to identify such topologies. The first uses an "inclusive" approach evaluating the impact parameters of the charged-particle track. The term "inclusive" is used because of the non-complete displaced vertices reconstruction. It contemporarily uses two impact parameters: the transverse d_0 and the longitudinal one z_0. The transverse impact parameter d_0 is the distance of closest approach of the tracks to the primary vertex point, in the $r - \phi$ projection. The longitudinal impact parameter z_0 is the difference between the z coordinates of the primary vertex and of the closest approach tracks in $r - \phi$ projection. The tracks from b-hadron decay products tend to have large impact parameters which can be distinguished from tracks stemming from the primary vertex. Two tagging algorithms exploiting these properties are discussed: JetProb, used mostly for early data, and IP3D for high-performance tagging [91]. The second class of algorithm, called "exclusive" approach, reconstructs explicitly the displaced vertices. Two algorithms make use of this technique: the SV (Second Vertex) algorithm attempts to reconstruct the secondary vertex; while the JetFitter [92] algorithm aims at reconstructing the complete b-hadron decay chain. Finally, the results of several of these algorithms are combined in a multivariate algorithm (called MV1) to improve the light-flavour-jet rejection and to increase the range of b-jet tagging efficiency for which the algorithms can be applied.

Approach	Tagging algorithm	Method
inclusive	JetProb	reconstruction of the impact parameters (early data)
inclusive	IP3D	reconstruction of the impact parameters (improved performances)
exclusive	SV	reconstruction of secondary vertex
exclusive	JetFitter	reconstruction of the complete b-hadron decay chain

Table 4.1: Different tagging algorithms used in the b-tagging chain.

The MV1 is based on a neural network approach [90], used widely in ATLAS physics analyses, during Run-1, to discriminate b-jets from light (u, d, s-quark or gluon jets) and c-jets. The MVA training sample is based on two simulated samples of b-jets (signal hypothesis) and light-flavour jets (background hypothesis). Most of the jets are obtained from simulated $t\bar{t}$ events and their average transverse momentum is around 60 GeV. To provide jets with higher p_{T} for the training, simulated dijet events with jets in the 200 GeV $< p_{\mathrm{T}} <$ 500 GeV range are also included.

The MV2 algorithm [93][94] constitutes a significant revision of the MV1, based on a Boosted Decision Tree algorithm (BDT, explained in details in Chapter 5). The new approach not only improves the performance, as will be shown in the following, but also significantly simplifies the algorithm by directly using the variables from the basic algorithms, omitting the additional intermediate multivariate tools.

The input variables, both for MV1 and MV2, are shown in table 4.2, obtained from the three basic algorithms (IP3D, SV and JetFitter). The training is performed on a set of approximately five millions $t\bar{t}$ events. Three MV2 variants were released, MV2c00, MV2c10 and MV2c20, where the names of the taggers indicate the c-jet fraction in the training, e.g. in MV2c20, the background sample is composed of 20% (80%) c- (light-flavour) jets.

Variable	Description
$p_{\mathrm{T}}(\text{jet})$	Jet transverse momentum
$\eta\,(\text{jet})$	Jet pseudo-rapidity
$\log(P_b/P_{light})$	Likelihood ratio between the b- and light-jet hypotheses
$\log(P_b/P_c)$	Likelihood ratio between the b- and c-jet hypotheses
$\log(P_c/P_{light})$	Likelihood ratio between the c- and light-jet hypotheses
$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion masses
$f_E(\text{SV})$	Fraction of the charged jet energy in the secondary vertex
$N_{TrkAtVtx}(\text{SV})$	Number of tracks used in the secondary vertex
$N_{2TrkVtx}(\text{SV})$	Number of 2-tracks vertex candidates
$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertices
$L_{xyz}(\text{SV})$	Distance between the primary and secondary vertices
$S_{xyz}(\text{SV})$	Distance between the primary and secondary vertices divided by its uncertainty
$\Delta R(jet, SV)$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex
$N_{2TrkVtx}(\text{JF})$	Number of 2-track vertex candidates (prior to decay chain fit)
$m(\text{JF})$	Invariant mass of tracks from displaced vertices assuming pion masses
$S_{xyz}(\text{JF})$	Significance of the average distance between the primary and displaced vertices
$f_E(\text{JF})$	Fraction of the charged jet energy in the secondary vertices
$N_{1-trkvertices}(\text{JF})$	Number of displaced vertices with one track
$N_{\geq 2-trkvertices}(\text{JF})$	Number of displaced vertices with more than one track
$N_{TrkAtVtx}(\text{JF})$	Number of tracks from displaced vertices with at least two tracks
$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices

Table 4.2: The 24 input variables used by the MV1 and MV2 b-tagging algorithms. [94]

The performance of the optimised MV2c00, MV2c10 and MV2c20 b-tagging algorithms is shown in figures 4.7a and 4.7b respectively for the light and c-jet rejection as a function of the b-jet efficiency in comparison to the 2015 MV2c20 configuration. The rejection is defined as the reciprocal of the fraction of light and c-jets sample that pass the b-tagging algorithm requirements in a background sample. It is noted that the current MV2c10 (2016 configuration) discriminant provides a similar light-flavour jet rejection (improvement of approximately 4% at 77% b-jet efficiency) to the 2015 MV2c20 configuration, but a significantly larger c-jet rejection (+40%). The 2016 MV2c20 set-up provides even better charm rejection, but at the expense of a reduced light-flavour jet rejection. The MV2c10 tagger algorithm has therefore been chosen as the standard b-tagging discriminant for 2016 analyses.

In figure 4.8, the MV2c10 BDT output is shown for b-, c- and light-flavour jets. A BDT output, as widely
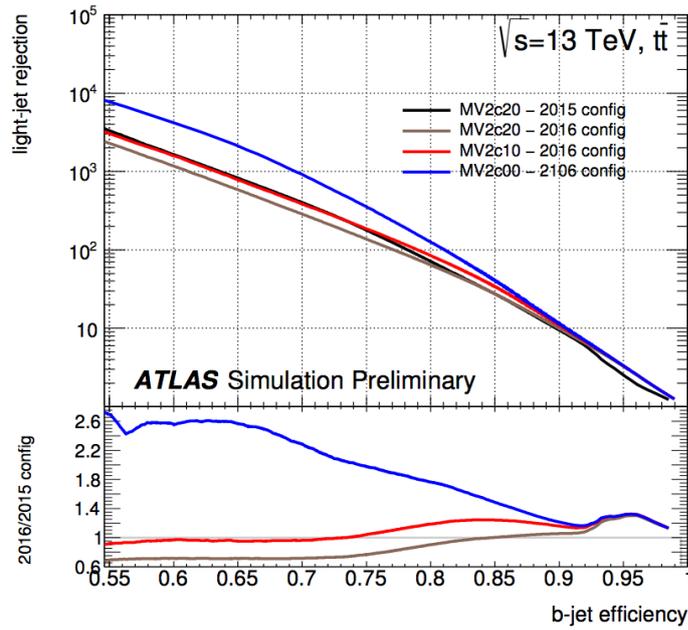
(a)
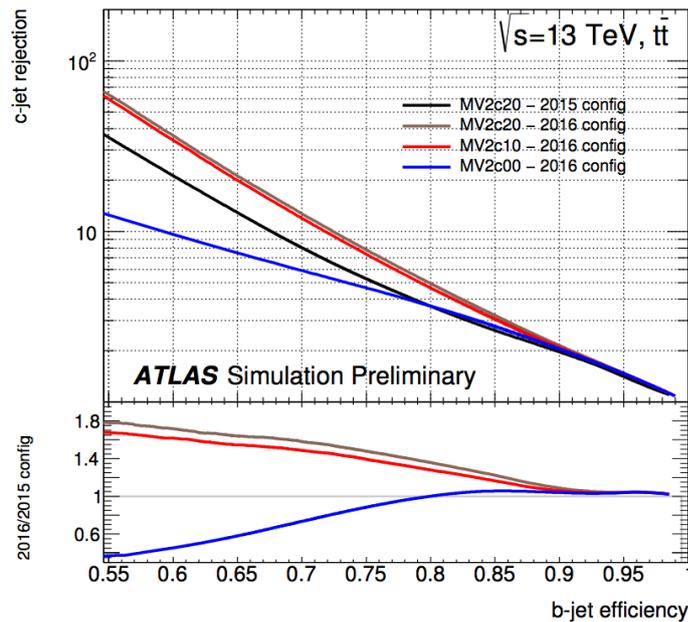
(b)

Figure 4.7: Light-flavour jet (a) and c-jet (b) rejection versus b-jet efficiency for the previous (2015 config) and the current configuration (2016 config) of the MV2 b-tagging algorithm evaluated on $t\bar{t}$ events [93]. The ratio is calculated between the 2016 and the 2015 configuration.

explained in Chapter 5, has a range from -1 to 1 and is a very good discriminant the b-jets recognition with respect to the c- and light-jets.

The efficiency calibrations of the b-tagging algorithm are performed with fixed thresholds (or "work-
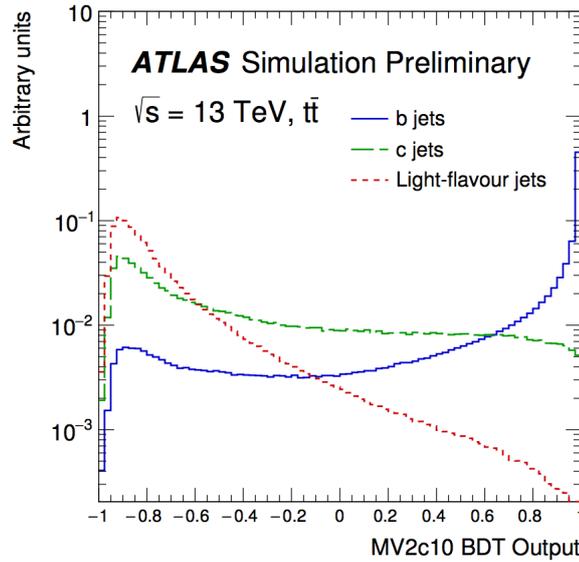
Figure 4.8: The MV2c10 [93] output for b- (solid blue), c- (dashed green) and light-flavour (dotted red) jets in $t\bar{t}$ events.

ing points", WP) of the tag weights computed by the b-tagging algorithms. These WP are defined by a single cut value on the MV2 output distribution and are chosen to provide specified b-jet efficiencies based on the inclusive p_T and η spectra of jets from an inclusive $t\bar{t}$ sample. Table 4.3 shows the four operating points defined for the MV2c10 (2016 configuration) b-tagging algorithms. Within this thesis, emphasis will be placed on WP tuned to an average 70% and 77% b-tagging efficiency, but for the purpose of the analysis the 70% working point will be used.

BDT cut value	b-jet efficiency [%]	c-jets rejection	light-jets rejection	τ rejection
0.9349	60	34	1538	184
0.8244	70	12	381	55
0.6459	77	6	134	22
0.1758	85	3.1	33	8.2

Table 4.3: Working points for the MV2c10 b-tagging algorithm, including benchmark numbers for the efficiency and rejections rates. These values have been extracted from $t\bar{t}$ events, the main requirement being jet p_T above 20 GeV.

Performance enhancement from detector and algorithmic improvements

It is fundamental to determine the total improvement in the b-tagging performance achieved between Run-1 and Run-2 due to the addition of the IBL (as discussed in Chapter 2) and the algorithmic updates. Figures 4.9a and 4.9b show such a comparison between the performance of the default b-tagging algorithm with improved c-jet rejection for Run-1, MV1c [95], with the Run-1 detector and reconstruction software, compared to the Run-2 b-tagging algorithm, MV2c20 (now the default is MV2c10, even better than Mv2c20, as explained above), with the Run-2 detector and reconstruction software. Comparing the

two algorithms, the light-flavour jet rejection is improved by a factor of about 4 and the c-jet rejection by a factor of between 1.5-2 for a 70% b-jet efficiency.
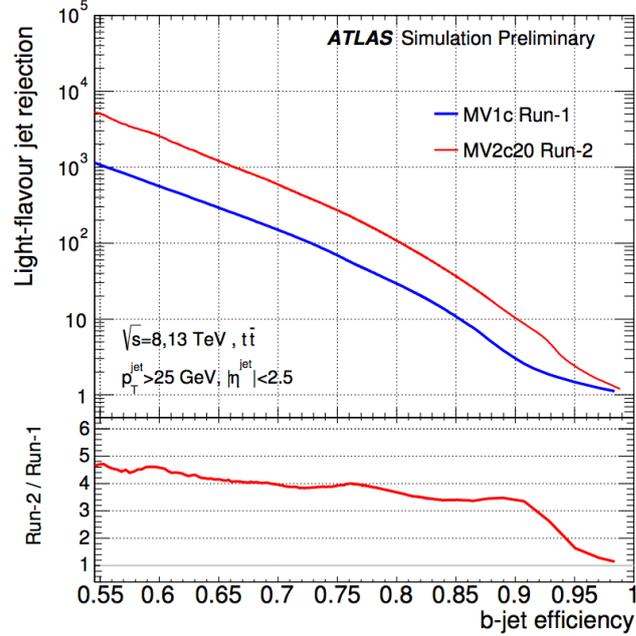
(a)
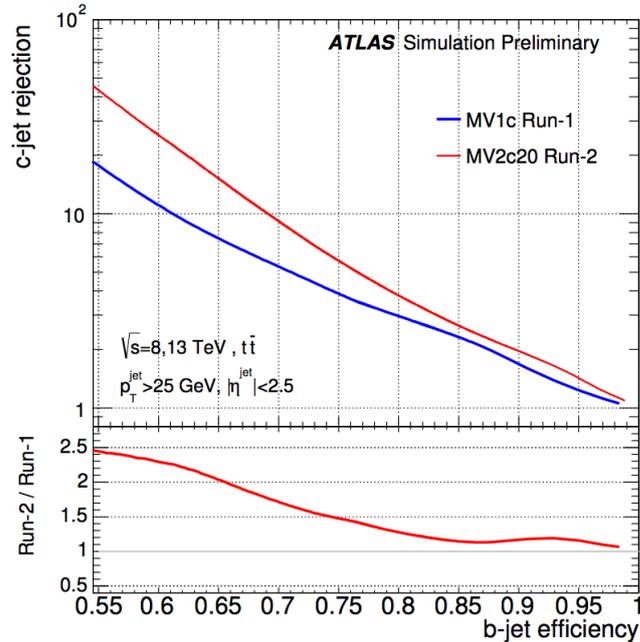
(b)

Figure 4.9: The light (a) and c-jet rejection (b) versus b-jet efficiency for the MV1c b-tagging algorithm using the Run-1 detector and reconstruction software (blue) compared to the MV2c20 b-tagging algorithm using the Run-2 setup (red). [93]

The IBL improves the impact parameter resolution of tracks mostly at p_T up to 5-10 GeV, and thus the improvement is concentrated in the low to medium jet p_T region (see Chapter 2), while at high jet p_T most of the improvement comes from the new algorithms.

4.5   Boosted objects

During the Run-2, LHC is exploring a completely new physics regime where the available center-of-mass energy far exceeds the masses of known standard model particles. At such energies, heavy particles such as W, Z and H bosons and top quarks are often produced with large transverse momentum (*boosted particles*) that implies large Lorentz boost for their decay products. The property of boosted object decays is that they are collimated to the momentum direction of the boosted mother particle in the detector rest frame. Figure 4.10 shows the true angular separation between the W and b decay products of a top quark in simulated $Z' \to t\bar{t}$ ($m_{Z'}$ = 1.6 TeV) events (where Z' is a new heavy gauge boson), as well as the separation between the light quarks of the subsequent hadronically-decaying W. In each case, the angular separation of the decay products is approximately

$$\Delta R \simeq 2m/p_\mathrm{T}, \tag{4.5.9}$$

where, $\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2}$ and p_T and m are the transverse momentum and mass of the decaying particle, respectively. For $p_\mathrm{T}{}^{W} > 200$ GeV, the ability to resolve the individual hadronic decay products using standard narrow-cone jet algorithms begins to degrade, and above $p_\mathrm{T}{}^{top} > 350$ GeV, the decay products of the hadronic top quark tend to have a separation $\Delta R < 1.0$.
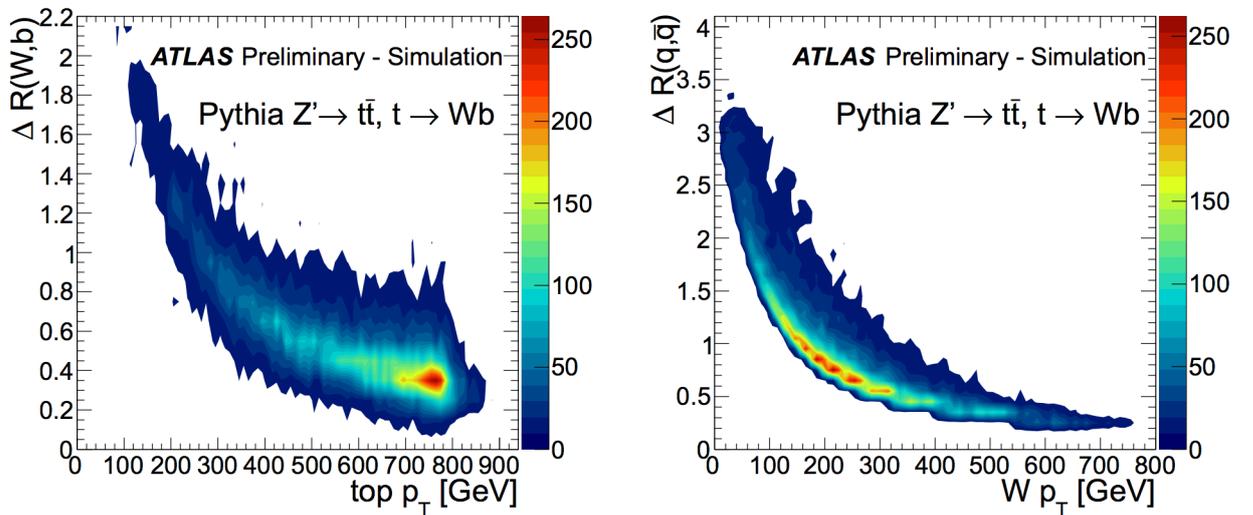
Figure 4.10: On the left, the opening angle between the W and b in top decays, $t \to Wb$, as a function of the top p_T in simulated Pythia $Z' \to t\bar{t}$ ($m_{Z'}$ = 1.6 TeV) events. On the right, The opening angle of the $W \to q\bar{q}$ system from $t \to Wb$ decays as a function of the W p_T.

As a consequence, the traditional reconstruction algorithms loose significantly the efficiency, due to the overlapping of the jets coming from an hadronic decay of a mother particle. In the 2015 and 2016 data analysis, the large integrated luminosity collected at $\sqrt{s}$= 13 TeV allows to explore the high-p_T region for events with unprecedent sensitivity. Consequently, at high p_T, the decay products of a hadronically

decaying object merge into a single, energetic and large radius jet (*large-R jet*) with a characteristic substructure different from those initiated by a single parton.

In this analysis, the jets of each events are reconstructed twice: in both cases the anti-k_t algorithm is used, but the first time with the standard radius R fixed to 0.4 and successively with R set to 1, to have both the standard jets and the large-R jets reconstructed. This simplifies the problem of the combinatorics in the final state event, but increases the probability to have energy deposit contributions from pile-up sources into the large-R jets. In the mean time, due to the peculiar topology considered, it reduces the contribution from the various background sources (other jets, pile-up, etc.) in the event. Figure 4.11 shows the schematic representation of the jet configuration coming from a low-p_T top quark hadronic decay (on the left), that is the "resolved regime" (see section 1.5.2) with respect to one coming from a high-p_T top quark (in the center); on the right, the last configuration shows how it appears in a single, large-R jet reconstruction.
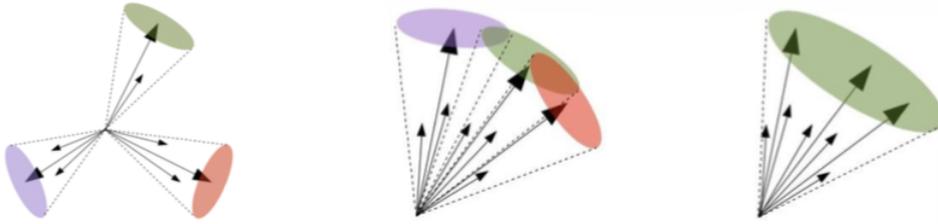
Figure 4.11: Graphical representation of the jet produced in top quark hadronic decays, in case of low (left) and high (center) values of top p_T. The picture on the right show the same high top p_T configuration as before using a large-R jet reconstruction.

A single large-R jet that contains all of the decay products of a massive particle will have significantly different properties than a single large-R jet of the same p_T originating from a single light-quark or gluon. The characteristic two-body or three-body decays of a vector boson or top quark result in a hard substructure (absent from the light-quark and gluon jets) that can be more resolved by removing soft radiation from jets. This selective removal of soft radiation during the process of iterative recombination in jet reconstruction is generally referred to as jet "grooming" (see paragraph 4.5.2).

In the latest years new techniques, called *tagging algorithms*, have been proposed in order to recognize large-R jets originated by massive particle decays with the aim to increase efficiency and purity in high energy analyses. Such techniques involve the study of the substructure of large-R jets both via direct comparison of the decay signature and by using a selection based on substructure variables [96].

In this chapter an overview of the most useful quantities for top and Higgs tagging and the principal grooming techniques used in ATLAS are presented.

4.5.1 Jet substructure observables

The principal jet substructure variables used by ATLAS tagging algorithms are jet mass, splitting scale and N-subjettiness.

- jet mass, m_{jet},

- N-subjettiness, τ_{32},

- energy correlation, D_2,

- splitting scale, $\sqrt{d_{12}}$.

Jet mass

The jet mass m_{jet} is calculated from the energies and momenta of the constituents of the jet

$$m_{jet} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i p_i\right)^2} \tag{4.5.10}$$

where Ei and p_i are respectively the energy and momentum's absolute value of the ith constituent of the large-R jet, namely the *topo-clusters* and the tracks. The jet mass is a powerful discriminant between boosted high mass particles and low ones coming from background, used to tag the top quark in the analysis of this thesis. In figure 4.12 the distribution of the large-R jet mass for the 2015 data is shown, compared with the Monte Carlo simulations [97]. In the large-R jets considered the contamination from pile-up, multiple parton interactions and initial-state radiation has been removed (see paragraph 4.5.2).
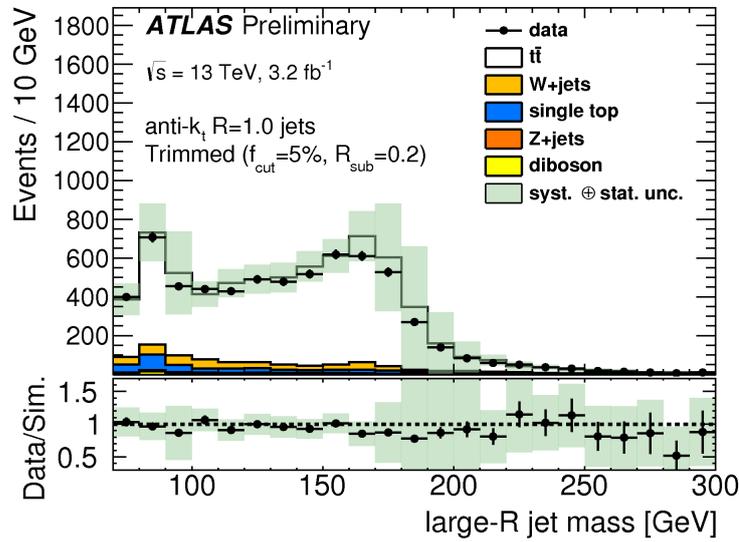
Figure 4.12: Mass spectrum for the leading-p_T jet in 13 TeV data and MC simulation using anti-k_t R = 1.0 jets calibrated at the LCW+JES scale, with $p_T > 300$ GeV. These jets are also required to contain at least one b-tagged R = 0.2 track jet, where the tag is defined at the 70% efficiency point of the MV2c20 algorithm. [97]

In this analysis, only large-R jets with $m_{jet} > 50$ GeV have been considered.

N-subjettiness

N-subjettiness variables τ_N are observables related to the pseudo-jet multiplicity. The τ_N variable is calculated by clustering with the k_t algorithm the constituents of the jet requiring exactly N pseudo-jets to be found. The k_t algorithm, in this case, interrupts the clusterization process when there are exactly N pseudo-jets remaining. The τ_N variables are then defined as the sum over all k constituents of the jet:

$$\tau_N = \frac{1}{d_0} \sum_k p_{Tk} \times \min(\delta R_{1k}, \delta R_{2k}, ..., \delta R_{Nk})$$
$$d_0 \equiv \sum_k p_{Tk} \times R \tag{4.5.11}$$

where R is the jet radius parameter (fixed to 0.4 as the standard anti-k_t jet reconstruction), p_{T_k} is the p_T of the kth constituent and δR_{ik} is the distance between the ith pseudo-jet to k^{th} constituent. From this definition, τ_N indicates how well the large-R jet can be described as containing N or fewer pseudo-jets, discriminating by how constituents are localized close to the pseudo-jet axes. The ratios $\tau 2/\tau 1$ and $\tau 3/\tau 2$ can be used to provide discrimination between jets formed from the parton shower of light quarks or gluons and jets containing two or three hadronic decay products from Z bosons, for example, or from top quarks. These ratios will be referred to as τ_{21} and τ_{32} respectively. For example, $\tau_{21} \approx 1$ corresponds to a jet that is very well described by a single pseudo-jet whereas a lower value implies a jet that is much better described by two pseudo-jets than one.

In order to discriminate a large-R jet derived from a boosted top quark with respect to one originated by a Higgs boson, the ratio τ_{32} is used in this analysis. In figure 4.13 the distribution of the large-R jet τ_{32} for the 2015 data is shown, being compared with the Monte Carlo simulation [97]. In order to obtain an observable independent of the recoil of soft radiation inside the jet, the pseudo-jets are reconstructed with a "winner-take-all" (wta) recombination scheme [98]: during the k_t reconstruction algorithm, each pseudo-jet formed from two pseudo-jets 1 and 2 is defined to be the massless four-vector of energy $E_r = E_1 + E_2$ and of three-momentum $\vec{p}_r = E_r \hat{n}_r$, where $\hat{n}_r$ is the unit-normalized three-momentum of the highest transverse momentum proto-jet.
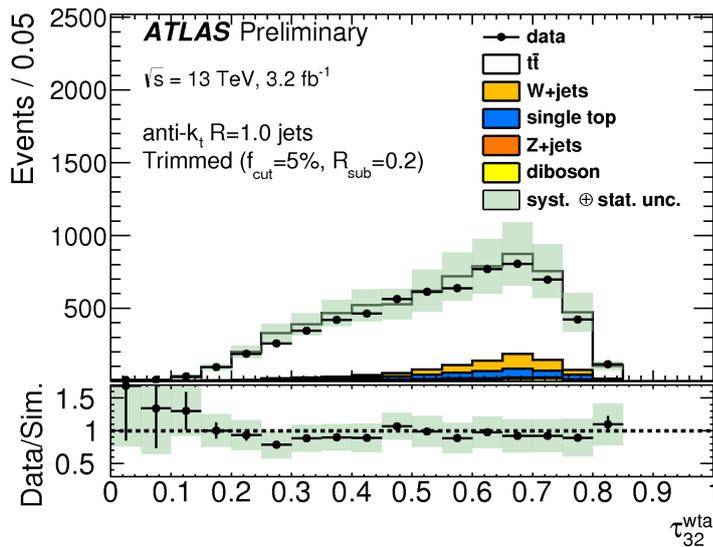
Figure 4.13: Distribution of N-subjettiness (with the "wta" axes definition) for the leading-p_T jet in 13 TeV data and MC simulation using anti-k_t R = 1.0 jets calibrated at the LCW+JES scale, with $p_T > 300$ GeV. These jets are also required to contain at least one b-tagged R = 0.2 track jet, where the tag is defined at the 70% efficiency point of the MV2c20 algorithm. [97]

In this analysis, the τ_{32} observable is used in the top-tagging algorithm, as explained in section 4.6, in which a p_T-dependent cut is applied on it.

Energy correlation

The goal of boosted boson discrimination is to distinguish between one-prong jets, arisen from a gluon or any QCD products, and two-prong jets, characteristic of the hadronic boson decay. The energy and angular correlation between the constituents of a jet is a powerful approach for the boson tagging. Unlike

the previous jet substructure methods, these correlation functions do not require the explicit identification of pseudo-jet regions. The method defines N+1 point correlation functions sensitive to identify N-prong jet substructure through the energies and pair wise angles of the constituents within a jet. In order to tag the two-prong structure of a jet, the 2- and 3-point energy correlation functions ($e_2^{(\beta)}$ and $e_3^{(\beta)}$ respectively) [99][100] are defined as:

$$
\begin{aligned}
e_2^{(\beta)} &= \frac{1}{p_{\mathrm{T}\,\mathrm{jet}}^2} \sum_{i\in\mathrm{jet}}^{n_c} \sum_{j<i\in\mathrm{jet}}^{n_c} p_{\mathrm{T}i} p_{\mathrm{T}j} \Delta R_{ij}^{\beta} \\
e_3^{(\beta)} &= \frac{1}{p_{\mathrm{T}\,\mathrm{jet}}^3} \sum_{i\in\mathrm{jet}}^{n_c} \sum_{j<i\in\mathrm{jet}}^{n_c} \sum_{k<j\in\mathrm{jet}}^{n_c} p_{\mathrm{T}i} p_{\mathrm{T}j} p_{\mathrm{T}k} \Delta R_{ij}^{\beta} \Delta R_{ik}^{\beta} \Delta R_{jk}^{\beta}
\end{aligned}
\tag{4.5.12}
$$

where the summations run over all the constituents of the jet (making all the combinations without duplication of the terms), $p_{\mathrm{T}\,\mathrm{jet}}$ is the transverse momentum of the jet, $p_{\mathrm{T}i(j,k)}$ is the transverse momentum of the constituent $i(j,k)$, n_c is the number of constituents contained in the jet, β is an angular exponent to be adjusted to optimize the discrimination power but greater than 0 to assure the infrared and collinear (IRC) safety. The boost-invariant angle $\Delta R_{ij}^2 = (\phi_i - \phi_j)^2 + (y_i - y_j)^2$ is defined as the distance in the azimuth-rapidity plane.

The key characteristic of the method is that the (N+1)-point correlation function goes to zero if there are only N pseudo-jets in the jet; more generally, if a system has N pseudo-jets, the (N+1)-point should be significantly smaller than the N-point correlation function.

A dimensionless variable useful to determine if a jet is composed by two pseudo-jets is the ratio

$$
D_2^{(\alpha,\beta)} = \frac{e_3^{(\alpha)}}{(e_2^{(\beta)})^{3\alpha/\beta}}
\tag{4.5.13}
$$

where in general the angular exponents α and β are different (as for the e^+e^- collisions), but in this analysis they have been both fixed to 1 ($D_2^{(\beta=1)}$). The distribution of $D_2^{(\beta=1)}$ is shown in figure 4.14 for a background sample of dijet qcd.

Splitting scale

The splitting scale variables are evaluated during each step of the algorithm reconstruction as the k_t-distance of the two pseudo-jets:

$$
\sqrt{d_{ij}} = \min(p_{T,i}, p_{T,j}) \cdot \Delta R_{ij}
\tag{4.5.14}
$$

where ΔR_{ij} is the angular distance between the two pseudo-jets, previously defined. The most useful splitting scales for boosted tagging algorithms are obtained at the last and second to the last reclustering step, namely the $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ variables. Because of the k_t algorithm combines the harder constituents last and that $\sqrt{d_{ij}}$ uses the minimum p_{T} between the i and j pseudo-jets, the parameters $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$ can be used to distinguish heavy particle decays, which tend to be reasonably symmetric with respect to the largely asymmetric splittings in light quark or gluon jets. The expected value for a two-body heavy particle decay is approximately $\sqrt{d_{12}} \approx m^{jet}/2$, whereas jets from the parton shower of light quarks and gluons will tend to exhibit a steeply falling spectrum for both $\sqrt{d_{12}}$ and $\sqrt{d_{23}}$. Due to the large correlation of these variables and the mass of the large-R jet, it has been decided to not use at the moment this method for this analysis, but to study it for further improvements.
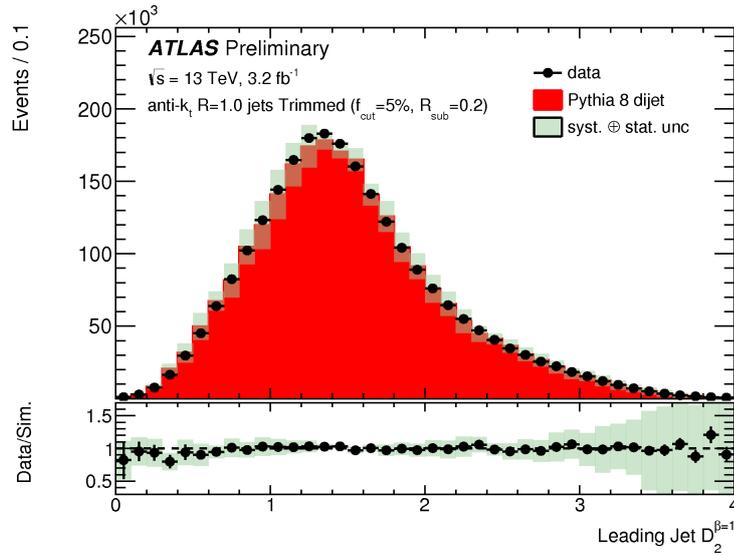
Figure 4.14: Distribution of $D_2^{(\beta=1)}$ for the leading-p_T jet in 13 TeV data and MC simulation using anti-k_t R = 1.0 jets calibrated at the LCW+JES scale, with $p_T > 300$ GeV. These jets are also required to contain at least one b-tagged R = 0.2 track jet, where the tag is defined at the 70% efficiency point of the MV2c20 algorithm. [97]

4.5.2 Soft radiation removal algorithms

The methods for the soft radiation removal inside jets are generally referred as "jet grooming algorithms" and can be divided in three different categories: mass-drop filtering, pruning and trimming. In this analysis the trimming algorithm is used and it will be discussed in details in this section. A brief description of the other two methods will be given as possible improvement for the future analyses.

Trimming

The trimming algorithm [101] removes contamination from pile-up, multiple parton interactions (MPI), and initial-state radiation (ISR) that are often much softer than hard-scattering partons products. The selection criteria used is based on the p_T ratio of the jet constituents. The trimming procedure reconstructs the large-R jet with the k_t algorithm finding the pseudo-jets constituents with a smaller radius (R_{sub} fixed to 0.2 in this analysis) and putting on them the p_T constraint $p_{T,i}/p_{T,large-Rjet} < f_{cut}$, where $p_{T,i}$ is the transverse momentum of the ith pseudo-jet, and f_{cut} is a parameter of the method, fixed to 0.05 in this analysis. The surviving constituents form the *trimmed jet*. This procedure is illustrated in figure 4.15.

In figures 4.16a, 4.16b and 4.16c the effect of the trimming algorithm on distributions of mass, splitting scales and N-subjettiness is shown, referred to the leading-p_T jet in the range $600 \leqslant p_T^{large-Rjet} < 800$ GeV. For these studies a $Z' \to t\bar{t}$ Monte Carlo sample ($m_{Z'}$ = 1.6 TeV) has been considered for signal-like events (red lines), compared with a MC multijets background (black lines). After the procedure, low-mass jets from a light quark or gluon usually loose 30-50% of their mass, while jets containing the decay products of a boosted object lose only a few percent of mass, mainly removing pile-up contribution; this is due to the large-R jet internal structure that is more uniform in the case of light quarks and gluons production. The effect of grooming increases the separation between signal and background distributions for all the substructure variables considered, helping the discrimination based on these quantities.
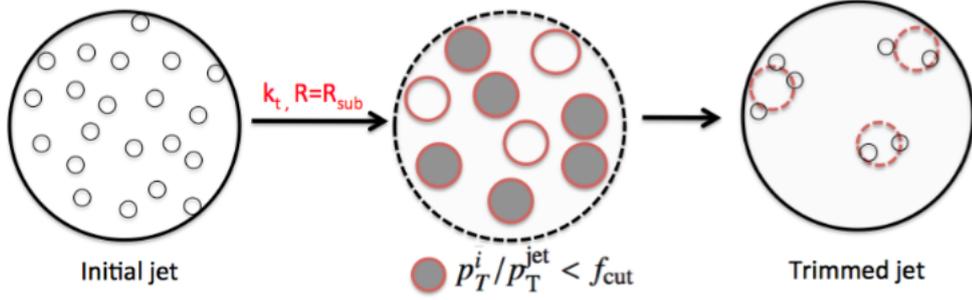
Figure 4.15: A scheme depicting the jet trimming procedure. [96]

Mass-drop Filtering

This procedure identifies relatively symmetric sub jets, each with a mass significantly smaller than the one of the whole fat jet; only the three most energetic pseudo-jets are conserved. This technique was developed to be used with C/A jets reconstruction. It is applied only to C/A jets since each clustering step of the algorithm combines the two widest angle proto-jets at that point in the shower history. Therefore, the structure of the C/A jet provides an angular-ordered description of substructure, which tends to be one of the most useful properties when searching for hard splittings within a jet.

Pruning

The pruning algorithm [102] is similar to trimming because it removes soft constituents from the large-R jet, but it adds a wide-angle radiation veto. The constituents of large-R jet are used to reconstruct again the jet, using either a C/A or k_t algorithm; at each pseudo-jet recombination step the following pruning cuts are placed

$$
\frac{p_{\mathrm{T}}^{j_2}}{p_{\mathrm{T}}^{j_1+j_2}} > z_{cut}
$$

$$
\Delta R_{j_1,j_2} < R_{cut} \cdot \frac{2m^{jet}}{p_{\mathrm{T}}^{jet}}
$$

(4.5.15)

where j_1, j_2 are the pseudo-jets considered in the current step ordered $p_{\mathrm{T}}^{j_1} > p_{\mathrm{T}}^{j_2}$, R_{cut} and z_{cut} are parameters of the tagger (usually fixed to 1.0 and 0.15 respectively). It is important to remark that these requirements are not directly related to the original large-R jet but to the pseudo-jets formed in the new reconstruction process.

4.6 Top Tagging algorithm based on substructure variables

In Run-2 a larger fraction of top-pair production ($t\bar{t}$) events is expected to produce boosted top quarks (see figure 4.10), and consequently top-tagging will be even more important. This section presents a simple top-tagging algorithm used in this analysis. The presented algorithm is intended to have strong and reliable performance in different event topologies, for top-quark transverse momentum higher than 200 GeV. An algorithm using only two jet substructure variables (the jet mass and τ_{32}) has been chosen. To verify the performance of the tagger, signal jets are considered to arise from hadronically decaying

(a)

(b)

(c)
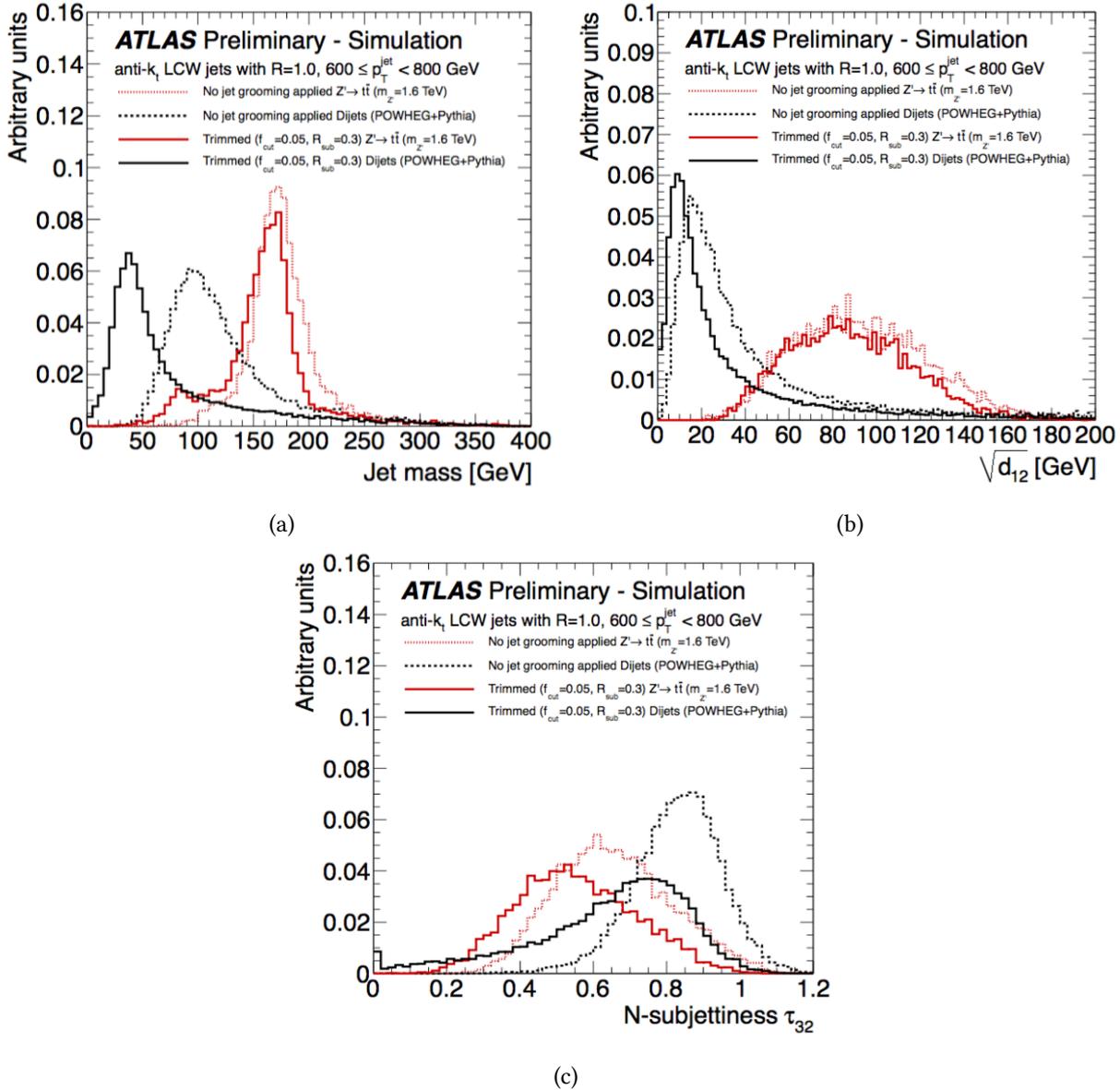
Figure 4.16: (a) Leading-p_T jet mass, (b) splitting scale $\sqrt{d_{12}}$ and (c) N-subjettiness comparing $Z' \to t\bar{t}$ ($m_{Z'}$ = 1.6 TeV) signal to POWHEG multi-jet background for jets in the range $600 \leqslant p_T^{large-Rjet} < 800$ GeV. The dotted (solid) lines show the leading-p_T jet distribution without (with) the application of the trimming algorithm (f_{cut} = 0.05, R_{sub} = 0.3). The distributions after the application of the trimming algorithm are normalized with respect to the ungroomed distributions, which are themselves normalized to unity. [96]

top quarks. Using simulated $t\bar{t}$ and $Z' \to t\bar{t}$ events and background jets from simulated QCD multi-jet production events the performance of the algorithm in the busier event topology characteristic of top-pair associated Higgs boson production is also studied [103].

4.6.1   Algorithm design

The aim is to build an algorithm capable to tag the large-R jet arising from a top quark, using its substructure information, described in section 4.5.1.

A set of substructure variables, including the calibrated m_{jet}, N-subjettiness ratios (τ_{21}, τ_{32}), splitting scale variables ($\sqrt{d_{12}}$, $\sqrt{d_{23}}$) and the minimum dijet mass from the three pseudo-jets (Q_w) [104], has been used. Two working points (50% and 80%) have been chosen in order to study pairs of variables and to chose the one providing the best performances, in terms of background rejection and correlation. Figures 4.17a and 4.17b show the background rejection of pairs of variables for both the working points, for two different p_T ranges. It can be seen that the promising pairs are τ_{32}-m_{jet}, τ_{32}-$\sqrt{d_{12}}$, τ_{32}-$\sqrt{d_{23}}$ and Q_w-τ_{32}.

Because of high p_T jets are more correlated, some substructure variables have different performances in the low and high p_T regions. The correlation between pairs of variables is given in figures 4.18a and 4.18b, it can be seen that the τ_{ij} variables are uncorrelated with respect to the mass and energy scale variables.

The choice of tagging variables is motivated by their lack of correlated behaviour, the strong performance of this combination and its robustness across the p_T range under study. For these reasons, the two chosen variables are m_{jet} and τ_{32}.

Two working points are provided: 50% signal efficiency, for analyses with very large reducible background (i.e. processes that do not include hadronically decaying top quarks), and 80% signal efficiency, for analyses dominated by signal and/or irreducible background (i.e. processes containing hadronically decaying top quarks). Optimisation was performed by testing all possible thresholds on τ_{32} and m_{jet} in bins of jet p_T, followed by a regularised interpolation between the points to obtain jet the requirements at a given p_T. For the working point at 50% (80%) efficiency, the upper allowed value of τ_{32} varies from 0.75 (0.85) at p_T = 200 GeV to 0.57 (0.7) for $p_\mathrm{T} \geqslant$ 1600 GeV. The lower threshold on m_{jet} varies from 85 GeV (70 GeV) at p_T = 200 GeV to 140 GeV (135 GeV) for $p_\mathrm{T} \geqslant$ 1600 GeV.

4.6.2   Algorithm performance

The performance of the top-tagging algorithm is quantified by evaluating the efficiency for jets arising from the decay products of top quarks and rejecting other jets. The signal efficiency is defined as the fraction of reconstructed jets in $Z' \rightarrow t\bar{t}$ simulated samples, matching an hadronic top and passing the top-tagging algorithm. The background rejection is the reciprocal of the fraction of jets from the simulated QCD dijet sample, that pass the top-tagging algorithm.

The jet matching criterion requires that the reconstructed jet lies within $\Delta R < 0.75$ of a hadronically decaying top quark with transverse momentum larger than 200 GeV.

