

**Search for charged Higgs bosons decaying  
into top and bottom quarks with  
single-lepton final states using  $pp$  collisions  
collected at a centre-of-mass energy of  
13 TeV by the ATLAS detector**

DISSERTATION

zur Erlangung des akademischen Grades  
doctor rerum naturalium  
(Dr. rer. nat.)  
im Fach: Physik  
Spezialisierung: Experimentalphysik

eingereicht an der  
Mathematisch-Naturwissenschaftlichen Fakultät  
der Humboldt-Universität zu Berlin

von

**M.Sc. Francesco Peri**

Präsident der Humboldt-Universität zu Berlin:  
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:  
Prof. Dr. Elmar Kulke

Gutachter: 1. Prof. Dr. Thomas Lohse  
2. Prof. Dr. Johannes Haller  
3. Prof. Dr. Heiko Lacker

Tag der mündlichen Prüfung: 27.02.2019



## Erklärung

Ich erkläre, dass ich die Dissertation selbständig und nur unter Verwendung der von mir gemäß §7 Abs. 3 der Promotionsordnung der Mathematisch-Naturwissenschaftlichen Fakultät, veröffentlicht im Amtlichen Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 126/2014 am 18.11.2014 angegebenen Hilfsmittel angefertigt habe.

I declare that I have produced this doctors thesis independently using only the tools I have specified, in accordance with section 7 para. 3 of the Faculty of Mathematics and Natural Sciences PhD regulations, published in the Official Gazette of Humboldt-Universitat zu Berlin (Amtliches Mitteilungsblatt) no. 126/2014 on 18/11/2014.

Berlin, 04.12.2018

Francesco Peri





For my parents.



# Abstract

This thesis presents a search for charged Higgs bosons produced in proton–proton collisions at a centre-of-mass energy of 13 TeV, using  $36.1 \text{ fb}^{-1}$  of data collected by the ATLAS detector at the LHC in 2015 and 2016. The existence of charged Higgs bosons is predicted by various theories Beyond the Standard Model and it is motivated by the inadequacy of the Standard Model to explain some observed experimental phenomena. The work focuses on charged Higgs bosons heavier than the top quark and decaying via  $H^\pm \rightarrow tb$ . The production in association with a top and a bottom quark ( $pp \rightarrow tbH^\pm$ ) is investigated in the mass range between 200 and 2000 GeV. A final state containing one charged lepton and jets is considered. Multiple kinematic variables are combined using a boosted decision tree (BDT) in order to separate signal and background. The output of the BDT is used to perform a profile likelihood fit of the Monte Carlo predictions to the observed data. No significant excess of events above the expected Standard Model background is observed, therefore upper limits are set for the cross-section of the charged Higgs boson production times the branching fraction of its decay. Limits are also provided for the  $\tan\beta$  parameter of the MSSM, in the  $m_h^{\text{mod-}}$  and hMSSM benchmark scenarios. The work improves the reach of all previous searches, including for the first time masses ranging up to 2000 GeV.



## Zusammenfassung

In dieser Arbeit wird die Suche nach geladenen Higgs Bosonen ( $H^\pm$ ) vorgestellt. Dafür wurden Proton–Proton Kollisionen, die bei einer Schwerpunktsenergie von 13 TeV in den Jahren 2015 und 2016 mit einer integrierten Luminosität von  $36.1 \text{ fb}^{-1}$  mit dem ATLAS-Experiment produziert wurden, untersucht. Die Existenz solcher geladener Higgs Bosonen wird in verschiedenen Modellen jenseits des Standardmodells vorhergesagt und ist auch dadurch motiviert, dass das Standardmodell nicht immer eine Erklärung für verschiedenste beobachtete Phänomene liefern kann. Diese Arbeit konzentriert sich auf geladene Higgs Bosonen, die eine höhere Masse als das top Quark besitzen und über  $H^\pm \rightarrow tb$  zerfallen. Die  $H^\pm$  Produktion, in Verbindung mit einem top und einem bottom Quark  $pp \rightarrow tbH^\pm$ , wird im Massenbereich von 200 bis 2000 GeV untersucht. Die Suche nach  $H^\pm$  in dieser Arbeit beschränkt sich auf Endzustände mit einem geladenen Lepton und mehreren Jets. Mit Hilfe eines boosted decision trees werden verschiedenste kinematische Variablen miteinander kombiniert, um dadurch das  $H^\pm$  Signal besser von dem Standardmodell Untergrund unterscheiden zu können. Es wurde kein signifikanter Unterschied zum vorhergesagten Standardmodell-Untergrund gemessen und deshalb Ausschlussgrenzen für den Produktionswirkungsquerschnitt mal Verzweigungsverhältnis dieses  $H^\pm$  Zerfalls berechnet. Weitere Ausschlussgrenzen wurden für den  $\tan\beta$  Parameter des MSSM Modells für die  $m_h^{\text{mod-}}$  und hMSSM Benchmark-Szenarien bestimmt. Die Ergebnisse dieser Arbeit übertreffen alle bisherigen Ausschlussgrenzen - insbesondere werden zum ersten Mal Erkenntnisse über den Massenbereich bis zu 2000 GeV gewonnen.