The signal efficiency and background rejection are shown in figures 4.19a and 4.19b. Both the 50% and 80% efficiency working points present a flat efficiency for p_T above 400 GeV, where the majority of top-quarks are fully contained within the large-R jet. The efficiency rises in the 200 GeV $< p_\mathrm{T} <$ 400 GeV region as the fraction of fully contained top-quarks increases. The 50% (80%) working point reduces the background by at least a factor of 6 (2) across the full p_T range. For both working points, rejection deteriorates at high p_T as the background mass peak shifts towards higher values.
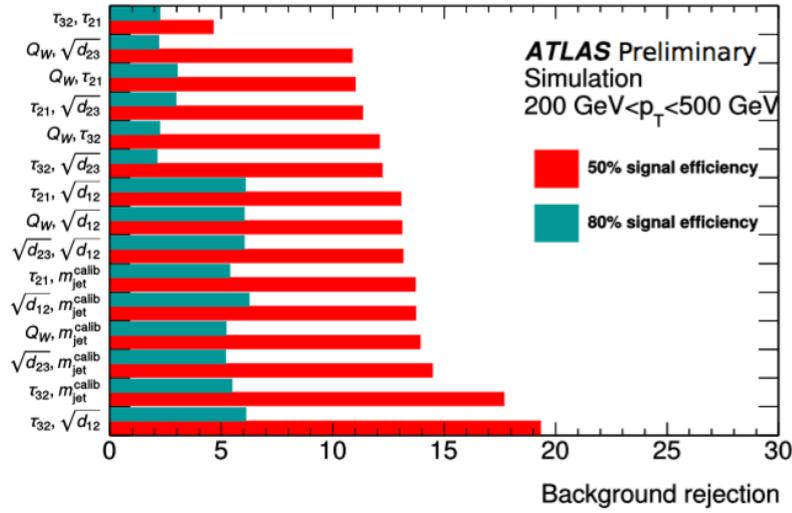
(a)
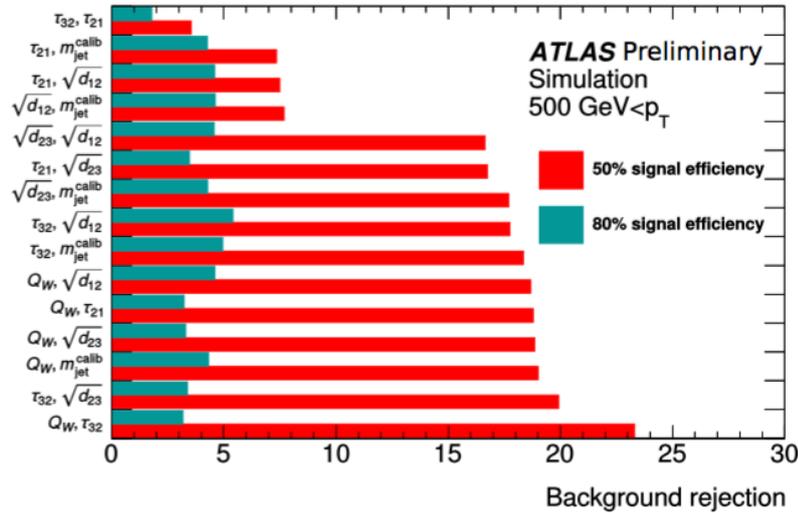
(b)

Figure 4.17: Background obtained for two signal efficiency points (50% and 80%), for two p_{T} regions. The statistical uncertainties are around 1 unit. [103]

4.6.3   Application in the associated production of Higgs boson with a top quark pair

With respect to $Z' \rightarrow t\bar{t}$ or SM $t\bar{t}$ processes described in the previous paragraphs, the $t\bar{t}H$ process leads to a particularly busy final state, due to the presence of one charged lepton, one neutrino, four b quarks and two additional quarks (from the hadronic W decay). The presence of the Higgs boson decay products can affect the reconstruction of the large-R jet corresponding to the top-quark decay, resulting in a large-R jet containing a mixture of contributions from the hadronically decaying top and from the Higgs boson. This has an impact on the reconstructed p_{T} and on the substructure variables, that could reduce the efficiency of the proposed top-tagging algorithm to identify reconstructed large-R jets induced by hadronically
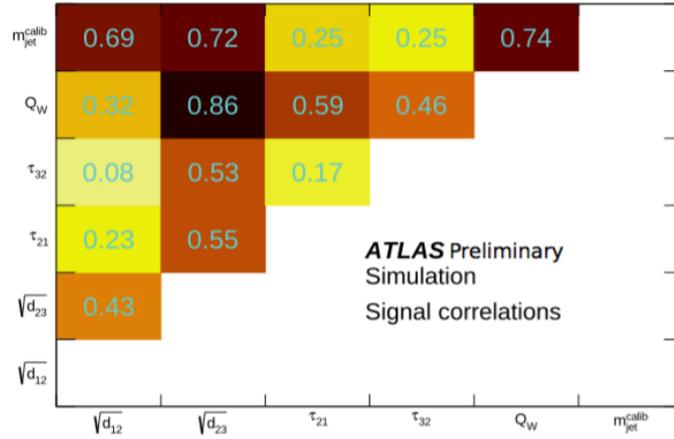
(a)
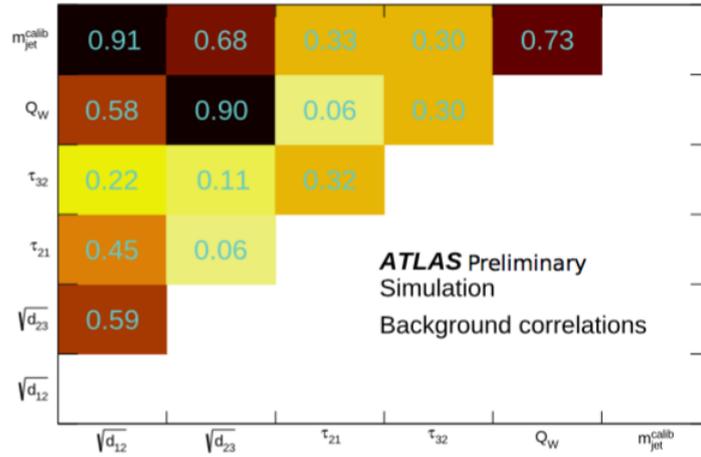
(b)

Figure 4.18: Correlation matrices for the different studied variables, for signal (left) and background (right) events. [103]

decaying top quarks.

In order to verify if the results obtained for the $t\bar{t}$ events (see paragraph 4.6.1) are consistent with those of the $t\bar{t}H$ one, the ratio of the mass and the τ_{32} has been evaluated generating both $t\bar{t}H$ and $t\bar{t}$ events. As shown in figures 4.20a and 4.20b, the mass response is similar in $t\bar{t}H$ and $t\bar{t}$, while, for the τ_{32} response, discrepancies are observed for low p_T due to the dependence of the τ_{32} response on the jet mass and p_T distributions within the samples. The top-tagging efficiencies, using a generator-level-matching identical to that used for $t\bar{t}$, was also produced for $t\bar{t}H$ events, and the ratio of those efficiencies with respect to the ones from $t\bar{t}$ events is given in figures 4.21a and 4.21b, for the two working points. For both working points the $t\bar{t}H$ efficiency is higher in the low p_T region and become consistent with the performance in $t\bar{t}$ events for higher p_T. This efficiency increase is also present when generator-level jet p_T is used and is more visible in jets matched with top quark for which the Higgs boson is geometrically close. The large-R jet selection performed by the top-tagging algorithm favours higher mass and characteristics of a more multi-prong substructure. The contamination from the Higgs boson decay on the large-R jets in $t\bar{t}H$
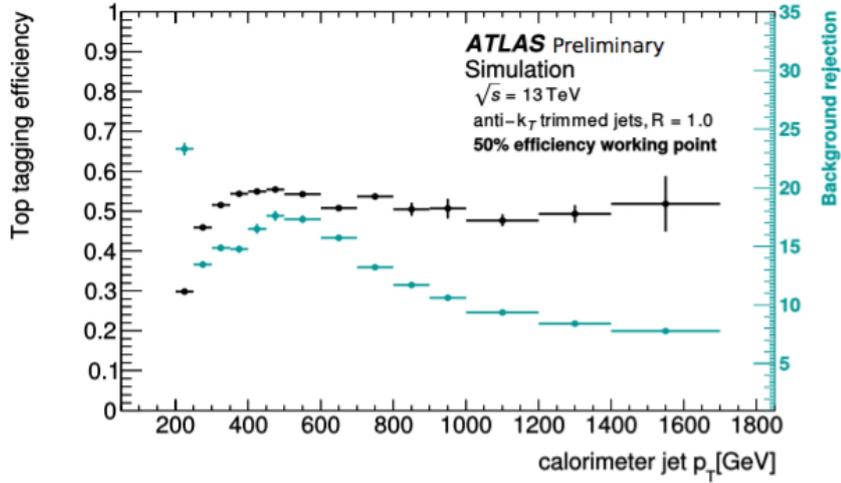
(a)
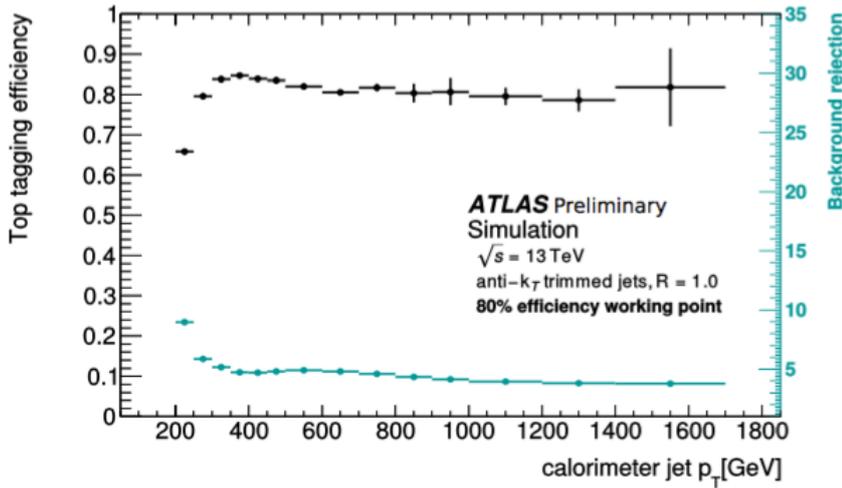
(b)

Figure 4.19: Efficiency and rejection of the top-tagging algorithm versus jet transverse momentum p_T for simulated $Z' \to t\bar{t}$ events and for simulated QCD dijet production: (a) 50% efficiency working point and (b) 80% efficiency working point. [103]

shifts jets in the favoured directions for these variables and therefore leads to an increase in the tagging efficiency. This shows that the capacity of the proposed top-tagger to identify a large-R jet induced by hadronically decaying top quark is not decreased when the reconstructed large-R is contaminated by other decay products in a busy environment.

4.7   Boson Tagging algorithm

One of the goal of the Run-2 analyses is to have a well-understood method for tagging boosted bosons. The studies undertaken with the Run-1 [105] were extended to explore the relative efficiency and background rejection of several variations on the pruning, trimming and mass-drop filtering techniques [106]. The

(a)   (b)
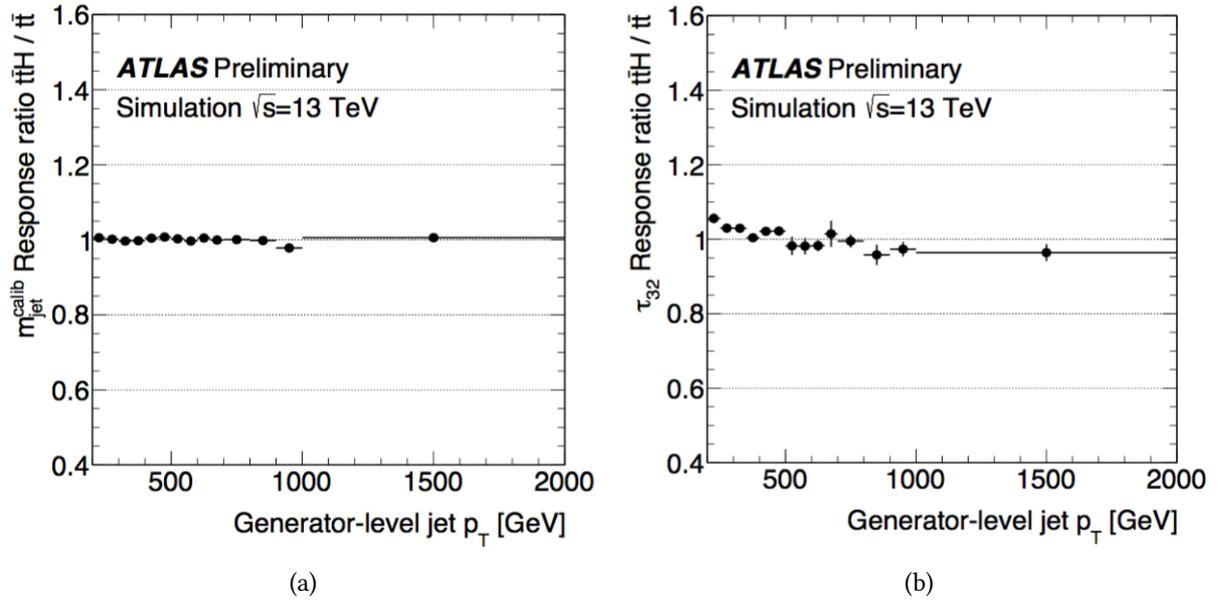
Figure 4.20: Ratio of the jet response for $t\bar{t}H$ and $t\bar{t}$ events, for (a) m_{jet}^{calib} and (b) τ_{32}. The error bars correspond to jet statistical uncertainties. [103]

(a)   (b)
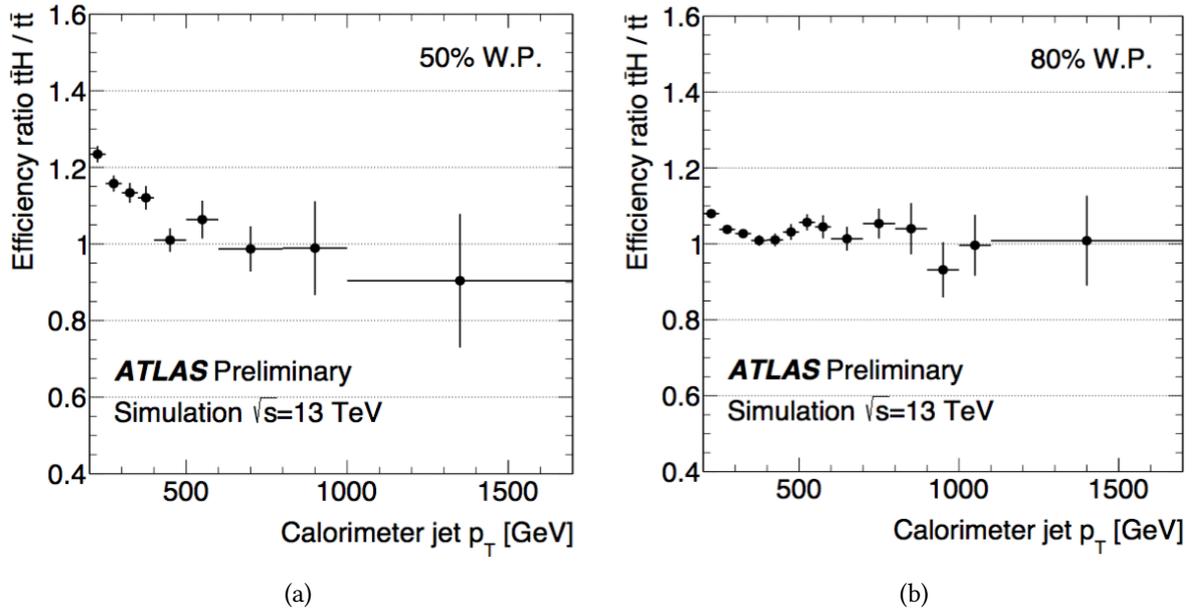
Figure 4.21: Ratio of the top-tagging efficiency for $t\bar{t}H$ and $t\bar{t}$ events, for (a) the 50% working point and (b) the 80% working point. The error bars correspond to statistical uncertainties. [103]

preliminary conclusions from the Run-1 studies provide a subset of substructure variables and jet grooming algorithm configurations, found to be efficient to identify W-jets in the hadronic decay(from both exotic high-mass resonances and Standard Model $t\bar{t}$ production) while rejecting multi-jet and combinatorial backgrounds.

The W boson tagging algorithm in the hadronic decay is a future improvement of the analysis sensi-

tivity, being used in two possible ways: both to tag the W boson arising from the hadronic top, in addition to the top-tagging, and as a starting point for the Higgs tagging (decaying in a pair of b-quarks), after the application of appropriate approximations.

4.7.1 W boson tagging algorithm and performances

The tagging algorithm of the boosted W decaying in the hadronic channel [107] has been built in order to tag the large-R jet arising from the W bosons, using its substructure information.

The simulated signal sample of W has been obtained from the hypothetical process $W^{'} \rightarrow WZ$, where both the W and Z bosons decay in a pair of jets. This sample is combined and weighted such that the p_{T} distribution of the leading (highest p_{T}) jet matches that of the background sample consisting of high-p_{T} multi-jets initiated by light quarks and gluons.

Three jet reconstruction algorithms, in four different configurations for the radiation removal, have been considered and summarized here:

- jet reconstructed with anti-k_t, R = 1.0, and radiation removal with trimming (R_{sub} = 0.2, f_{cut} = 5%);

- jet reconstructed with C/A, R = 1.0, and radiation removal with pruning (z_{cut} = 15%, $R_{cut} = \frac{1}{2}$);

- jet reconstructed with C/A, R = 1.2, and radiation removal with mass-drop filter (R_{sub} = 0.3, $y_{cut} >$ 15%);

- jet reconstructed with C/A, R = 1.2, and radiation removal with mass-drop filter (R_{sub} = 0.3, $y_{cut} >$ 4%).

For each configuration the optimization procedure has been tested on the m_{jet}, $D_2^{\beta=1}$, τ_{21} and C_2 (the ratio between the two and three point energy correlation functions) variables.

Two working points are provided: 25% ("tight") and 50% ("medium") signal efficiency, depending on the $D_2^{\beta=1}$ cut, optimized on the p_{T} of the particle, while the mass window is fixed to $m_W \pm 15$ GeV. The multi-jet background rejections achievable for the "medium" signal efficiency working points $\epsilon_W^{G\&T}$ = 50% are shown in figures 4.22a and 4.22b, for the leading reconstructed jets in events with $200 < p_{\mathrm{T}}^{truth} < 350$ GeV and $1500 < p_{\mathrm{T}}^{truth} < 2000$ GeV respectively; the figures show that independently from the different methods to reconstruct the large-R jet and to remove the pile-up contribution, the better background rejection is obtained by the contemporary use of the m_{jet} and $D_2^{\beta=1}$ variables. These are the two chosen variables for the tagger, tested to be the more appropriate from a wider set.

The performance of the W-tagging algorithm is quantified by evaluating the tagging efficiency on large-R jets arising from the decay products of W bosons and rejecting other jets. As usual, the background rejection has been defined as the reciprocal of the fraction of jets from the simulated multi+jets sample, that pass the W-tagging algorithm. The jet matching criterion requires that the reconstructed leading jet lies within $\Delta R < 0.75$ of a hadronically decaying W boson. The efficiency versus rejection curves for W-jets versus multi-jets are shown in figures 4.23a and 4.23b for the leading reconstructed jet in events with $200 < p_{\mathrm{T}}^{Truth} < 350$ GeV and $1500 < p_{\mathrm{T}}^{Truth} < 2000$ GeV respectively.

4.7.2 Higgs boson tagging algorithm and performances

The Higgs-tagging algorithm is inspired to the boosted W boson tagger and uses the same set of variables: m_{jet} (calibrated as described in section 4.5.1), $D_2^{\beta=1}$, C_2, $\sqrt{d_{12}}$ and τ_{21}. Figures from 4.24a to 4.24e show the variable distributions for both signal and background processes.

(a)

(b)
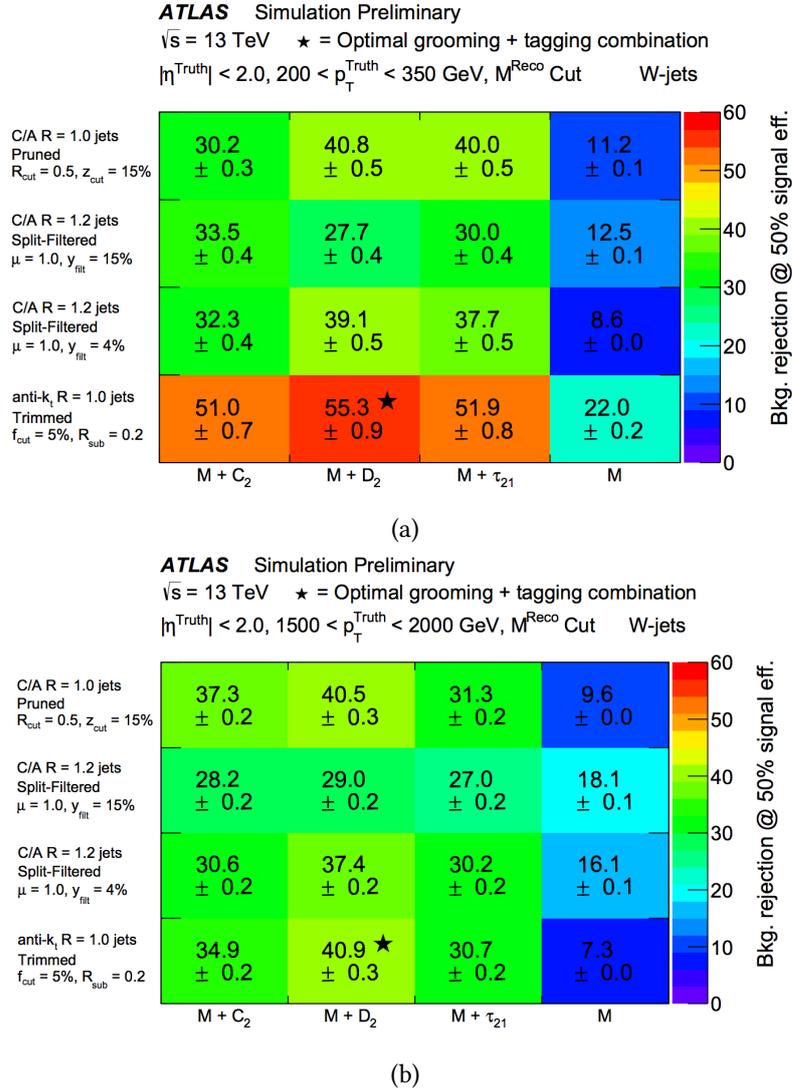
Figure 4.22: Matrix showing the background rejections achievable when tagging W-jets with $\epsilon_W^{G\&T}$ = 50% for different combinations of a mass window criteria on the groomed jet and a substructure variable requirement, and for mass-only, in events with the leading ungroomed C/A jet in the range $200 < p_T^{truth} < 350$ GeV (a) and $1500 < p_T^{truth} < 2000$ GeV (b). The uncertainties are statistical errors only. [107]

The configuration used for the jet reconstruction algorithm is: anti-k_t, R = 1.0, trimmed with R_{sub} = 0.2, f_{cut} = 5% and a truth matching between the large-R jets (with $p_T > 200$ GeV) and the truth hadronically decaying Higgs parton ($H \rightarrow b\bar{b}$) has been required, with a $\Delta R < 0.75$. The optimization is the same as for the top tagging algorithm: it is a two-dimensional scan testing all possible thresholds of a pair of variables in bins of jet p_T, followed by a regularised interpolation between the points to obtain the given WP.

The simulated signal sample has been obtained from the $t\bar{t}H$ process, where the $H \rightarrow b\bar{b}$ and the $t\bar{t}$ system decays semi-leptonically, while the background sample derives from $t\bar{t}$+jets events. This sample, as for the W tagging algorithm, is combined and weighted such that the p_T distribution of the leading jet matches that of the backgroung sample $t\bar{t}$+jets.
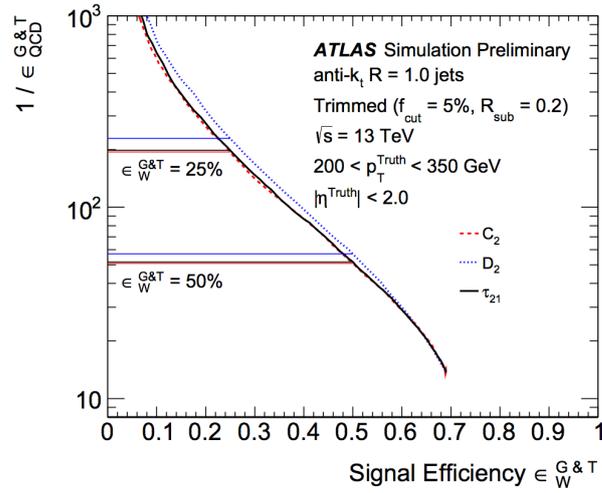
(a)
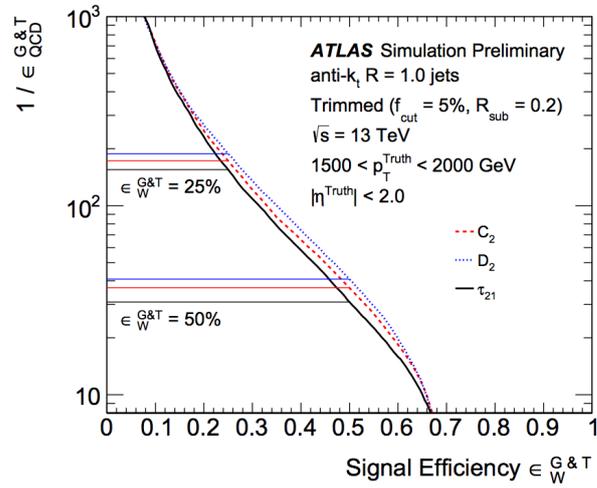
(b)

Figure 4.23: The efficiency versus rejection curves for W-jets versus multi-jets for the leading reconstructed jet in events with (a) $200 < p_{\mathrm{T}}^{Truth} < 350$ GeV and (b) $1500 < p_{\mathrm{T}}^{Truth} < 2000$ GeV. [107]

It has been chosen an efficiency working point of 50% and a tagger composed by two variables, selecting the pair that gives the best background rejection with the lowest correlation. Figures 4.25a and 4.25b show the correlations between all the variables for both the background and signal samples.

As usual, the signal efficiency is defined as the fraction of reconstructed large-R jets in $t\bar{t}H$ simulated samples, matching a hadronic Higgs ($H \to b\bar{b}$) and passing the Higgs-tagging algorithm, while the background rejection is the reciprocal of the fraction of jets from the simulated $t\bar{t}$+jets sample, that pass the Higgs-tagging algorithm. The signal efficiencies and background rejections are shown in figures 4.26a and 4.26b respectively, for all the possible pairs of variables.

Among the different pairs of variables, the performances are not so different ranging up to a maximum of a factor two, but only at high p_T values. Some pairs show the best signal efficiency (up to 80%): m_{jet}-D_2 and m_{jet}-τ_{21}, that have also a very similar background rejection, and m_{jet}-$\sqrt{d_{12}}$, but they are strongly correlated, as shown in figures 4.25a and 4.25b, and have also the lowest background rejection.

Due to the high efficiency at high p_{T}, the good background rejection and no strong correlation, the m_{jet}-D_2 and m_{jet}-τ_{21} pairs will be considered for the application of the Higgs-tagging into the analysis and their efficiencies and rejections are shown singularly in figures 4.27a and 4.27b.

At present, the Higgs tagging can be considered an interesting extension of the requests that can be applied to the analysis samples in order to select Higgs-rich events. In the following, due to the low overall efficiency, this tagging will not be applied and the event selection will be optimized for the top tagging alone. In the future, the cut flow and event selection will be optimized for the contemporary requests of the Higgs and top taggings.
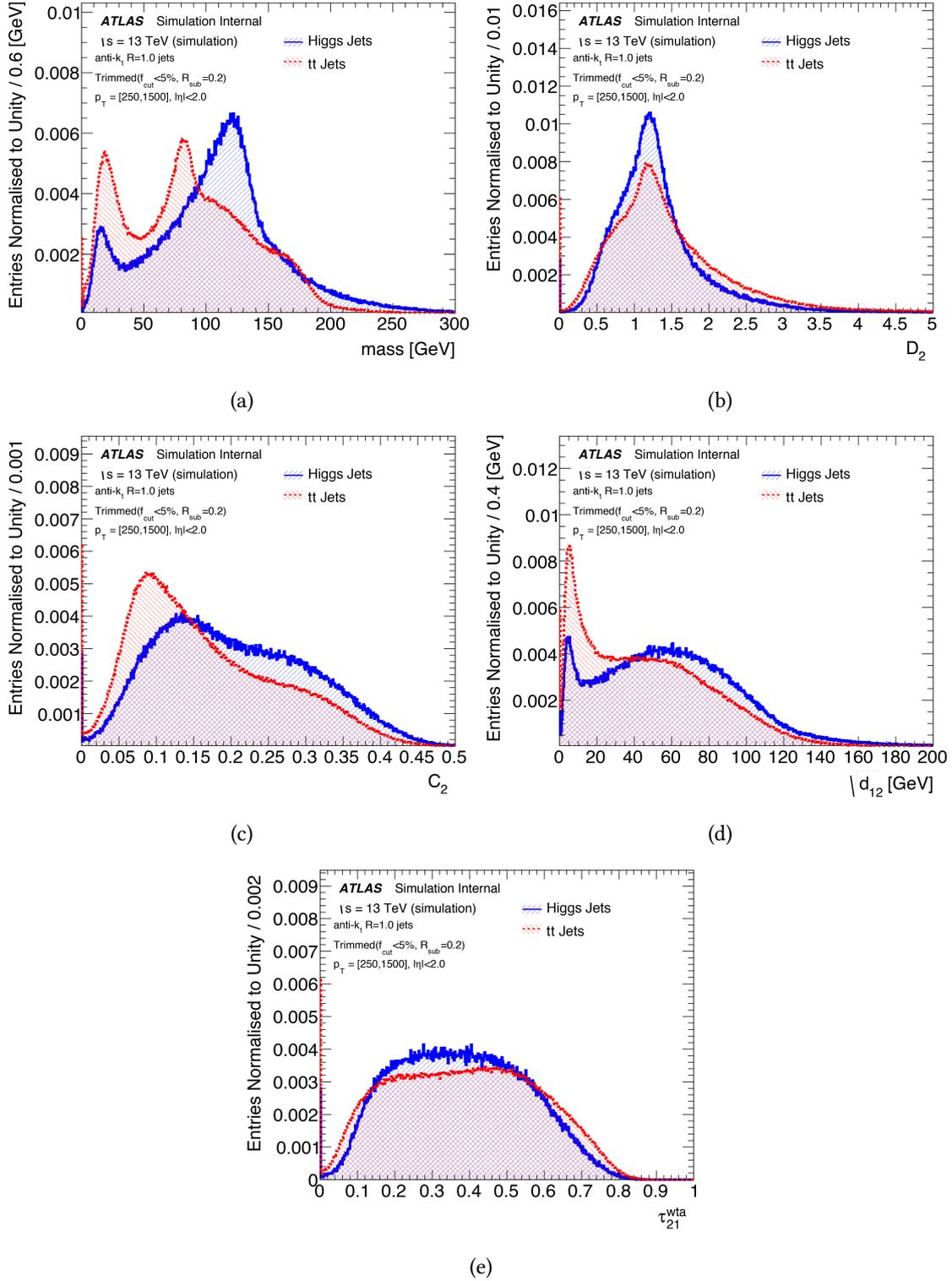
Figure 4.24: Distributions of the substructure variables studied for the Higgs-tagging alrogithm for both signal and background. The large-R jets considered in these distributions have been required to have p_{T} = [250,1500] GeV and $|\eta|$ < 2.0.
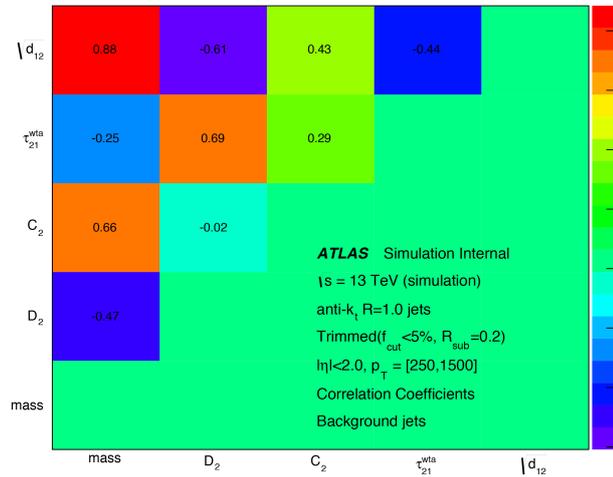
(a)
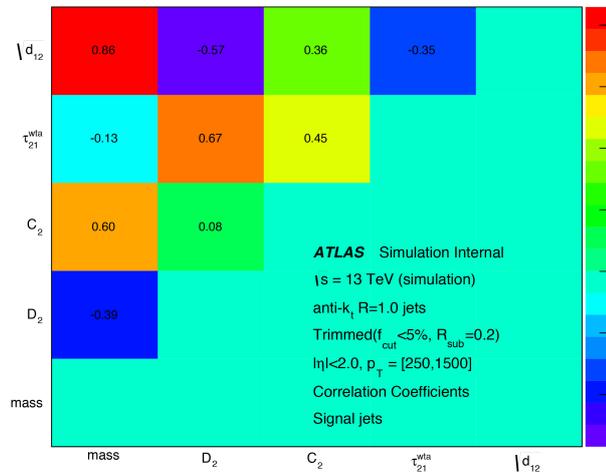
(b)

Figure 4.25: Matrices of correlations (a) in the background and (b) in the signal sample, between the substructure variables studied for the Higgs-tagging alrogithm.
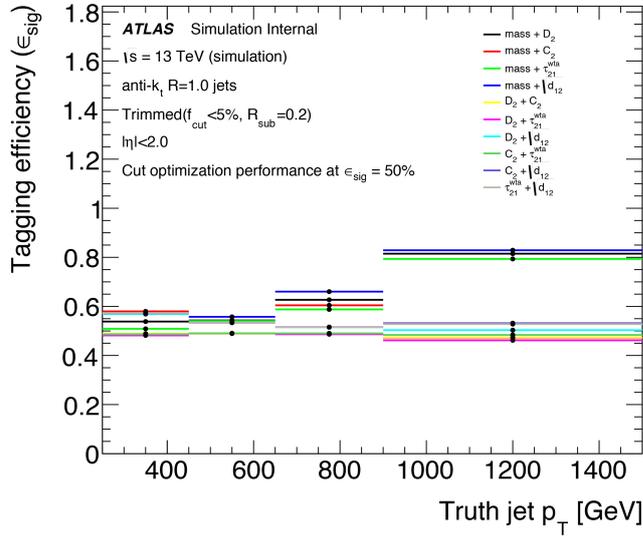
(a)
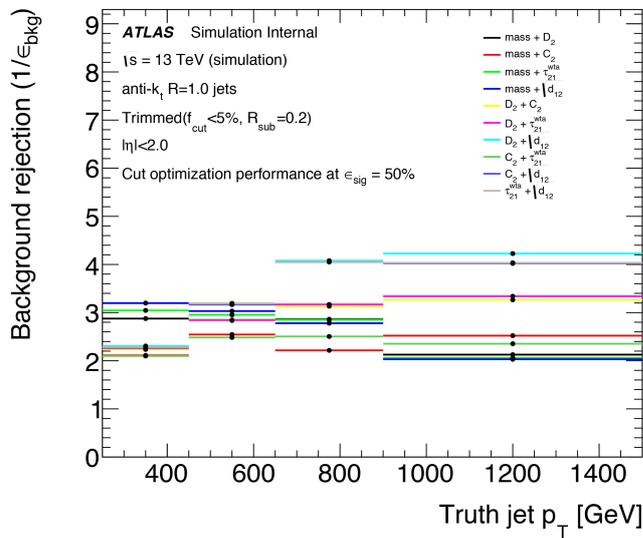
(b)

Figure 4.26: (a) Signal efficiency and (b) background rejection as a function of the truth jet p_{T}, for the 50% WP. All the possible pairs of variables have been used.

(a)   (b)
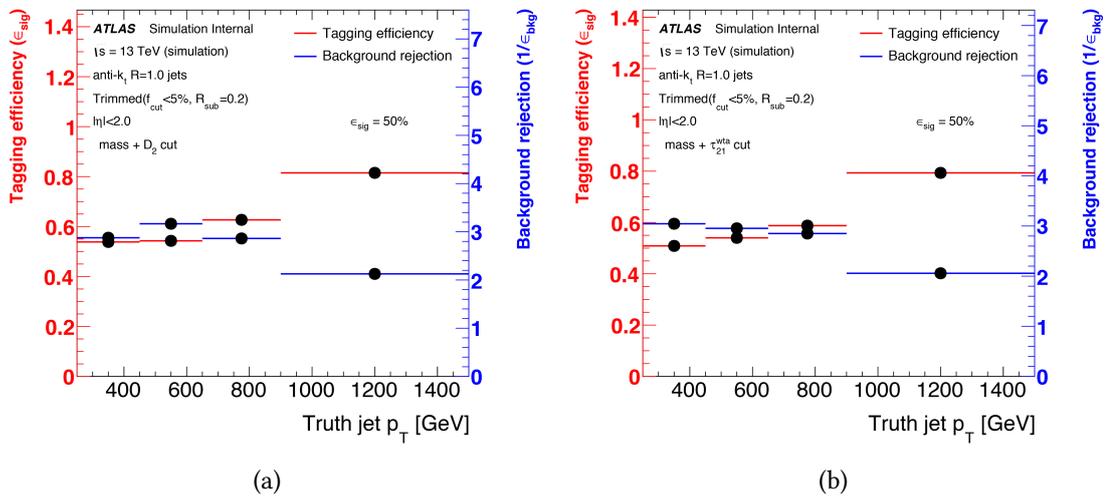
Figure 4.27: Signal efficiency and background rejection of the pairs m_{jet}-D_2 (a) and m_{jet}-τ_{21} (b), for 50% WP. The performances are very similar, the main difference in terms of background rejection is in the first p_T bin.

Chapter 5

Analysis strategy and techniques

The final goal of the analysis described in this thesis is the ratio ("μ") of the $t\bar{t}H$ cross section measurement with respect to the one predicted by the SM. The analysis exploits the full 2015 and 2016 data sets collected at $\sqrt{s}$ = 13 TeV, with an integrated luminosity of 3.2 fb^{-1}and 33.3 fb^{-1}, respectively. At this energy, the top or/and the Higgs could be produced in a resolved (p_{T} < 250 GeV) or boosted (p_{T} > 250 GeV) regime; in order to exploit the largest p_{T} range of the $t\bar{t}H$ final state a combined fit of the resolved and boosted analyses (see section 5.4.2) is performed avoiding obviously double counting events (see section 5.1.2).

The analysis foresees the definition of the exclusive kinematic regions based on the number of jets and b-tagged jets, characterized by events containing only background ("control regions") or background plus signal ("signal regions") accordingly. These regions have been exploited by a likelihood fit to evaluate μ. Both the analyses have been optimized to be sensitive to the Higgs boson decay to a pair of b quarks and to the semileptonic decay of the $t\bar{t}$ pair.

Since in the boosted regime the top quarks and/or Higgs boson decay products are collimated, the event signatures cannot be efficiently identified with the standard algorithms used in the resolved analysis. For this reason in the boosted analysis, substructure observables and specific tagging algorithms will be used, as described in Chapter 4.

5.1 Event Selection

To extract the signal from the huge background environment, particular care has been taken in the event selection, which is based on standard trigger and object requirements already described in Chapter 3 and on additional requests that will be described in this chapter. The specific event selection for the boosted and resolved regimes will be described separately in the following.

5.1.1   Resolved analysis

In this regime, a typical $t\bar{t}H$ event is composed by one lepton, missing energy, due to the presence of one neutrino, and six standard jets (reconstructed with the anti-k_t algorithm with R = 0.4) at least four of which b-tagged (see figure 1.26a). The events have been recorded using a logical OR combination of a single-lepton trigger with low p_T threshold and isolation requirements and one with higher p_T threshold but without isolation. For muons, the lowest p_T threshold was 20 (24) GeV in 2015 (2016), while the higher one was 40 (50) GeV. For electrons, isolated triggers with a p_T threshold of 24 GeV and non-isolated triggers with p_T threshold of 60 GeV in both years are used together with a 120 (140) GeV trigger which also uses looser identification criteria (see Chapter 3).

The event selection further requires the following cuts:

- events must belong to the so called good run list of events acquired when all detectors work properly;

- exactly one good reconstructed electron/muon with $p_T > 27$ GeV matched with the triggered one;

- at least four good reconstructed standard jets and no bad reconstructed jets with $p_T > 25$ GeV and $|\eta| < 2.5$;

- at least two b-tagged jets.

Jets are tagged as "good jets" if they pass some general quality criteria (avoiding events with hardware problems, cosmic rays, beam-gas interactions, and so on) explained in Chapter 4.

In order to take advantage of the higher jet and b-jet multiplicity of the $t\bar{t}H$ signal process, the events are classified into exclusive "regions" based on the number of jets and the number of b-tagged jets. A region with m jets and n b-tagged jets is labelled as (mj, nb). The regions where the signal-to-background ratio (S/B) is larger than 1% and $S/\sqrt{B}$ is larger than 0.3 are referred to as "signal regions". In these regions a two-stage multivariate technique is used to separate the signal from the background (see section 5.2). The remaining regions are considered as "control regions": no further attempt is made to separate the signal from the background. They allow the evaluation of the background and systematic parameters included in a combined fit with the signal regions.

Events are divided into four, five, or at least six jets, and two, three, or at least four b-tagged jets, as illustrated in figures 5.3a and 5.3b; the signal regions result to be (5j, $\geq$ 4b), ($\geq$ 6j, 3b), and ($\geq$ 6j, $\geq$ 4b). The background is dominated by $t\bar{t}$+light jets events in the majority of the control regions, while $t\bar{t}$+heavy-flavour jets are especially important in the signal regions. Figure 5.4 shows the number of events of real data compared to the background prediction in each of these regions, including also the boosted one, described in the next section.

5.1.2   Boosted analysis

In this p_T regime a typical $t\bar{t}H$ event is composed by one lepton, missing energy, at least one large-R jet (reconstructed with the anti-k_t algorithm with R = 1) and at least three jets, at least two of which coming from b quarks (see figure 1.26b).

Events in the boosted regime have been recorded using the same single-lepton triggers of the resolved analysis, described in section 5.1.1.

The event selection further requires the following cuts:

- events must belong to the so called good run list of events acquired when all detectors work properly;

- exactly one good reconstructed electron/muon with $p_T > 27$ GeV matched with the triggered one;

- at least one large-R jet top-tagged at the 80% working point (see section 4.6) with $p_T > 250$ GeV and not overlapped to an electron;

- at least three good reconstructed jets and none bad reconstructed jets in the events, with $p_T > 25$ GeV and $|\eta| < 2.5$ and not overlapped to any top-tagged jet in the event;

- at least two b-tagged jets which do not overlap to any top-tagged large-R jet.

The overlap between the large-R jet and the jets or the electron has been avoided by requiring that each of these objects satisfy the condition $\Delta R > 1.0$ with the axis of all top-tagged jets in the event. The overlap removal between the large-R jet and the electron has been required because in the electron channel the fraction of events with two top-tagged large-R jets (one of which is due to a bakground contamination) is higher (3.4%, see figure 5.1a) than the one in the muon channel (1.7%, see figure 5.1b). As shown in figures 5.2a the electron direction is both opposite to the large-R jet (the peak at π) as expected and superimposed to it (the peak at 0) giving the possibility to have a false top-tagging due to the presence of the electron cluster. This possible false top-tagging is not present in the muon channel (figure 5.2b) where the muon and the top-tagged large-R jets are mainly back-to-back. Requiring the overlap removal, the analysis sensitivity increases by 20%.

(a)   (b)
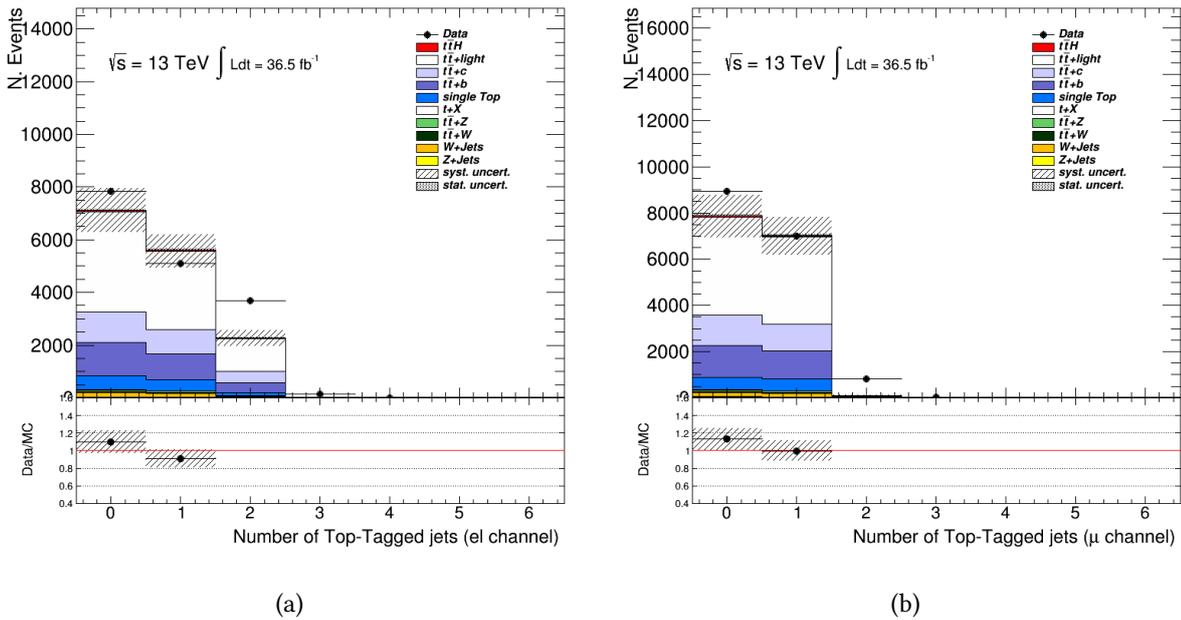
Figure 5.1: Comparison of the distribution of top-tagged jets multiplicity in electron (a) and muon (b) channels. The selection used in this study is the baseline one used in the analysis.

Differently from the resolved analysis, the boosted one has only one signal region, as illustrated in figures 5.3a and 5.3b, called 3211 because it requires at least three small-R jets (3), two b-tagged jets (2) outside any top-tagged large-R jet, at least one top-tagged large-R jet (1) and exactly one lepton (1). This signal region has been chosen as a compromise between the involved statistics, the minimum superimposition with the resolved selection and the maximum S/B ratio. In this region, the background is dominated by $t\bar{t}$+light jets events, followed by the $t\bar{t}$+heavy-flavour jets events. The number of events of real data
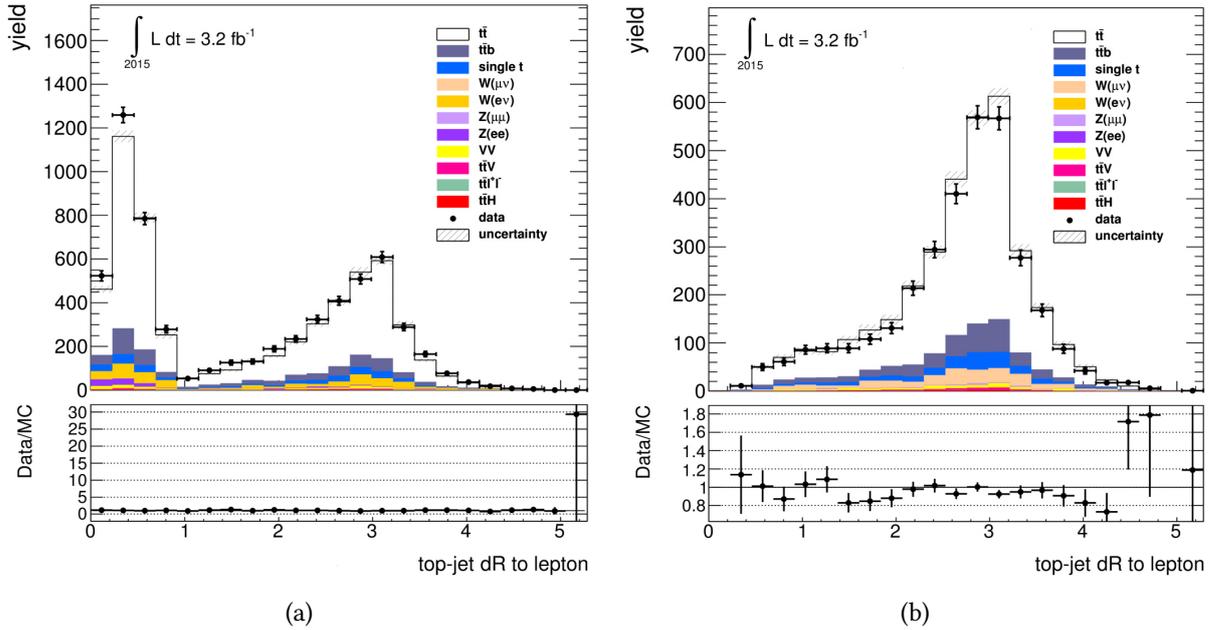
Figure 5.2: Comparison of the distribution of ΔR between top-tagged jets and the electron (a) and (b) muon in the event. The selection used in this study requires at least one b-tagged jet, instead of two as the boosted signal region requires.

compared to the background prediction in the boosted region, together with all the resolved regions, is shown in figure 5.4.

5.2 MultiVariate Analysis technique

The most challenging task of this analysis is to identify events that are both rare and overwhelmed by a wide variety of processes that mimic the signal and for which the conventional approach by using cuts on individual kinematic variables can be far from optimal. As the identification process of the signal on the real data becomes more challenging, the multivariate techniques can be used with increasing success and efficiency with respect to cut-based techniques.

The variables characterizing an object, or an event, can be represented by a n-dimensional vector x, referred to as *feature variable* (or *n-dimensional feature space*). An example can be the four-vectors of particles, the energy deposited in calorimeter cells, the deduced kinematic quantities of the physics object and the global event characteristics. Generally these *feature variables* are correlated then it is necessary to treat them in a fully multivariate way to extract results with maximum precision. When correlations exist the effective dimensionality of the problem is smaller than n.

The reason to apply MultiVariate Analysis (MVA) methods is, in most cases, the lack of knowledge about the mathematical dependence of the quantity of interest on the relevant measured variables. Either there is no mathematical model at all and an exhaustive search is the only possibility of finding the correct dependence, or the known models are insufficient and statistical training provides a better description of data. A typical list of problems addressed by a multivariate analysis technique is:

- signal-to-background discrimination;
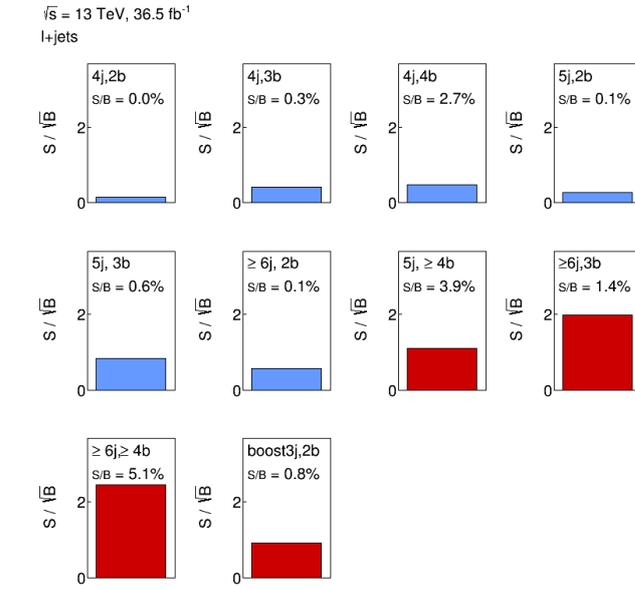
(a)
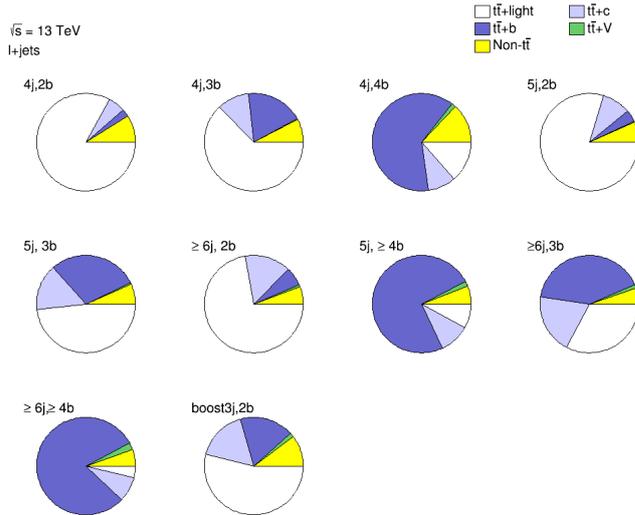
(b)

Figure 5.3: (a) The S/B and $S/\sqrt{B}$ ratios for the regions are shown. Signal regions are shaded in red, while the control regions are shown in blue. (b) The fractional contributions of the various backgrounds to the total background prediction in the region is shown. The $t\bar{t}$ background is classified as described in Chapter 3.

- variable selection (e.g., finding variables with the maximum signal/background discrimination);

- dimensionality reduction of the multiple quantities characterizing an object or an event (called feature space) and simplification (by reducing the number of variables);

- finding regions of interest in data.

An MVA analysis is based on three fundamental steps: *training*, *testing* and *classification*.
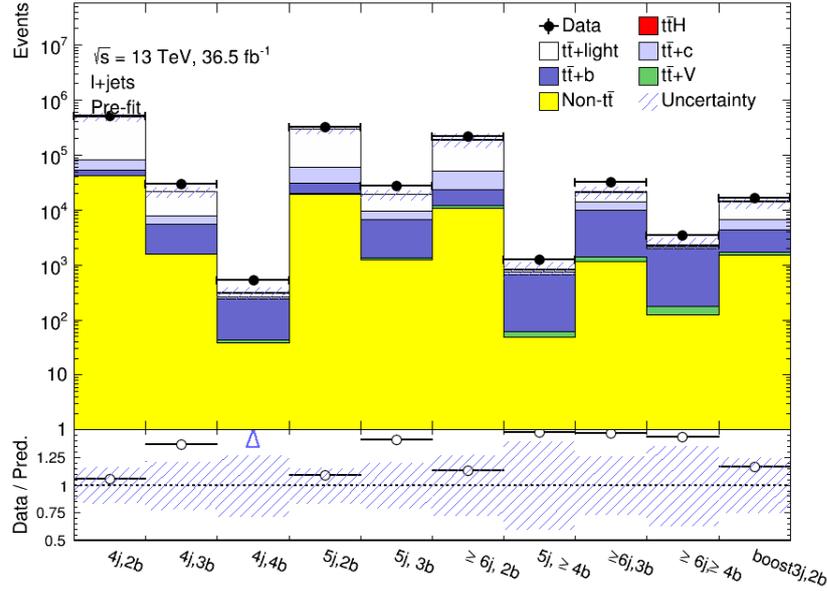
Figure 5.4: Comparison of predicted and observed event yields in the resolved and boosted single-lepton channels. The $t\bar{t}$ background is divided as described in Chapter 3, while the "non-$t\bar{t}$" background category includes contributions from single top, W/Z+jets, diboson and fakes. The $t\bar{t}H$ signal is shown as a filled red area stacked on the backgrounds. The hashed area corresponds to the total uncertainty on the prediction, not including an uncertainty on the normalisation of $t\bar{t}+\geq 1b$ or $t\bar{t}+\geq 1c$.

The *training* (or *learning*) process takes as input a set of events, characterized by the feature variables, in order to define a function (*classifier*) that will be used in the *classification* step, to identify each real data event belonging to the signal or to the background category. Training algorithms have been designed to "learn from data" in order to be able to respond correctly to future data. These algorithms can be divided in two main groups:

- **supervised training:** where a set of training events with correct category association is given;

- **unsupervised training:** where no "a priori" categories are given and the algorithm has to find them by itself.

Only the supervised training will be discussed further on, because it is the method chosen for this analysis.

The classifiers, resulting from the training step, are divided in linear and non-linear:

- **Linear Classifier:** the most common classifier to discriminate signal from background events is, in the simplest case, the application of a group of rectangular cuts on selected variables. This classifier is not actually a multivariate analyser but a sequence of univariate ones, because no combination of the variables is achieved and a cut on a variable does not depend on another one. In many cases a cut-based selection is not the best option as for example illustrated in figure 5.5a where two test variables (x_1 and x_2) present a correlation; a cut-based selection (showed in figure 5.5d) produces a separation of the two variables with large overlap, while a more effective cut can be obtained with a linear combination of the two variables:

$$\alpha x_1 + \beta x_2 < \gamma, \tag{5.2.1}$$

where α, β and γ are three optimized parameters. By generalizing to N variables, a linear combi-
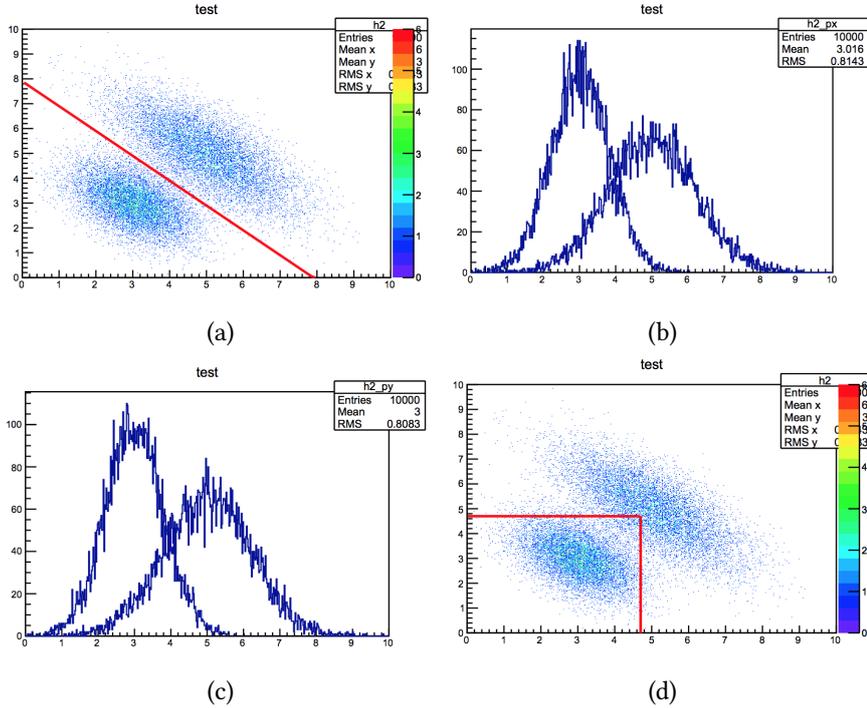
Figure 5.5: Example of 2-dimensional plot with two populations to be discriminated. (a) The scatter-plot is shown with a diagonal cut. (b)(c) The x and y projections are shown, illustrating how smaller is the discrimination capability in case of a 1-dimensional cut. Finally (d) the effect of the combination of 2 independent cuts on the same 2-dimensional plot. This example shows the benefit of the most simple multivariate selection.

nation t can be applied:

$$t = \sum_{i=1}^{N} \alpha_i x_i < t_{cut} \qquad (5.2.2)$$

where the α_i parameters have to be defined by optimizing the separation between signal and background and t_{cut} is the final cut to apply to the t variable.

- **Non-linear Classifier:** it uses a "non-linear" function, meaning that a single cut on a variable depends simultaneously on all the other variables cuts not necessarily in a linear way. The main non-linear classifiers are:

  - **Neural Networks and Multi Layer Perceptron:** the development of Artificial Neural Networks (NN) was inspired by the research on the central nervous system and the neurons (axions, dendrites and synapses) which constitute their information processing elements. Currently, the approach stimulated by biological research has been extended to an approach based on statistics, mathematics and optimization theory. Neural Networks are efficient models for statistical pattern recognition [108]. The main idea is to find a non-linear function $f : x \to y$ that relates input variables belonging to a n-dimensional space R_n with discriminating output variables in a m-dimensional space R_m. The word "network" arises because the function f is a composition of other functions g_i which can also be compositions of other functions h_j and so on. This structure is represented as a network in which each function is a node and the arrows are the dependences between functions;

– **Boosted Decision Tree (BDT):** a decision tree is a binary tree structured classifier. Since it is the method used in this analysis, it will be described in section 5.2.2.

In the second step, (*testing*), the discriminant variable distributions are obtained from other additional MC signal and background samples, statistically independent from those used in the training phase, and compared to those of the training test.

A good agreement between training and testing distributions is crucial because it assures that the definition of the discriminating variables is not due to a specific features of the training sample (for instance a statistical fluctuation). A possible inconsistency between the training and the test distributions (*overtraining*) suggests that the definition of the discriminant variables rely on features of the particular sample used to train the classifier rather than on a general feature of the kind of events to be selected. Overtraining occurs when a machine learning problem has too few degrees of freedom, because too many model parameters of an algorithm were adjusted to too few data points. Overtraining leads to a false increase in performance over the objectively achievable one, if measured on the training sample, and to an effective performance decrease when measured in an independent test sample. The sensitivity to overtraining therefore depends on the MVA method. For example, without the appropriate counter measures, boosted decision trees (see 5.2.2) usually suffer from at least partial overtraining, owing to their large number of nodes. Various method-specific solutions exist to counteract overtraining.

The last step is the *classification* which is the process of assigning objects or events to one of the possible discrete classes (e.g. signal and background) by the classifier found in the training step. After this, the events of real data are split into signal or background classes.

5.2.1   Classification performance evaluation

For the classification of a set of events, several different methods can be applied, having different performances depending on the data set. Before the decision of which classification technique is to be used, it is recommended that several different benchmark quantities are evaluated on independent test sample to assess the performance and the discriminatory power of each method.

The main benchmark quantities are:

- the **signal efficiency at three representative background rejections** (the efficiency, in this particular case, is equal to 1 - rejection) obtained from a cut on the classifier output. Also important is the area of the background rejection versus signal efficiency function (the larger the area the better the performance);

- the **separation** $< S^2 >$ of a classifier y, defined by the integral

$$< S^2 > = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy, \tag{5.2.3}$$

where $\hat{y}_S$ and $\hat{y}_B$ are the signal and background probability density functions of y, respectively. The separation is zero for identical signal and background shapes, and it is one for shapes with no overlap;

- the discrimination **significance** of a classifier, defined by the difference between the classifier means for signal and background divided by the quadratic sum of their root-mean-squares;

- the **correlation** between two random variables X and Y is usually measured with the correlation coefficient ρ, defined by

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}, \tag{5.2.4}$$

where $cov(X, Y)$ is the covariance between the two variables and σ_X (σ_Y) is the variance of the variable X (Y). The correlation coefficient is symmetric in X and Y, lies within the interval [-1,1], and quantifies by definition a linear relationship. Thus $\rho = 0$ holds for independent variables, but the contrary is not true in general. In particular, higher order functional or non-functional relationships may not, or only marginally, be reflected in the value of ρ.

- The **correlation ratio** $\eta^2(Y|X)$ is defined by

$$\eta^2(Y|X) = \frac{\sigma_{E(Y|X)}}{\sigma_Y}, \tag{5.2.5}$$

where X and Y are two random variables and

$$E(Y|X) = \int y \; P(y|x) \; \mathrm{d}y, \tag{5.2.6}$$

is the conditional expectation of Y given X with the associated conditional probability density function $P(Y|X)$. The correlation ratio η^2 is in general asymmetric and its value lies within [0,1], according to how well the data points can be fitted with a linear or non-linear regression curve.

- The **mutual information** $I(X, Y)$ allows to detect any predictable relationship between two random variables, be it of functional or non-functional form. It is defined by

$$I(X, Y) = \sum_{X, Y} P(X, Y) \ln \frac{P(X, Y)}{P(X)P(Y)}, \tag{5.2.7}$$

where $P(X, Y)$ is the joint probability density function of the random variables X and Y, and $P(X)$, $P(Y)$ are the corresponding marginal probabilities. Mutual information is symmetric and takes positive absolute values. In the case of two completely independent variables $I(X, Y)$ is zero.

Some characteristic examples and their corresponding values for η^2, ρ^2 and I are shown in figure 5.6.

5.2.2   Boosted Decision Tree

A decision tree is a binary tree structured classifier. Repeated binary (yes/no) decisions are taken on one single variable at a time, until a stop criterion is fulfilled. The phase space is thus split into many regions eventually classified as signal or background, depending on the majority of training events that end up in the final leaf node. In figure 5.7 a schematic view of a decision tree is presented; a sequence of binary splits is applied to the data, using discriminating variables. Each split uses a cut off variable which should give the best separation between signal and background. At the end of the BDT the leaves are labelled as signal or background, depending on the majority of events in the respective nodes.

The *boosting* of a decision tree extends this concept from one tree to several trees which form a "forest". The trees are derived from the same training ensemble by reweighting events and finally combined into a single classifier which is given by an average of the individual decision trees. Boosting makes the response of the decision trees, with respect to fluctuations in the training sample, more stable thus enhancing the performance with respect to a single tree.
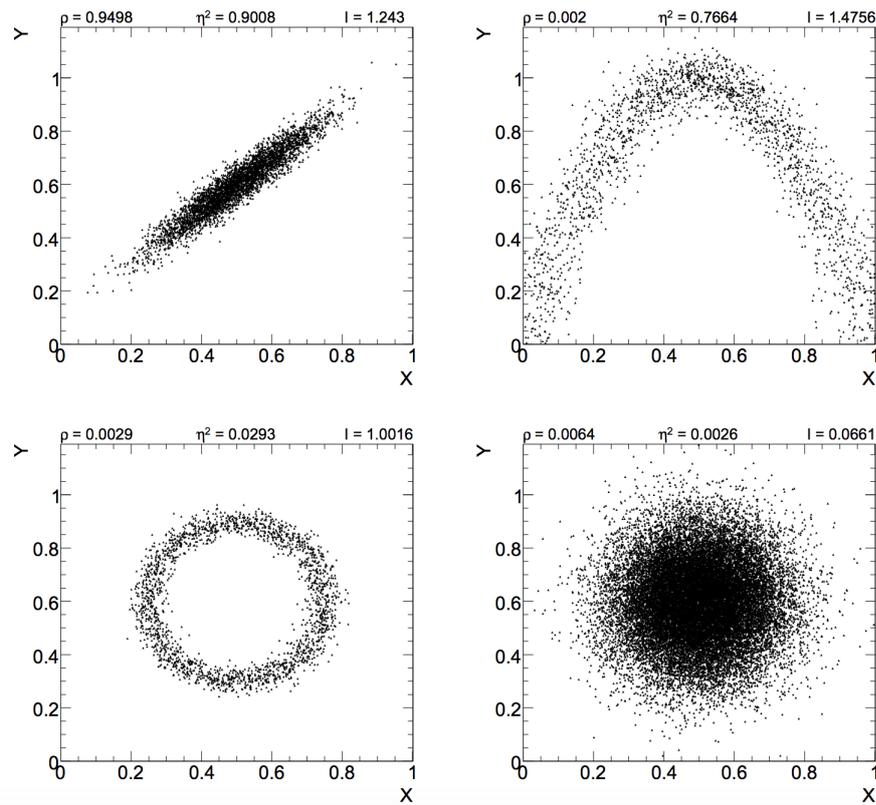
Figure 5.6: Different types of correlations between two random variables and their corresponding values for the correlation coefficient ρ, the correlation ratio η, and mutual information I. Linear relationship (upper left), functional relationship (upper right), non-functional relationship (lower left), and independent variables (lower right). [110]

Description and implementation

Decision trees allow a straightforward interpretation as they can be visualized by a simple two-dimensional tree structure. They are in this respect similar to rectangular cuts. However, whereas a cut-based analysis is able to select only one hypercube as the region of phase space, the decision tree is able to split the phase space into a large number of hypercubes, each of which is identified as either signal-like or background-like. The path along the tree represents an individual cut sequence that selects signal or background, depending on the type of the leaf node.

A shortcoming of decision trees is their instability with respect to statistical fluctuations in the training sample from which the tree structure is derived. For example, if two input variables exhibit similar separation power, a fluctuation in the training sample may cause the tree growing algorithm to decide to split on one variable, while the other variable could have been selected without that fluctuation. In such a case the whole tree structure is altered below this node, possibly resulting in a substantially different classifier response and leading to overtraining issues. An example of overtraining is shown: in 5.8a an extremely flexible classifier has managed to enclose all of the signal events and exclude all the backgrounds. However, if that decision boundary is applied to a statistically independent data sample, the contortions that led to good performances on the training sample will not work so well, as visible in figure 5.8b. The error rate calculated from the same set of events used to train the classifier underestimates the rate on a statistically independent sample.

This problem is overcome by constructing a forest of decision trees and classifying an event on a
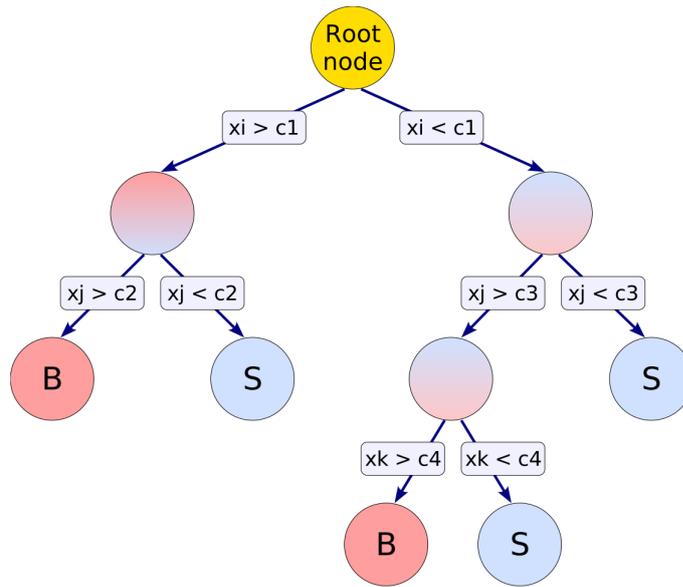
Figure 5.7: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is applied to the data. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled "S" for signal and "B" for background depending on the majority of events that end up in the respective nodes. For regression trees, the node splitting is performed on the variable that gives the maximum decrease in the average squared error when attributing a constant value of the target variable as output of the node, given by the average of the training events in the corresponding (leaf) node. [110]

majority vote of the classifications done by each tree in the forest. All trees in the forest are derived from the same training sample, with the events being subsequently subjected to so-called *boosting* (see in the sub-paragraph "Boosting"), a procedure which modifies their weights in the sample.

BDT suffers from at least a partial overtraining, due to the large number of nodes. Various methods exist to counteract the overtraining as to reduce the number of nodes removing the insignificant ones ("tree pruning"). Nevertheless, the general structure of the selection can already be understood by looking at a limited number of individual trees. In many cases, the boosting performances are better if applied to trees (classifiers) that, taken individually, have not much classification power. They are the so-called "weak classifiers", small trees limited in depth. Boosting almost completely eliminate the tendency of overtraining for simple decision trees which are usually grown to a large depth and then applied a procedure to decrease the number of nodes ("tree pruning").

Training a decision tree

The training (or *growing*) of a decision tree is the process that defines the splitting criteria for each node. The training starts with the root node, where an initial splitting criterion for the full training sample is determined. The split results in two subsets of training events, each going through the same algorithm determining the splitting criteria of the next nodes. This procedure is repeated until the whole tree is built. At each node, the split is determined by finding the variable and the corresponding cut value that provides the best separation between signal and background events that reach that node. The cut value is optimised by scanning over the variable range, therefore the bins granularity places an important role in

(a)   (b)
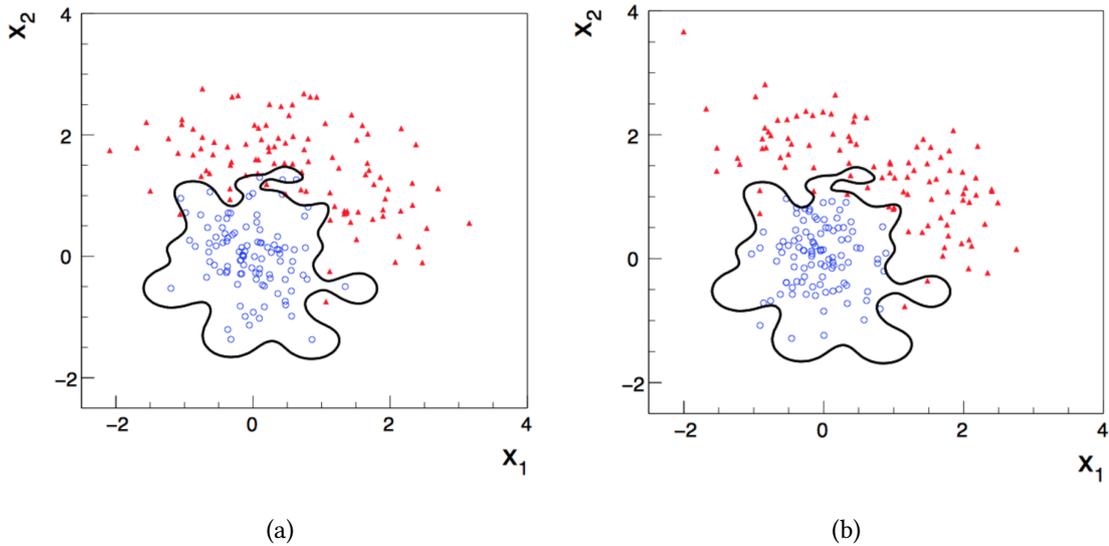
Figure 5.8: Scatter plot of two input variables for events classified as signal (blue circles) and background (red triangles). The decision boundary determined by a particularly flexible classifier is shown as a black line. It is applied on the training sample (a) and on a statistically independent data sample (b), in which the contortions that led to good performances on the training sample will not work so well. The error rate calculated from the same set of events used to train the classifier underestimates the rate on a statistically independent sample.

the training step and should be considered case by case. The addition of nodes to the tree stops once the number of events that should be split is below a threshold which is specified in the BDT configuration. The leaf nodes are classified as signal or background according to the class the majority of events belongs to.

Several separation criteria can be employed to assess the performance of a node in terms of the variable and the cut requirement used. Because a cut that selects predominantly background is as valuable as one that selects signal, the criteria are symmetric with respect to the event classes. All separation criteria have a maximum where the samples are fully mixed, i.e., at purity $p = 0.5$, and fall off to zero when the sample consists of one event class only. The purity of a node is given by the ratio of signal events to all events in that node. Hence pure background nodes have zero purity. Tests have revealed no significant performance disparity of a node trained using one of the following separation criteria, making them all equivalent:

- *Gini Index* (default), defined by $p \cdot (1 - p)$;

- *Cross entropy*, defined by $-p \cdot \ln(p) - (1 - p) \cdot \ln(1 - p)$;

- *Misclassification error*, defined by $1 - \max(p, 1 - p)$;

- *Statistical significance*, defined by $S/\sqrt{S + B}$.

In principle, the splitting could continue until each leaf node contains only signal or only background events, which could suggest that perfect discrimination is achievable. However, such a decision tree would be strongly overtrained. To avoid overtraining the application of a pruning procedure is necessary.

Pruning a decision tree

The pruning is the process of cutting back a tree from the bottom up after it has been built to its maximum size. Its purpose is to remove statistically insignificant nodes and thus reduce the overtraining of the tree. It has been found to be beneficial to first grow the tree to its maximum size and then cut back, rather than interrupting the node splitting at an earlier stage, because apparently insignificant splits can nevertheless lead to good splits further down the tree. Two tree pruning algorithms are possible:

- all leaf nodes for which the statistical error of the parent nodes are smaller than the combined statistical error of their daughter nodes are recursively deleted. The statistical error estimate of each node is calculated using the binomial error $\sqrt{p \cdot (1-p)/N}$, where N is the number of training events in the node and p its purity;

- the algorithm (named *Cost-complexity pruning*) relates the number of nodes in a sub-tree below a node to the gain, in terms of misclassified training events, in the full sub-tree against the one of the node itself with no further splitting. The cost estimate R chosen for the misclassification of training events is given by the misclassification rate $1 - max(p, 1-p)$ in a node, where p is the node purity. The cost-complexity ρ for this node is then defined by

$$\rho = \frac{R(\text{node}) - R(\text{sub} - \text{tree below that node})}{\text{number of nodes(sub} - \text{tree below that node}) - 1}. \tag{5.2.8}$$

The node with the smallest ρ value in the tree is recursively cut away as long as ρ is smaller than a certain value (the *pruning strength*) that should be decided case by case. If the *pruning strength* is set to a negative value, an algorithm attempts to automatically detect the optimal strength parameter. The training sample is divided into two sub-samples, of which only one is actually used for training, while the other one serves for validation. The tree is cut sequentially starting from the node which has the smallest value of the cost-complexity in the tree. After each pruning step the performance of the tree is assessed using the validation sample. This process is repeated until the ROOT node would be cut. As optimal prune strength for this tree it is chosen the value which corresponds to the best performing tree using the validation sample.

Boosting

The boosting procedure inside the BDT method consists in building a sequential list of trees each reweighted depending on the performances of the previous tree. This increases the statistical stability of the classifier and typically also improves the separation performance compared to a single decision tree of typically weak MVA methods by sequentially applying an MVA algorithm to reweighted (boosted) versions of the training data and then taking a weighted majority vote of the sequence of MVA algorithms thus produced. However, the advantage of the direct interpretation of the decision tree is lost. While it is possible to interpret the training result at a limited number of trees, it is hard to do so for hundreds of trees in a forest.

Among the main boosting algorithms there are the *AdaBoost*, *Gradient Boost*, *Bagging* and *Randomized Trees* [110].

The type of boosting chosen for this analysis is the Adaptive Boost (*AdaBoost*) where misclassified events during the training of a decision tree are given a smaller event weight in the training of the following tree. Starting with the original event weights while training the first decision tree, the subsequent tree is trained using a modified event sample where the weights of previously misclassified

events are multiplied by a common *boost weight* factor α, derived from the misclassification rate, err ($err = \text{misclassified events}/\text{total events}$), of the previous tree:

$$\alpha = \frac{1 - err}{err}. \tag{5.2.9}$$

The weights of the entire event sample are then renormalized such that the sum of weights remains constant. For the data set $\boldsymbol{x}$ used for training, the result of an individual classifier is $t_i(\boldsymbol{x})$ = +1 for signal and -1 for background; the boosted event classification $y_{boost}(\boldsymbol{x})$ is then given by:

$$y_{boost}(\boldsymbol{x}) = \frac{1}{N_{trees}} \sum_{i}^{N_{trees}} \ln\left(\alpha_i\right) \cdot t_i(\boldsymbol{x}) \tag{5.2.10}$$

where the sum runs over all the trees in the collection. Small values of $y_{boost}(\boldsymbol{x})$ thus indicate a background-like events, while large values a signal-like event.

Importance ranking

The ranking of each BDT input variable measures the importance of the variable and it is derived by two points:

- by evaluating the number of times the variables are used to split decision tree nodes;

- by weighting each split occurrence (performed using the same variable) by the separation achieved and by the number of events in the splitting node.

This ranking definition can be used for a single decision tree as well as for a forest.

5.3   MVA on $t\bar{t}H$ channel

The $t\bar{t}H$ production covers about 1% of the total inclusive Higgs production cross-section and, even if, the hadronic decay chosen for the Higgs boson ($H \rightarrow b\bar{b}$) is the highest possible ($\sim 60\%$), the expected signal is much smaller with respect to the background. Besides, the final state of the chosen channel is extremely complex. For these reasons, the application of a MultiVariate Analysis is the best solution to discriminate the signal from the background.

Due to the availability of MC events for signal and background, a supervised learning Boosted Decision Tree technique is used because it is easier to tune, does not require any adaption on the input quantities. The BDT is defined and tuned using an implementation in the MVA Toolkit (TMVA) [110], that can be integrated in the standard ROOT analysis code. Both in the resolved and boosted channels the BDT has been used as discriminating variable. Depending on the specific signal regions, the whole MVA procedure is slightly different in the number and the definition of the variables. Moreover, a specifically tuned "reconstruction BDT" is used in the resolved analysis for the object reconstruction (see in the following). In this chapter, the details of the BDT procedure relative to the boosted channel are described and only a brief overview of the resolved procedure is presented for completeness.

In the boosted analysis, the training has been performed with an initial set of variables (up to 17) that individually show a separation between the signal and background processes and take into account the topology of the events, the top tagging discrimination and the substructure of the large-R jets in each event. The following variables are considered:

- substructure observables related to the reconstruction of the hadronically decaying top in the event:

  – m_{top}^{lead} : mass of the leading top-tagged large-R jet in the event;

  – $\tau_{32,top}^{lead}$: N-subjettiness ratio τ_{32} (see Chapter 4) of the leading top-tagged large-R jet in the event;

  – $\sqrt{d_{23,top}^{lead}}$: splitting scale $\sqrt{d_{23}}$ (see Chapter 4) of the leading top-tagged large-R jet in the event;

- observables related to the reconstruction of the Higgs boson:

  – ΔR_{bb}^{min} : minimum ΔR between any b-jets in the event;

  – ΔR_{bb}^{avg} : average ΔR between any b-jets in the event;

  – $\Delta R_{(add)bb}^{min}$: minimum ΔR between b-jets that do not overlap with any top-tagged large-R jet in the event ("add" means additional);

  – $\Delta R_{top(add)b}^{avg}$: average ΔR between the top-tagged large-R jets and b-jets that do not overlap in the event;

  – $\Delta R_{top(add)j}^{avg}$: average ΔR between the top-tagged large-R jets and jets that do not overlap in the event;

  – m_{bb}^{max} : maximum mass of a pair of b-jets in the event;

  – $m_{(add)bb}^{H}$: invariant mass of a pair of b-jets (that do not overlap with any top-tagged large-R jet) that is closest to the mass of the Higgs boson ($m_H \equiv 125$ GeV);

  – $m_{(add)bj}^{H}$: invariant mass of a pair of one b-jet and one jet (that do not overlap with any top-tagged large-R jet) that is closest to the mass of the Higgs boson ($m_H \equiv 125$ GeV);

  – $m_{(add)jj}^{H}$: invariant mass of a pair of jets (that do not overlap with any top-tagged large-R jet) that is closest to the mass of the Higgs boson ($m_H \equiv 125$ GeV);

  – $m_{H,bb}^{reco}$: invariant mass of a pair of closest (minimum ΔR) b-jets in the event;

- observables related to global event topology:

  – N_j^{40} : number of jets with $p_T > 40$ GeV in the events;

  – $N_{(add)j}^{40}$: number of jets with $p_T > 40$ GeV that do not overlap with any top-tagged large-R jet in the events;

  – H_T^{jet} : scalar sum of the p_T of all the jets in the event;

  – $H_T^{(add)jet}$: scalar sum of the p_T of all top-tagged large-R jets and all jets that do not overlap with the top-tagged large-R jets in the event.

In figures 5.9a-5.9c the distribution (signal and background shapes) of all the 17 variables is shown in order to evaluate the discrimination power.

The signal and background samples have been split in two samples, with even and odd events, respectively. In this case, the weights obtained from the training on the even events are applied to the odd events and viceversa. This way, all the statistics can be used in both the training and the testing steps. The output of the training is composed by a set of weights to be applied to the events and by the response of the BDT method. In order to check the stability of this splitting method, the BDT responses of the two different training have been compared, as shown in figure 5.10: the two distributions result compatible

(a)

(b)

(c)
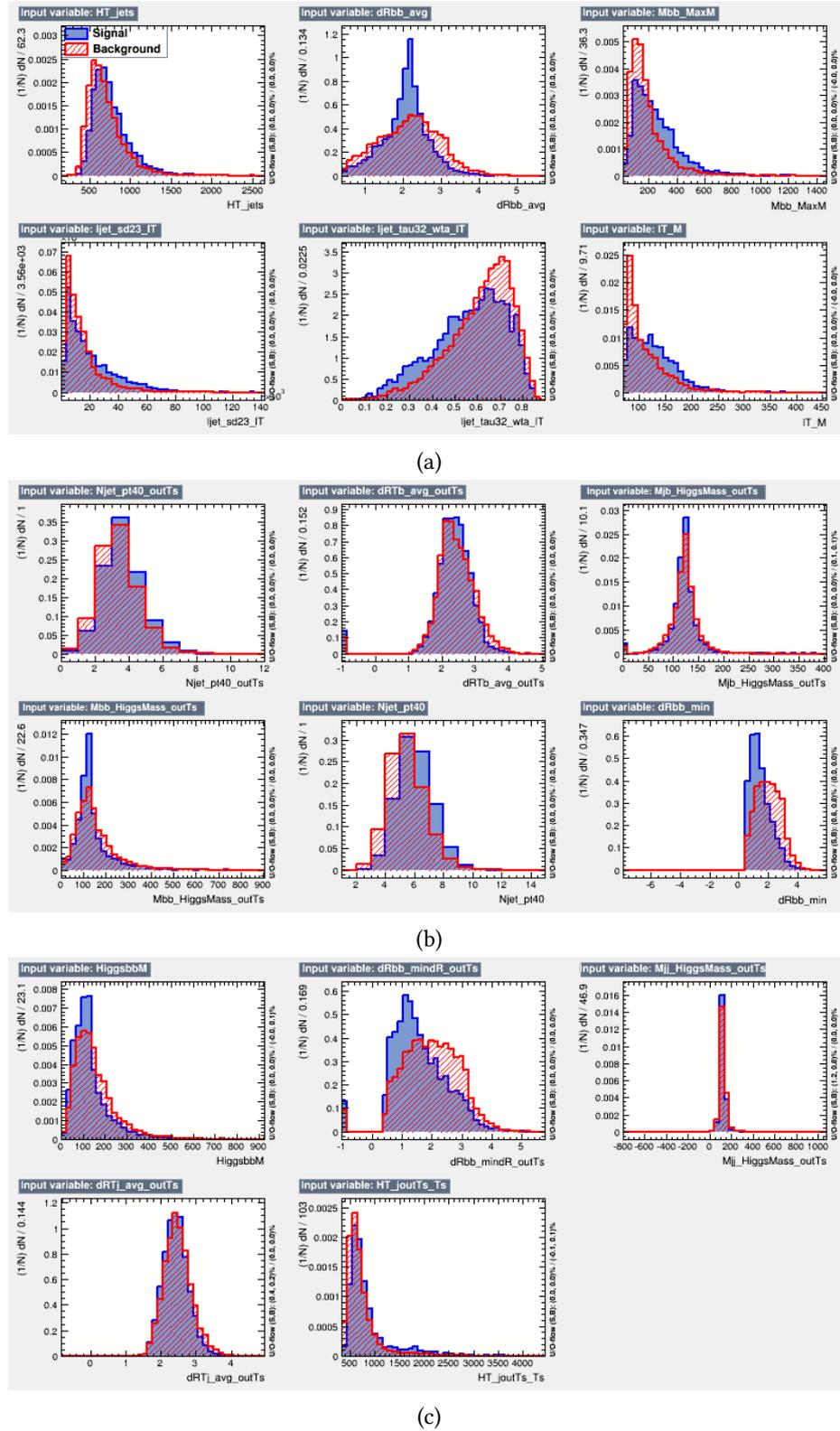
Figure 5.9: Signal (blue) and background (red) shapes of all the variables used in the first step of the BDT training. The signal sample is composed by the $t\bar{t}H$ process (considering all the three $t\bar{t}$ system decay modes: semileptonic, di-leptonic and hadronic). The background is composed by all the contributions for this channel (see Chapter 3). The two samples are normalized to the same number.

Ranking	Variable	Separation power		Ranking	Variable	Importance
1	$\mathbf{\Delta R}_{bb}^{min}$	0.0798		1	$\mathbf{\Delta R}_{bb}^{min}$	0.1003
2	$\mathbf{\Delta R}_{bb}^{avg}$	0.0675		2	$\mathbf{\Delta R}_{bb}^{avg}$	0.0864
3	m_{top}^{lead}	0.0588		3	m_{top}^{lead}	0.0787
4	m_{bb}^{max}	0.0450		4	N_j^{40}	0.0717
5	N_j^{40}	0.0471		5	$\mathbf{\Delta R}_{(add)bb}^{min}$	0.0704
6	$\mathbf{\Delta R}_{(add)bb}^{min}$	0.0413		6	m_{bb}^{max}	0.0615
7	$\sqrt{d_{23,top}^{lead}}$	0.0401		7	$m_{H,bb}^{reco}$	0.0608
8	$m_{(add)bb}^{H}$	0.0361		8	$\mathbf{\Delta R}_{top(add)b}^{avg}$	0.0604
9	H_T^{jet}	0.0340		9	$\mathbf{\Delta R}_{top(add)j}^{avg}$	0.0578
10	$m_{H,bb}^{reco}$	0.0267		10	$m_{(add)bb}^{H}$	0.0577
11	$H_T^{(add)jet}$	0.0266		11	$\sqrt{d_{23,top}^{lead}}$	0.0548
12	$\tau_{32,top}^{lead}$	0.0226		12	H_T^{jet}	0.0527
13	$\mathbf{\Delta R}_{top(add)b}^{avg}$	0.0176		13	$\tau_{32,top}^{lead}$	0.0476
14	$m_{(add)bj}^{H}$	0.0139		14	$H_T^{(add)jet}$	0.0404
15	$\mathbf{\Delta R}_{top(add)j}^{avg}$	0.0137		15	$m_{(add)bj}^{H}$	0.0396
16	$N_{(add)j}^{40}$	0.0136		16	$N_{(add)j}^{40}$	0.0317
17	$m_{(add)jj}^{H}$	0.0087		17	$m_{(add)jj}^{H}$	0.0275

Table 5.1: Separation power (left) and importance (right) of the initial set input variables.
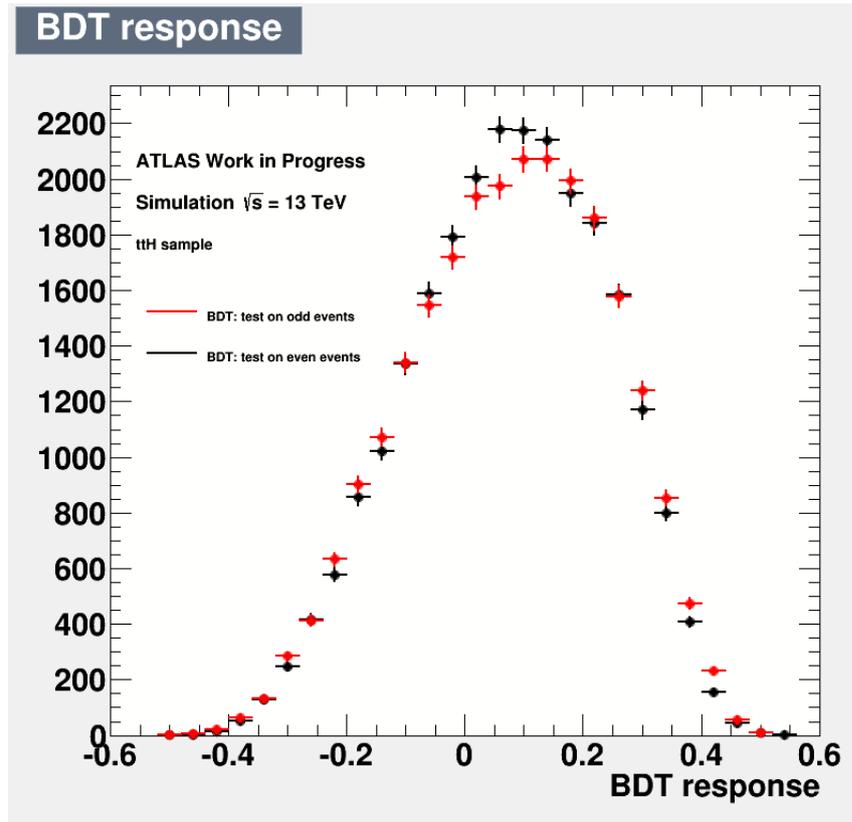
Figure 5.10: Comparison between the BDT responses from the even events (red) and the odd events (black), using the set of 17 variables. Only the signal BDT ($t\bar{t}H$ sample) is taken into account in this test.

and coherent. Therefore the procedure has been adopted in all the following steps of the analysis applying the "even" weights to the even events and the "odd" weights to the odd events.

The correlations among the variables (eq. 5.2.5) for signal and background (shown in figures 5.11a and 5.11b), the separation power (eq. 5.2.3) and the importance ranking (see tables 5.1) are the factors taken into account in order to choose the final set of 10 variables. The highly correlated variables with the lowest ranking and poorest separation have been removed. The table 5.2 reports the separation power and the importance ranking for the second iteration of the BDT with the 10 variables. The final BDT distribution is shown in figure 5.12.

The BDT output distribution on real events will be submitted to a fit procedure to estimate the signal strength $\mu = \sigma_{obs}/\sigma_{SM}$ and its upper limit.

Regarding the analysis relative to the resolved channel, in each of the signal regions, a sequence of BDTs is employed. The first, known as the "reconstruction BDT", is trained to match the reconstructed jets to the partons emitted from top and Higgs decays in simulation. For this purpose only $t\bar{t}H$ simulation is used, with correct jet assignations trained against the incorrect ones. Many kinematic quantities of the event are used, taking into account the topological information from the $t\bar{t}$ system and from the Higgs boson. The number of variables in each region is limited to avoid the effects of overtraining and the variables are chosen to achieve optimal performance. The best possible efficiency can be obtained by including information related to the Higgs boson, such as the candidate Higgs boson invariant mass. However, this biases the background distribution for these Higgs-related variables to be closer to the signal expectation, reducing their discriminating power. For this reason, two versions of the reconstruction
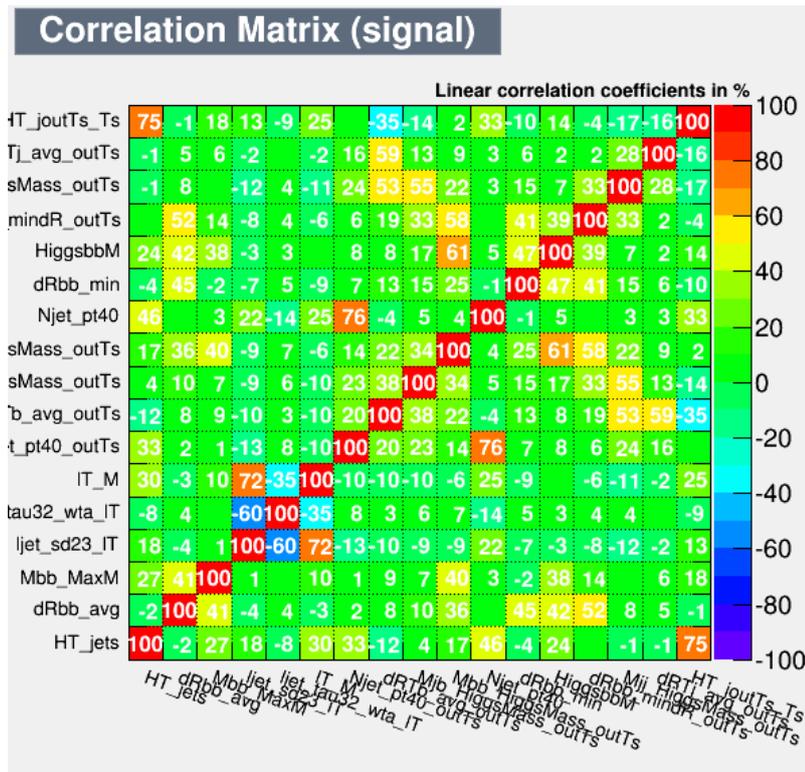
(a)
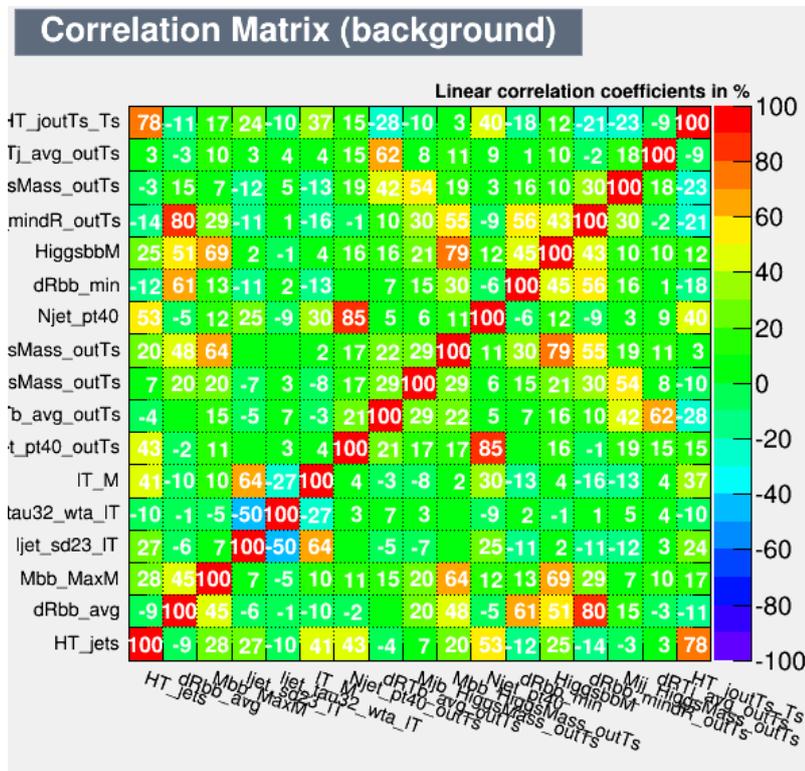
(b)

Figure 5.11: Correlations amongst input variables in the initial set for the BDT training for signal (a) and background (b).