# Contents

<b>List of figures</b>	<b>xiii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Theoretical basis</b>	<b>3</b>
1.1 The Standard Model	3
1.2 Beyond the Standard Model, a supersymmetric approach	10
1.3 Searching for charged Higgs bosons	15
1.4 Monte Carlo simulations	24
<b>2 Statistical methods</b>	<b>29</b>
2.1 Hypothesis testing	29
2.2 Boosted decision trees	32
<b>3 The LHC and the ATLAS experiment</b>	<b>37</b>
3.1 The Large Hadron Collider	37
3.2 The ATLAS experiment	40
<b>4 Object reconstruction in ATLAS</b>	<b>49</b>
4.1 Base-object reconstruction: tracks and vertices	49
4.2 Electrons	51
4.3 Muons	53
4.4 Additional requirements on leptons	54
4.5 Jets	55
4.6 Overlap removal	58
4.7 Missing transverse momentum	58
<b>5 Signal and background modelling</b>	<b>59</b>
5.1 Signal modelling	61
5.2 Background modelling	62
<b>6 Search for <math>\bar{t}bH^+(H^+ \rightarrow t\bar{b})</math></b>	<b>71</b>
6.1 Event selection and categorisation	72
6.2 Multivariate methods	77
6.3 Systematic uncertainties	85

6.4	Statistical analysis	93
6.5	Results	105
<b>7</b>	<b>Future prospects</b>	<b>111</b>
7.1	Low-mass range and correlation	111
7.2	Improving the discrimination power	116
<b>8</b>	<b>Summary and conclusions</b>	<b>127</b>
	<b>Bibliography</b>	<b>129</b>
	<b>Additional material</b>	<b>143</b>
<b>A</b>	<b>Glossary of abbreviations</b>	<b>143</b>
<b>B</b>	<b>Fake leptons</b>	<b>147</b>
<b>C</b>	<b>Fit validation</b>	<b>149</b>
<b>D</b>	<b>Additional studies on the BDT input variables</b>	<b>167</b>
<b>E</b>	<b>Systematic uncertainties and adversarial networks</b>	<b>171</b>



## List of figures

1.1	The elementary particles of the Standard Model.	5
1.2	One-loop correction to the Higgs mass.	10
1.3	The running constants in the SM and MSSM.	11
1.4	Feynman diagrams for the $H^+$ production.	16
1.5	Branching ratios of the $H^+$ decay.	17
1.6	Cross-section for the production of charged Higgs bosons.	18
1.7	Expected and observed limits on the cross-section of the $H^+ \rightarrow t\bar{b}$ production obtained by ATLAS and CMS at 8 TeV.	20
1.8	Expected and observed limits on $\tan\beta$ as a function of the $H^+$ mass obtained by ATLAS at 8 TeV.	21
1.9	Exclusion limits on $\tan\beta$ and $m_{H^+}$ obtained by flavour physics constraints.	22
1.10	Feynman diagram for the $H^+$ decay.	23
1.11	Schematic representation of an hadronic interaction.	24
2.1	$p$ -value and significance.	30
2.2	Example of a decision tree.	33
2.3	Example of a ROC curve.	35
3.1	The LHC pre-accelerators and the detector complex.	39
3.2	The ATLAS experiment and its sub-detectors.	40
3.3	The ATLAS coordinate system.	41
3.4	The ATLAS inner detectors.	42
3.5	The ATLAS calorimeters.	44
3.6	The ATLAS muon system.	45
3.7	The data acquisition system of ATLAS.	47
3.8	Integrated luminosity recorded by ATLAS.	48
4.1	The track parametrisation in ATLAS.	50
4.2	Electron reconstruction efficiency.	51
4.3	Electron identification efficiency.	52
4.4	Muon reconstruction and identification efficiency.	54
4.5	Output of the tagging-BDT for different jet flavours.	57
5.1	Examples of Feynman diagrams for the most important backgrounds.	60
5.2	ME and PS modelling of the $t\bar{t} + \text{jets}$ background.	65