Ranking	Variable	Separation power
1	ΔR_{bb}^{min}	0.0906
2	ΔR_{bb}^{avg}	0.0615
3	m_{top}^{lead}	0.0610
4	N_j^{40}	0.0483
5	m_{bb}^{max}	0.0476
6	$\Delta R_{(add)bb}^{min}$	0.0456
7	$m_{(add)bb}^{H}$	0.0362
8	H_T^{jet}	0.0311
9	$\tau_{32,top}^{lead}$	0.0261
10	$\Delta R_{top(add)b}^{avg}$	0.0180

Ranking	Variable	Importance
1	ΔR_{bb}^{min}	0.1335
2	ΔR_{bb}^{avg}	0.1257
3	m_{top}^{lead}	0.1105
4	$\Delta R_{top(add)b}^{avg}$	0.1103
5	$m_{(add)bb}^{H}$	0.1011
6	$\Delta R_{(add)bb}^{min}$	0.0947
7	N_j^{40}	0.0867
8	m_{bb}^{max}	0.0808
9	$\tau_{32,top}^{lead}$	0.0794
10	H_T^{jet}	0.0775

Table 5.2: Separation power (left) and importance (right) of the final set input variables.
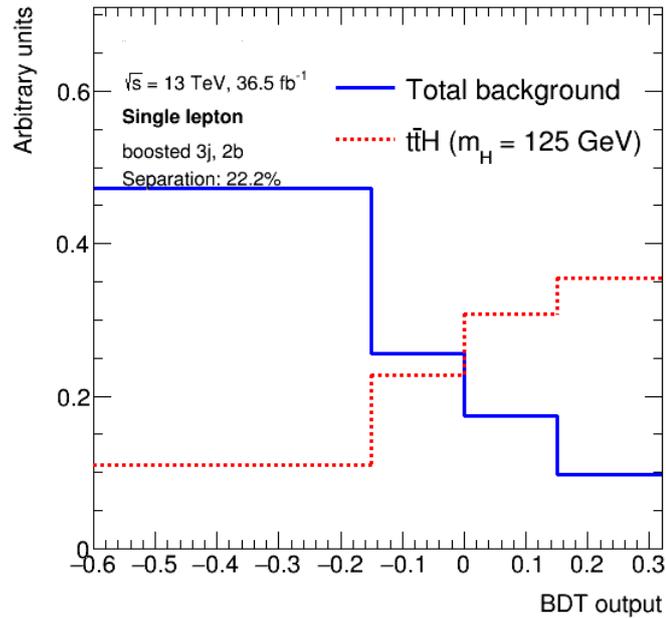
Figure 5.12: BDT response for signal (red, normalized to the background events) and background (blue) samples, with the relative separation power. The background sample includes all the background contributions (see Chapter 3) for the studied channel. The selection is the baseline signal region requirement, labeled as *boosted 3j,2b*, and the set of 10 variables has been used for the training. Only the MC simulations are shown in this distribution.

BDT are used, either with or without the Higgs boson information. Both the results are used in the next step.

For each signal region, information from the output of the reconstruction BDTs are combined with other kinematic variables in a "classification BDT", which classifies events as more signal- or background-like. The variables used in these BDTs are listed in table 5.3. The distributions of the classification BDTs, shown in figures 5.13a, 5.13b and 5.13c, are used as the final discriminants in the fit to data described in Chapter 6.

Variable	Definition	Regions		
		$\leq$ 6j, $\leq$ 4b	$\leq$ 6j, 3b	5j, $\leq$ 4b
ΔR_{bb}^{avg}	Average ΔR for all b-tagged jet pairs	✓	✓	✓
$\Delta R_{bb}^{\max p_{\mathrm{T}}}$	ΔR between the two b-tagged jets with the largest vector sum p_{T}	✓	-	-
$\Delta \eta_{jj}^{\max}$	Maximum $\Delta\eta$ between any two jets	✓	✓	✓
$m_{bb}^{\min \Delta R}$	Mass of the combination of the two b-tagged jets with the smallest ΔR	✓	✓	-
$m_{jj}^{\min \Delta R}$	Mass of the combination of any two jets with the smallest ΔR	-	-	✓
$m_{bj}^{\max p_{\mathrm{T}}}$	Mass of the combination of a b-tagged jet and any jet with the largest vector sum p_{T}	-	✓	-
p_{T}^{jet5}	p_{T} of the fifth leading jet	✓	✓	✓
$N_{bb}^{Higgs\ 30}$	Number of b-jet pairs with invariant mass within 30 GeV of the Higgs boson mass	✓	-	✓
N_{40}^{jet}	Number of jets with $p_{\mathrm{T}} \leq 40$ GeV	-	✓	-
$H_{\mathrm{T}}^{\mathrm{had}}$	Scalar sum of jet p_{T}	-	✓	✓
$\Delta R_{lep-bb}^{\min \Delta R}$	ΔR between the lepton and the combination of the two b-tagged jets with the smallest ΔR	-	-	✓
Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with all jets	✓	✓	✓
Centrality	Scalar sum of the p_{T} divided by sum of the E for all jets and the lepton	✓	✓	✓
H_1	Second Foxâ€“Wolfram moment computed using all jets and the lepton	✓	✓	✓
RecoBDT output	Output form the reconstruction BDT	✓*	✓*	✓*
m_H	Higgs boson mass	✓	✓	✓
$m_{H,b_{\mathrm{lep\ top}}}$	Mass of Higgs boson and b-jet from leptonic top	✓	-	-
$\Delta R_{H\ bb}$	ΔR between b-jets from the Higgs boson	✓	✓	✓
$\Delta R_{H\ t\bar{t}}$	ΔR between Higgs boson and $t\bar{t}$ system	✓*	✓*	✓*
$\Delta R_{H\ lep\ top}$	ΔR between Higgs boson and leptonic top	✓	-	-
$\Delta R_{H,b_{\mathrm{had\ top}}}$	ΔR between Higgs boson and b-jet from hadronic top	-	✓*	✓*

Table 5.3: Definition of the variables used in the classification BDT for the signal regions in the resolved channel. For the variables from the reconstruction BDT (second bunch), those with a * are from the BDT using Higgs boson information, while those with no * are from the BDT without Higgs boson information.

(a)

(b)

(c)
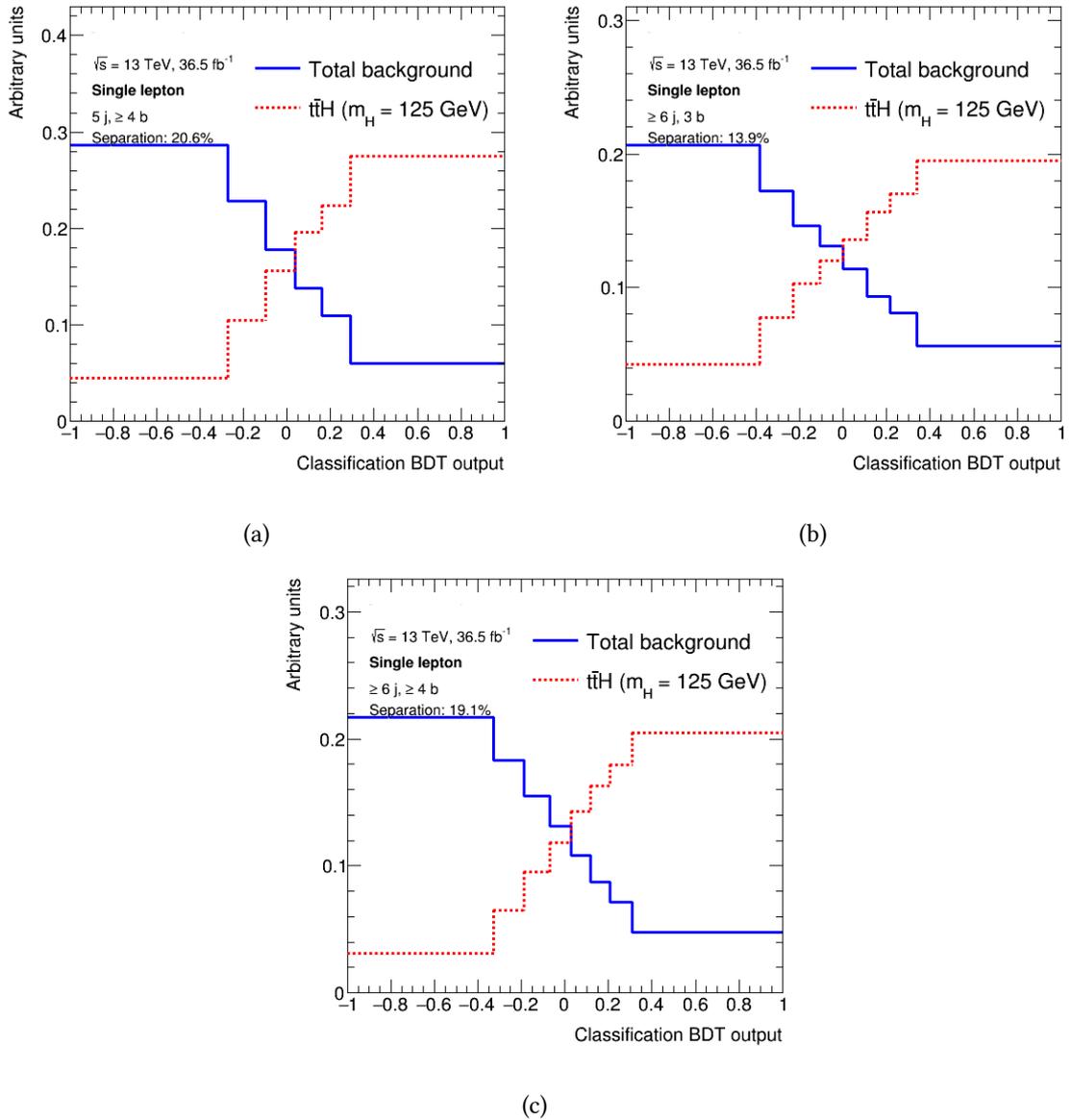
Figure 5.13: BDT responses for signal (red, normalized to the number of background events) and background (blue) samples, with the relative separation power for the three resolved signal regions. The background sample includes all the background contributions (see Chapter 3) for the studied channel. Only the MC simulations are shown in this distribution.

5.4 Signal extraction technique

Since the signal is very low with respect to the background, a complex statistical analysis is needed. It is based on an iterative procedure of maximum likelihood fit of the distributions in the signal and control regions in order to test for the presence of a signal. The profile likelihood ratio and the confidence level (CL) method are used to report the results in a frequentist way. Section 5.4.1 describes these fundamental concepts used to search for a new process.

5.4.1 Likelihood-based test

The standard procedure to discover a new signal process starts with the definition of the null hypothesys H_0, describing only known processes, here labelled as background, against the alternative signal hypothesys H_1, which includes both background and signal.

A widely used procedure to establish discovery (or exclusion) in particle physics is based on a significance test using a *profile likelihood ratio* as a test statistic [111]. For the success of the method, it is necessary that the model predictions for data distributions represent accurately the underlying theory being tested, meaning that any errors due to approximations (e.g. in detector modelling or in methods used to relate observable quantities to the fundamental theories) should be negligible in the full parameter space. By including additional parameters to the model (accounting for systematics effects) it is possible to approach this ideal situation more closely, but resulting in a loss in sensitivity.

For each event selected in the signal sample, a variable x of a certain kinematic quantity can be measured and these values can be used to construct a histogram $\boldsymbol{n} = (n_1, ..., n_N)$ of N bins. The expectation value of n_i can be written

$$E[n_i] = \mu s_i + b_i \tag{5.4.11}$$

where the mean number of entries in the ith bin from signal and background are

$$s_i = s_{tot} \int_{\text{bin } i} f_s(x; \boldsymbol{\theta}_s) dx,$$
$$b_i = b_{tot} \int_{\text{bin } i} f_b(x; \boldsymbol{\theta}_b) dx. \tag{5.4.12}$$

The parameter μ determines the strength of the signal process, with $\mu = 0$ corresponding to the background only hypothesis and $\mu = 1$ being the nominal (predicted by the SM) signal hypothesis. The functions $f_s(x; \boldsymbol{\theta}_s)$ and $f_b(x; \boldsymbol{\theta}_b)$ are the probability density functions (pdfs) of the variable x for signal and background events respectively, and $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_b$ represent parameters ("nuisance parameters") describing in general unknown systematic effect, whose contributions must be fitted from the data. The quantities s_{tot} and b_{tot} are the total signal and background events and the integrals represent the probabilities for an event to be found in bin i only. The notation $\boldsymbol{\theta} = (\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, b_{tot})$ will be used in the following to denote all the nuisance parameters and the background contribution. The signal normalization s_{tot} is a parameter fixed to the value predicted by the nominal signal model. In addition to the measured histogram $\boldsymbol{n}$, further subsidiary measurements are often performed to help constraining the nuisance parameters. This can be done by evaluating some chosen kinematic variables in a control region in order to construct a new histogram. This gives a set of values $\boldsymbol{m} = (m_1, ..., m_M)$ for the number of entries in each of the M bins and the expectation value of m_i can be written

$$E[m_i] = u_i(\boldsymbol{\theta}), \tag{5.4.13}$$

where the u_i are calculable quantities depending on the parameters $\boldsymbol{\theta}$. This measurement provides information on the background normalization parameter b_{tot} and also possibly on the signal and background shape parameters.

The likelihood function is defined as the product of Poisson probabilities for all bins:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}. \tag{5.4.14}$$

To test a hypothesized value of μ, the *profile likelihood ratio* is considered

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{5.4.15}$$

where the numerator of this ratio is the *profile likelihood function*. The quantity $\hat{\hat{\boldsymbol{\theta}}}$ denotes the value of $\boldsymbol{\theta}$ that maximizes $\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})$ for the specified μ and represents the conditional maximum-likelihood (ML) estimator of $\boldsymbol{\theta}$ (and thus is a function of μ). The denominator is the maximized (unconditional) likelihood function, i.e., $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are their ML estimators. The profile likelihood ratio $\lambda(\mu)$ assumes values between 0 and 1 (at $\mu = \hat{\mu}$), with λ close to 1 implies good agreement between data and the hypothesized value of μ. The presence of the nuisance parameters broadens the profile likelihood as a function of μ relative to what one would have if their values were fixed. This reflects the loss of information about μ due to the systematic uncertainties.

Test statistic q_μ for upper limits

For the purpose of establishing an upper limit on the strength parameter μ, the test statistics q_μ is defined as

$$q_\mu = \begin{cases} 0, & \mu < \hat{\mu} \\ -2\ln\lambda(\mu), & \mu \geq \hat{\mu}, \end{cases} \tag{5.4.16}$$

where $\lambda(\mu)$ is the *profile likelihood ratio* as defined in eq. 5.4.15. Higher values of q_μ represent greater incompatibility between the data and the hypothesized value of μ. To clarify the formula and the results, it is possible to consider what happens in the conditions of the central limit theorem where, given enough statistics, the $\lambda(\mu) \approx exp(-\chi^2/2)$ and therefore $q(\mu) = \chi^2(\mu)$ for $\mu > \hat{\mu}$. In these conditions, high values of $q(\mu)$ are equivalent to high values of a χ^2, implying incompatibility between the data and the test hypothesis. The reason for setting $q_\mu = 0$ for $\mu < \hat{\mu}$ when finding an upper limit is to not represent the values with $\mu < \hat{\mu}$ with less compatibility with respect to the μ obtained from the data, that is, the upper limit is obtained by testing μ against the alternative hypothesis consisting of lower values of μ.

The level of agreement between the data and the hypothesized μ is quantified with the p-value, that is the probability, under the assumption of H, of finding data of equal or greater incompatibility with the predictions of H. For an observed value $q_{\mu,obs}$, it has been defined as

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu | \mu) dq_\mu, \tag{5.4.17}$$

where $f(q_\mu | \mu)$ is the pdf of q_μ assuming the hypothesis μ. In this notation, in $f(q_\mu | \mu)$, the subscript of q refers to the hypothesis being tested and the second argument gives the value of μ assumed in the distribution of the data. When considering upper limits, it is quoted the value of μ for which the median p-value is equal to 0.05, as this gives the median upper limit on μ at 95% confidence level.

Together with the p-value it is convenient to define the significance Z. If x is a Gaussian distributed variable with mean m_x, $\hat{x}$ ($\hat{x} > m_x$) is defined as the value of x which has an upper-tail probability equal to the p-value p. The significance Z is defined as the number of standard deviations of $\hat{x}$ with respect to m_x:

$$Z = \Phi^{-1}(1 - p), \tag{5.4.18}$$

where Φ^{-1} is the quantile (that is the inverse of the cumulative distribution) of the standard Gaussian. For a signal process such as the Higgs boson, the particle physics community has tended to regard rejection of the background hypothesis with a significance of at least $Z = 5$, that corresponds to a p-value $= 2.87 \times 10^{-7}$, as an appropriate level to constitute a physics discovery. For the purpose of excluding a signal hypothesis, a threshold p-value of 0.05 (i.e., 95% confidence level, CL) is often used, which corresponds to $Z = 1.64$.

The sensitivity of an experiment is characterized not only in the significance obtained from a single data set, but rather in the expected median significance to reject different values of μ. In this analysis, this estimator is evaluated by using the so called "Asimov" data set [111], that replaces the ensemble of real data and that is generated depending on the distribution of the MC samples, according to the chosen test statistic. For the specific case of setting exclusion limits, the sensitivity is characterized by the median significance, assuming data generated using the $\mu = 0$ hypothesis, rejecting a non-zero value of μ (usually $\mu = 1$ is of greatest interest). The sensitivity of an experiment is illustrated in figure 5.14, which shows the pdf for q_μ assuming both a strength parameter μ and a different value μ'. The distribution $f(q_\mu|\mu')$ is shifted to higher value of q_μ, corresponding on average to lower p-values. The sensitivity of an experiment can be characterized by giving the p-value corresponding to the median q_μ assuming the alternative value μ'. As the p-value is a monotonic function of q_μ, this is equal to the median p-value assuming μ'.
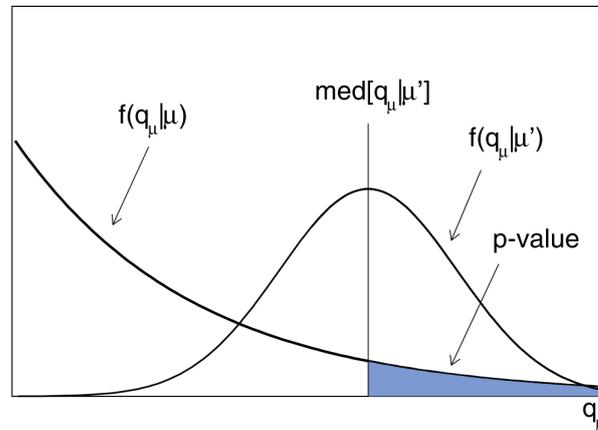
Figure 5.14: Illustration of the p-value corresponding to the median of q_μ assuming a strength parameter μ. [111]

Figure 5.15a shows the distributions $f(q_\mu|0)$ (red) and $f(q_\mu|\mu)$ (blue), that are the distributions for the value of μ that gave $p_\mu = 0.05$, corresponding to the 95% CL upper limit. The vertical line gives the median value of q_μ assuming a strength parameter $\mu' = 0$. The area to the right of this line under the curve of $f(q_\mu|\mu)$ (shown shaded in green) gives the p-value of the hypothesized μ. The upper limit on μ at a confidence level CL $= 1 - \alpha$ is the value of μ for which the p-value is $p_\mu = \alpha$.

By simulating the experiment many times with Monte Carlo, it is possible to obtain a histogram of the upper limits on μ at 95% CL, as shown in figure 5.15b. The $\pm 1\sigma$ (green) and $\pm 2\sigma$ (yellow) error bands

(a)   (b)
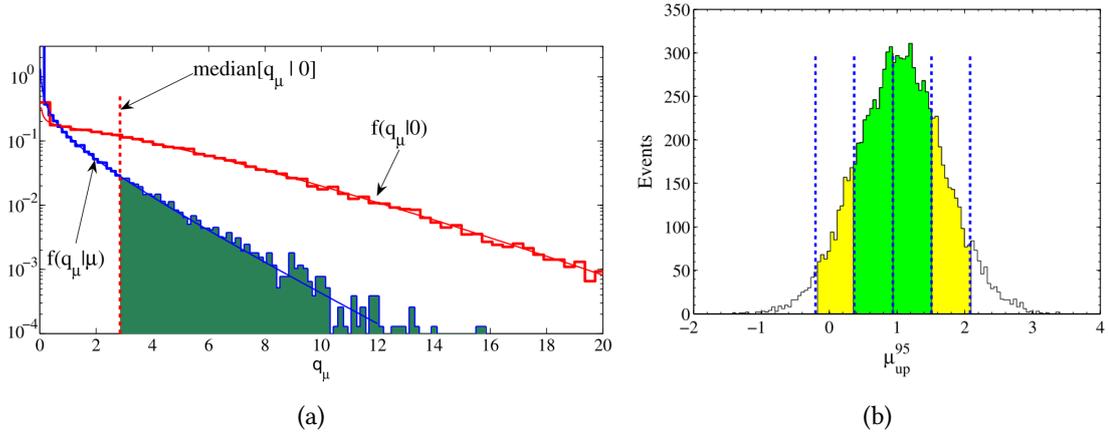
Figure 5.15: Comparison between the distributions $f(q_\mu|0)$ (red) and $f(q_\mu|\mu)$ (blue). (b) Distribution of the upper limit on μ at 95% CL, assuming data corresponding to the background-only hypothesis. [111]

are obtained from the MC pseudo-experiments. The vertical lines indicate the error bands as estimated directly without Monte Carlo simulation. As can be seen from the figure 5.15a, the agreement between the formulae and MC predictions is excellent. This is the procedure to find an upper limit on μ for a given value of the signal strength. In a search for a signal of unknown strength, the procedure would be repeated (in small steps) for a range of signal strength.

The Modified Frequentist CL_s method

The method used in this analysis is the modified frequentist confidence level $\mathrm{CL_s}$ [112] [113]. The confidence level for excluding the possibility of signal on top of background (the s+b hypothesis), can be defined as

$$\alpha_{\mathrm{s+b}} = P_{\mathrm{s+b}}(q_\mu \leq q_{\mu,obs}), \tag{5.4.19}$$

that is the probability, assuming the presence of both signal and background at their hypothesized levels, that the test statistic would be less than or equal to that observed in the data.

The confidence level (1-$\alpha_{\mathrm{s+b}}$) may be used to quote exclusion limits. Although it has the disturbing property that if too few candidates are observed to account for the estimated background, then any signal, and even the background itself, may be excluded at a high confidence level. It nonetheless provides exclusion of the signal at exactly the confidence level computed. A typical limit computation involves also computing the confidence level for the background alone,

$$\alpha_{\mathrm{b}} = P_{\mathrm{b}}(q_\mu \leq q_{\mu,obs}), \tag{5.4.20}$$

where the probability assumes the presence of the background only. This confidence level has been suggested to quantify the confidence of a potential discovery, as it expresses the probability that background processes would give a number of events smaller than or equal to the number of observed candidates. The Modified Frequentist confidence level $\mathrm{CL_s}$ is then computed as the ratio

$$CL_{\mathrm{s}} = \alpha_{\mathrm{s+b}}/\alpha_{\mathrm{b}}. \tag{5.4.21}$$

This is the method used to set the upper limit of the $t\bar{t}H$ production cross-section in this analysis.

5.4.2   Statistical analysis procedure in the boosted and resolved analyses

In the boosted $t\bar{t}H$ channel, there is only one signal region and the discriminant distribution chosen to be fitted is the BDT output (since it has been built as the most discriminant variable between signal and background), shown in figure 5.16.
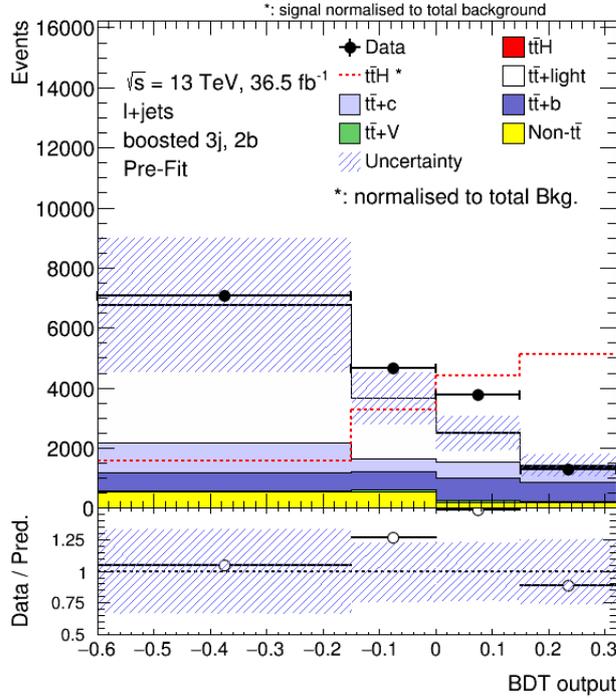
Figure 5.16: Distribution of the discriminating variable BDT output in the boosted signal region, before the fit described below. The dotted red line represents the signal $t\bar{t}H$ normalized to the toal background events.

In order to increase the $t\bar{t}H$ phase space and to improve the sensitivity of the analysis, the final result has been obtained by the combination of the boosted and the resolved channels. The fit procedure has been performed using the BDT discriminant in the signal regions of both channels and the H_T^{had} variables in the control regions (of the only resolved one). A veto has been put on events in the resolved signal and control regions that pass also the boosted selection. The BDT output and the H_T^{had} distributions are shown in figures 5.17 and figures 5.18 respectively.

The distributions of the discriminants from each of the channels and regions considered are combined to test for the presence of a signal, assuming a Higgs boson mass of $m_H = 125$ GeV. The statistical analysis is based on the binned likelihood function $\mathcal{L}(\mu, \theta)$ described in section 5.4.1. The likelihood function depends on the signal-strength parameter μ, defined as the ratio of the observed/expected cross-section to the SM cross section, and the nuisance parameters θ, that encode the effects of systematic uncertainties on the signal and background expectations. Therefore, the total number of expected events in a given bin depends on μ and θ. This procedure allows the impact of systematic uncertainties on the search sensitivity to be reduced by taking advantage of the highly populated background-dominated control regions included in the likelihood fit. It requires a good understanding of the systematic effects affecting the shapes of the discriminant distributions. The test statistic q_μ is used to measure the compatibility of the observed data with the background-only hypothesis (i.e. for $\mu = 0$), and to make statistical inferences about μ, such as upper limits using the CL_s method (see section 5.4.1) as implemented in the RooFit

(a)   (b)

(c)
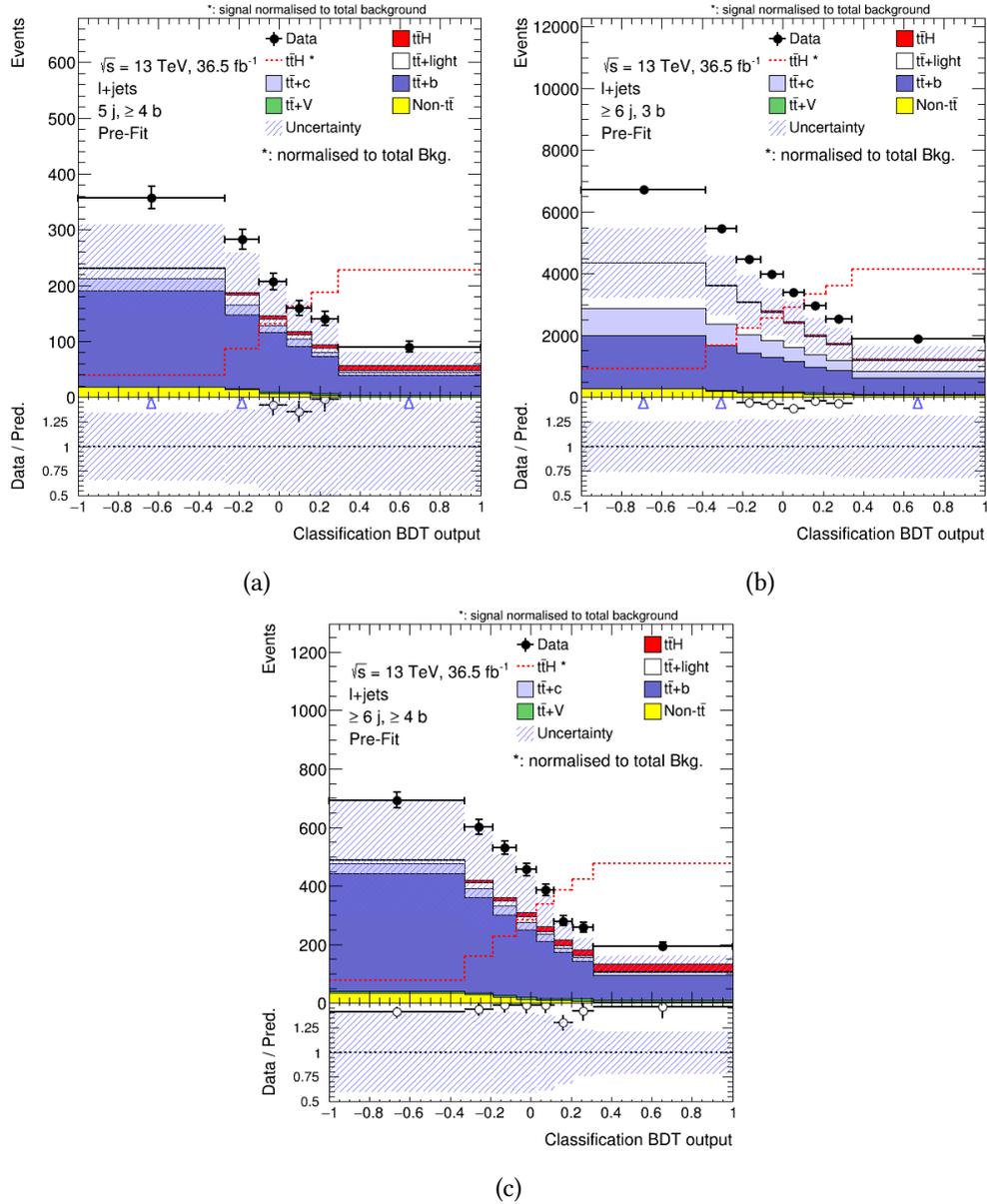
Figure 5.17: Distributions of the discriminating variables BDT output in the resolved signal regions, before the fit procedure.

package [114] [115].

To obtain the final result, a simultaneous fit to the data is performed on the distributions of the discriminants in 10 regions: nine in the resolved and one in the boosted channel. The fits are performed under the s+b hypothesis, where the signal-strength parameter μ is the parameter of interest of the fit and is allowed to float freely, but requiring to be the same in all 10 regions. The normalisation of each background contribution is determined by the fit, simultaneously with μ. Contributions from $t\bar{t}$, W/Z+jets production, single top, diboson and $t\bar{t}+V$ backgrounds are constrained by the uncertainties on the respective theoretical calculations, the uncertainty on the luminosity and on the detector. Statistical uncertainties in each bin of the discriminant distributions are taken into account by dedicated parameters in the fit, in
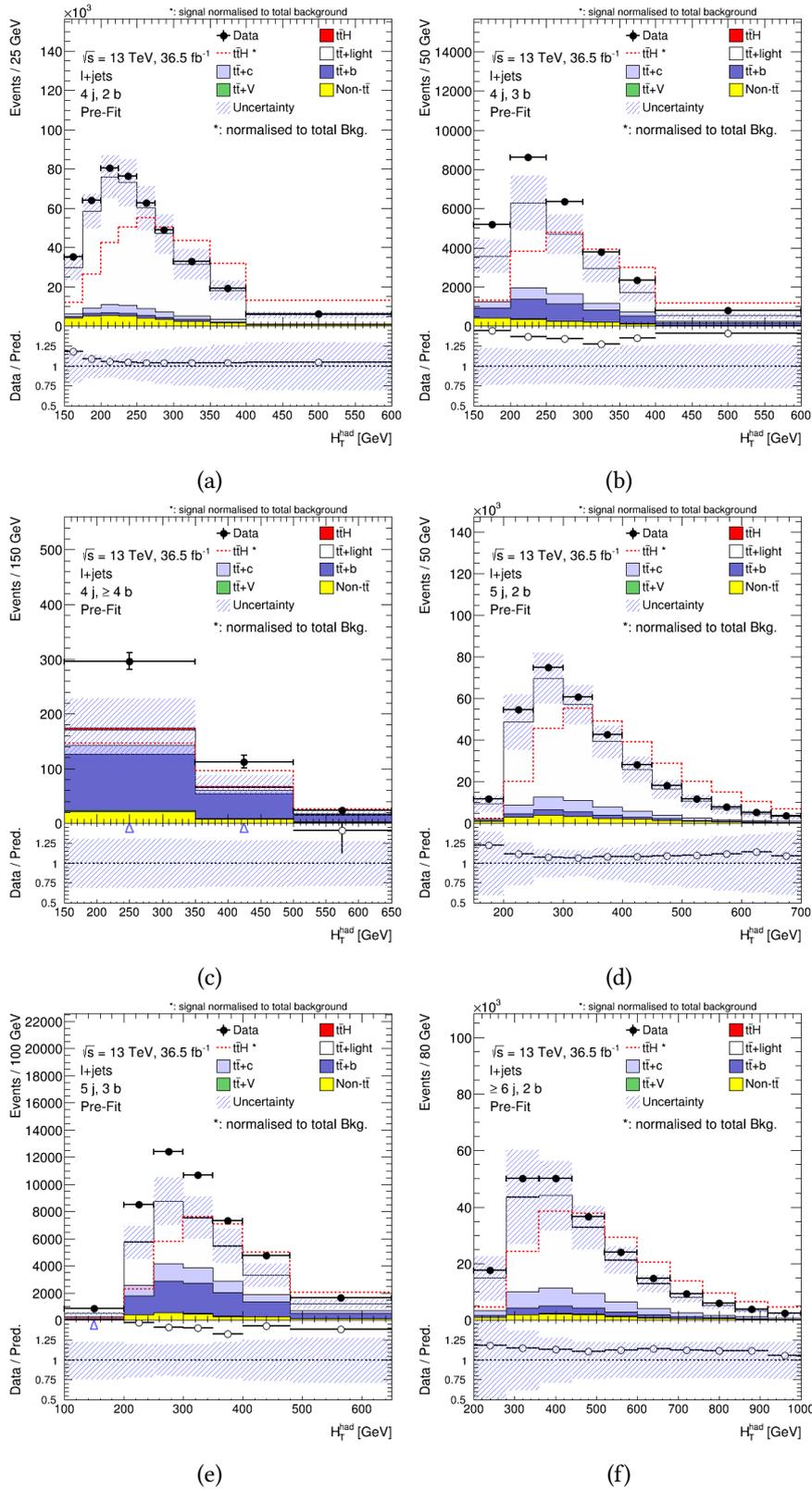
Figure 5.18: Distributions of the discriminating variables H_T^{had} in the resolved single-lepton control regions, before the fit described below.

order to understand the different components and sources. All the results will be shown in Chapter 6.

5.5 b-tagging algorithm optimization

The boosted analysis is the most innovative and the most delicate due to the very low statistics and the busy final state. It is based on event cuts, taggers (for the boosted objects) and working points (WPs) optimized during specific studies on specific samples, that could be not optimal for the $t\bar{t}H$ channel. For this reason, the simultaneous application of cuts and taggers could be not necessarily the most efficient.

Many optimization studies on different quantities have been done in order to maximize the sensibility of the boosted analysis. The overlap removal between the electron and the large-R jets in each event has been studied accurately (section 5.1.2), different selections have been tested, studying the influence of the top-tagging algorithm and the number of jets and b-jets to include within. The Higgs-tagging algorithm has been considered in these studies as well.

After defining the baseline selection of the analysis, the effect of the b-tagging algorithm on the analysis sensitivity has been studied and it resulted to be the very sensitive to its working point. In the ATLAS experiment there are different WPs available for the b-tagging, as explained in Chapter 4, that differ for the purity of the selection, the tagging efficiency and, consequently, for the background rejection. The WPs available for this analysis, using the algorithm mv2c10, are listed in the table 5.4, with the relative efficiency and purity.

Working point (%)	b-jet efficiency (%)	Purity (%)
77	76.9	95.2
70	69.9	97.5
60	60.0	99.0

Table 5.4: Working points of the b-tagging algorithm mv2c10, with the relative efficiency and purity.

In order to motivate the choice of a certain WP, some considerations are needed: first of all the statistics of the survived events needs to be enough to ensure a low statistical error with respect the systematic one, then the evaluation of the background rejection and the determination of the effect on the shape of the MVA discriminant variable. For these reasons the MVA training must be optimized for each WP and this likely changes the set of the input variables and their separation and importance ranking. The BDT response (see figure 5.12), consequently, changes its shape and its separation power, affecting the significance, the background composition, the results of the fit procedure and the limit values.

As expected, the yields of the signal region (see table 5.5) decreases dependently on the tightness of the WP selection, while the S/B increases and the S/$\sqrt{\text{B}}$ slightly decreases, but remains compatible. This leads to different performances in terms of signal efficiency and background rejection, as shown in figure 5.19. It is clear that the 60% WP is the best choice, looking at the performance improvement with respect to the 70%, but the full chain of the analysis needs to be completed in order to figure out if the tighter WP is actually the one allowing a better analysis sensitivity.

Depending on the WP, the input variable set changes because, as shown in tables 5.1 and 5.2, many variables are strongly related to the b-tagging WP.

The limit on the signal-strength μ has been calculated for both the WP (70% and 60%). At the moment,

Working point (%)	Background yield	Signal yield	S/B (%)	S/$\sqrt{\text{B}}$
70	5467	50	0.92	0.68
60	3778	39	1.03	0.63

Table 5.5: Signal and background yields and significance values for different b-tagging working points, in the boosted signal region. The yields values are scaled for the 13.2 fb^{-1} luminosity.
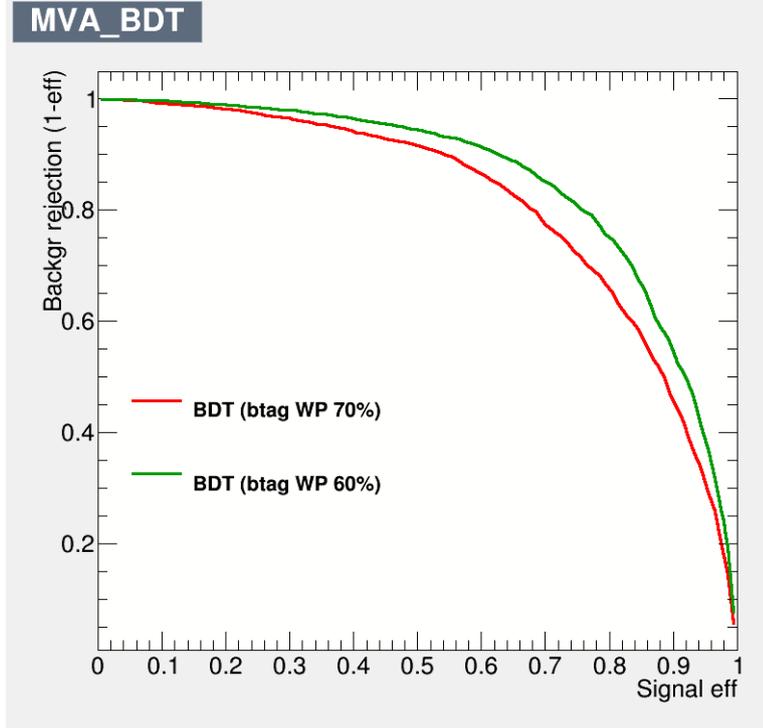
Figure 5.19: Background rejection versus signal efficiency of the BDT, trained with variables calculated with the 70% (red) and 60% (green) b-tagging working point.

for this optimization study, it has been determined the limit values for the boosted channel only (in the fit only the boosted signal region has been included), in order to check if there is an improvement in the performance of the fit method. This study provides a good scenario about how the response of the fit method (in terms of limit settings) changes depending on the b-tagging WP and the luminosity increase (from 13.2 fb^{-1} to 30 fb^{-1}), as shown in table 5.6.

The table shows that increasing the statistics and tightening the WP the limit value decreases. Despite the fact that the 60% WP looks better in the boosted channel, it has been decided to use the 70%, the same used by the resolved channel, in order to combine the two analyses without a further increase in the systematics due to using two different working points. In this step, indeed, it is very important to include as many systematics uncertainties as possible, in order to take into account the whole broadening effect of them on the measure. For a detailed description of all the systematic uncertainties and their treatment in this analysis, see Chapter 6.

Luminosity	Working point (%)	$\mu/\mu_{\max}$
13.2 fb^{-1}	70	1
	60	0.9
30.0 fb^{-1}	70	0.8
	60	0.7

Table 5.6: Variations of the 95% CL upper limit on the signal-strength, using 70% and 60% *b*-tagging working points, Asimov data sets and luminosity of 13.2 and 30.0 fb^{-1}. The $\mu_{\max}$ is the baseline configuration used in the analysis: 70% WP and 13.2 fb^{-1}.

Chapter 6

Signal strength measurement

The measurement of the observed signal strength of the $t\bar{t}H$ production cross section is presented in this chapter, together with its statistical limit, for the resolved-only analysis and the combined (resolved plus boosted) one. Both the results are obtained from the analyses on the 2015 and 2016 data samples collected with the ATLAS detector. The techniques used in order to estimate the statistical and systematic uncertainties affecting this measurement are described in sections 6.1 and 6.2; final results are shown in section 6.3. The comparisons with the Run-1 and the early Run-2 results (section 6.4.1) and with the CMS results (section 6.4.2) are presented, for the resolved-only analysis. At the end, future perspectives of the boosted $t\bar{t}H$ analysis are shown in section 6.5.

6.1 Statistical uncertainties

The statistical uncertainty is due to the finite number of both selected data events and simulated MC events. Since the number of real data events after all the selections is around 15000, while the MC events can be produce as much as needed, the data statistics will affect the measurement precision more than the MC statistics.

The statistical uncertainty on the fitted distribution (H_T^{had} for the control and BDT output for the signal regions) in each bin is calculated as the squared root of the observed number of events. It will be summed in quadrature with the total systematic uncertainty, since the two have to be considered uncorrelated.

The analysed channel is affected by a dominant background in practically all signal regions: this reduces significantly the accuracy with which it is possible to measure the signal strength. The final value and the uncertainty of the signal strength is determined directly by the fit procedure described in Chapter 5.

6.2   Systematic uncertainties

Systematic uncertainties are originated by both an imperfect knowledge of the detector, that entails an uncertainty on the parameters used in the event reconstruction, and by an approximate theoretical modeling of signal and background events. Usually each systematic uncertainty in the signal strength is evaluated by varying the corrisponding variable distribution by one standard deviation (σ) and reweighting accordingly all the events. This leads to two shifted distributions for each variable of interest representing the fluctuation ($\pm 1\sigma$) with respect to the nominal distribution. In cases where both shifted distributions produce an excess or a defect with respect to the nominal distribution, the resulting variation is assumed to be of the same size in both directions and is therefore symmetrized. Some systematic uncertainties can not be treated in this way and require a specific case dependent approach. The convention of the systematics (and their nuisance parameters) naming is explained in Appendix 9.

     The measurement is affected by many sources of systematic uncertainty that can be categorized in three different classes: the luminosity, the reconstruction of physics objects and the signal or background modelling, all described in the next paragraphs. Uncertainties may affect the normalisation of the samples, the shape of the final discriminants, or both. A summary of the systematic uncertainties with similar sources grouped together is given in table 6.1.

6.2.1   Luminosity

The preliminary uncertainty on the combined 2015+2016 integrated luminosity is 4.1%. It is derived, following the prescription detailed in refs [116] and [117], from a preliminary calibration of the luminosity scale using the $x - y$ beam-separation scans performed in August 2015 and November 2016.

6.2.2   Object reconstruction uncertainties

Jet Energy Scale

The Jet Energy Scale (JES) is a factor applied to the deposited energy by a jet to obtain its initial energy. The JES term represents one of the main contributions to the systematic uncertainty in jet-based analyses and its determination is a rather challenging task due to the difficult environment of hadron-hadron colliders. The jet energy scale depends on a variety of detectors and physic effects that includes the non-linearity in the calorimeter response, hardware problems of the detector and additional energy due to the underlying and pile-up event. The possible loss of energy during the jet reconstruction procedure is also included in this source of systematic uncertainty.

     The jet energy scale and its uncertainty are derived by combining information from test-beam data, LHC collision data, and simulation [118]. The uncertainties from these measurements are factorised into six independent sources. These are combined with additional uncertainties related to jet flavour (see figure 6.1a), pile-up treatment, η interpolation, and high-p_T jets, for a total of 18 sources. Although the uncertainties are not large, from 0.5% to 5.5% per jet, the effects are amplified by the large number of jets in the final state due to their correlation. Other jet-related uncertainties include the jet energy resolution (JER), shown in figure 6.1b, and the jet vertex tagger (JVT) efficiency.

Jet flavour tagging scale factor

The efficiency to correctly tag b-jets is measured using $t\bar{t}$ simulated events. The mis-tag rate for c-jets is measured using D^* mesons, while for light jets it is measured using jets with impact parameters and sec-

Uncertainty sources	Groups
Luminosity	Luminosity
Calorimeter response	
Hardware problems	
Jet flavour	
Pile-up treatment	JES and JER
η interpolation	
High-p_{T} jets	
JVT efficiency	
b-tag efficiency	
Mis-tag rate for c-jets	
Mis-tag rate for light jets	Jet Flavour Tagging
Extrapolation outside kinematic ranges	
Run-1 extrapolation for c- and light-tag	
Trigger and reconstruction efficiencies	
Identification and isolation efficiencies	Light leptons
Lepton momentum scale and resolution	
Data/MC disagreement	
Modelling	
Tracking efficiency	Large-R jets
Statistics of samples	
$t\bar{t}H$ cross-section	
Higgs boson BR	Signal modelling
QCD scale choice	
$t\bar{t}$ cross-section	
$t\bar{t}$ modelling	
W/Z cross-section	
W+HF flavour jets	Background modelling
Single-top cross-section	
Diboson cross-section	
$t\bar{t}+V$ NLO cross-section	

Table 6.1: Summary of the systematics groups and their different components used in the analysis. The $t\bar{t}$ modelling systematic uncertainties will be treated in more details in paragraph 6.2.3, since a specific study has been done for the HF classification of the sample and the relative many different sources of uncertainties.
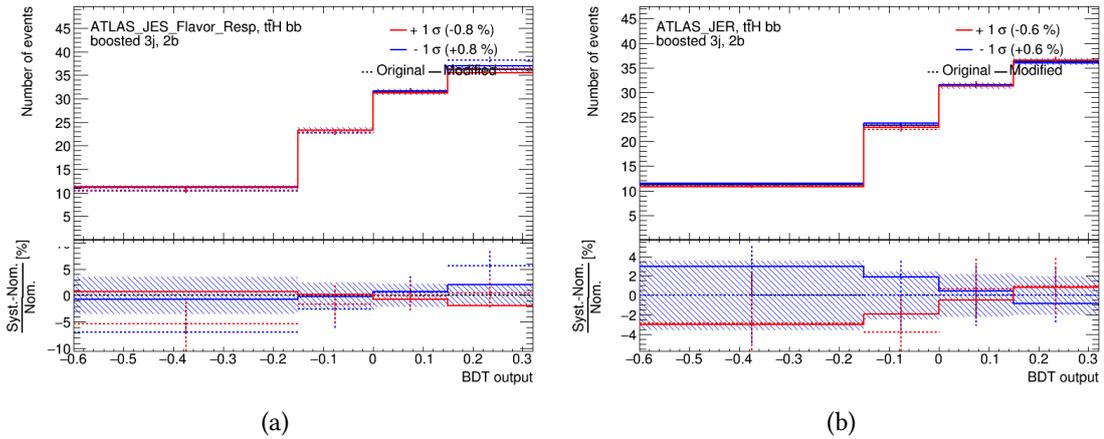
Figure 6.1: Variation of the number of events distribution depending on the systematics ($\pm 1\sigma$) due to the uncertainties related to the the JES (a) and the JER (c), in the boosted signal region for the $t\bar{t}H$ (with $H \to b\bar{b}$) process. For the JES systematics, as an example, it is shown the variations depending on a component only.

ondary vertices consistent with a negligible or at limit negative lifetime [91]. The uncertainties associated with these measurements are factorised into statistically independent sources, relative to b-jets, c-jets and light jets, shown in figure 6.2a, 6.2b and 6.2c respectively. An additional uncertainty is included for the extrapolation to jets outside the kinematic ranges covered by these measurements; this uncertainty should be considered totally correlated with the other components, independently from the jet flavour. Finally, the c- and light-jet measurements, which are performed with data from Run 1, have an uncertainty on the extrapolation to Run 2.

Lepton contribution

Uncertainties associated with leptons arise from the trigger, reconstruction, identification (see figure 6.3a), and isolation effciencies, as well as the lepton momentum scale (see figure 6.3b) and resolution. These are measured in data using leptons in $Z \to l^+l^-$ and $J/\psi \to l^+l^-$ events, as well as the energy and momentum of electrons in $W \to e\nu$. These uncertainties have a small impact on the results presented in this thesis.

Large-R jets related uncertainties

Uncertainties related to the large-R jets have been derived in situ using the R_{trk} procedure (using the information from the tracks that are meatched to the considered large-R jets) and they are all parametrised in terms of p_T and m_{jet}/p_T. The overall contribution dependes on 12 separate sources, grouped below in four categories:

- Baseline: evaluates the base difference between data and Pythia8 generator, figure 6.4a;

- Modelling: looks at the difference between Pythia and Herwig generators, without changing the shower simulator, figure 6.4b;

- Tracking: calculates the uncertainties on the tracks associated to the large-R jets, used as a reference, figure 6.4c. This takes into account three different effects:

  1 the tracking efficiency and related uncertainties;
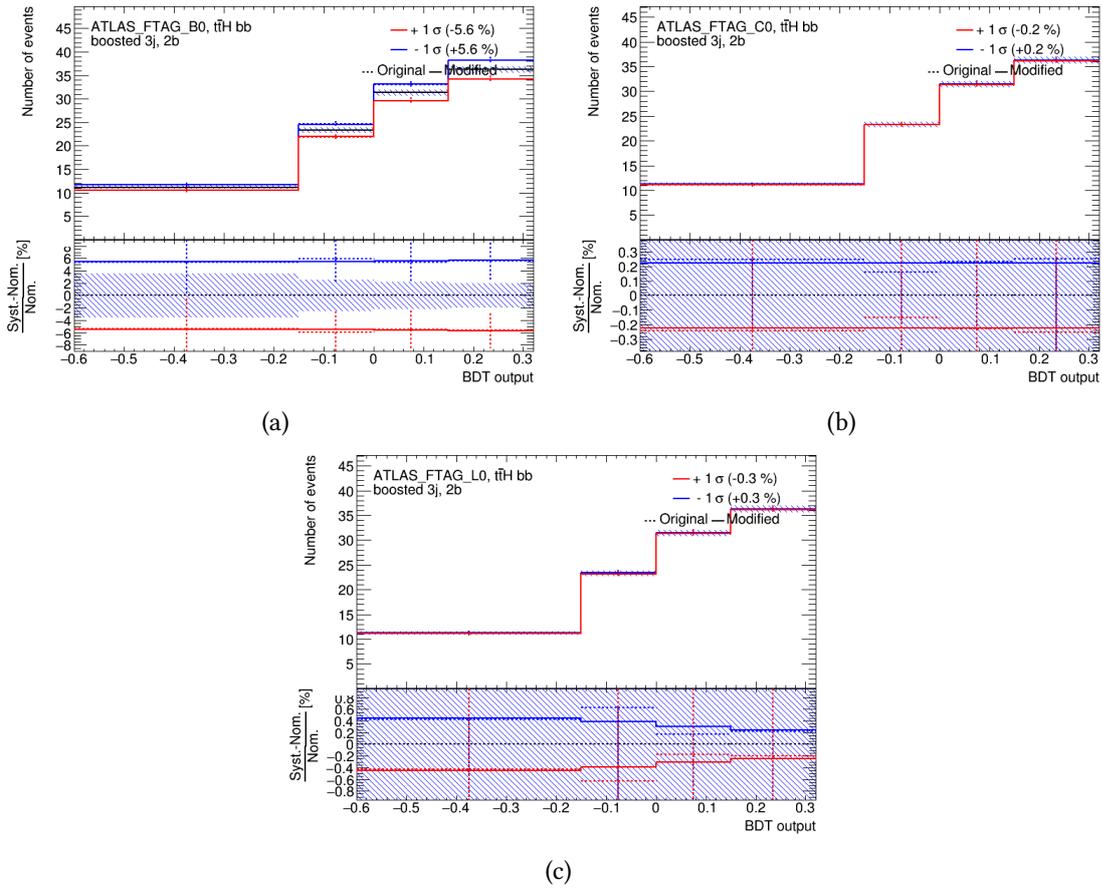
Figure 6.2: Variation of the number of events distributions depending on the systematics ($\pm 1\sigma$) due to the uncertainties related to the b- (a), c- (b) and the light-jets (c) mis-tag rate, in the boosted signal region for the $t\bar{t}H$ (with $H \to b\bar{b}$) process. Only an example component has been chosen for each of the systematics group.
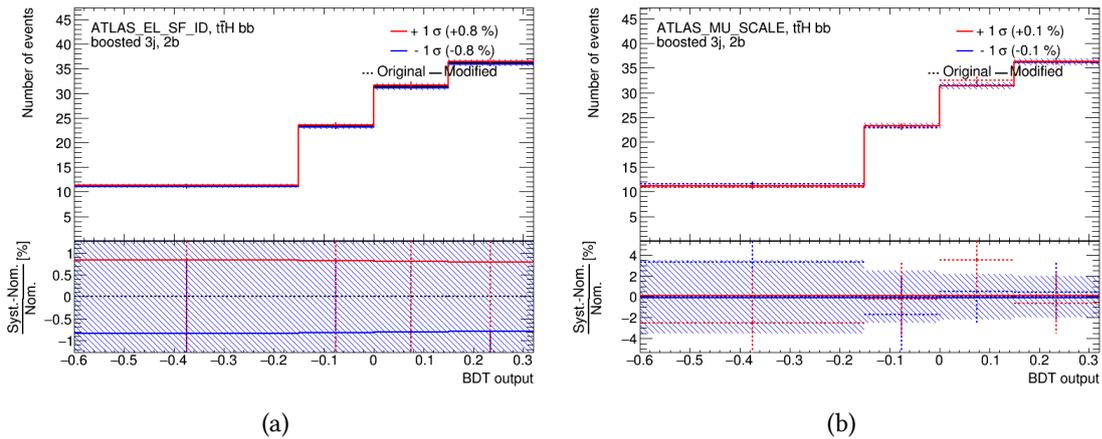
Figure 6.3: Variation of the number of events distributions depending on the systematics ($\pm 1\sigma$) due to the uncertainties related to the electron identification (a) and the muon momentum scale (b), in the boosted signal region for the $t\bar{t}H$ (with $H \to b\bar{b}$) process. Only an example component has been chosen for each of the systematics group.

2 the tracking fake rate uncertainties;

3 the tracking q/p_T bias uncertainties.

- TotalStat: the statistical uncertainty on the samples used in the steps above, both data and MC, figure 6.4d.

The uncertainties have been defined for anti-k_t (R = 1.0) reconstructed jets, which have passed the trimming process (with f_{cut} = 0.05 and R_{sub} = 0.2, see Chapter 4), and across a limited kinematic spectrum in terms of:

- p_T : from 150 to 3000 GeV;

- m/p_T : from 0 to 1.0;

- $|\eta|$: from 0 to 2.0.

Different assumptions on the correlation between the jet p_T, mass, and substructure (D_2/τ_{32}) scales (to be used since the top-tagging algorithm uses the τ_{32} variable) are considered, leading to a different combination of the systematic uncertainties components. These assumptions are strongly case dependent, therefore they should be tested to prove the specific analysis sensitivity to the correlations. Three configurations are considered:

- strong: the p_T, mass, and substructure (D_2/τ_{32}) scales are fully correlated. It results in four components;

- medium: the p_T and mass scales are fully correlated, D_2 and τ_{32} are fully correlated, but the first and second pair are uncorrelated with respect to each other. It results in eight components;

- weak: the p_T, mass and substructure (D_2/τ_{32}) scales are uncorrelated. It results in 12 components.

In this analysis, the medium correlation configuration has been chosen.

6.2.3 Modelling uncertainties

Signal modelling

The uncertainty on the signal modelling depends on different sources:

- the uncertainty on the $t\bar{t}H$ signal cross-section is $+10\% - 13\%$, including contributions from scale and PDF uncertainties, which are treated as uncorrelated [119]-[122];

- the uncertainties on the Higgs boson branching ratios amount to 1.2% for the $b\bar{b}$ decay mode [123];

- the effect of the QCD scale choice is evaluated by varying the renormalisation and factorisation scales up and down by a factor of two;

(a)

(b)

(c)

(d)
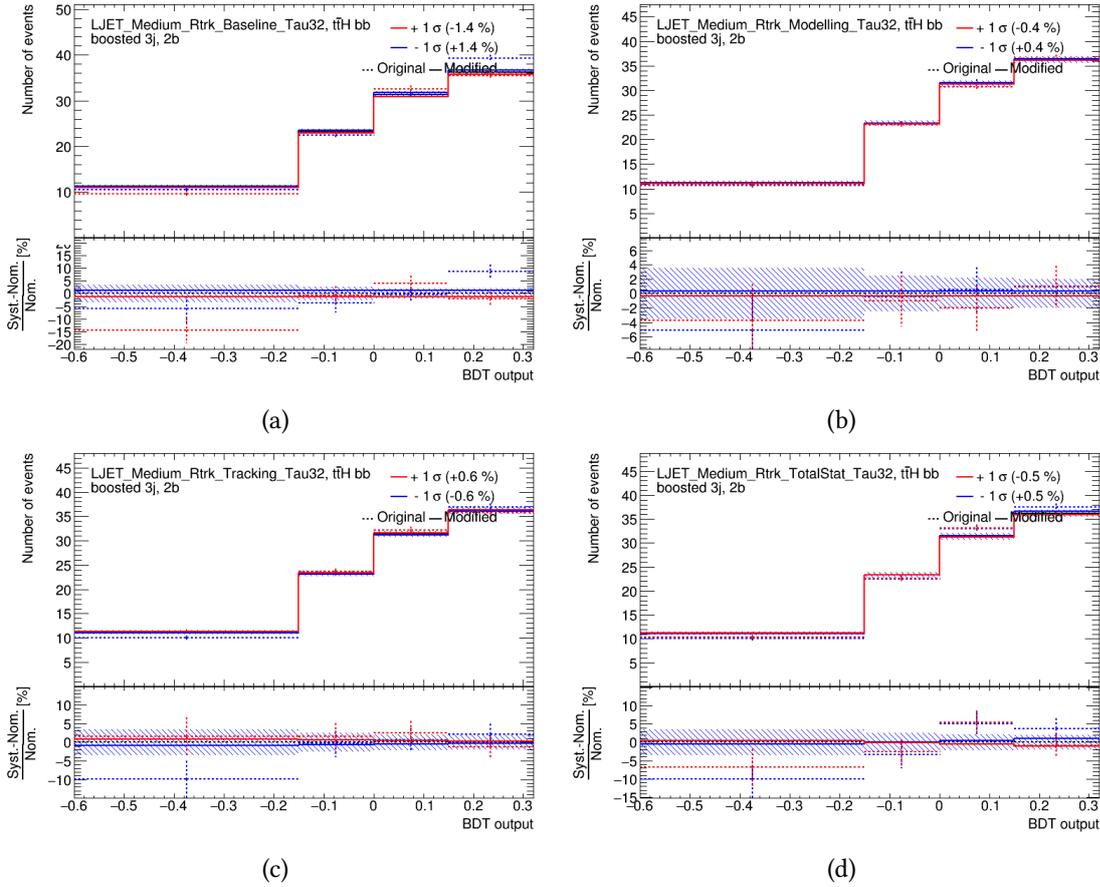
Figure 6.4: Variation of the number of events distributions depending on the systematics ($\pm 1\sigma$) due to the baseline (a), modelling (b), tracking (c) and statistical (d) uncertainties related to the large-R jets, in the boosted signal region for the $t\bar{t}H$ (with $H \to b\bar{b}$) process. Only an example component has been chosen for each of the systematics group.

$t\bar{t}$+jets modelling

A number of systematic uncertainties affecting the modelling of the $t\bar{t}$+jets background are considered, summarised in table 6.2 and described below.

An uncertainty of $\pm 6\%$ is assumed for the inclusive $t\bar{t}$ NNLO+NNLL production cross-section [125] (figure 6.5), from the theoretical calculations on Powheg.

An uncertainty associated with the choice of NLO generator is derived by comparing two alternative predictions, Powheg-Box and MG5_aMC, each of which is showered with Herwig++ 6.6a.

An uncertainty due to the choice of parton shower and hadronisation model is derived by comparing the prediction from Powheg-Box interfaced either to Pythia 6 or Herwig++ 6.6b.

The $t\bar{t}$+jets simulated sample is reweighted according to the NNLO calculation of the differential cross-section at 13 TeV [124]. As already done in the Run-1 analysis, events are reweighted in order to reproduce both the distributions of top (or anti-top) p_{T} and the $t\bar{t}$ system. The reweighteing is performed in a sequential way: the p_{T} ($t\bar{t}$) first, then the p_{T} (t). It is applied at generator level in order to make both distributions on the inclusive $t\bar{t}$ simulated sample (before any event selection) matching the normalized differential cross-section predictions. The comparison between the p_{T} distribution for t, with (fig. 6.7a) and without (fig. 6.7b) the sequential reweighting. The distributions are similar but in the case of no

Systematic source	How evaluated	$t\bar{t}$ categories
$t\bar{t}$ cross-section	$\pm 6\%$	All, correlated
NLO generator (res.)	Powheg-Box+Herwig++ vs. MG5_aMC+Herwig++	All, uncorrelated
Radiation (res.)	Variations of μ_R, μ_F and $hdamp$	All, uncorrelated
PS & hadronisation (res.)	Powheg-Box+Pythia6 vs. Powheg-Box+Herwig++	All, uncorrelated
NNLO top & $t\bar{t}$ p_{T}	Maximum variation from any NLO prediction	$t\bar{t}+\leq 1c$, $t\bar{t}$+light, uncorrelated
$t\bar{t}+b\bar{b}$ renorm. scale (rew.)	Up or down by a factor of two	$t\bar{t}+\leq 1b$
$t\bar{t}+b\bar{b}$ resumm. scale (rew.)	Vary μ_Q from $H_T/2$ to μ_{CMMPS}	$t\bar{t}+\leq 1b$
$t\bar{t}+b\bar{b}$ global scales (rew.)	Set μ_Q, μ_R and μ_F to μ_{CMMPS}	$t\bar{t}+\leq 1b$
$t\bar{t}+b\bar{b}$ shower recoils (rew.)	Alternative model scheme	$t\bar{t}+\leq 1b$
$t\bar{t}+b\bar{b}$ PDF (rew.)	CT10 vs. MSTW or NNPDF	$t\bar{t}+\leq 1b$
$t\bar{t}+b\bar{b}$ MPI	Up or down by 50%	$t\bar{t}+\leq 1b$
$t\bar{t}+b\bar{b}$ FSR	Radiation variation samples	$t\bar{t}+\leq 1b$

Table 6.2: Summary of the systematic uncertainties on the $t\bar{t}$+jets modelling. For the $t\bar{t}+\leq 1b$ background, the inclusive $t\bar{t}$ sample is reweighted to a NLO $t\bar{t}+b\bar{b}$ prediction; uncertainties on the inclusive sample are labelled *residual* (res.), while those on the NLO prediction are labelled *reweighting* (rew.).
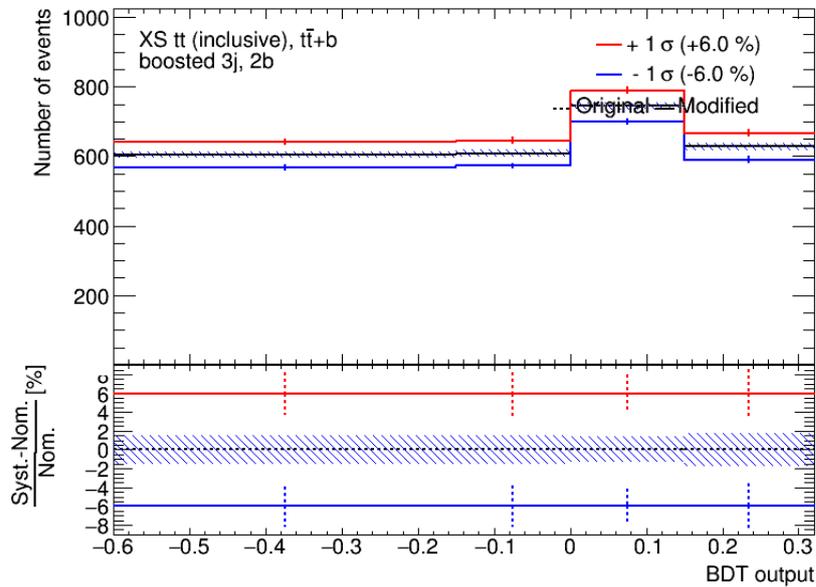
Figure 6.5: Variation of the $t\bar{t}+\geq 1b$ distribution due to the $t\bar{t}$ cross-section systematic uncertainty, in the boosted signal region.

reweighting the agreement between data and MC simulations is slightly worse.

The reweighting described above is applied in the analysis only for the $t\bar{t}$+light and $t\bar{t}+\leq 1c$ back-

(a)   (b)
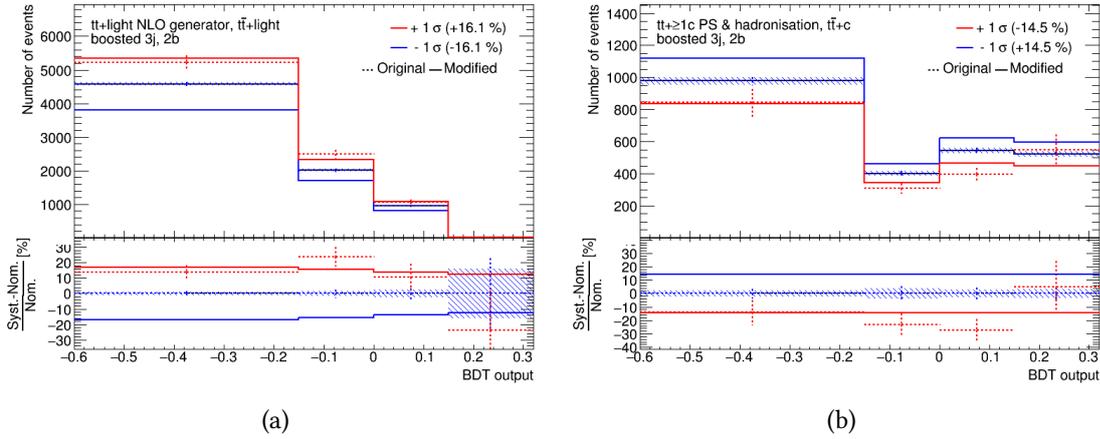
Figure 6.6: Variation of the $t\bar{t}$+light distribution due to the NLO generator choice (a) and variation of the $t\bar{t}+\geq 1c$ distribution due to the PS and hadronization model choice (b), in the boosted signal region.

(a)   (b)
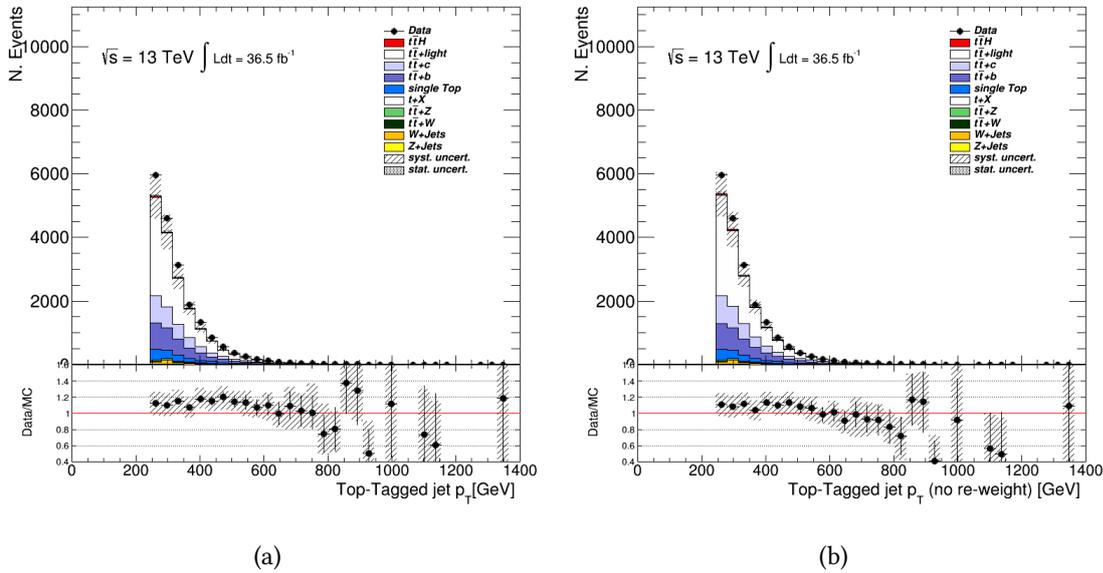
Figure 6.7: Comparison between the p_T distribution of the top-tagged large-R jet in the event, with (a) and without (b) the p_T and $t\bar{t}$ p_T reweighting.

ground components, as shown in figures 6.8a and 6.8b, since the $t\bar{t}+\leq 1b$ component is already corrected to a dedicated NLO $t\bar{t}+b\bar{b}$ prediction.

All uncertainties on $t\bar{t}$+jets background modelling, except the uncertainty on the inclusive cross-section, are considered to be uncorrelated among $t\bar{t}+\geq 1b$, $t\bar{t}+\geq 1c$ and $t\bar{t}$+light.

In the case of $t\bar{t}+\geq 1b$, all alternative samples described above are reweighted to the NLO SherpaOL prediction in the same way as the nominal, prior to evaluating the relevant uncertainty. Additional *reweighting* uncertainties on the NLO prediction are considered:

- varying the renormalisation scale up and down by a factor of two, changing the functional form of the resummation scale to μ_{CMMPS}, and adopting a global scale choice, $\mu_Q = \mu_R = \mu_F = \mu_{CMMPS}$ (see figure 6.9a);

154
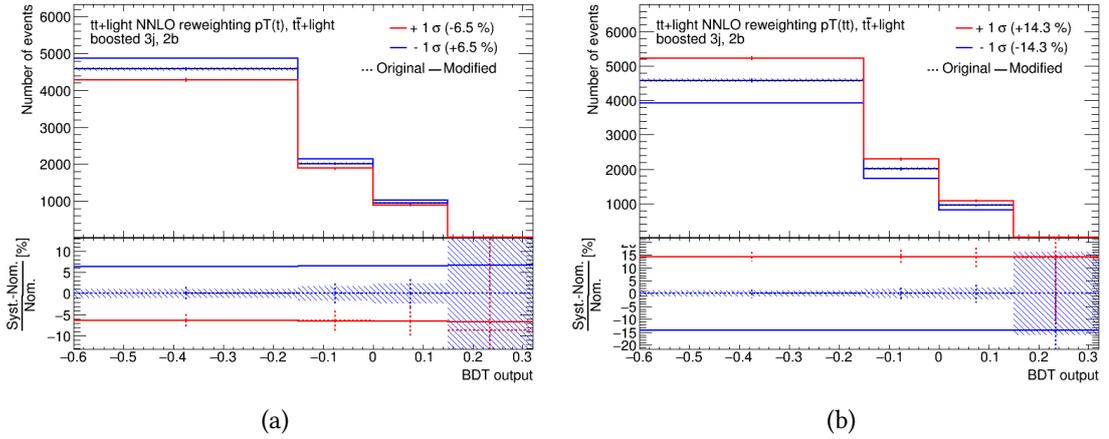
Figure 6.8: Variation of the $t\bar{t}$+light distribution due to the uncertainties related to the reweighting for the p_T of the top (or anti-top) (a) and of the $t\bar{t}$ system (b), in the boosted signal region.

- two alternative PDF sets, MSTW [126] and NNPDF (see figure 6.9b), are considered, as well as an alternative shower recoil scheme;

- separate uncertainties are applied to the $t\bar{t}+\leq 1b$ events not included in the SherpaOL prediction: a 50% uncertainty is assumed on the contribution from MPI (MultiParton Interactions), based on studies of different underlying event tunes, while the uncertainty on the FSR contribution is taken from the alternative radiation samples described above.
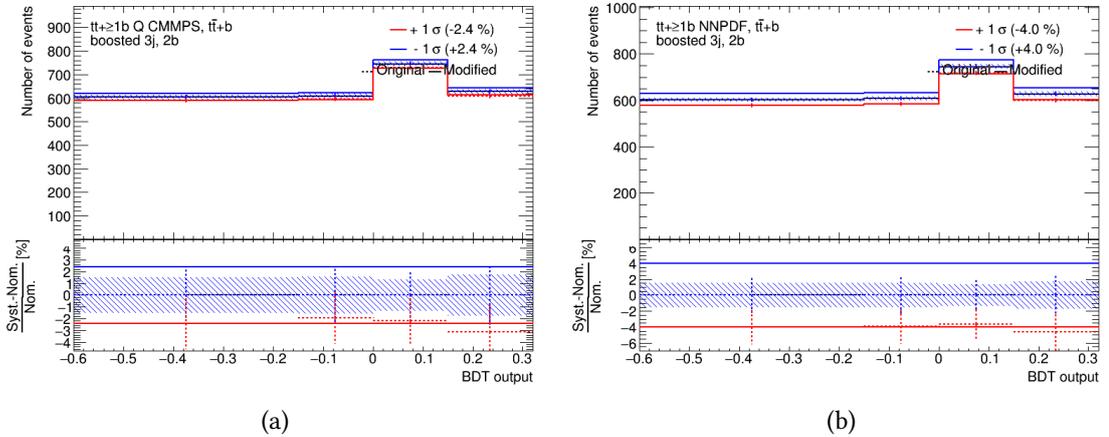
Figure 6.9: Variation of the $t\bar{t}+\geq 1b$ distribution due to the uncertainties related to the changed resummation scale μ_{CMMPS} (a) and the alternative NNPDF (b), in the boosted signal region.

As shown in figure 5.4, the data overshoot the predictions in the regions with large $t\bar{t}$+HF background components, which includes all resolved and boosted signal regions. The excess is still compatible with the prediction, given the large uncertainties associated with $t\bar{t}$+HF production [128],[129]. However, such a discrepancy could bias the fitted values of the uncertainties that affect the MVA discriminant shapes. To avoid this, the normalisation of $t\bar{t}+\leq 1b$ and $t\bar{t}+\leq 1c$ are conservatively allowed to float freely in the fit, with no prior uncertainty applied. The shape of these distributions, and the relative normalisation of the $t\bar{t}+\leq 1b$ and $t\bar{t}+\leq 1c$ sub-components, are constrained by the uncertainties described above.

Other backgrounds modelling

An uncertainty of 30% is assumed for the W/Z+jets cross-section (see figures 6.10a and 6.10b), decorrelated among jet bins for the resolved and inclusive for the boosted channel.

An additional 30% uncertainty due to W+heavy flavour jets is estimated in the resolved channel. These uncertainties are based on variations of the scales and matching parameters in Sherpa MC.

(a)   (b)
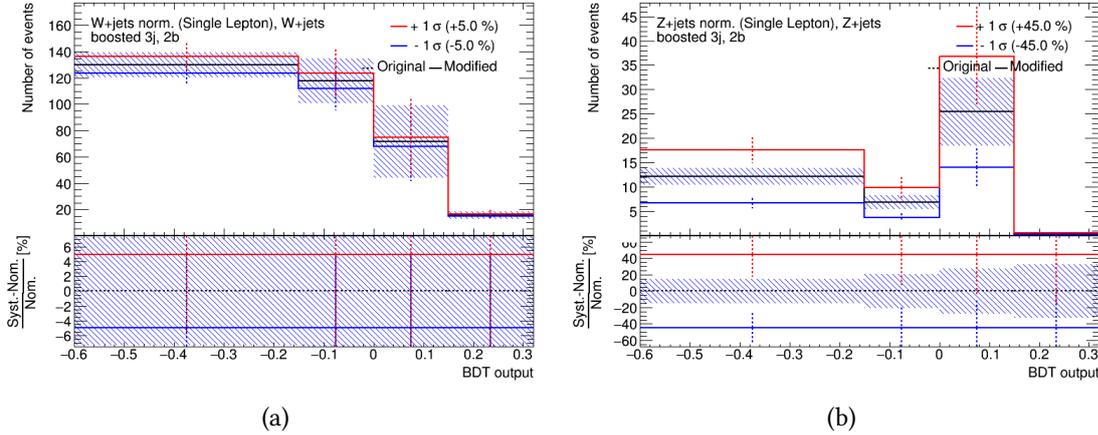
Figure 6.10: Variation of the W+jets (a) and Z+jets (b) distributions due to the uncertainties related to the relative cross-sections, in the boosted signal region.

An asymmetric uncertainty of +5% and -4% is used on the total cross-section for single-top production [130]-[132].

An additional uncertainty on initial and final-state radiation is evaluated separately in a manner similar to that used for $t\bar{t}$.

A 50% normalisation uncertainty on the diboson background is used, which includes uncertainties on the inclusive cross-section and additional jet production [134].

The uncertainty on the $t\bar{t}$+V NLO cross-section prediction is established at 15% [135].

6.3   Signal strength results

A simultaneous fit of all signal and background MC simulations and of all collected data in 2015 and 2016 (corresponding to 36.5 fb^{-1}) is performed on the discriminant distributions in the 10 control and signal regions of both resolved and boosted channels. All the details of the adopted fit procedure have already been explained in section 5.4, while the convention of all the nuisance parameters used in the fit is explained in appendix 9.

In order to build the Likelihood function, different factors are needed:

- the initial model hypothesis: S+B for this analysis;

- the composition of signal and background in each control and signal region for each fitted variable;

- the effects of the systematic uncertainties (see figures from 6.3a to 6.10b) for each nuisance parameter.

The most important free parameter of the fit procedure is the signal strength μ that is left freely floating, as the two normalization factors $k(t\bar{t}+\geq 1b)$ and $k(t\bar{t}+\geq 1c)$ that describe the contribution of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ background component, respectively. They have been left floating because of the lack of precision in the $t\bar{t}$ background simulation and the consequent mis-modelling affecting the distributions. Many other free parameters (nuisance parameters, θ) are included in the fit procedure (one for each systematics), that can vary in a bound region determined by 1σ variation at each systematic.

The minimization of the Likelihood $\mathcal{L}(\mu, \theta)$ function with respect to all its parameters provides the *best-fit* value of the measurement of the signal strength μ, crucial step for calculating the upper limit on it.

6.3.1   Signal strength on an Asimov test

As a further check, before applying the fit procedure to the real data, the fit performance is tested with Asimov data sets using all the same control and signal regions and the same MVA discriminants with the signal plus background hypothesis. This test is crucial to understand any possible mis-modelling of the dominant background contribution and problems with the systematic uncertainties implementation and provides important information on the final statistical uncertainty and on the sensitivity of the analysis.

The test has been performed with about 10^4 random realizations of pseudo-experimental data sets (for all the signal and control regions) with the same statistics of those selected from real data. A first group of results, providing the capability to put stringent limits on the signal strength $\mu = \sigma(t\bar{t}H)/\sigma_{SM}$, has been obtained using the hypothesis of null signal ($\mu = 0$). The results are shown in table 6.3, where the median 95% CL is below 1 for the resolved and combined analyses.

	Expected (μ = 0)		
	Median	$\pm 1\sigma$	$\pm 2\sigma$
Resolved-only	0.83	[0.60,1.18]	[0.45,1.64]
Combined	0.83	[0.60,1.17]	[0.44,1.63]

Table 6.3: Expected 95% CL upper limits on the signal strength, fitting on Asimov data sets.

A second group of test results has been obtained using the hypothesis of signal plus background ($\mu = 1$). The average best fit values for μ are 1.0 ± 0.4 for both resolved and combined data sets. This proves the capability of the fit to converge towards the known (input) signal strength and provides an estimation of the total uncertainty we can expect with the available data statistics and known systematic effects: $\sigma_\mu = 0.4$. The mean best fit values (the black dots) for each parameter and their uncertainty are shown in figures 6.11a (the so-called *pull plots*) and 6.11b for the resolved-only case and the combined one, respectively. As awaited in the Asimov test, the best-fit values of all the parameters are centered around the expected position (0 or 1 depending on the specific parameter), highlighting a correct fit procedure. An eventual shift of the best-fit value ("pulled" parameter) should be due to a compensation operated by the fit for some data/MC disagreements. Another important aspect is the evaluation of the uncertainty associated to the systematic effect parameters, that in the Asimov test should be close to 1. A lower uncertainty evidences a too large variation of that systematic with respect to the statistical power of data.

The effect on the best fit value μ with respect to the top 15 nuisance parameters is shown in figures 6.12a and 6.12b (the so-called *ranking plot*) for the resolved and the combined cases, respectively. The normalization factor $k(t\bar{t}+\geq 1b)$, the NLO generator and parton shower uncertainties and the QCD scale choice uncertainty have the largest effect on the measurement precision, in both the resolved and combined cases. Another common aspect of the ranking plots is that the systematic uncertainty related to the $t\bar{t}$+light parton shower and hadronization choice has a very large effect before the fit, while it has been very reduced during the fit procedure.
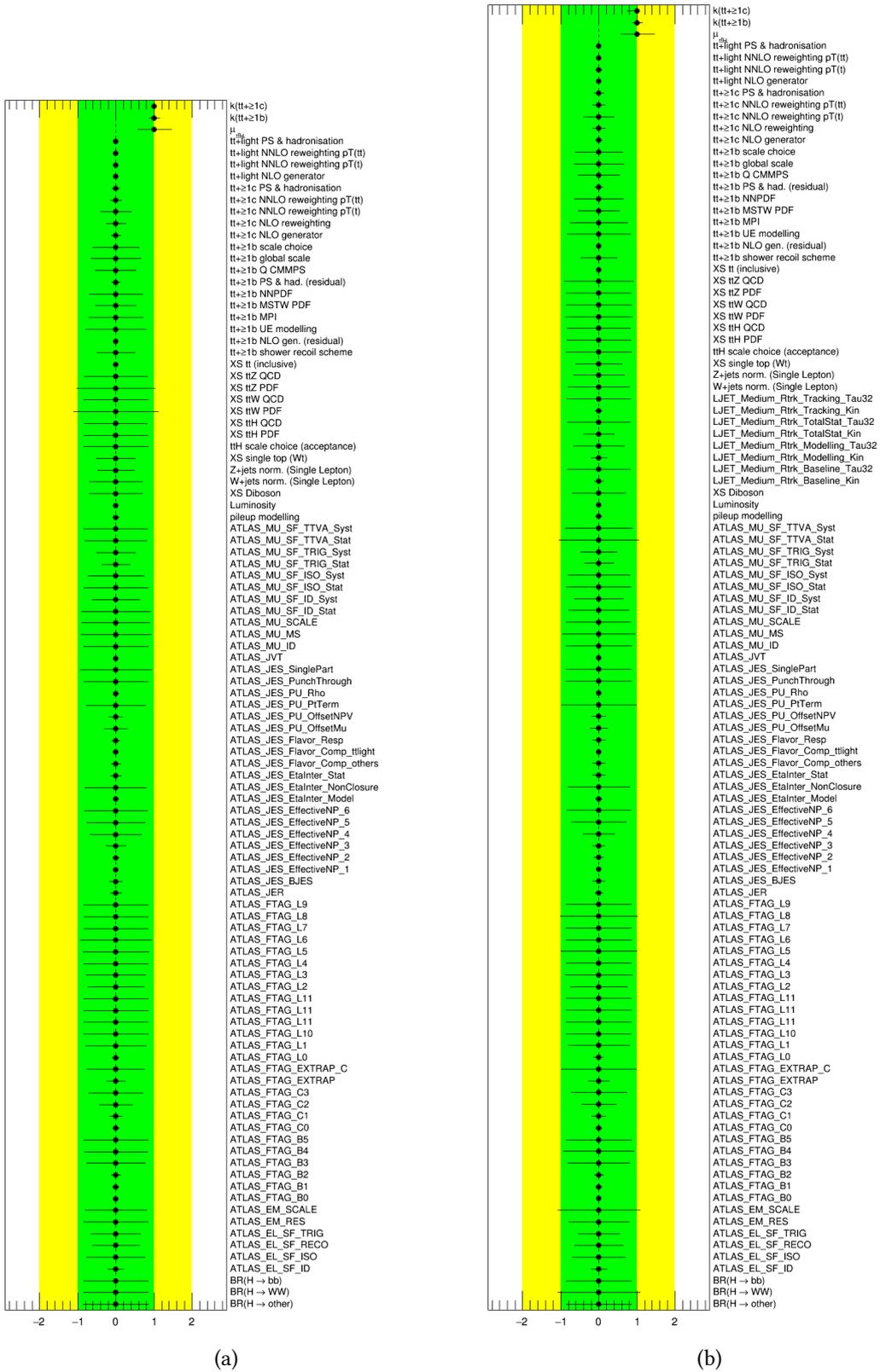
Figure 6.11: Fitted parameters (k factors, μ and nuisance parameters) using Asimov data set and signal plus background hypothesis in the resolved (a) and combined analysis (b).

Figure 6.12: Ranking of the nuisance parameters used in the fit according to their effect on $\hat{\mu}$ for the fit to Asimov data-set, for the resolved-only (a) and the combined (b) cases. The top 15 parameters are shown. The empty blue rectangles correspond to the pre-fit impact, while the filled blue ones to post-fit impact. The k parameters are the normalization factors (with respect to prediction) for the $t\bar{t}$HF (b and c) components.

6.3.2 Signal strength on the resolved analysis

The results performed by the fit procedure with a signal plus background hypothesis, without any blinding cut, are reported in this section for the analysis that includes only the resolved channel.

Figures 6.13 and 6.14 show the post-fit H_T^{had} and BDT distributions of data and MC simulations for the signal and control regions, respectively. In the signal regions, the shape of the $t\bar{t}H$ signal (normalized to the total background contribution) is significantly different with respect to the shape of the total background contribution.

(a)

(b)

(c)
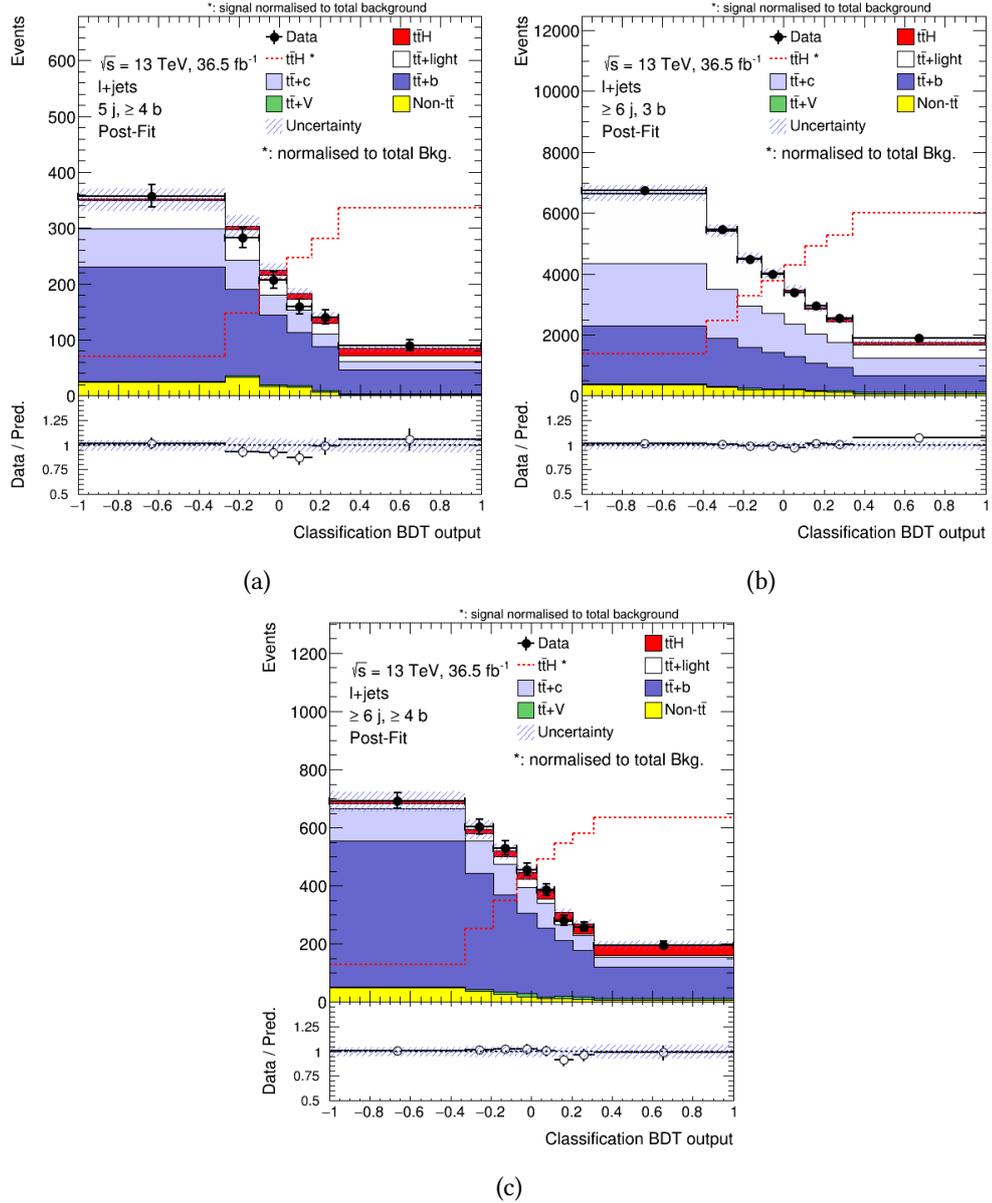
Figure 6.13: Distributions of the discriminating variables BDT output in the resolved (a), (b), (c) and boosted (d) signal regions, after the fit described below. The dotted red line represents the signal $t\bar{t}H$ normalized to the total background events.
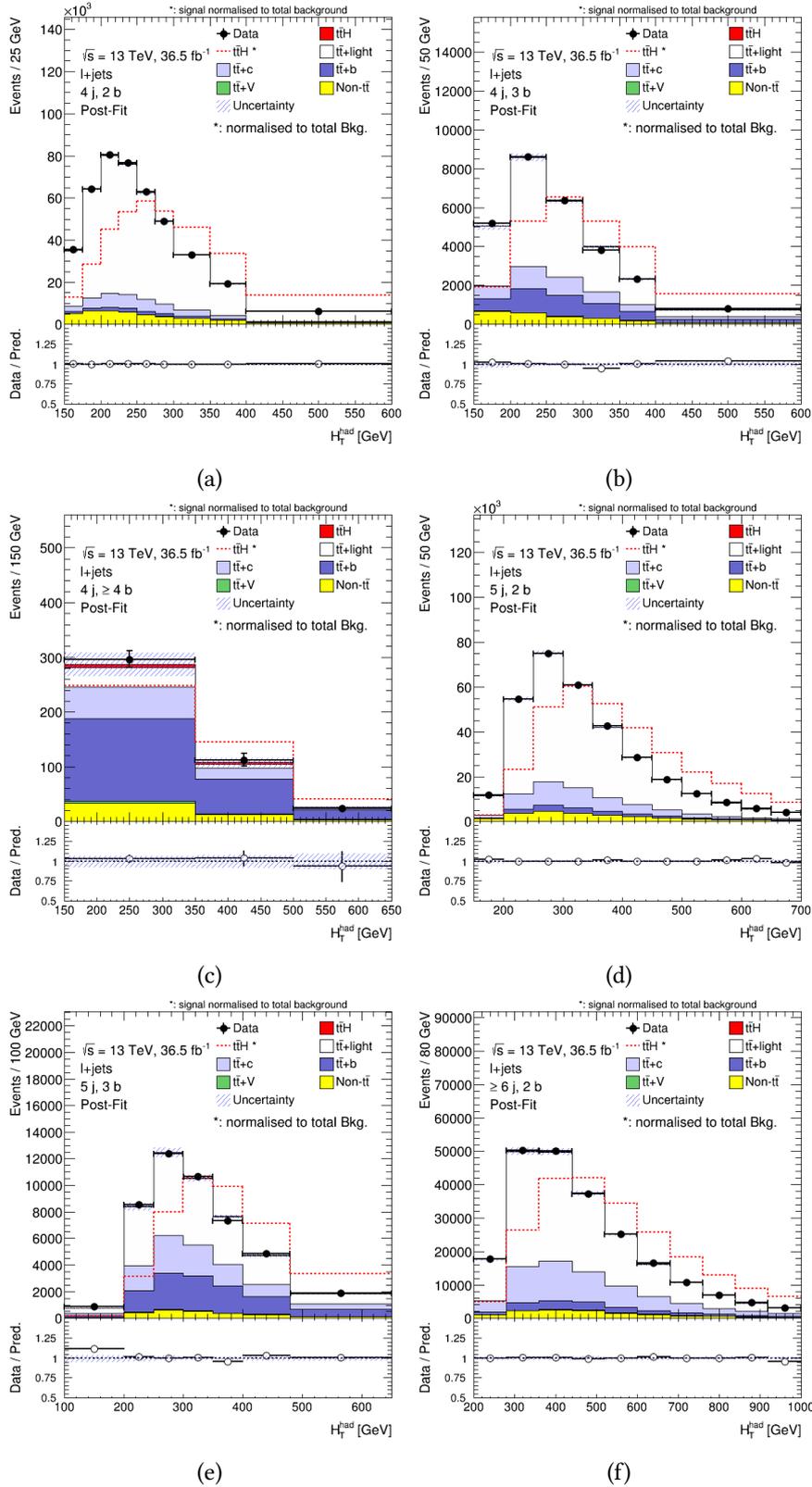
Figure 6.14: Distributions of the discriminating variables H_T^{had} in the resolved single-lepton control regions, after the fit described below.

Figure 6.15 shows the event yields of data and MC background contributions, in all the signal and control regions; after the fit procedure, the uncertainty band is significantly reduced and the data/MC agreement is improved with respect to the pre-fit situation (see figure 5.4). To facilitate the comprehension of the figure 6.15, it has to be noted that the integrated background contribution terminates with the color white, difficult to be distinguished by data, and that the red color of the $t\bar{t}H$ signal is not visible due to the small amount of events.
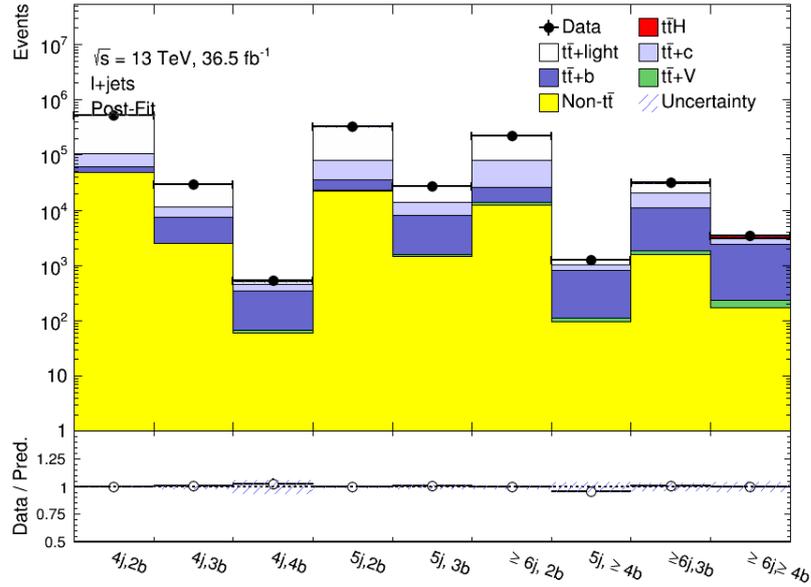
Figure 6.15: Comparison of predicted and observed event yields in the resolved single-lepton channel, after the fit procedure. The $t\bar{t}$ background is divided as described in Chapter 3, while the "non-$t\bar{t}$" background category includes contributions from single top, W/Z+jets and diboson. The hashed area corresponds to the total uncertainty on the prediction, not including an uncertainty on the normalisation of $t\bar{t}+\geq 1b$ or $t\bar{t}+\geq 1c$.

In any case, all the number of events, produced by the fit procedure, for data and MC simulation processes are reported in the tables 6.4 and 6.5, for the control and signal regions respectively.

The best-fit value for each parameter and its uncertainty are shown in figure 6.16; some systematics parameters (as, for example, $t\bar{t}+$ light NNLO reweighting $p_T(t)$ and the JER) have been shifted by the fit to adjust the data/MC agreement. Other systematics parameters (as, for example, the $t\bar{t}+$ light ($\geq 1c$) PS & hadronization and the $t\bar{t}+$ light ($\geq 1c$) NLO generator) present an uncertainty smaller than 1σ that results in a decrease of the relative systematic uncertainty. This happens when the initial systematic uncertainty is too large with respect to the data statistical error.

The effect on the best-fit value μ with respect to the top 15 nuisance parameters are shown in figure 6.17; some systematics, like the $t\bar{t}H$ QCD scale choice, the b-tagging and the $t\bar{t}$ NLO generator and PS & hadronization choice, have a large effect on the measurement precision.

The fitted values of the two free-floating normalization factors are:

$$k(t\bar{t}+ \geq 1b) = 1.1^{+0.2}_{-0.1}$$
$$k(t\bar{t}+ \geq 1c) = 1.6^{+0.3}_{-0.3}$$

(6.3.1)

found to be very close to the nominal value (expected to be 1), meaning that the $t\bar{t}$ distributions do not need special adjustments during the fit procedure.

	4j, 2b	4j, 3b	4j, ≥ 4b	5j, 2b	5j, 3b	≥ 6j, 2b
$t\bar{t}H$ ($H \rightarrow b\bar{b}$)	150 ± 50	90 ± 30	14 ± 4	210 ± 80	170 ± 60	350 ± 140
$t\bar{t}H$ ($H \rightarrow WW$)	39 ± 12	3 ± 1	0.01 ± 0.01	70 ± 20	5 ± 1	190 ± 70
$t\bar{t}H$ ($H \rightarrow other$)	44 ± 15	4 ± 1	0.07 ± 0.03	70 ± 20	7 ± 2	150 ± 50
$t\bar{t}$+light	416000 ± 10000	18000 ± 1000	50 ± 20	241000 ± 9000	13000 ± 1000	142000 ± 7000
$t\bar{t}$+≥1c	44000 ± 9000	4200 ± 700	100 ± 40	45000 ± 7000	5900 ± 700	53000 ± 5000
$t\bar{t}$+≥1b	13000 ± 2000	4900 ± 600	280 ± 20	12000 ± 2000	6600 ± 800	13000 ± 2000
$t\bar{t}$+W	270 ± 40	16 ± 2	0.3 ± 0.1	340 ± 50	26 ± 4	520 ± 70
$t\bar{t}$+Z	310 ± 40	44 ± 6	6 ± 1	390 ± 50	80 ± 10	700 ± 90
Single Top	28000 ± 2000	1400 ± 120	37 ± 7	13000 ± 1000	980 ± 90	7200 ± 700
Diboson	1400 ± 800	130 ± 80	17 ± 13	770 ± 470	40 ± 30	530 ± 330
W+jets	16000 ± 1000	790 ± 160	3 ± 2	7100 ± 800	360 ± 130	3900 ± 500
Z+jets	2900 ± 1200	160 ± 70	2 ± 1	1300 ± 500	90 ± 60	830 ± 340
t+X	94 ± 4	22 ± 1	1.8 ± 0.3	65 ± 3	12 ± 1	83 ± 5
Total	522000 ± 13000	28800 ± 1400	520 ± 50	321000 ± 1000	27000 ± 2000	222000 ± 9000
Data	521885	29417	530	321603	27419	222146

Table 6.4: Yields of the resolved control regions for each signal and background process and for data.

	5j, ≥ 4b	≥ 6j, 3b	≥ 6j, ≥ 4b
$t\bar{t}H$ ($H \rightarrow b\bar{b}$)	50 ± 16	420 ± 150	190 ± 60
$t\bar{t}H$ ($H \rightarrow WW$)	0.2 ± 0.1	28 ± 8	3 ± 1
$t\bar{t}H$ ($H \rightarrow other$)	0.7 ± 0.2	27 ± 8	4 ± 1
$t\bar{t}$+light	190 ± 70	10000 ± 1000	140 ± 60
$t\bar{t}$+≥1c	230 ± 30	9700 ± 900	640 ± 140
$t\bar{t}$+≥1b	700 ± 50	9300 ± 1100	2200 ± 150
$t\bar{t}$+W	0.7 ± 0.1	77 ± 11	5 ± 1
$t\bar{t}$+Z	15 ± 2	200 ± 30	57 ± 8
Single Top	42 ± 8	900 ± 100	104 ± 14
Diboson	2 ± 1	80 ± 50	11 ± 7
W+jets	50 ± 30	460 ± 70	36 ± 13
Z+jets	2 ± 2	60 ± 20	6 ± 3
t+X	2.5 ± 0.4	39 ± 2	14 ± 1
Total	1300 ± 100	31300 ± 1900	3400 ± 200
Data	1235	31401	3398

Table 6.5: Yields of the resolved signal regions for each signal and background process and for data.

The observed signal strength obtained in the resolved single lepton channel is:

$$\mu = 1.4^{+0.5}_{-0.5}. \tag{6.3.2}$$

A signal strength larger than 2.3 can be excluded at the 95% confidence level:

$$\mu < 2.3 \ @ \ 95\% \ \text{CL}. \tag{6.3.3}$$
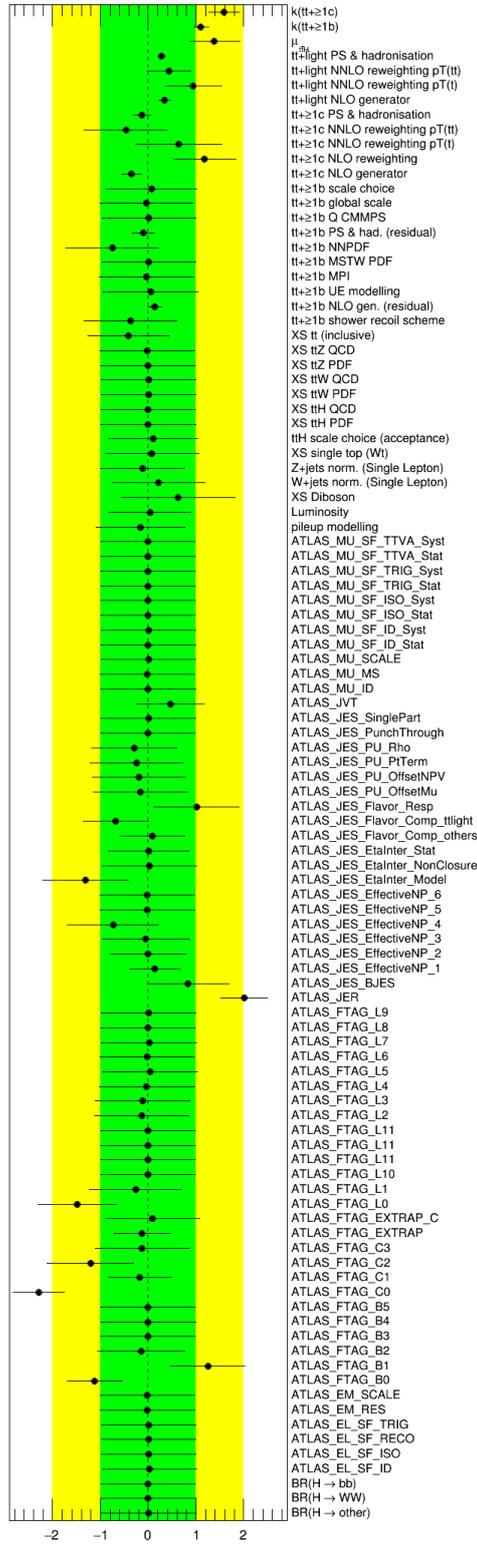
Figure 6.16: Fitted nuisance parameters using data and signal plus background hypothesis in the resolved-only analysis.
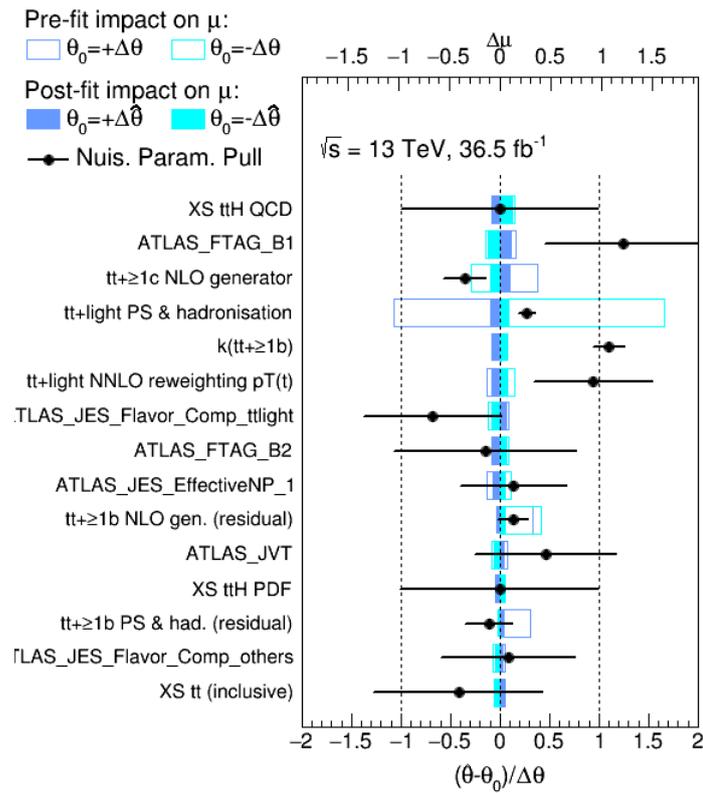
Figure 6.17: Ranking of the top 15 nuisance parameters used in the fit according to their effect on μ for the fit to data, for the resolved-only analysis. The empty blue rectangles correspond to the pre-fit impact, while the filled blue ones to post-fit impact. The k are the normalization factors (with respect to prediction) for the $t\bar{t}$HF (b and c) components.

6.3.3 Signal strength on the combined analysis

The results obtained by the fit procedure to data, with the signal plus background hypothesis, without any blinding cut, are reported in this section for the analysis that combines both the resolved and the boosted channels.

Figures 6.18 and 6.19 show the post-fit H_T^{had} and BDT distributions of data and MC simulations for the control and signal regions, respectively. After the fit procedure, the uncertainty band is significantly reduced and the data/MC agreement is better than the pre-fit one (see figures 5.17 and 5.18).
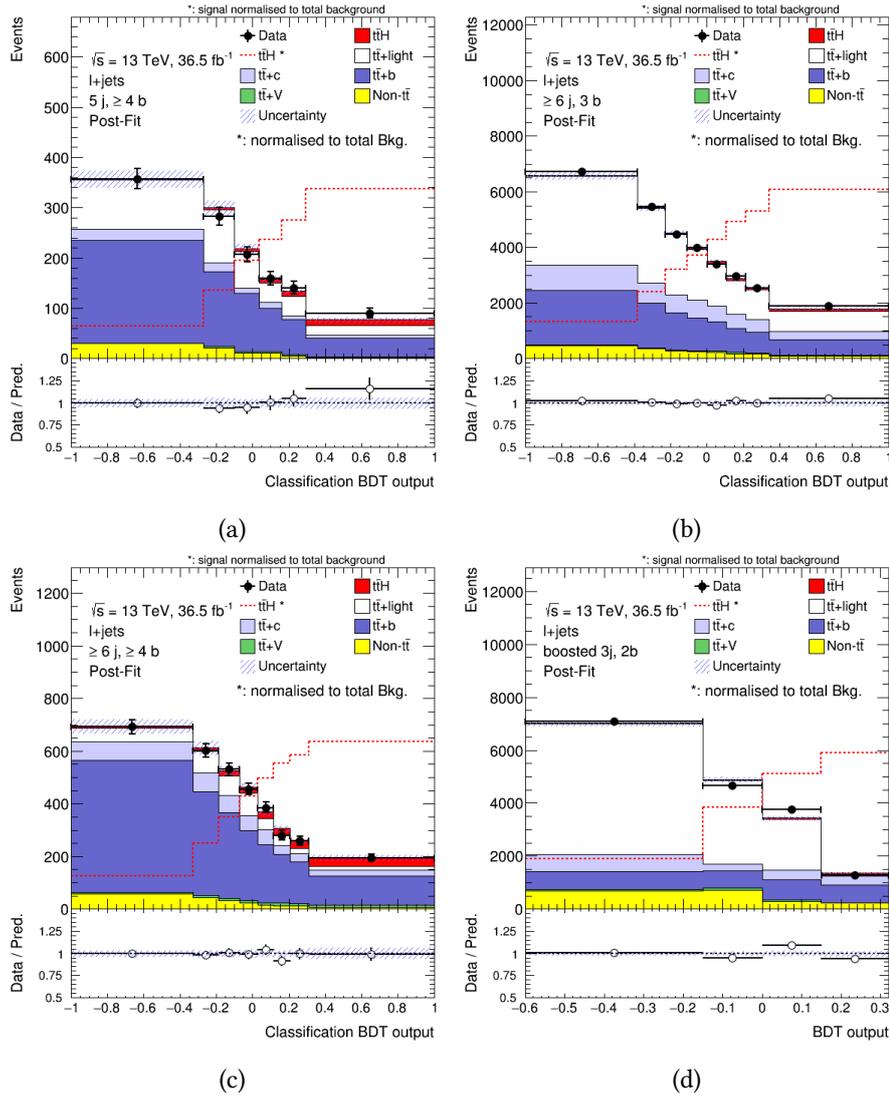
Figure 6.18: Distributions of the discriminating variables BDT output in the resolved (a), (b), (c) and boosted (d) signal regions, after the fit described below. The dotted red line represents the signal $t\bar{t}H$ normalized to the total background events.
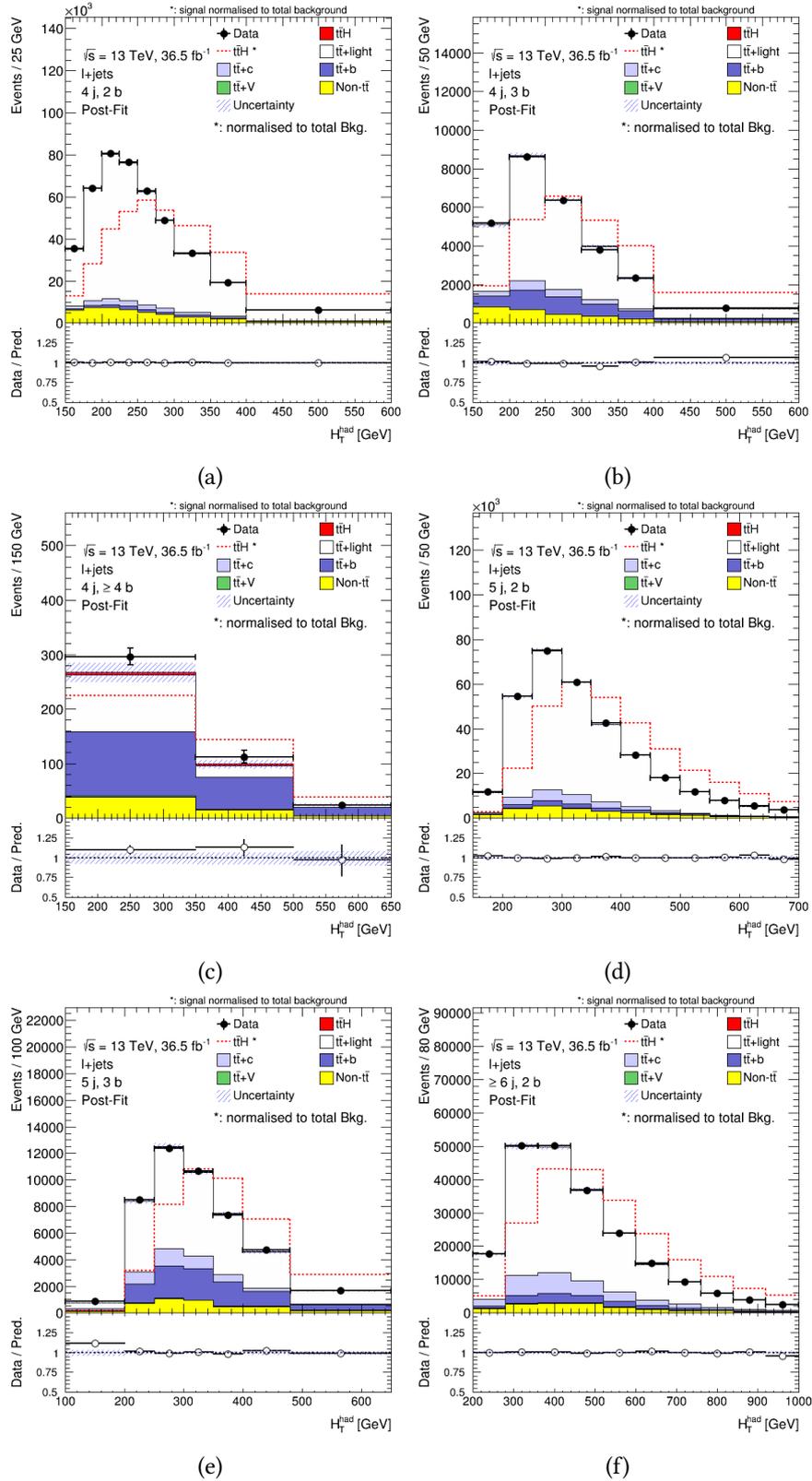
Figure 6.19: Distributions of the discriminating variables H_T^{had} in the resolved single-lepton control regions, after the fit described below.

Figure 6.20 shows the event yields of data and of background contributions, evaluated by the MC, in all the signal and control regions after the fit procedure. The last bin of the plot corresponds to the region where the boosted events have been selected.
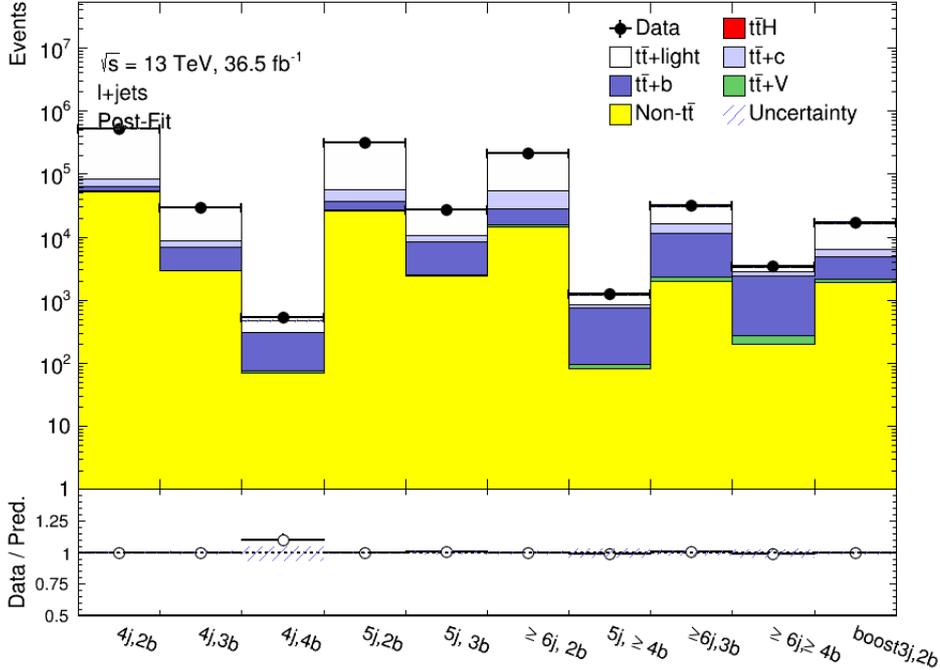
Figure 6.20: Comparison of predicted and observed event yields in the resolved and boosted single-lepton channels, after the fit procedure. The $t\bar{t}$ background is divided as described in Chapter 3, while the "non-$t\bar{t}$" background category includes contributions from single top, W/Z+jets and diboson. The hashed area corresponds to the total uncertainty on the prediction, not including an uncertainty on the normalisation of $t\bar{t}+\geq 1b$ or $t\bar{t}+\geq 1c$.

The number of events, estimated by the fit procedure, for data and MC simulation processes are reported in the tables 6.6 and 6.7, for the control and signal regions respectively.

Adding the four signal regions, the $t\bar{t}H(H \rightarrow b\bar{b})$ signal contains about 730 events, with an increament of about 10% with respect to the 650 events present in the resolved analysis.

The best-fit value for each parameter and its uncertainty are shown in figure 6.21; the situation is similar to the one obtained by the resolved analysis (see figure 6.16).

The effect on the best-fit value μ with respect to the top 15 nuisance parameters are shown in figure 6.22; some systematics, like the $t\bar{t}H$ QCD scale choice, the b-tagging and the $t\bar{t}$ NLO generator and PS & hadronization choice, have a large effect on the measurement precision. Differently from the resolved analysis, the $t\bar{t}+\geq 1b$ PS & hadronization has a larger effect than the $t\bar{t}$+light, which is absent in the combined case.

170

	4j, 2b	4j, 3b	4j, $\geq$ 4b	5j, 2b	5j, 3b	$\geq$ 6j, 2b
$t\bar{t}H$ ($H \to b\bar{b}$)	120 ± 50	70 ± 30	11 ± 4	170 ± 70	130 ± 50	290 ± 120
$t\bar{t}H$ ($H \to WW$)	30 ± 11	2 ± 1	0.02 ± 0.01	50 ± 20	4 ± 1	170 ± 60
$t\bar{t}H$ ($H \to other$)	35 ± 14	3 ± 1	0.08 ± 0.03	60 ± 20	6 ± 2	130 ± 50
$t\bar{t}$+light	439000 ± 6000	20700 ± 700	160 ± 30	262000 ± 5000	16200 ± 600	161000 ± 5000
$t\bar{t}$+$\geq$1c	18000 ± 6000	1600 ± 500	0 ± 6	19300 ± 5400	2300 ± 600	26000 ± 5000
$t\bar{t}$+$\geq$1b	10700 ± 1800	4000 ± 600	230 ± 20	10700 ± 1700	5700 ± 700	11900 ± 1800
$t\bar{t}$+W	250 ± 30	15 ± 2	0.4 ± 0.1	340 ± 40	27 ± 4	530 ± 70
$t\bar{t}$+Z	290 ± 30	40 ± 5	4 ± 1	370 ± 40	70 ± 9	690 ± 80
Single Top	28500 ± 1700	1500 ± 100	47 ± 7	14000 ± 1000	1110 ± 80	7900 ± 700
Diboson	1300 ± 600	110 ± 50	12 ± 9	800 ± 400	60 ± 30	600 ± 300
W+jets	18800 ± 1400	1110 ± 140	6 ± 2	8800 ± 800	880 ± 170	4600 ± 500
Z+jets	4700 ± 2100	200 ± 90	2 ± 1	2000 ± 1000	380 ± 190	1400 ± 600
t+X	96 ± 4	22 ± 1	2.2 ± 0.3	68 ± 3	13 ± 1	84 ± 5
Total	522000 ± 9000	29000 ± 1000	480 ± 40	319000 ± 8000	27000 ± 1000	215000 ± 7000
Data	521749	29398	530	318964	26905	214822

Table 6.6: Yields of the resolved control regions for each signal and background process and for data.

	5j, $\geq$ 4b	$\geq$ 6j, 3b	$\geq$ 6j, $\geq$ 4b	boosted 3j, 2b
$t\bar{t}H$ ($H \to b\bar{b}$)	42 ± 15	380 ± 140	170 ± 60	130 ± 50
$t\bar{t}H$ ($H \to WW$)	0.10 ± 0.04	28 ± 8	3 ± 1	18 ± 6
$t\bar{t}H$ ($H \to other$)	0.5 ± 0.2	26 ± 8	4 ± 1	17 ± 6
$t\bar{t}$+light	370 ± 70	14500 ± 800	400 ± 80	10100 ± 300
$t\bar{t}$+$\geq$1c	80 ± 30	4700 ± 900	400 ± 100	1600 ± 500
$t\bar{t}$+$\geq$1b	660 ± 50	9330 ± 970	2160 ± 140	2800 ± 300
$t\bar{t}$+W	0.8 ± 0.1	90 ± 13	7 ± 1	80 ± 10
$t\bar{t}$+Z	15 ± 2	220 ± 30	63 ± 9	151 ± 19
Single Top	64 ± 9	1190 ± 110	128 ± 14	1170 ± 90
Diboson	2 ± 1	80 ± 40	10 ± 5	70 ± 30
W+jets	10 ± 5	580 ± 80	40 ± 13	540 ± 60
Z+jets	2 ± 2	90 ± 40	9 ± 4	90 ± 40
t+X	3.1 ± 0.4	42 ± 2	15 ± 1	30 ± 2
Total	1250 ± 90	31200 ± 1500	3400 ± 200	16700 ± 700
Data	1235	31401	3398	16763

Table 6.7: Yields of the resolved and boosted signal regions for each signal and background process and for data.

171
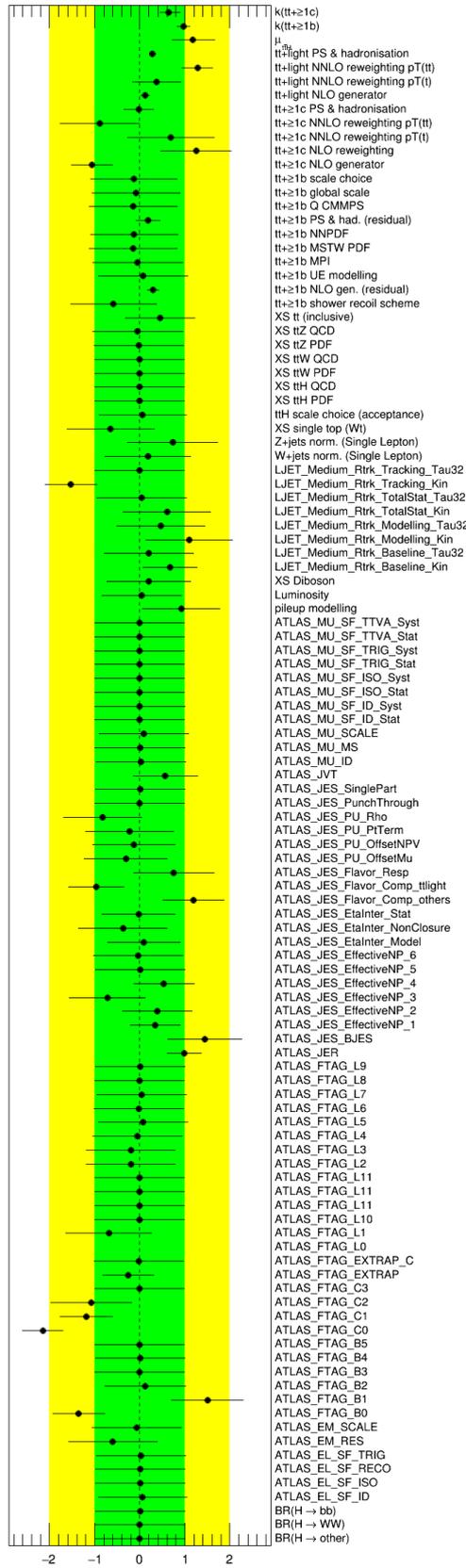
Figure 6.21: Fitted parameters (k factors, μ and nuisance parameters) using data set and signal plus background hypothesis in the combined analysis.
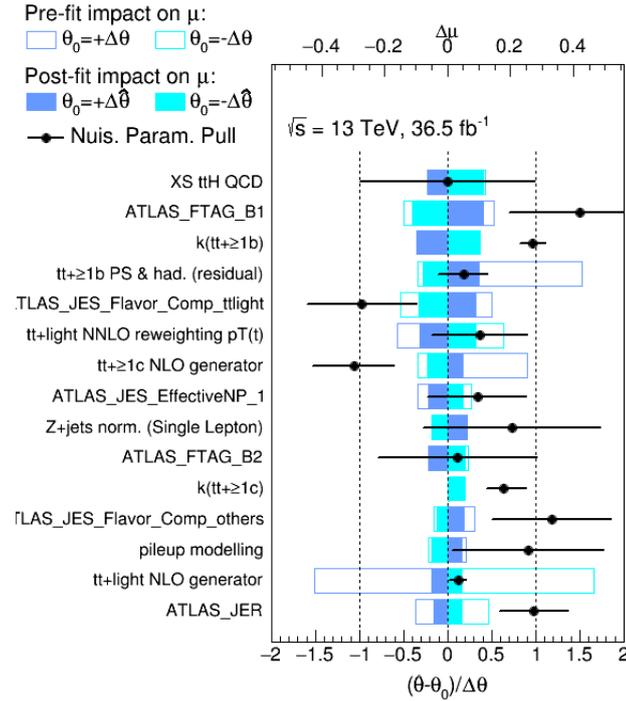
Figure 6.22: Ranking of the nuisance parameters used in the fit according to their effect on $\hat{\mu}$ for the fit to data, for the combined analysis. The top 15 parameters are shown. The empty blue rectangles correspond to the pre-fit impact, while the filled blue ones to post-fit impact. The k are the normalization factors (with respect to prediction) for the $t\bar{t}$HF (b and c) components.

The fitted values of the two free-floating normalization factors $k(t\bar{t}+\geq 1b)$ and $k(t\bar{t}+\geq 1c)$ are listed in table 6.8, in comparison with the ones obtained by the resolved analysis. As shown, the values are compatible with the expected prediction inside 1σ and 2σ for $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$, respectively.

	$k(t\bar{t}+\geq 1b)$	$k(t\bar{t}+\geq 1c)$
Resolved-only	$1.1^{+0.2}_{-0.1}$	$1.6^{+0.3}_{-0.3}$
Combined	$0.9^{+0.1}_{-0.1}$	$0.6^{+0.2}_{-0.2}$

Table 6.8: Fitted values of the free-floating normalization factors for the $t\bar{t}$HF (b and c) components, for the resolved-only and combined analyses.

The observed signal strength obtained by the fit combining both the resolved and the boosted data is

$$\mu = 1.2^{+0.5}_{-0.5} \tag{6.3.4}$$

as shown in figure 6.23.

A signal strength larger than 2.0 can be excluded at the 95% confidence level, as seen in figure 6.24 and table 6.9, in comparison with the values obtained in the resolved analysis.

Both the resolved and the combined results are compatible with the prediction of the SM $t\bar{t}H$, but do not have enough sensitivity to exclude the null signal hypothesis. Nevertheless, the addition of the boosted channel, sensitive to a significantly different kinematic region due to the higher p_{T}, constrains
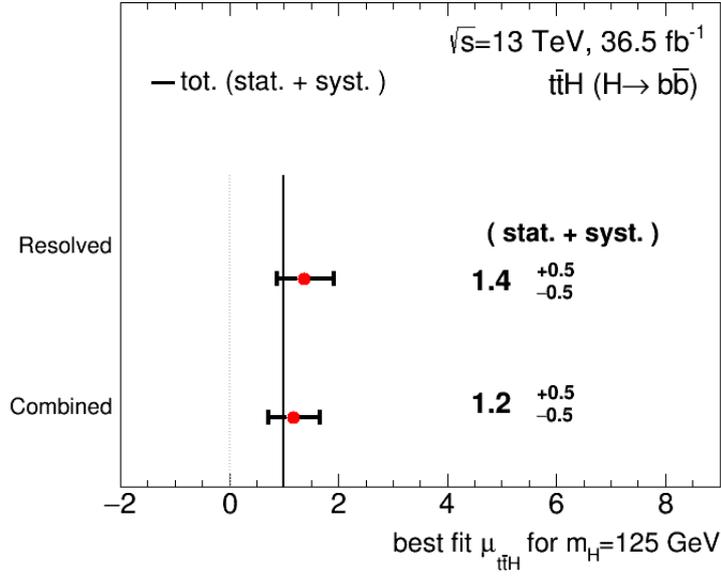
Figure 6.23: Summary of the signal strength measurements in the resolved-only and combined channels.

	Observed	Expected ($\mu = 0$)		
	(data)	Median	$\pm 1\sigma$	$\pm 2\sigma$
Resolved	2.3	1.0	[0.7,1.4]	[0.5,1.9]
Combined	2.0	0.9	[0.7,1.3]	[0.5,1.9]

Table 6.9: Observed and expected 95% CL upper limits on the signal strength.

in a stronger way the upper limit on μ to values that are closer to the SM predictions. Another important aspect to be noted in the combined analysis is that, despite the introduction of new reconstruction algorithms (large-R reconstruction and top-tagging), with the consequently inclusion of new related systematic contributions, the uncertainty of the best-fit value and the upper limit does not increase with respect to the one obtained in the resolved analysis. These reasons, therefore, encourage this analysis strategy, leading to more improvements and new refinements.
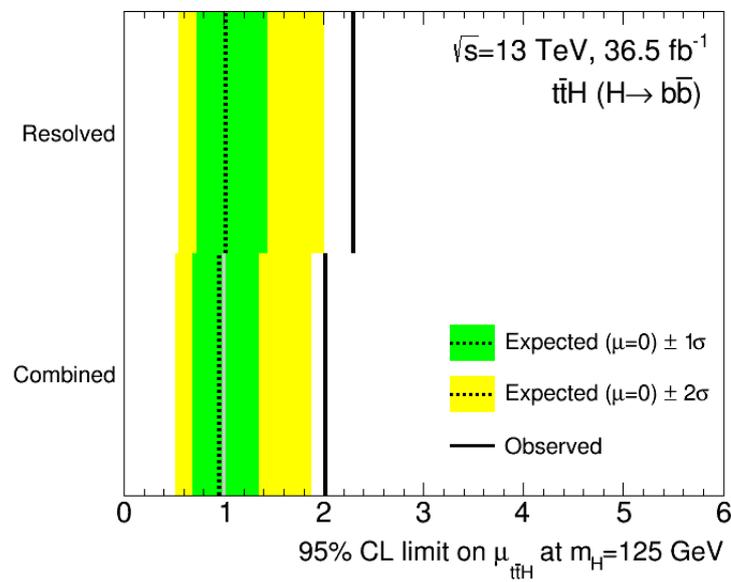
Figure 6.24: 95% CL upper limits on $\sigma_{t\bar{t}H}$ relative to the SM prediction, for the resolved and the combined analyses. The observed limits (solid lines) are compared to the expected (median) limits under the background-only hypothesis and under the signal-plus-background hypothesis assuming the SM prediction for $\sigma_{t\bar{t}H}$ and pre-fit prediction for the background. The surrounding shaded bands correspond to the 68% and 95% confidence intervals around the expected limits under the background-only hypothesis, denoted by $\pm 1\sigma$ and $\pm 2\sigma$, respectively.

6.4   Comparison with previous results

6.4.1   ATLAS results

In the ATLAS experiment, the $t\bar{t}H$ signal strength and its upper limit have been evaluated both in Run-1 (8 TeV) [36] and in the early Run-2 (13 TeV) [37], with an integrated luminosity of 20.3 fb^{-1} and 13.2 fb^{-1}, respectively for the resolved channel only. During both the analyses, the single-lepton and the dilepton channels have been studied individually and, at the end, combined, differently from the approach followed in this thesis that considers only the single-lepton channel.

Figures 6.25a and 6.25b give a summary of the results obtained in the Run-1: a signal 3.4 times larger than predicted by the SM is excluded at 95% CL using the CLs method. The observed signal strength μ is 2.8±2.0 fitting the dilepton data alone, and 1.2±1.3 fitting the single-lepton data; the combined value is 1.5±1.1.

(a)   (b)
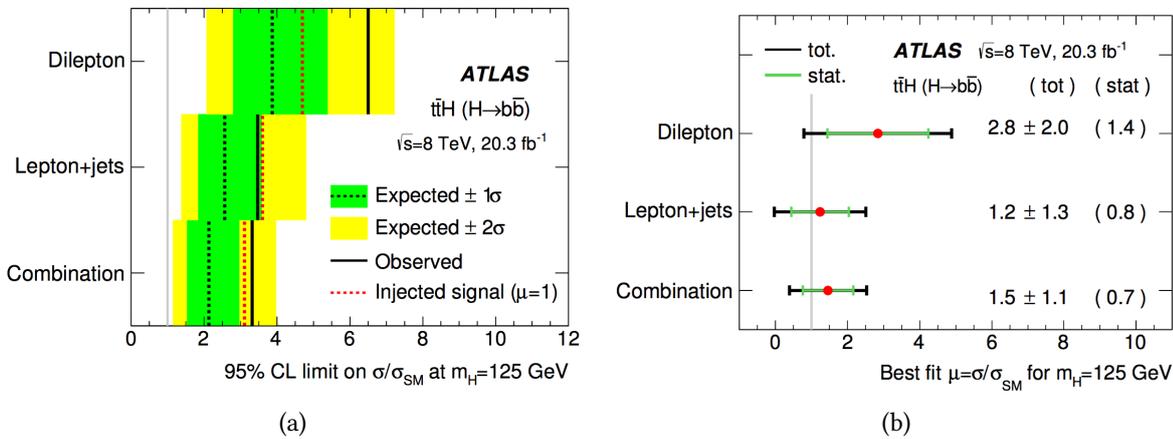
Figure 6.25: (a) 95% CL upper limits on $\sigma_{t\bar{t}H}$ relative to the SM prediction, for the individual channels as well as their combination, in the Run-1. (b) The fitted values of the signal strength and their uncertainties for the individual channels and their combination, in the Run-1. [36]

Figures 6.26a and 6.26b give a summary of the results obtained in the Run-2: a signal larger than 4.0 can be excluded at the 95% confidence level. The observed signal strength μ is $4.6^{+2.9}_{-2.3}$ fitting the dilepton data alone, and $1.6^{+1.1}_{-1.1}$ fitting the single-lepton data; the combined value is $2.1^{+1.0}_{-0.9}$.

In both the Run-1 and the Run-2 analysis, the combination of the two channels provides a more sensitive exclusion limit, even if the uncertainties (both statistical and systematics) effect on the measurement precision has not been significanlty improved. This is due to some systematic uncertainty sources (for example $t\bar{t}+\geq 1b$ background modelling) that affect the measurement in both Run-1 and Run-2 analyses and the same strategy used in the analysis.

Comparing these two results with the one reported in this thesis, corresponding to a luminosity of 36.5 fb^{-1} (figure 6.24), it is evident that the exclusion limit is much more sensitive. In fact, the last result provides a limit that is outside the 95% CL band, differently with respect to the two previous limits, which are still inside the band. Moreover, the uncertainty on the measurement precision (figure 6.23) is almost a factor 2 better than the early Run-2 result (and consequently also the Run-1 value). This effect is principally due to the increase of a factor 3 of the statistics (with respect to the Run-2 result).
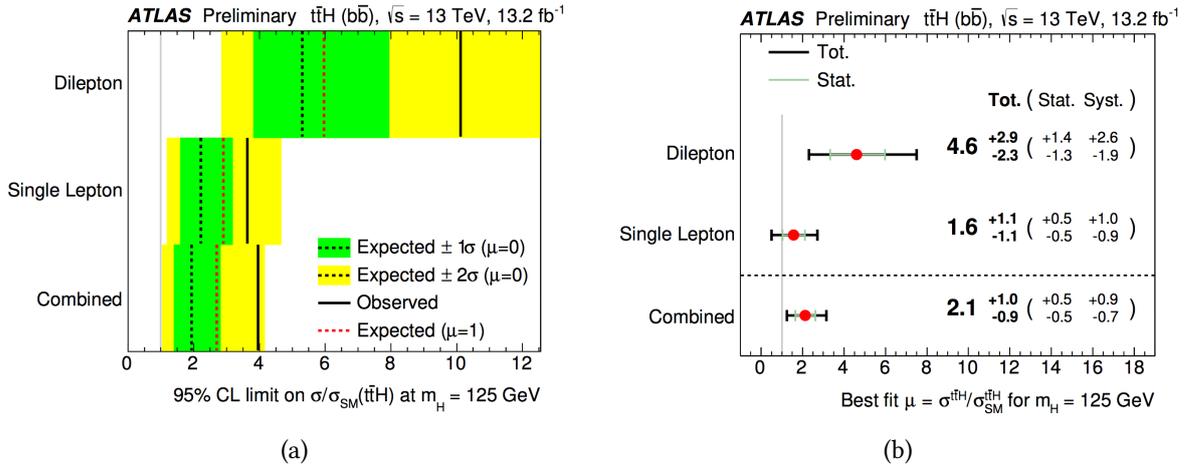
Figure 6.26: (a) 95% CL upper limits on $\sigma_{t\bar{t}H}$ relative to the SM prediction, for the individual channels as well as their combination, in the Run-2. (b) The fitted values of the signal strength and their uncertainties for the individual channels and their combination, in the Run-2. [37]

6.4.2 CMS results

The results of the search for the associated production $t\bar{t}H$ in proton-proton collisions at a center-of-mass energy of $\sqrt{s}$ = 13 TeV have been provided also by the CMS collaboration, with data corresponding to an integrated luminosity of 12.9 fb^{-1} [136]. Since the studied channel is the same as the resolved one, it is very important to compare two totally independent measurements of the same observables (the upper limit and the best-fit of the signal strength).

Moreover, it is important to compare two parallel measurements, obtained through different analysis strategies, looking at the similarities and the differences between the two processes. In order to increase the sensitivity of the search, in both ATLAS and CMS analyses, the selected events are split into several categories with different expected signal and background rates. Differently, CMS has only four regions for the single-lepton channel and three for the dilepton one. In the CMS analysis, the BDT method has been used to separate events in a low-BDT and a high-BDT categories; in each sub-categories the Matrix Element Method (MEM) has been used as the final discriminant in the fit procedure.

Another important difference between the two analyses is that the CMS Collaboration used the same $t\bar{t}H$ and $t\bar{t}$ MC simulation samples but with a new specific tuning in order to improve the data/MC agreement for the jet multiplicity. Since $t\bar{t}H$ events typically show a high jet multiplicity, this helps in reducing the background mismodelling which has a very strong effect on the ATLAS result (especially the effect related to the $t\bar{t}+\geq 1b$ component).

A combined fit of the discriminant distributions in all categories results in an observed (expected) upper limit of $\mu < 1.5$ (1.7) at the 95% confidence level, as shown in figure 6.27a. A best fit value of $\mu = -0.19^{+0.45}_{-0.44}$(stat.) $^{+0.66}_{-0.68}$(syst.), shown in figure 6.27b.

The results obtained in this thesis are below the CMS upper limit and compatible within 2 σ from the CMS point value which sits in the unphysical region ($\mu < 0$).
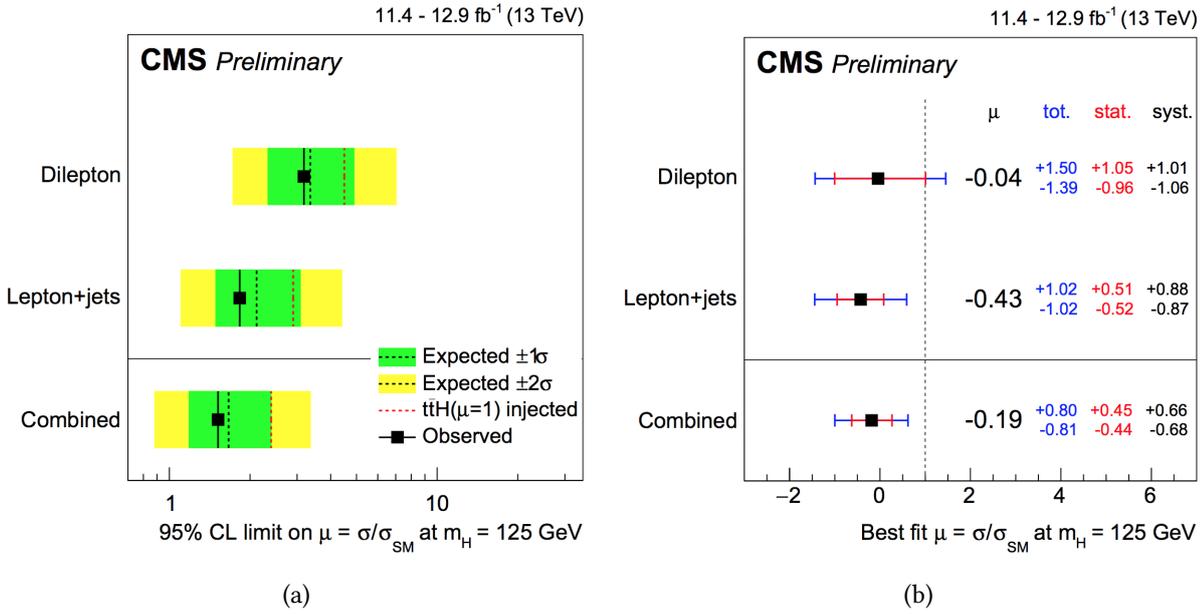
Figure 6.27: (a) Median expected and observed 95% CL upper limits on μ. The expected limits are displayed together with $\pm 1\sigma$ and $\pm 2\sigma$ confidence intervals. Also shown are the limits in case of an injected signal of $\mu = 1$. (b) Best-fit values of the signal strength modifiers μ with their $\pm 1\sigma$ confidence intervals, also split into their statistical and systematic components. [136]

6.5 Future perspectives

The strength of the Yukawa interaction between the top quark and the Higgs boson has played a fundamental role in the recent discovery of the Higgs boson. Thanks to its large value, the Higgs production in gluon fusion (ggF), which mostly proceeds through a top-quark loop in the SM (see section 1.3.3), has already provided the necessary statistics for a discovery with a modest integrated luminosity acquired during the Run-1 of LHC. For this reason, it is crucial to set the best determination of the upper limit on the strength of the Yukawa interaction ($\mu_{t\bar{t}H}$) in the whole kinematic range (including resolved and boosted regions) as performed in this thesis. Beyond the evaluation of the strength, it is fundamental to determine the structure and the properties of such interaction, in particular whether the Higgs-top-quark coupling is CP violating, i.e. the Higgs interacts with quarks and leptons through a scalar or pseudoscalar coupling. This is equivalent to test if the Higgs boson is a scalar particle 0^{+} as predicted by the SM (and in this case the Higgs-top-quark coupling is not CP violating), or if it is a pseudo-scalar 0^{-} or a mixed $0^{\pm}$ boson. In this context, it is important to stress that so far all experimental determinations of the Higgs CP properties [137]-[139] have been obtained from the $H \rightarrow VV \rightarrow 4l$ decay mode and therefore only constrain the HVV interactions. The determination of the CP properties of the top-quark Yukawa interaction is difficult because, there is no decay mode of the Higgs to or through top quarks that can be effectively studied at the LHC and consequently only Higgs production can be considered. In addition, even if different couplings (either scalar, pseudoscalar or mixed) have an impact on the production rates [140] and can also be bound by indirect measurements [141], only specially designed observables can provide direct evidence of CP-violating effects at hadron colliders. In the $t\bar{t}H$ channel, the information on the CP nature of the top-quark coupling is encoded in the correlations between the top-antitop decay products. For this reason, the $t\bar{t}H$ production plays a crucial role in the study of the CP nature of the Yukawa coupling. It is interesting to compare low- and high-p_{T} regimes because some CP-sensitive variables have different

dependence on the p_T of the Higgs particle. Jet substructure variables and boosted techniques can help in this effort.

The production of a Higgs-like spin-0 particle (X_0) with CP-mixed coupling to the top quark, in gluon-gluon fusion and in association with a top-quark pair has been simulated in a aMC@NLO framework, including parton-shower effects [142]. In table 6.10 the total $t\bar{t}X_0$ cross section at NLO is reported for the three different CP states; the production rate for the pseudoscalar case is unequivocally larger than that for the scalar case. Such a difference is proportional to the top-quark mass, as the amplitudes for the scalar and pseudoscalar interactions are identical in the limit where the Yukawa coupling is kept constant and the quark mass is neglected. Also for this reason it is clear that a precise determination of the $t\bar{t}H$ cross section could give important information on the Higgs parity.

CP scenario	σ_{NLO} (fb)
0^+	$525.1^{+5.7}_{-8.7} \pm 2.1\%$
0^-	$224.3^{+6.8}_{-10.5} \pm 3.2\%$
$0^\pm$	$374.1^{+6.0}_{-9.3} \pm 2.5\%$

Table 6.10: NLO cross-sections for $t\bar{t}+X_0$ at the 13 TeV LHC, for the three different CP scenarios. The quoted errors contain the fractional scale (left) and PDF ($+\alpha_S$) (right) uncertainties.

The differential cross sections for $t\bar{t}X_0$ production at the 13-TeV LHC as a function of the transverse momentum of the resonance $p_T(X_0)$ is shown in figure 6.28. The difference between the various scenarios is significant in the low-p_T region, while the high-p_T tail of the distributions, featuring exactly the same shape, are not sensitive to the CP mixing. The final experimental answer on which CP state better describes the prediction will arrive only when the acquired statistics will allow the determination of the $t\bar{t}H$ differential cross section. At the moment a preliminary result can be reached by evaluating the $t\bar{t}H$ cross section in a low and in a high p_T range of the Higgs (that is in the resolved and in the boosted regimes), measuring their ratio (which is independent by global underestimation or overestimation of the cross section) and evaluating which one better fits the expectation. This underlines the importance of the boosted analyses presented in this thesis regarding the study of the properties of the Higgs-top-quark coupling.

Many different CP-sensitive observables has been studied theoretically [142] to check if they are also sensitive to the boosted regimes, as shown in figures 6.29a and 6.29b. Compared to the SM, a CP-odd X_0 tends to be produced more centrally, while the accompanying top quarks are more forward. The most sensitive distribution to CP mixing is the rapidity difference between the top and antitop $\Delta\eta(t,\bar{t}) \equiv \eta(t) - \eta(\bar{t})$. This observable is hardly affected by the $p_T(X_0) > 200$ GeV cut, thus the correlations among the top-antitop decay products provide a good CP-discriminating power also in the boosted regime. This provides another good motivation for a future study using the boosted technique.

Many other interesting future developments of this research are possible. Future data, to be acquired in the next years, corresponding to an integrated luminosity 150 fb^{-1} at the end of 2018 and 300 fb^{-1} at the end of 2023 (see Chapter 2), will allow to verify the most recent theoretical predictions like the QCD corrections at NLO to the $t\bar{t}H$ cross-section [143].

With a very high integrated luminosity (3000 fb^{-1} or, at least, 300 fb^{-1}) it will be possible to probe the Higgs self coupling via a single Higgs production, looking at the trilinear Higgs self coupling that is an alternative to the direct measurement of Higgs pair production total cross sections and differential
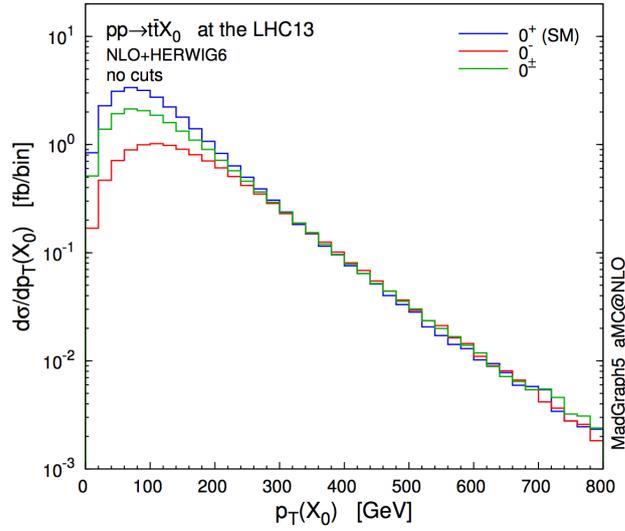
Figure 6.28: Distribution of the transverse momentum of X_0 in at $pp \to t\bar{t}X_0$ the 13 TeV LHC. [142]
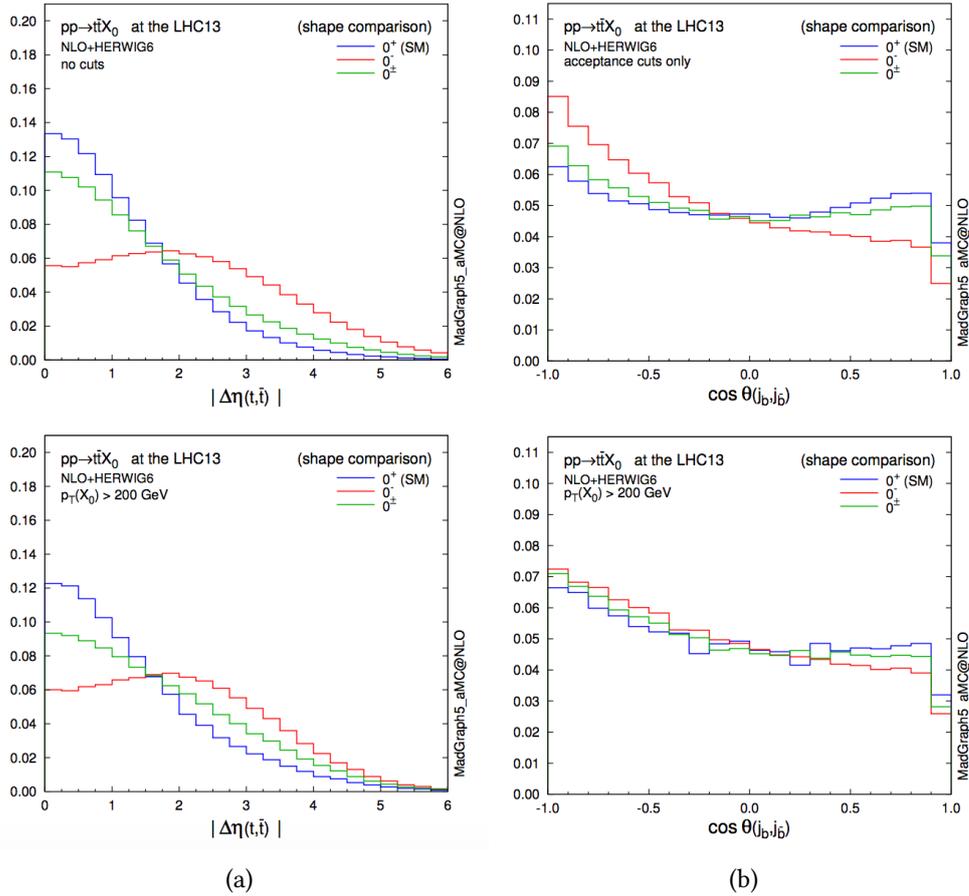
Figure 6.29: Shape comparisons of normalized distributions for the $pp \to t\bar{t}X_0$ process without cuts (top), while with the $p_T(X_0) > 200$ GeV cut (bottom).

180

distributions [144]. In this study, it has been shown that the $t\bar{t}H$ production poses strong constraints on the coupling parameters and can allow to determine them with high precision. For this reason, this could be a possible future development after at least 300 fb^{-1} luminosity of collected data by LHC, since the theoretical calculations require $\sim 10\%$ (or less) uncertainty on the measurement. As shown in this chapter, increasing the luminosity to 36.5 fb^{-1} important improvements have been reached with respect the previous results, paving the way for these kind of studies.

Conclusions

The production of a Higgs boson in association with a top quark pair has a particular importance among the main Higgs boson production modes. At LHC, the $pp \rightarrow t\bar{t}H + X$ cross section is predicted to increase with respect to the centre-of-mass energy of the pp collision faster than other Higgs production channels. This allows a better determination of its signal strength ($\mu = \sigma_{obs}/\sigma_{t\bar{t}H_{SM}}$) or a lower value of its upper limit with respect to the previous data taking at 8 TeV. The $t\bar{t}H$ production is the only channel that allows a direct measurement of the quark top-Higgs Yukawa coupling (expected to be close to one), allowing to test the SM prediction and to constrain models on new physics phenomena in particle interactions. The determination of the $t\bar{t}H$ production cross section is also essential in the understanding of the dynamics of the Higgs boson and its intrinsic characteristics. For example, recent theoretical calculations foresee a dependence of the $t\bar{t}H$ production cross-section, evaluated in different kinematic ranges, on the Higgs CP state or $(t\bar{t}) - H$ coupling.

The measurements of the $t\bar{t}H$ signal strength and its upper limit at a center-of-mass energy of pp collision of 13 TeV are presented in this thesis. The pp data collected in 2015 and 2016 with the ATLAS detector, corresponding to an integrated luminosity of 36.5 fb^{-1} have been submitted to a detailed analysis in search for this reaction. The analysis focuses on the $t\bar{t}H$ channel in which the Higgs boson decays in a pair of b quarks and the $t\bar{t}$ system decays semileptonically ($t\bar{t} \rightarrow l\nu bq\bar{q}'\bar{b}$).

The $t\bar{t}H$ signal strength and its upper limit have been measured using two different analysis approaches. In the first (namely "resolved") the procedure used in the previous published papers is applied, through standard identification and reconstruction algorithms. In the second (namely "combined"), the events are separated in two exclusive regimes, including events containing the hadronically decaying top quark with a low transverse momentum ($p_T < 250$ GeV) and the boosted one with the opposite requirement. The decay products of the boosted particles are almost totally collimated, with an overlapping of the decay jets. In this situation, the standard jet reconstruction algorithms loose their efficiency and new innovative techniques are needed. The *boosted techniques* are widely used in this analysis, including the substructure determination of the objects (the "large-R jets") for tagging purposes. Classification and selection tools have been specifically developed and made available to the ATLAS collaboration: tagging techniques and their optimization for the boosted hadronic top and the application for the boosted Higgs boson. The low signal to noise ratio observed in these channels required the use and development of advanced statistical tools such as the MVA analysis and fitting tools to extract the signal strength value or upper limit. Therefore, particular attention has been paid in optimizing the selection criterias and in the evaluation of the systematic uncertainties.

The final results on the signal strength and on its upper limit in the resolved analysis are:

$$\mu_{\mathrm{res}} = 1.4 \pm 0.5 \qquad \text{and} \qquad \mu_{\mathrm{res}} < 2.3 \text{ at } 95\% \text{ CL}.$$

In the combined analysis (resolved and boosted sample, with no events in common), the results are:

$$\mu_{\mathrm{combined}} = 1.2 \pm 0.5 \qquad \text{and} \qquad \mu_{\mathrm{combined}} < 2.0 \text{ at } 95\% \text{ CL}.$$

Both the resolved and the combined results are compatible with the prediction of the SM ($\mu = 1$). Nevertheless, the addition of the boosted channel, sensitive to a significantly different kinematic region, constrains

in a stronger way the μ upper limit to values that are closer to the SM predictions. This underlines the importance to include the boosted regime in the analysis, to provide encouraging improvements in the limit on the signal strength measurement.

The knowledge, experience and tools developed during this study will allow at the end of 2018, when the LHC is expected to record an integrated luminosity of 300 fb^{-1}, more detailed studies on the $t\bar{t}H$ production mechanism. At that time it will also be possible to perform the first studies of differential cross section, giving the final instrument for the research of new physics and for the determination of $(t\bar{t})$-Higgs coupling.

Appendix

A   Upgrade of the ATLAS Muon Barrel Trigger

The current ATLAS muon trigger in the barrel region is based on three layers of RPC (Resistive Plate Chambers, see figure 2.17), two in the middle station (RPC1 and RPC2) and one in the outer station (RPC3). Each chamber consists of two sensitive gas layers, read out by orthogonal η and ϕ strips. Actually the trigger for single high-p_{T} muons is issued when hits are found in coincidence in all three stations in both η and ϕ views. During Run-1, the system has demonstrated high selectivity and high reliability [145], but it has suffered from limited redundancy, because of any efficiency loss in one of the three layers turned out as an important reduction of the final trigger efficiency. The three-layers coincidence limits the trigger acceptance to the regions with full coverage by all the three layers, which corresponds to only 73% for $|\eta| < 1.05$ (the barrel region covered by the RPC). The main sources of acceptance losses are the presence in the middle layer of RPC of the barrel toroid ribs and the hole for the calorimeter services around $|\eta| = 0$.

The ATLAS RPC have been certified for 10 years of operation at a peak luminosity of 10^{34} cm^{-2} s^{-1} and for an integrated charge up to 0.3 C/cm^2, corresponding to a counting rate of 100 Hz/cm^2 [146], which is the design limit rate. At the High Luminosity LHC (HL-LHC), the so-called Phase-2 (the run period from 2026), the expected rates for the chambers will reach up to 340 Hz/cm^2 [147]. The RPC system will run for 25 years instead of 10, absorbing an integrated dose that exceeds the design specifications by a factor three, but the estimations foresee, even in this case, a stable overall performance of the chambers. The ATLAS muon collaboration proposes an upgrade of the system by installing another inner layer (RPC0), fig. A.1, of a new generation RPCs during the LHC shutdown (LS3), foreseen from 2023 to 2026; this will increase the system redundancy in the $|\eta| < 1.05$ region and will allow operation with better efficiency and selectivity. The insertion of this new layer will increase the total geometrical acceptance in the barrel region from 73% to 90% and will improve the resolution of the muon momentum. The trigger electronics will be upgraded to operate at a design peak luminosity larger from 5 to 10 $\times$ 10^{34} cm^{-2} s^{-1} and at a bunch crossing rate of 40 MHz (the Phase-2 condition).

My contribution to this study [148] concerns the performance evaluation of different trigger settings after the insertion of this new layer with the HL-LHC experimental setup configuration.
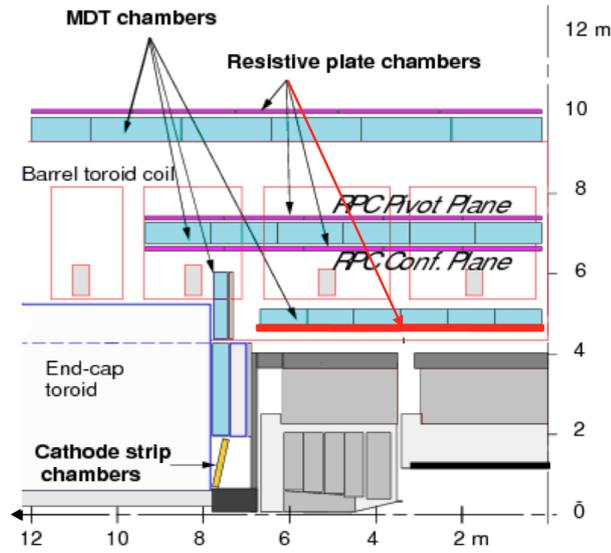
Figure A.1: Scheme of the upgrade of the Muon Spectrometer for the Phase-2. The inner layer, RPC0, is the red one under the inner MDT chamber in the barrel region.

A.1 Upgrade motivations

RPC trigger coverage

The studied triggers are of two types:

- *low-p_T trigger*: requires the coincidence between the hits on RPC1 and RPC2 and $p_\mathrm{T} > 10$ GeV of the muon candidates;

- *high-p_T trigger*: requires the coincidence between the hits on RPC1 and RPC2 and RPC3 and $p_\mathrm{T} > 11$ GeV of the muon candidates.

The actual *high-p_T trigger* geometrical coverage is limited to 73% because of the ATLAS toroid mecchanical support structure, as shown in all η-ϕ regions in figure A.2a. In part (b) of the same figure, the projection η (integrating on ϕ) of both *low-* and *high-p_T trigger* is shown, using real (2012 data) and simulated muons. The main inefficiencies are in the Barrel Medium region (the Medium chambers in $6 < z < 8$ m, see fig. A.1), because of the toroid ribs, and in the "feet" regions ($0.3 < \eta < 0.65$ and $\phi = 240°$ and $300°$), where the detector support structure ("feet") are present.

RPC trigger rates and extrapolations

Several studies are made to estimate the RPC trigger rates in the Phase-2 conditions and to test the RPCs aging issue [146]. The figure A.3a shows the *high-p_T trigger* (with a further request of $p_\mathrm{T} > 20$ GeV) rate as a function of the instantaneous luminosity using 2012 data. The total rate ranges from 200 to 600 Hz, with a constant percentage of fake around 15%, due to secondary particles, like protons produced in dense materials as the magnets. The figure A.3b shows, for each η-ϕ region of the RPC, the hit rate per cm^2 extrapolated to the Phase-2 condition, meaning luminosity of 7.5×10^{34} cm^{-2} s^{-1} and a center of mass energy of 14 TeV. In the region of large η inside the RPC coverage, the average trigger rate is > 200 Hz/cm^2, above a factor 2 over the design limit fixed, as already anticipated, at 100 Hz/cm^2.
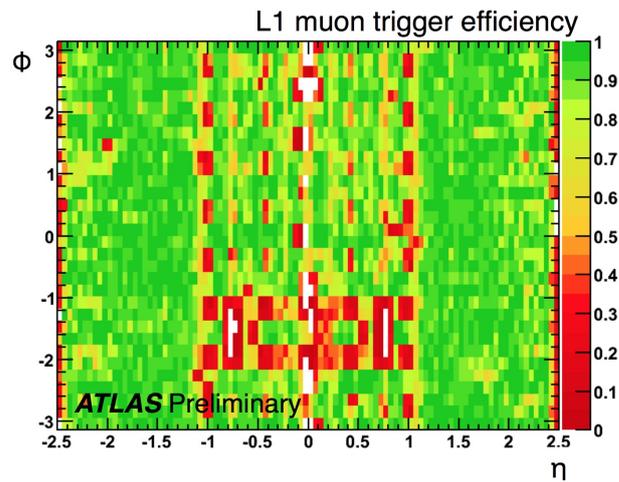
(a)
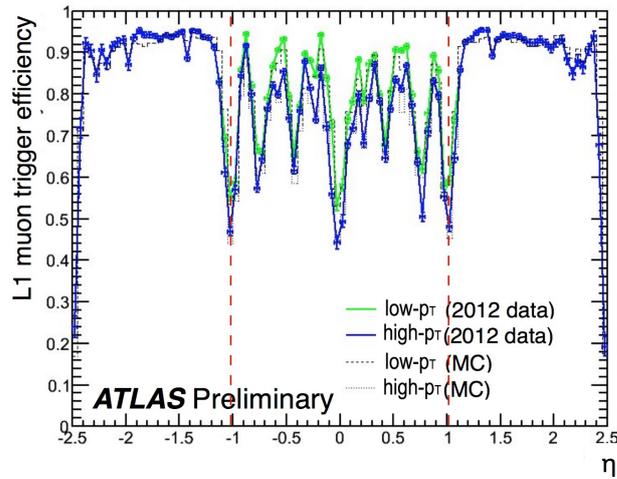
(b)

Figure A.2: RPC trigger efficiency: a) efficiency map of the *high-p_T trigger*, evaluated with real data, and b) efficiency as a function of η for both *low-* and *high-p_T trigger* [149]. The RPC system covers only the $|\eta| < 1.05$ region.

Rate [Hz]

slope (total)= 8.5 ± 0.4
slope (fake)= 1.4 ± 0.1

ATLAS Preliminary
Data 2012 $\sqrt{s}$ = 8 TeV

● high-p_{T} trigger
▼ high-p_{T} trigger fake
— Linear Fit

Inst. Lum. [10^{32} cm^{-2}s^{-1}]

(a)
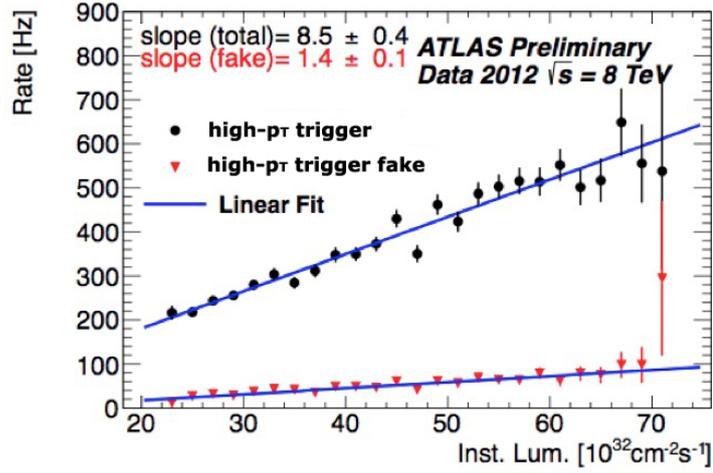
Sector Φ Id.	RPC unit Id. along Z direction																				Average
	-6.2	-6.1	-5.0	-4.0	-3.2	-3.1	-2.2	-2.1	-1.2	-1.1	1.1	1.2	2.1	2.2	3.1	3.2	4.0	5.0	6.1	6.2	
01.01	342	280	301	225	145	114	128	101		61	61	71	113	122	127	136	214	276	269	285	180
01.02	293	281	303	218	159	133	129	143		76	71	75	127	143	140	148	215	295	278	297	188
2	168	204	188	138	109	90	77	63	55	56	45	48	61	79	94	104	140	207	196	152	115
03.01	297	296	281	198	148	128	119	119	68	67	65	71	131	125	114	125	207	329	268	290	177
03.02	300	243	277	210	151	129	155	122	85	75	75	70	122	127	152	132	207	315	243	299	179
4	112	166	158	151	101	83	65	77	41	53	46	41	73	68	92	108	160	196	175	112	101
05.01	171	173	263	138	105	102	140	127	68	60	60	69	124	177	102	137	185	290	173	171	149
05.02	227	198	237	158	109	105	136	143	77	61	63	71	111	136	108	141	200	267	255	282	159
6	175	186	208	163	105	95	90	95	77	59	52	59	84	81	106	124	189	200	227	167	131
07.01	305	263	288	191	154	129	131	114		76	78		122	139	124	148	185	261	268	305	183
07.02	327	258	216	203	141	112	129	108		74	77		105	114	112	152	184	278	276	279	175
8	146	196	195	161	103	85	80	70	50	57	54	54	67	74	85	108	168	196	194	156	118
09.01	319	246	301	206	155	117	149	119		46	64		106	134	124	135	197	283	262	297	181
09.02	347	258	287	205	143	95	107	103		58	67		99	112	95	137	188	285	265	292	174
10	174	201	207	147	99	86	68	71	46	43	41	50	64	69	80	103	148	193	201	170	115
11.01	308	244	237	157	97	84	81	87		40	43		83	94	92	94	148	227	215	278	132
11.02	196	193	157	105	78	66	55	57		33	31		50	62	58	71	98	151	160	185	98
12						80	81	66	51	36	36	51	75	87	80						64
13.01	291	278	253		140	102	96	84	43	41	47	50	87	95	99	123		249	263	319	149
13.02	299	264	262		104	97	105	86	49	48	50	56	93	103	97	110		252	227	294	146
14					142	68	64	52	41	40	49	63	68	136							76
15.01	196	221	148	113	76	71	67	49		38	36		50	59	86	87	104	156	173	196	104
15.02	183	159	246	164	116	98	103	75		44	43		75	106	112	107	158	248	159	183	133
16	173	214	216	173	108	89	54	75	56	59	50	50	77	54	87	103	177	209	208	154	124
Average	229	223	234	167	118	101	96	88	56	52	51	56	86	97	103	118	171	240	221	221	137

(b)

Figure A.3: *high-p_{T} trigger* rates: (a) using 2012 data [150] and (b) extrapolating to the Phase-2 condition with a peak luminosity of $\mathcal{L}$= 7.5×10^{34} cm^{-2} s^{-1}. The plot shows rates in Hz/cm^2 as a function of the ϕ sector and of the station number along z [147].

A.2 RPC trigger upgrade proposal for Phase-2

The insertion of a new RPC inner layer could allow to increase the current detector efficiency up to 90%. The proposal is to have four concentric RPC chambers (RPC0, RPC1, RPC2 and RPC3), for a total of nine or ten layers (at the moment RPC0 is not decided to be composed of three or four layers yet). The system redundancy will help to face with the high RPC hit rates foreseen with the HL-LHC. In this condition, the trigger can be performed using a 2/4 majority (the coincidence of at least two hit layers between the four RPCs) and 3/4 (the coincidence of at least three hit layers between the four RPCs). The first one is more robust than the other one, especially in case of RPC inefficiencies. The probability to record at least one hit in at least one of the two strip panels (η or ϕ) will be consistently reduced without affecting the trigger efficiency. In this new configuration, the trigger robustness is increased and the longer lever arm allows for a sharper momentum threshold for identified muon tracks.

Trigger study and analysis strategy

The aim of the study is the comparison between different RPC trigger requirements (see figure A.4) in order to estimate the acceptances and the rates in a Phase-2 condition in terms of luminosity and of detector layout.
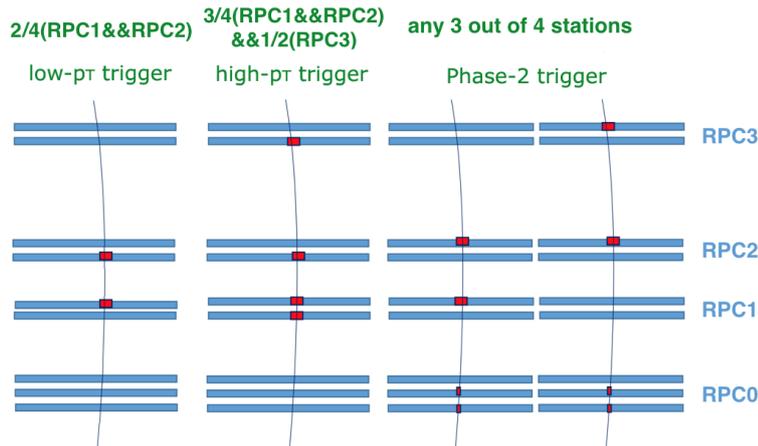
Figure A.4: Different RPC trigger configurations. The first configuration (first column), *low-p_T trigger*, requiring a coincidence amongst hits on RPC1 and on RPC2 with a reconstructed momentum $p_\mathrm{T} > 10$ GeV of the muon candidates; the second (second column), *high-p_T trigger*, requires a coincidence amongst hits on two chambers (RPC1, RPC2) and on RPC3 with a reconstructed momentum $p_\mathrm{T} > 11$ GeV of the muon candidates; the third (third and fourth columns) is the *Phase-2 trigger* that requires the coincidence of any three layers of the four RPCs.

The study concerns the comparison between old configurations (including only the RPC1, RPC2 and RPC3 layers) and new ones, including the RPC0. Three different trigger configurations have been simulated for the Phase-2 condition:

- *Phase-2 trigger*: requires the coincidence of any three layers of the four RPCs and $p_\mathrm{T} > 20$ GeV of the muon candidates (third and fourth column of fig. A.4);

- *Phase-2 trigger test1*: requires the coincidence of RPC0 and three hit layers of RPC1, RPC2 and RPC3, and $p_\mathrm{T} > 20$ GeV of the muon candidates;

- *Phase-2 trigger test2*: requires the coincidence of RPC0 and the coincidence on RPC1+RPC2 and the coincidence of RPC3, and $p_T > 20$ GeV of the muon candidates;

The trigger acceptance has been evaluated simulating events of single muons with p_T = 25 GeV and $|\eta| < 1.05$ (in order to be above all the p_T trigger thresholds and inside the RPC coverage). At the moment, the RPC0 is simulated using the signal coming from the MDT chambers in the Barrel Inner region (see fig. A.1). The muon reconstruction has been obtained requiring a $\Delta\eta \leq 0.016$ in the association between the RPC hits and the MDT segments. The algorithm reconstructs the space point (which indicates a possible muon candidate) looking at the intersection between one strip η and one strip ϕ in each layer; the best muon candidate has been chosen as the one with the highest number of hits, as shown in fig. A.5.
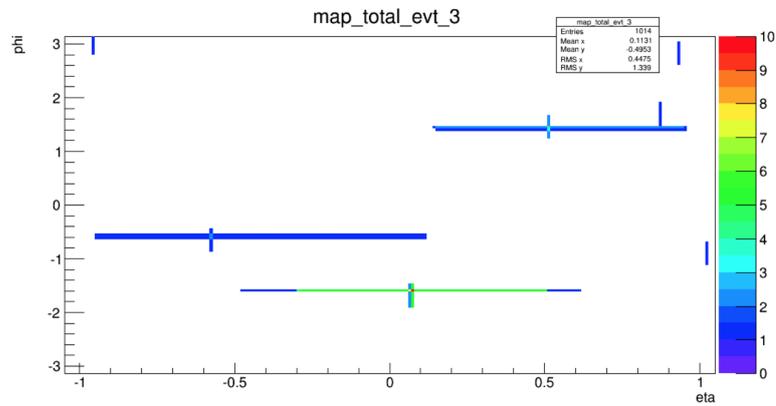
Figure A.5: Example of the hits of a reconstructed muon candidate (in green) in a single event. The algorithm searches for the intersection between the strip η and the strip ϕ with the highest number of hits from each RPC layer.

Trigger acceptance

For a proper understanding of the trigger efficiencies in different conditions, it is needed to disentangle the effects of the RPC hit efficiencies from the algorithms chosen for defining the trigger. In these studies, all the RPC hit efficiencies are assumed equal. The trigger efficiencies evaluated at 100% hit efficiency is therefore geometrical and it will be called here shortly "trigger acceptance".

The comparison of the trigger efficiency of the *high-p_T trigger* and the *Phase-2 trigger* has been obtained as a function of the RPC hit efficiency and shown in fig. A.6. The *high-p_T trigger* efficiency increases significantly from 44.2%, with an 70% hit efficiency, to 73%, with a 100% hit efficiency, while the *Phase-2 trigger* shows an efficiency already high of 76.7%, with the 70% hit efficiency, that further increases to 90% when the RPCs are totally efficient. The trigger acceptance map for the *high-p_T trigger* and *Phase-2 trigger* is shown in fig. A.7a and fig. A.7b, respectively. The trigger acceptance (trigger efficiency at 100% hit efficiency) improvement is due to the insertion of the RPC0, that allows to trigger events without requiring the coincidence of all the RPCs.

The trigger efficiency as a function of RPC hit efficiency has been calculated also for other trigger configurations, with and wthout the RPC0, in the Phase-2 condition and reported in the table A.1.
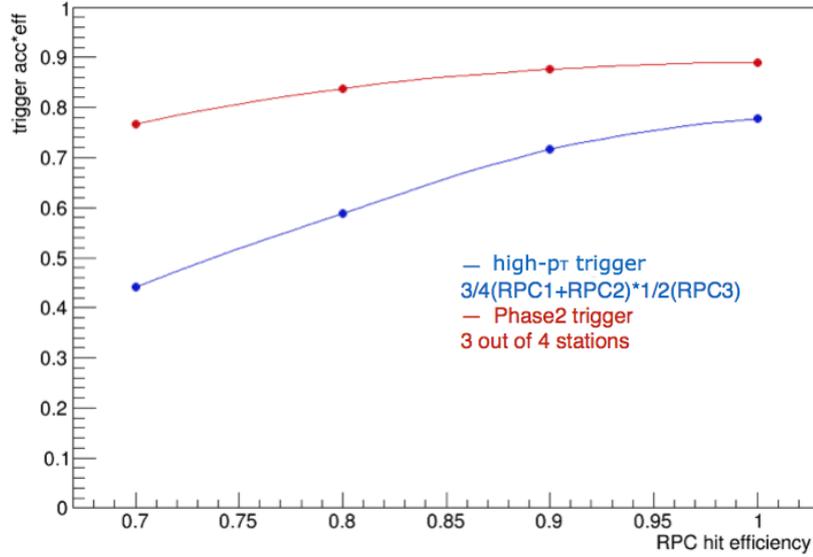
Figure A.6: Trigger efficiency of the *high-p_T trigger* and *Phase-2 trigger* as a function of the RPC hit efficiency.

RPC requirements	Trigger	Trigger efficiency (%)			
		70% eff.	80% eff.	90% eff.	100% eff.
3/4 (RPC1+RPC2)	*low-p_T trigger*	51.1	64.7	76.0	81.8
3/4 (RPC1+RPC2) AND 1/2 (RPC3)	*high-p_T trigger*	44.2	58.8	71.6	73.0
1/2 (RPC1) AND 1/2 (RPC2) AND 1/2 (RPC3)	Run-1 trigger test1	57.6	67.7	75.2	78.0
2/4 (RPC1+RPC2) AND 1/2 (RPC3)	Run-1 trigger test2	67.2	76.1	82.3	85.2
4/6 (RPC1+RPC2+RPC3)	Run-1 trigger test3	58.6	72.7	83.0	88.1
3/6 (RPC1+RPC2+RPC3)	Run-1 trigger test4	79.2	85.7	89.0	89.9
any 3 out of 4 stations (all RPCs)	*Phase-2 trigger*	76.7	83.9	87.6	90.0
RPC0 AND 3/6 (RPC1+RPC2+RPC3)	Phase-2 trigger test1	73.2	79.3	82.2	84.9
RPC0 AND (RPC1 OR RPC2) AND RPC3	Phase-2 trigger test2	66.7	73.9	78.1	81.3

Table A.1: Trigger efficiencies and acceptances (last column) of nine different trigger requirements studied. In the first six rows tests on the trigger configuration using only RPC1, RPC2 and RPC3 are presented, instead in the last three rows the Phase-2 configuration is shown. The RPC hit efficiency varies from 70% to 100%.

Trigger rate estimations

The rates estimation is performed using the run 216432, acquired in December 2012 with a bunch crossing (BC) every 25 ns and short bunch trains. The run peak luminosity was 4.5×10^{32} cm^{-2} s^{-1} with a μ mean value of 7.5. The aim of this study is to estimate the trigger rate extrapolating to the Phase-2 condition, meaning a peak luminosity of 5×10^{34} cm^{-2} s^{-1}, an inelastic cross section σ_{inel} = 80 mb and

(a)

(b)
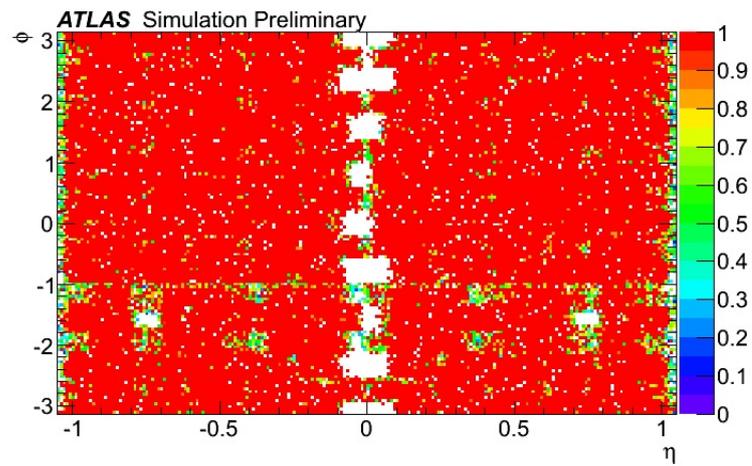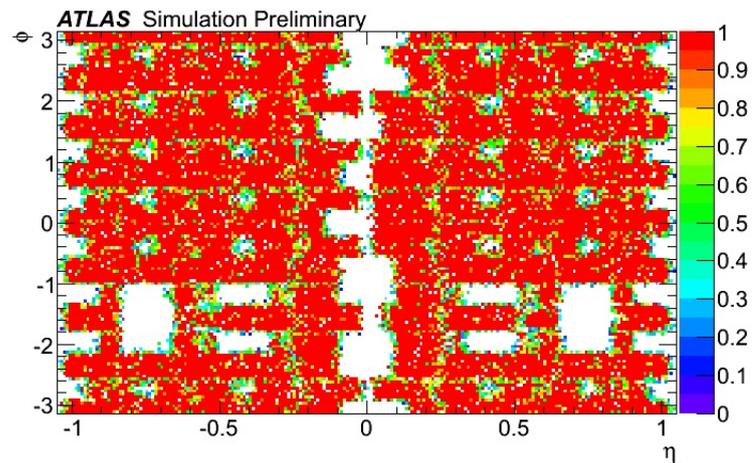
Figure A.7: Trigger acceptance for (a) *high-p_T trigger* in Run-1 condition and (b) the *Phase-2 trigger* in Phase-2 condition [151].

a bunch crossing frequency f = 40 MHz. Since the mean value of the number of interactions per BC is $<\mu> = L\sigma_{inel}/f$, the Phase-2 require a $<\mu> \simeq 150$. Two kinds of dataset are used: the ZeroBias sample, in which are collected all the collision BCs, and the Empty one, containing only the empty BCs of the run.

In order to reproduce a $<\mu> \simeq 150$, an *Overlaid* dataset has been created overlapping to each event 20 random events, containing 7.5 collisions each, from the ZeroBias BCs datasets and 150 random triggers from the Empty BCs to reproduce the cavern background (see fig. A.8).
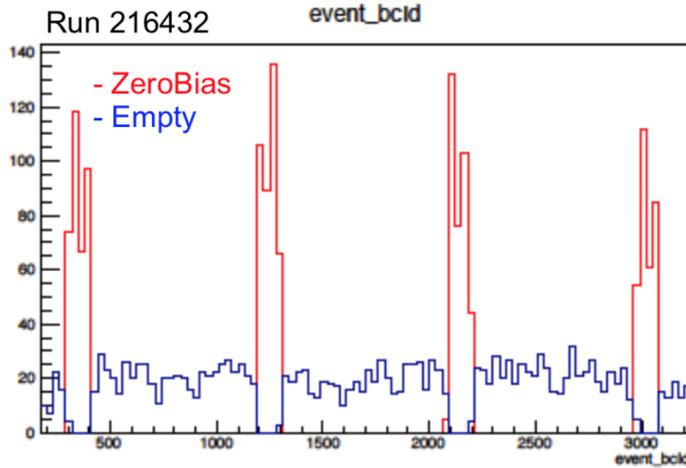
Figure A.8: Distribution of the event BCid in the dataset of run 216432 in which each event is overlaid with 20 random ZeroBias events and 150 random Empty events, to reproduce the Phase-2 condition.

For this rate estimation two specific options are used:

- the number of the MDT hits is chosen to be at least four to be similar to the RPC ones;

- in the MDT-RPC association, the muon candidate is choosen in a $|\Delta\eta| \leq 0.01$ region (the obtained results are completely compatible with the ones obtained with $|\Delta\eta| \leq 0.016$).

As shown in figure A.9, the *Phase-2 trigger* (seventh bin), shows a much higher rate with respect to the *high-p_T trigger* (second bin), at a level that will be unmanageable by the TDAQ system. The present study aims to keep a similar rate of Run-1, by a proper selection of the most interesting muon tracks. The high rate mostly derives from low-p_T tracks in RPC0, RPC1 and RPC2 and this component could be reduced with a $\Delta\eta$ optimization between the hits-segments association in the muon reconstruction. It is important to underline that in the Run-1 condition (without RPC0), the triggers with higher acceptance ($> 85\%$, relatively to *Run-1 trigger test 1-4* of table A.2) have an acceptance similar to the *Phase-2 trigger* configuration but with a higher rate from a factor 5 to 25 and are not acceptable in sight of the HL-LHC. In order to decrease the still high rate of the *Phase-2 trigger*, two new configurations (*Phase-2 trigger test1* and *Phase-2 trigger test2*, previously defined) requiring more stringent coincidence between the RPCs have been tested. The result (see table A.2) confirms the lower rate especially for the *Phase-2 trigger test2* that shows a rate decrease by about a factor three and a lower acceptance of only $\sim 10\%$ with respect to the *Phase-2 trigger*.
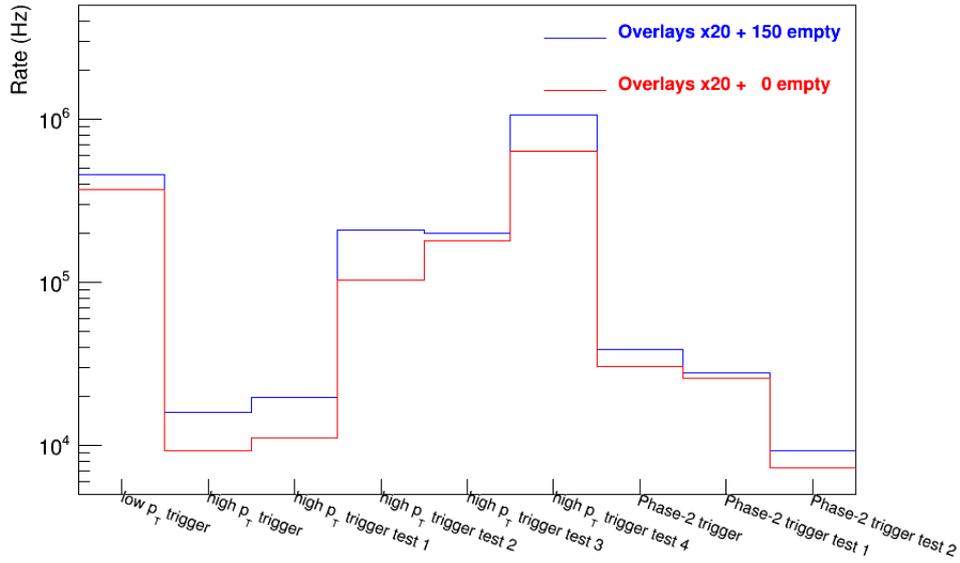
Figure A.9: Muon trigger rates expected in Phase-2 for the 9 different studied configurations, calculated for an Overlaid dataset as described above and for a dataset with ZeroBias overlaid events and no Empty triggers. The second bin represents the present *high-p_T trigger*, while the seventh is the *Phase-2 trigger*.

RPC requirements	Trigger	Total acceptance (%)	Overlay20 Rate (kHz)	Overlay20+150Empty Rate (kHz)
3/4(RPC1+RPC2)	*low-p_T trigger*	81.8	373.2	461.2
3/4(RPC1+RPC2) AND 1/2(RPC3)	*high-p_T trigger*	73.0	9.3	15.9
1/2(RPC1) AND 1/2(RPC2) AND 1/2(RPC3)	Run-1 trigger test1	78.0	11.2	19.8
2/4(RPC1+RPC2) AND 1/2(RPC3)	Run-1 trigger test2	85.2	102.6	211.1
4/6(RPC1+RPC2+RPC3)	Run-1 trigger test3	88.1	179.3	199.2
3/6(RPC1+RPC2+RPC3)	Run-1 trigger test4	89.9	635.9	1069.3
any 3 out of 4 stations (all RPCs)	*Phase-2 trigger*	90.0	30.4	39.0
RPC0 AND 3/6(RPC1+RPC2+RPC3)	Phase-2 trigger test1	84.9	25.8	27.8
RPC0 AND (RPC1 OR RPC2) AND RPC3	Phase-2 trigger test2	81.3	7.3	9.3

Table A.2: Muon trigger rates and trigger acceptances for the 9 different studied configurations. The RPC hit efficiency for this table is 100%. Two kind of samples are used: one with events obtained by overlaying 20 random collision BCs events per each event; the other sample is obtained by overlaying 20 random collision BCs and 150 empty events per each event of the file.

A.3 Conclusions

The insertion of a new RPC inner layer in the inner barrel region (RPC0) has been studied as a proposal for the Phase-2 upgrade of the ATLAS detector. This new generation PRC chamber will be installed during

the LHC shutdown, foreseen for the year 2023 till 2026, to increase the system redundancy and to allow operation with better efficiency and selectivity during the HL-LHC phase.

Preliminary results on the acceptances and rates estimations in a Phase-2 experimental setup have been evaluated using the segments from the MDT chambers as the hits of the RPC0.

The study demonstrates the acceptance improvement, passing from a 73%, obtained with the current high-p_T trigger, to 90%, with the Phase-2 trigger requirement that includes the RPC0.

The estimated rate grows from $\sim$ 15 kHz, with the *high-p_T trigger*, to $\sim$ 40 kHz, with the *Phase-2 trigger*, a rate not acceptable in sight of the Phase-2 condition. For this reason, different trigger requirements for the Phase-2 are being studied obtaining a lower rate by a factor three with a loss of only 10% of acceptance, that provides a good hint for the future steps.

This study is preliminary and the rate estimation needs a proper RPC0 simulation (instead of using the MDT segments) and a precise optimization of the muon candidate reconstruction. The estimated rate is just an upper limit of the Phase-2 rates.

Appendix

B   Data/MC comparison plots

In this appendix, more plots of the data/MC agreement are shown. The disturbutions are chosen according to the importance of the variable in the analysis. The most of them have been used for the MVA analysis studies as a potential set of discriminating variables.

All the distributions show shapes and systematics bands without any modifications due to the fit procedure.

(a)   (b)
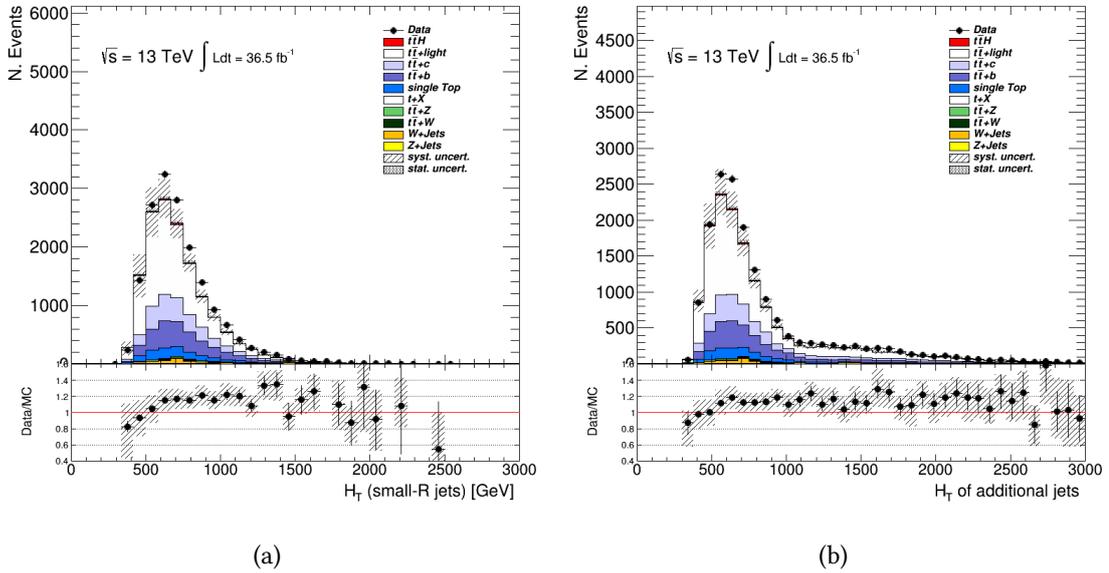
Figure B.1: Distributions of the H_T variable, defined as the scalar sum of the p_T of all the small-R jets (a) or of the p_T of all the small-R jets outside the top-tagged large-R jet (b) in the event, for MC and data.

(a)   (b)
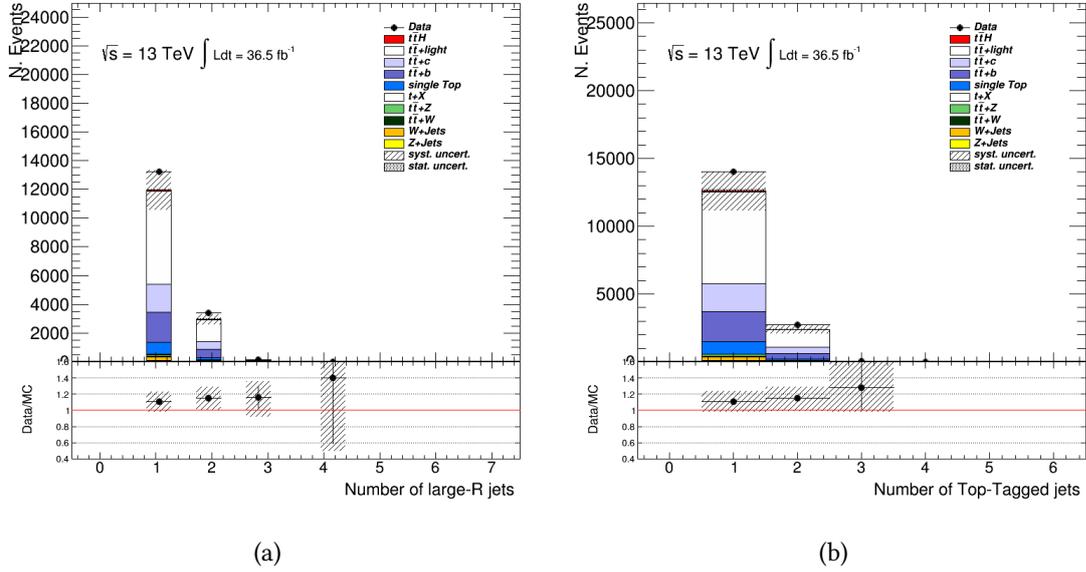
Figure B.2: Distributions of the number of large-R jets (a) and of top-tagged large-R jets (b) in the event, for MC and data.

(a)   (b)
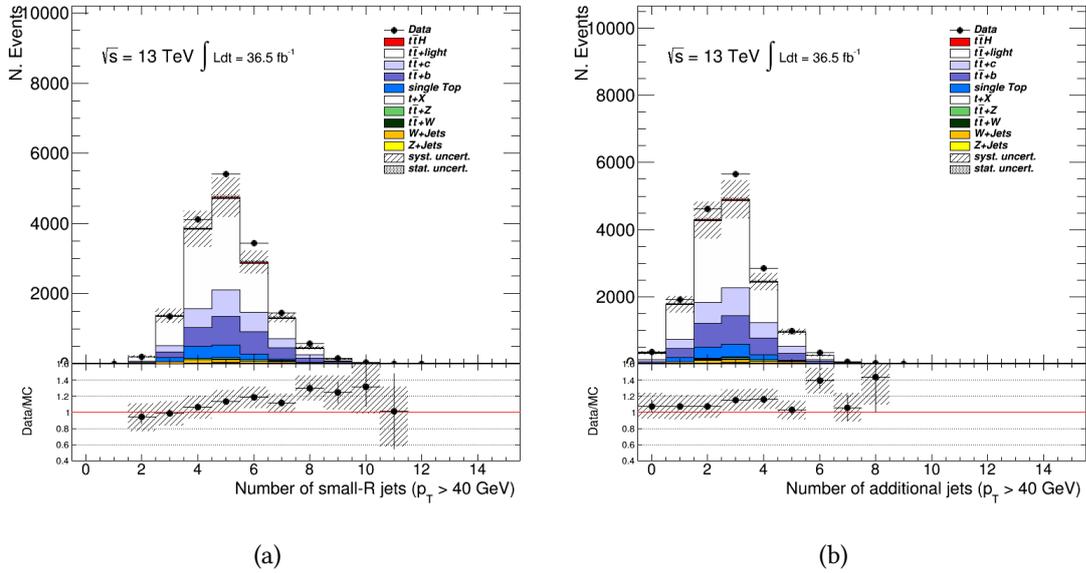
Figure B.3: Distributions of the number of small-R jets with $p_T > 40$ GeV (a) and the number of small-R jets outside any top-tagged large-R jet with $p_T > 40$ GeV (b) in the event, for MC and data.
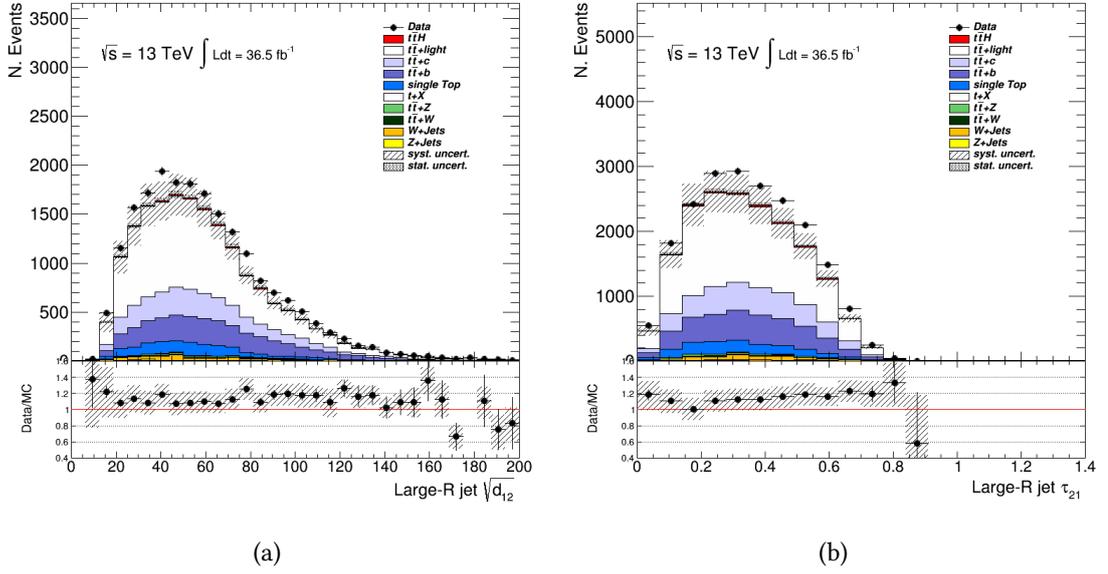
Figure B.4: Distributions of the $\sqrt{d_{12}}$ (a) and τ_{21} (b) of large-R jets in the event, for MC and data.
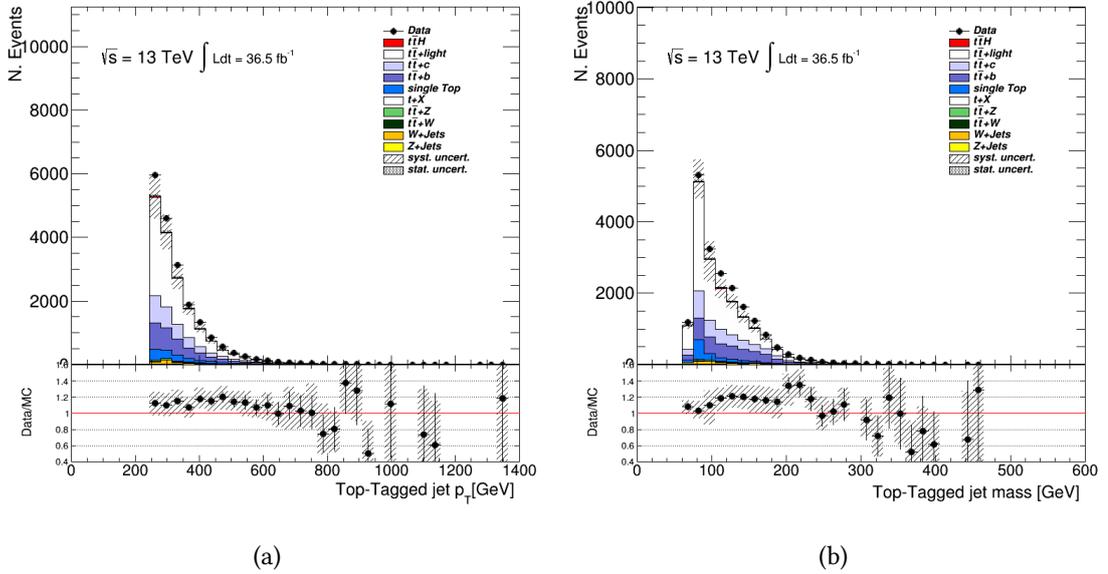
Figure B.5: Distributions of the p_T (a) and mass (b) of top-tagged large-R jets in the event, for MC and data.
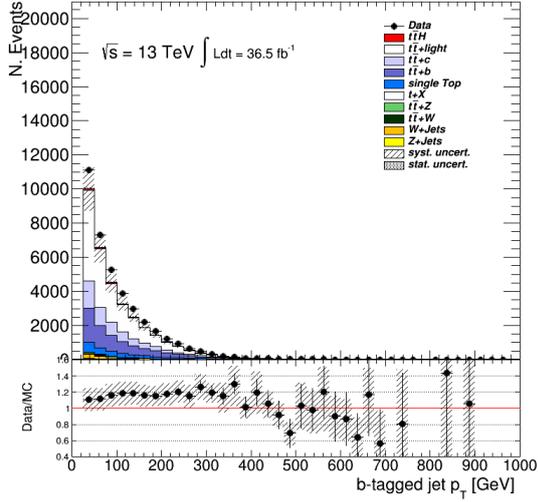
(a)
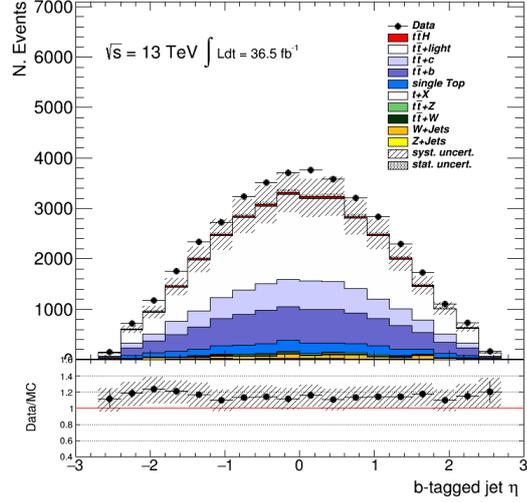
(b)
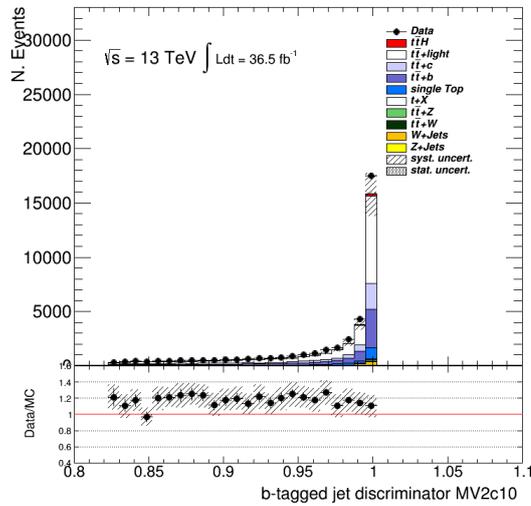
(c)

Figure B.6: Distributions of the p_T (a), η (b) and b-tagging discriminator variable MV2C10 (c) of b-tagged small-R jets in the event, for MC and data.

(a)   (b)

(c)   (d)
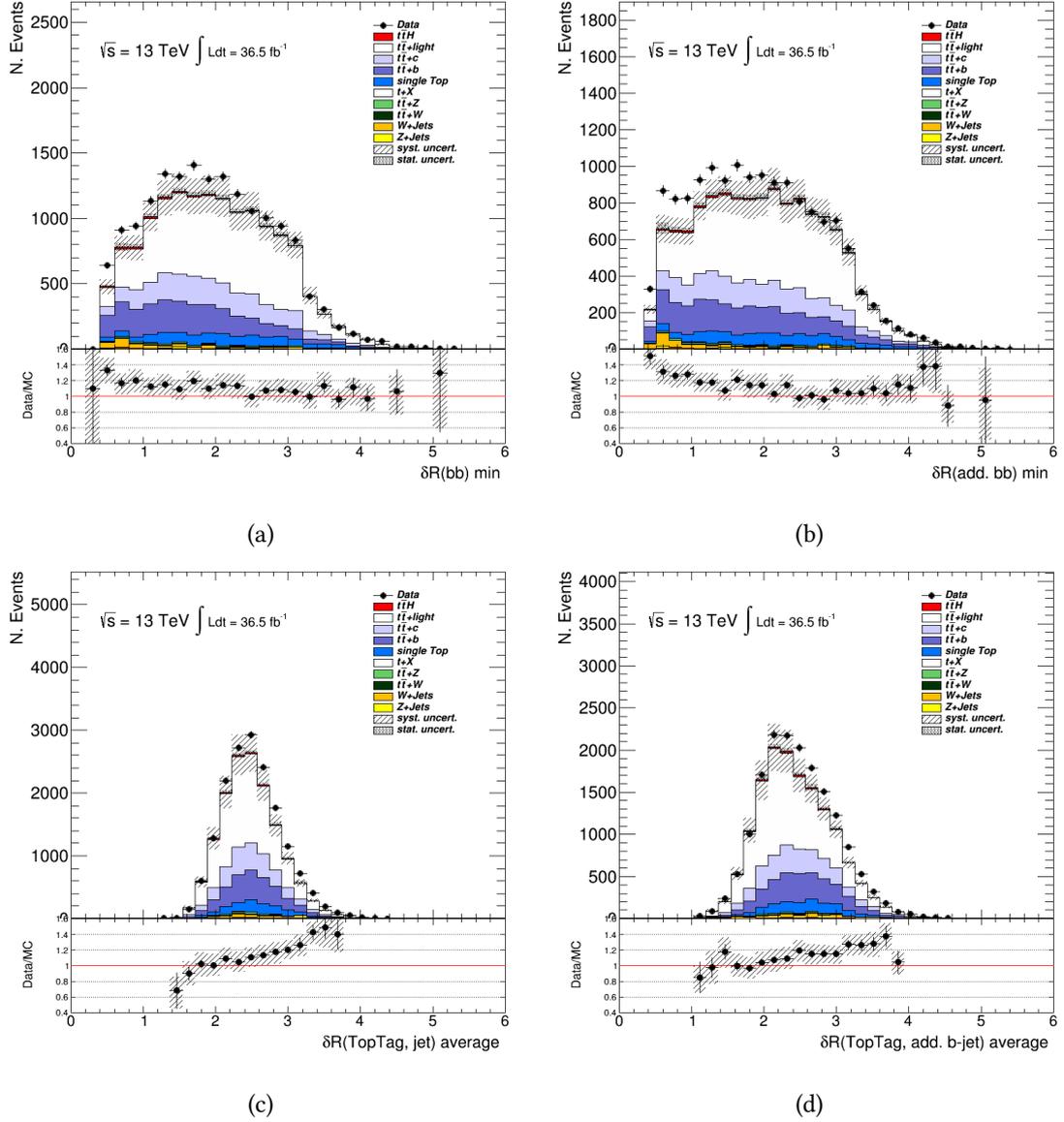
Figure B.7: Distributions of the ΔR_{bb}^{min} (a), $\Delta R_{(add)bb}^{min}$ (b), $\Delta R_{top(add)j}^{avg}$ (c) and $\Delta R_{top(add)b}^{avg}$ (d) variables, for MC and data. For the specific definition of all these variables, used in the MVA process, see Chapter 5.
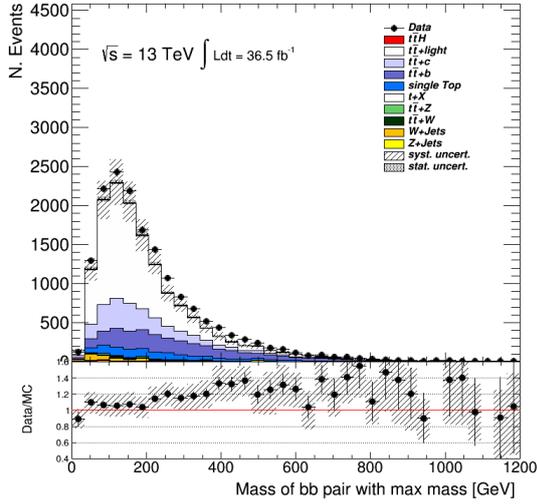
(a)
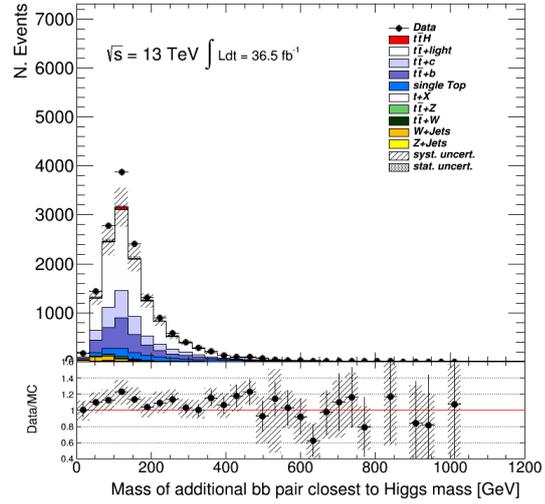
(b)
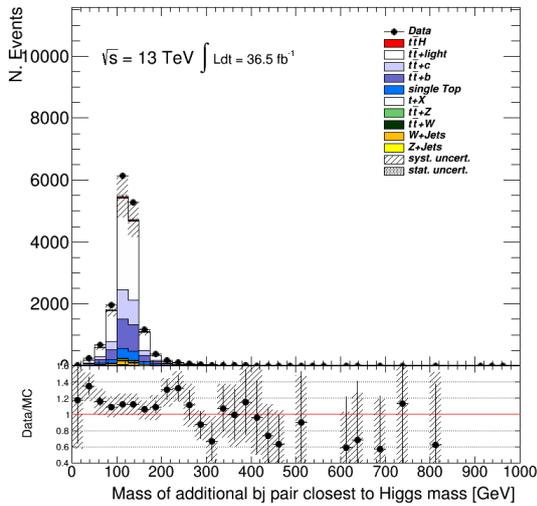
(c)
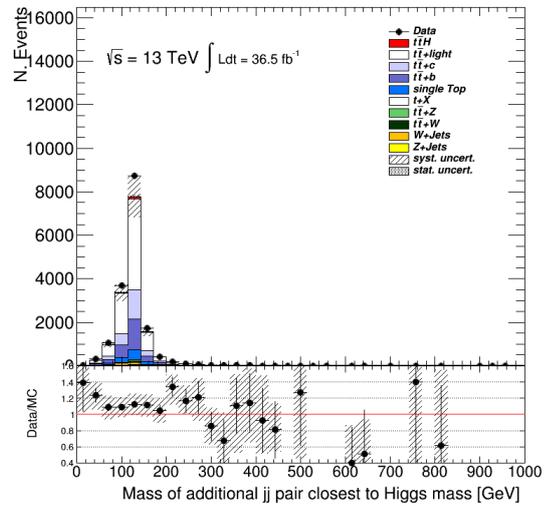
(d)

Figure B.8: Distributions of the m_{bb}^{max} (a), $m_{(add)bb}^{H}$ (b), $m_{(add)bj}^{H}$ (c) and $m_{(add)jj}^{H}$ (d) variables, for MC and data. For the specific definition of all these variables, used in the MVA process, see Chapter 5.

Appendix

C   Naming of the nuisance parameters

Free-floating normalisation factors:

- $\mu_{t\bar{t}H}$: signal strenght for $t\bar{t}H$, the parameter of interest in the fit;

- $k(t\bar{t}+\geq 1b)$: normalisation of the $t\bar{t}+\geq 1b$ background component;

- $k(t\bar{t}+\geq 1c)$: normalisation of the $t\bar{t}+\geq 1c$ background component.

Signal modelling uncertainties:

- XS $t\bar{t}H$ QCD: signal cross-section uncertainty from scale variations;

- XS $t\bar{t}H$ PDF: signal cross-section uncertainty from PDF choice;

- $t\bar{t}H$ PS & hadronisation: derived comparing $t\bar{t}H$ generated with MG5_aMC + Herwig++ and MG5_aMC + Pythia 8;

- BR($H \rightarrow b\bar{b}$): uncertainty on the Higgs decay rate to $b\bar{b}$;

- BR($H \rightarrow WW$): uncertainty on the Higgs decay rate to WW;

- BR($H \rightarrow \tau\tau$): uncertainty on the Higgs decay rate to $\tau\tau$;

- BR($H \rightarrow other$): uncertainty on the Higgs decay rate to other final states;

- $t\bar{t}H$ scale choice (acceptance): derived from the $t\bar{t}H$ scale variation weights inside the MG5_aMC + Pythia 8 sample, keeing only the acceptance effect and removing the cross-section effect (i.e normalising to the same cross section as the nominal signal template).

$t\bar{t}$ background modelling uncertainties:

- XS $t\bar{t}$ (inclusive): total $t\bar{t}$ cross-section uncertainty ($\pm 6\%$);

- $t\bar{t}+\geq 1b$ scale choice: uncertainty on the scale choice on $t\bar{t}+\geq 1b$ reweighting, derived varying the default renormalisation scale by a factor of two up and down;

- $t\bar{t}+\geq 1b$ global scale: uncertainty on the global scale on $t\bar{t}+\geq 1b$ reweighting, derived using an aternative softer scale for both factorisation and resummation scales;

- $t\bar{t}+\geq 1b$ Q CMMPS: uncertainty on the scale choice on $t\bar{t}+\geq 1b$ reweighting, derived using an alternative softer scale only for the resummation scale;

- $t\bar{t}+\geq 1b$ shower recoil scheme: uncertainty from the shower recoil scheme in the $t\bar{t}+\geq 1b$ reweighting;

- $t\bar{t}+\geq 1b$ MSTW PDF: one of the two uncertainties from PDF choice on $t\bar{t}+\geq 1b$ reweighting;

- $t\bar{t}+\geq 1b$ NNPDF: one of the two uncertainties from PDF choice on $t\bar{t}+\geq 1b$ reweighting;

- $t\bar{t}+\geq 1b$ MPI: uncertainty on the normalisation of the $t\bar{t}+\geq 1b$ category from multi-parton-interactions (taken as Âś50%);

- $t\bar{t}+\geq 1b$ UE modelling: uncertainty on the modelling of the underliying event;

- $t\bar{t}+\geq 1b$ NLO generator (reweighting): derived comparing the nominal $t\bar{t}+\geq 1b$ reweighted to MG5_aMC + Pythia 8 ME prediction, symmetrised

- $t\bar{t}+\geq 1b$ PS & hadronisation (reweighting): derived comparing the nominal $t\bar{t}+\geq 1b$ reweighted to MG5_aMC + Herwig++ and to MG5_aMC + Pythia 8 ME prediction, symmetrised;

- $t\bar{t}+$light PS & hadronisation: derived comparing $t\bar{t}+$light generated with Powheg-Box + Pythia 6 and Powheg-Box + Herwig++, both fast simulation, after reweighting each sample to the NNLO predictions in $p_{\mathrm{T}}(t\bar{t})$ and $p_{\mathrm{T}}(t)$, symmetrised;

- $t\bar{t}+$light NLO generator: derived comparing $t\bar{t}+$light generated with MG5_aMC + Herwig++ and Powheg-Box + Herwig++, both fast simulation, after reweighting each sample to the NNLO predictions in $p_{\mathrm{T}}(t\bar{t})$ and $p_{\mathrm{T}}(t)$, symmetrised;

- $t\bar{t}+\geq 1c$ PS & hadronisation: derived comparing $t\bar{t}+\geq 1c+$light generated with Powheg-Box + Pythia 6 and Powheg-Box + Herwig++, both fast simulation, after reweighting each sample to the NNLO predictions in $p_{\mathrm{T}}(t\bar{t})$ and $p_{\mathrm{T}}(t)$, symmetrised;

- $t\bar{t}+\geq 1c$ NLO generator: derived comparing $t\bar{t}+\geq 1c+$light generated with MG5_aMC + Herwig++ and Powheg-Box + Herwig++, both fast simulation, after reweighting each sample to the NNLO predictions in $p_{\mathrm{T}}(t\bar{t})$ and $p_{\mathrm{T}}(t)$, symmetrised;

- $t\bar{t}+$light NNLO reweighting $p_{\mathrm{T}}(t)$: uncertainty on the $p_{\mathrm{T}}(t)$ reweighting for the $t\bar{t}+$light component, derived reweighting $p_{\mathrm{T}}(t)$ to the distribution in the "radHi" sample instead of to the NNLO one, symmetrised;

- $t\bar{t}+$light NNLO reweighting $p_{\mathrm{T}}(t\bar{t})$: uncertainty on the $p_{\mathrm{T}}(t\bar{t})$ reweighting for the $t\bar{t}+$light component, derived reweighting $p_{\mathrm{T}}(t\bar{t})$ to the distribution in the Powheg-Box + Herwig++ sample instead of to the NNLO one, symmetrised;

- $t\bar{t}+\geq 1c$ NNLO reweighting $p_{\mathrm{T}}(t)$: uncertainty on the $p_{\mathrm{T}}(t)$ reweighting for the $t\bar{t}+\geq 1c$ component, derived reweighting $p_{\mathrm{T}}(t)$ to the distribution in the "radHi" sample instead of to the NNLO one, symmetrised;

- $t\bar{t}+\geq 1c$ NNLO reweighting $p_{\mathrm{T}}(t\bar{t})$: uncertainty on the $p_{\mathrm{T}}(t\bar{t})$ reweighting for the $t\bar{t}+\geq 1c$ component, derived reweighting $p_{\mathrm{T}}(t\bar{t})$ to the distribution in the Powheg-Box + Herwig++ sample instead of to the NNLO one, symmetrised;

- $t\bar{t}+\geq 1c$ NLO reweighting: uncertainty derived comparing the $t\bar{t}+\geq 1c$ reweighted to the NLO ME prediction from MG5_aMC (3FS) with the inclusive prediction from MG5_aMC (5FS).

Other background modelling uncertainties:

- XS single top (Wt): inclusive cross-section uncertainty on the single-top Wt-channel process ($\pm 5\%$);

- XS single top (t- & s-chan.): inclusive normalisation uncertainty on the single-top non-Wt-channel processes ($\pm 5\%$);

- W+jets norm. (Res. Single Lepton, $4j$): normalisation uncertainty on the W+jets background in the single lepton regions with 4 jets ($\pm 30\%$);

- W+jets norm. (Res. Single Lepton, $5j$): normalisation uncertainty on the W+jets background in the single lepton regions with 5 jets ($\pm 30\%$);

- W+jets norm. (Res. Single Lepton, $\geq 6j$): normalisation uncertainty on the W+jets background in the single lepton regions with ≥ 6 jets ($\pm 30\%$);

- W+HF (Res. Single Lepton, $2b$): uncertainty on the fraction of W+HF jets in the single lepton regions with 2 b-tagged jets ($\pm 30\%$);

- W+HF (Res. Single Lepton, $3b$): uncertainty on the fraction of W+HF jets in the single lepton regions with 3 b-tagged jets ($\pm 30\%$); W+HF (Res. Single Lepton, $\geq 4b$): uncertainty on the fraction of W+HF jets in the single lepton regions with ≥ 4 b-tagged jets ($\pm 30\%$);

- Z+jets norm. (Res. Single Lepton): inclusive uncertainty on the Z+jets background in the single lepton channel ($\pm 45\%$);

- XS $t\bar{t}+W$ QCD: $t\bar{t}+W$ cross-section uncertainty from scale variations;

- XS $t\bar{t}+W$ PDF: $t\bar{t}+W$ cross-section uncertainty from PDF choice;

- XS $t\bar{t}+Z$ QCD: $t\bar{t}+Z$ cross-section uncertainty from scale variations;

- XS $t\bar{t}+Z$ PDF: $t\bar{t}+Z$ cross-section uncertainty from PDF choice;

- $t\bar{t}+W$ generator: derived comparing the LO sample with the new NLO one (symmetrised);

- $t\bar{t}+Z$ generator: derived comparing the LO sample with the new NLO one (symmetrised).

Detector and beam condition related (instrumental) uncertainties:

- luminosity;

- b-tag Eigenvar. [0-4];

- c-tag Eigenvar. [0-3];

- light-tag Eigenvar. [0-13];

- b-tag high p_T extrapolation;

- b-tag $c \to \tau$ extrapolation;

- electron energy resolution;

- electron energy scale;

- electron ID efficiency;

- electron isolation efficiency;

- electron reconstruction efficiency;

- electron trigger efficiency;

- muon energy resolution (ID);

- muon energy resolution (MS);

- muon energy scale;

- muon ID efficiency (stat);

- muon ID efficiency (syst);

- muon isol efficiency (stat);

- muon isol efficiency (syst);

- muon trig efficiency (stat);

- muon trig efficiency (syst);

- muon TTVA efficiency (stat);

- muon TTVA efficiency (syst);

- JES BJES;

- JES effective NP [1-6];

- JES η intercalibration modelling;

- JES η intercalibration total stat;

- JES flavour composition (splitted into JES flavour composition "ttlight" and virgothers);

- JES flavour response;

- JES pileup offset μ;

- JES pileup offset NPV;

- JES pileup p_T term;

- JES pileup ρ topology;

- JES punchthrough;

- JES single particle (high-p_T);

- JES η intercalibration non-closure;

- Jet energy resolution;

- Jet vertex tagger efficiency;

- pileup modelling.

Bibliography

[1] K. A. Olive *et al.* (Particle Data Group), Chin. Phys. **C 38** (2014) 090001 and 2015 update, http://pdg.lbl.gov/.

[2] F. Halzen, A. D. Martin, *Quarks and Leptons: An introduction course in modern particle physics* (1984), John Wiley & Sons.

[3] S.L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579-588.

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

[5] A. Salam, *Elementary Particle Theory*, Ed. N. Svarholm (1968).

[6] ATLAS Collaboration, *Combination of ATLAS and CMS top-quark pair cross-section measurements using proton-proton collisions at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2014-054.

[7] F. Abe *et al.*, (CDF collaboration), Phys. Rev. Lett. **74** (1995) 2626.

[8] S. Abachi *et al.*, (D0 collaboration), Phys. Rev. Lett. **74** (1995) 2632.

[9] F. P. Schilling, *Top Quark Physics at the LHC: A Review of the First Two Years*, Int. J. Mod. Phys. A **27** (2012) 1230016.

[10] ATLAS Collaboration, *Measurement of t-Channel Single Top-Quark Production in pp Collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, ATLAS-CONF-2012-132, 2012.

[11] CMS Collaboration, *Measurement of the t-channel single-top-quark production cross section and of the $|V_{tb}|$ CKM matrix element in pp collisions at $\sqrt{s}$ = 8 TeV*, JHEP **6** (2014) 1-43.

[12] C. Helsen on behalf of the ATLAS Collaboration, *Top Physics at the LHC, 9th Rencontres du Vietnam: Windows on the Universe,* Quy Nhon, Vietnam, 11 - 17 Aug 2013, 301-305.

[13] ATLAS Collaboration, *Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector using the Matrix Element Method*, Phys. Lett. **B** (2016) 228-246.

[14] CDF and D0 Collaborations, *Observation of s-channel production of single top quarks at the Tevatron*, Phys. Rev. Lett. **112** (2014) 231803.

[15] Hurng-Chun Lee, *Single Top Quark production at LHC*, PhD thesis, National Institute for Subatomic Physics NIKHEF (2013).

[16] A. D. Martin, W. J. Stirling, R. S. Thorne, G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C 63** (2009) 189-285.

[17] D0 collaboration, *Determination of the width of the top quark*, D0 Note 6034-CONF (2010).

[18] P. W. Higgs, Phys. Rev. Lett. **13** (1964) 508.

[19] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321.

[20] D. H. Perkins, *Introduction to High Energy Physics*, 2000, Cambridge University Press.

[21] ATLAS Collaboration, Phys. Lett. **B 716** (2012) 1-29.

[22] CMS Collaboration, Phys. Lett. **B 716** (2012) 30.

[23] ATLAS Collaboration, Phys. Lett. **B 726** (2013) 88.

[24] CMS Collaboration, *Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV*, CMS-PAS-HIG-13-005.

[25] LHC Higgs Cross Section Working Group,
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG .

[26] CERN Yellow Report Page At 13 14 TeV,
https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV2014 (2014).

[27] S. Heinemeyer *et al.*, LHC Higgs Cross Section Working Group, arXiv:1307.1347[hep-ph] (2013).

[28] S. Catani, D. de Florian, M. Grazzini and P. Nason, JHEP **307** (2003) 28.

[29] M. Ciccolini, A. Denner, S. Dittmaier, *Electroweak and QCD corrections to Higgs boson production in vector-boson fusion at LHC*, Phys. Rev. **D 77** (2008) 013002.

[30] G. Ferrera, M. Grazzini, and F. Tramontano, Phys. Rev. Lett. **107** (2011) 152003.

[31] S. Heinemeyer *et al.*, LHC Higgs Cross Section Working Group, arXiv:1307.1347[hepph] (2013).

[32] A. Djouadi, Int. J. Mod. Phys. A **10** (1995) 1-64.

[33] ATLAS Collaboration, Phys. Lett. **B 749** (2015) 519-541.

[34] ATLAS Collaboration, ATLAS-CONF-2016-058.

[35] ATLAS Collaboration, ATLAS-CONF-2016-067.

[36] ATLAS Collaboration, Eur. Phys. J. **C 75** (2015) 349.

[37] ATLAS Collaboration, ATLAS-CONF-2016-080.

[38] ATLAS Collaboration, ATLAS-CONF-2016-068.

[39] F. Maltoni, E. Vryonidoua and C. Zhangb, *Higgs production in association with a top-antitop pair in the Standard Model Effective Field Theory at NLO in QCD*, arXiv:1607.05330v2 [hep-ph] (2016).

[40] W.J. Stirling, private communication.

[41] High Luminosity LHC project, http://hilumilhc.web.cern.ch/about/hl-lhc-project

[42] CERN Courier, *The LHC first long run: High-quality beam from the injectors and full exploitation of options in the collider underpinned the LHC's performance in 2010-2013*, 19 August 2013.

[43] Luminosity Public Results Run-1,
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[44] Luminosity Public Results Run-2,
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[45] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, 2008 JINST 3 S08003.

[46] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$= 13 TeV.*, Eur. Phys. J. **C 76** (2016) 292.

[47] ATLAS Collaboration, *Insertable B-Layer Technical Design Report*, ATLAS-TDR-019.

[48] F. Hugging, *The ATLAS Pixel Insertable B-Layer (IBL)*, arxiv:1012.2742v1.

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

[50] ATLAS Collaboration, *The ATLAS Data Acquisition and High Level Trigger system*, JINST 11 (2016) P06008.

[51] ATLAS Collaboration, *Electron and photon reconstruction and identification in ATLAS: expected performance at high energy and results at 900 GeV*, ATLAS-CONF-2010-005.

[52] ATLAS Collaboration, *Identification of muon candidates in pp collisions at $\sqrt{s}$= 900 GeV with the ATLAS detector*, ATLAS-CONF-2010-015.

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

[54] ATLAS Collaboration, *ATLAS tunes of PYTHIA6 and PYTHIA8 for MC11*, ATL-PHYS-PUB-2011-009 (2011).

[55] F. Krauss (Sherpa Collaboration), *Sketch of a $t\bar{t}H$ event.* Available at http://www.opensciencegrid.org/wp-content/uploads/2014/05/event.jpg.

[56] T. Sjöstrand, S. Mrenna, P. Skands, *A brief introduction to* PYTHIA 8.1, arXiv:0710.3820v1 [hep-ph] (2008).

[57] M. Bähr *et al.*, *Herwig++ Physics and Manual*, Eur. Phys. J. **C 58** (2008) 639.

[58] T. Gleisberg *et al.*, *Event generation with* SHERPA 1.1, JHEP **02** (2009) 007.

[59] G. Altarelli, G. Parisi, Nucl. Phys. **B 126** (1977) 298.

[60] J. Alwall *et al.*, *MadGraph 5: Going Beyond*, JHEP **06** (2011) 128.

[61] S. Frixione, F. Stoeckli, P.Torrielli, B. R. Webber, C. D. White, *The* MC@NLO 4.0 *Event Generator*, arXiv:1010.0819 [hep-ph] (2010).

[62] V. Hirschi, R. Frederix, S. Frixione, M. V. Garzelli, F. Maltoni, R. Pittau, *Automation of one-loop QCD corrections*, JHEP **1105** (2011) 044.

[63] S. Alioli *et al.*, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043.

[64] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070.

[65] ATLAS Collaboration, *Further studies on simulation of top-quark production for the ATLAS experiment at $\sqrt{s}$ = 13 TeV*, ATL-COM-PHYS-2016-979.

[66] ATLAS Collaboration, *Studies of $t\bar{t}+c\bar{c}$ production with MadGraph5_aMC@NLO and Herwig++ for the ATLAS experiment*, ATL-PHYS-PUB-2016-011.

[67] B. Andersson, G. Gustafson, G. Ingelman, T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys.Rept. 97 (1983) 31-145.

[68] B. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. **B 238** (1984) 492.

[69] G. Marchesini, B. Webber, *Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation*, Nucl. Phys. **B 310** (1988) 461.

[70] S. Agostinelli *et al.*, Nucl. Instr. Methods **A 506** (2003) 250-303.

[71] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, Journal of Physics: Conference Series **396** (2012) 022031.

[72] T. Yamanaka (ATLAS Collaboration), *The ATLAS calorimeter simulation FastCaloSim*, Journal of Physics: Conference Series **331** (2011) 032053.

[73] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, M. Ubiali, *Parton distributions for the LHC Run II*, JHEP **1504** (2015) 040.

[74] H.-L. Lai *et al.*, *New parton distributions for collider physics*, Phys. Rev. **D 82** (2010) 074024.

[75] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instr. Methods **A 462** (2001) 152.

[76] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s}$ = 7 and 8 TeV in the ATLAS experiment*, Eur. Phys. J. **C 76** (2016) 6.

[77] J. Adelman *et al.*, *Search for the Standard Model Higgs boson produced in association with top quarks in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector at the LHC*, ATL-COM-PHYS-2014-1471 (2014).

[78] 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.

[79] F. Cascioli *et al.*, *NLO matching for $t\bar{t}+b\bar{b}$ production with massive b-quarks*, Phys. Lett. **B 734** (2014) 210-214.

[80] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007.

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

[82] D0 Collaboration, *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV using kinematic characteristics of lepton+jets events*, Phys. Rev. **D 76** (2007) 092007.

[83] ALTAS 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.

[84] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024.

[85] J. Illingworth and J. Kittler, *A survey of the Hough transform*, Computer Vision, Graphics, and Image Processing 44 (1988) 87-116.

[86] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$= 13 TeV*, Eur. Phys. J. **C 76** (2016) 292.

[87] 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) 17.

[88] M. Cacciari, G. P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) 63.

[89] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s}$= 8 TeV using the ATLAS detector*, CERN-PH-EP-2015-206 (submitted to EPJC, to be updated later).

[90] A. Hoecker *et al.*, *Toolkit for Multivariate Data Analysis with ROOT*, arXiv:physics/0703039 [Data Analysis, Statistics and Probability].

[91] ATLAS Collaboration, *Performance of b-jet Identification in the ATLAS Experiment*, JINST 11 (2016) 04008.

[92] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].

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

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

[95] ATLAS Collaboration, *Search for the bb decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, JHEP **01** (2015) 069.

[96] ATLAS Collaboration, *Performance of large-R jets and jet substructure reconstruction with the ATLAS detector*, ATLAS-CONF-2012-065.

[97] ATLAS Collaboration, Public plots: Large-radius jet performance plots using full 13 TeV 2015 dataset URL: http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2015-004/

[98] A. J. Larkoski, D. Neill, and J. Thaler, *Jet Shapes with the Broadening Axis*, JHEP **04** (2014) 017.

[99] A. J. Larkoski, I. Moult, D. Neill, *Analytic Boosted Boson Discrimination*, arXiv:1507.03018v5 .

[100] A. J. Larkoski, I. Moult, D. Neill, *Power Counting to Better Jet Observables*, arXiv:1409.6298v1 .

[101] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, JHEP **02** (2010) 084.

[102] S. D. Ellis, C. K. Vermilion, J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, Phys. Rev. **D 81** (2010) 094023.

[103] ATLAS Collaboration, *Boosted hadronic top identification at ATLAS for early 13 TeV data*, ATL-PHYS-PUB-2015-053.

[104] J. Thaler, L.-T. Wang, *Strategies to identify boosted tops*, JHEP **07** (2008) 092.

[105] ATLAS Collaboration, *Performance of Boosted W Boson Identification with the ATLAS Detector*, ATL-PHYS-PUB-2014-004.

[106] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparison with ATLAS data taken at $\sqrt{s}$= 8 TeV.*, Eur. Phys. J. **C 76** (2016) 1-47.

[107] ATLAS Collaboration, *Identification of Boosted, Hadronically-Decaying W and Z Bosons in $\sqrt{s}$= 13 TeV Monte Carlo Simulations for ATLAS*, ATL-PHYS-PUB-2015-033.

[108] M. Wolter, *MultiVariate Analysis methods in Physics*, Physics of Particles and Nuclei **38** (2007) 255-268.

[109] C. Bini, *Data analysis in experimental elementary particle physics.*, http://www.roma1.infn.it/people/bini/StatEPP.pdf.

[110] A. Hoeker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, CERN-OPEN-2007-007.

[111] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C 71** (2011) 1554.

[112] T. Junk, *Confidence level computation for combining searches with small statistics*, Nucl. Instr. Methods **A 434** (1999) 435.

[113] A. L. Read, *Presentation of search results: The CL(s) technique*, J. Phys. G **28** (2002) 2693.

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

[115] W. Verkerke and D. Kirkby, *RooFit Users Manual*, http://roofit.sourceforge.net/.

[116] ATLAS Collaration, *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.

[117] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s}$ = 8 TeV using the ATLAS detector at the LHC*, CERN-EP-2016-117.

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

[119] W. Beenakker *et al.*, *NLO QCD corrections to $t\bar{t}H$ production in hadron collisions*, Nucl. Phys. **B 653** (2003) 151.

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

[121] 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. **B 738** (2014) 1.

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

[123] A. Djouadi, J. Kalinowski and M. Spira, *HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension*, Comp. Phys. Comm. **108** (1998) 56.

[124] M. Czakon, D. Heymes and A. Mitov, *High-Precision Differential Predictions for Top-Quark Pairs at the LHC*, Phys. Rev. Lett. **116** (2016) 082003.

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

[126] A. D. Martin *et al.*, *Parton distributions for the LHC*, Eur. Phys. J. **C 63** (2009) 189.

[127] ATLAS Collaboration, *Studies of tt + cc production with MadGraph5_aMC@NLO and Herwig++ for the ATLAS experiment*, ATL-PHYS-PUB-2016-011 (2016).

[128] 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. **C 76** (2016) 11.

[129] CMS Collaboration, *Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s}$ = 8 TeV*, Eur. Phys. J. **C 76** (2016) 379.

[130] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W or H*, Phys. Rev. **D 82** (2010) 054018.

[131] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D 81** (2010) 054028.

[132] 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) 091503.

[133] S. Frixione *et al.*, *Single-top hadroproduction in association with a W boson*, JHEP **0807** (2008) 029.

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

[135] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^{\pm}$ production and decay at NLO*, JHEP **1207** (2012) 052.

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

[137] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. **B 726** (2013) 120-144.

[138] ATLAS Collaboration, *Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data*, ATLAS-CONF-2014-009.

[139] CMS Collaboration, *Measurement of the properties of a Higgs boson in the four-lepton final state*, Phys. Rev. **D 89** (2014) 092007.

[140] A. Freitas and P. Schwaller, *Higgs CP Properties From Early LHC Data*, Phys. Rev. **D 87** (2013) 055014.

[141] J. Brod, U. Haisch, and J. Zupan, *Constraints on CP-violating Higgs couplings to the third generation*, JHEP **1311** (2013) 180.

[142] F. Demartin, F. Maltoni, K. Mawatari, B. Page and M. Zaro, *Higgs characterisation at NLO in QCD: CP properties of the top-quark Yukawa interaction*, arXiv:1407.5089 .

[143] F. Maltoni, E. Vryonidou and C. Zhang, *Higgs production in association with a top-antitop pair in the Standard Model Effective Field Theory at NLO in QCD*, JHEP **10** (2016) 123.

[144] G. Degrassi, P. P. Giardino, F. Maltoni and D. Pagani, *Probing the Higgs self coupling via single Higgs production at the LHC*, JHEP **080** (2016) 1612.

[145] ATLAS Collaboration, *Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s}$ = 8 TeV*, Eur. Phys. J. **C 75** (2015) 120.

[146] G. Aielli, M. Alviggi, V. Ammosov, M. Bianco, M. Biglietti *et al.*, *New results on ATLAS RPC's aging at CERN's GIF*, IEEE Trans.Nucl.Sci. **53** (2006) 567-571.

[147] ATLAS Collaboration, *ATLAS Phase-2 Upgrade Scoping Document*, CERN-LHCC-2015-020 (2015).

[148] S. Biondi, *Upgrade of the ATLAS Muon Barrel Trigger for HL-LHC*, PoS EPS-HEP (2015) 289.

[149] L1 Barrel Muon Trigger Efficiency with 2012 Data,
https://cds.cern.ch/record/1647612.

[150] ATLAS Muon Trigger Public Results web page,
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonTriggerPublicResults.

[151] ATLAS Muon Public Results web page,
https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPublicResults.

List of Figures

218

221

222

List of Tables