5.3	Reweighting of the $t\bar{t} + \geq 1b$ categories.	66
5.4	Reweighting of the $t\bar{t} + \text{jets}$ leading-jet $p_T$ .	66
6.1	Signal over background ratio (S/B) as a function of the $H^+$ mass.	74
6.2	Background composition of control and signal regions.	75
6.3	Pre-fit comparison of the predicted and observed event yields in control and signal regions.	76
6.4	The output distributions of the LHD variable for the 200 GeV signal hypothesis.	81
6.5	Ranking of the BDT input variables.	83
6.6	The output distributions of the classification BDTs created using the 200 GeV and 800 GeV signal hypotheses.	84
6.7	Impact of the energy scale variations on the acceptance of $H^+$ events.	89
6.8	Post-fit comparison of the predicted and observed event yields in control and signal regions.	94
6.9	Pre-fit and post-fit distributions of the classification BDT output in the 5 jets signal regions, for the fit performed with the 200 GeV signal hypothesis.	95
6.10	Pre-fit and post-fit distributions of the classification BDT output in the $\geq 6$ -jets signal regions, for the fit performed with the 200 GeV signal hypothesis.	96
6.11	Pre-fit and post-fit distributions of the classification BDT output in the 5 jets signal regions, for the fit performed with the 800 GeV signal hypothesis.	97
6.12	Pre-fit and post-fit distributions of the classification BDT output in the $\geq 6$ -jets signal regions, for the fit performed with the 800 GeV signal hypothesis.	98
6.13	Impact of the systematic uncertainties on the observed signal strength at 200 GeV.	101
6.14	Impact of the systematic uncertainties on the observed signal strength at 800 GeV.	102
6.15	Evolution of the signal strength and the $t\bar{t} + \text{HF}$ normalisation factors as a function of the $H^+$ mass.	104
6.16	Expected and observed limits on the cross-section of the $H^+ \rightarrow t\bar{b}$ production for the single-lepton channel.	107
6.17	Expected and observed limits on $\tan\beta$ as a function of the $H^+$ mass for the single-lepton channel.	107
6.18	Expected and observed limits on the cross-section of the $H^+ \rightarrow t\bar{b}$ production for the combination between the single-lepton and the dilepton channels.	108
6.19	Expected and observed limits on $\tan\beta$ as a function of the $H^+$ mass for the combination between the single-lepton and the dilepton channels.	108
6.20	Superimposition of the exclusion limits of the $t\bar{b}$ and $\tau^+\nu_\tau$ final states.	109
7.1	The $t\bar{t} + \geq 1b$ PS and had. systematic uncertainty for the distribution of the BDT trained at 200 GeV.	113

7.2	Background composition of control and signal regions tagged with different efficiencies.	115
7.3	Performance of the classification BDT trained with the LHD.	118
7.4	Performance of the classification BDT trained with the reconstruction BDT.	121
7.5	Objects reconstructed with the Reco-BDT set 1.	122
7.6	Objects reconstructed with the Reco-BDT set 2.	123
7.7	Expected limits on the $H^+$ production cross-section for the single-lepton channel.	125
B.1	Effect of the smoothing on the distribution of fake electrons.	148
C.1	Post-fit distribution of the leading-jet $p_T$ in the signal regions, for the fit performed with the 200 GeV signal hypothesis.	150
C.2	Post-fit distribution of the leading-jet $p_T$ in the signal regions, for the fit performed with the 800 GeV signal hypothesis.	151
C.3	Post-fit distribution of the $M_{bb}^{\text{Max}(p_T)}$ variable in the signal regions, for the fit performed with the 200 GeV signal hypothesis.	152
C.4	Post-fit distribution of the $M_{bb}^{\text{Max}(p_T)}$ variable in the signal regions, for the fit performed with the 800 GeV signal hypothesis.	153
C.5	Post-fit distribution of the $H_1^{\text{all}}$ variable in the signal regions, for the fit performed with the 200 GeV signal hypothesis.	154
C.6	Post-fit distribution of the $H_1^{\text{all}}$ variable in the signal regions, for the fit performed with the 800 GeV signal hypothesis.	155
C.7	Post-fit distribution of the $H_T^{\text{jets}}$ variable in the signal regions, for the fit performed with the 200 GeV signal hypothesis.	156
C.8	Post-fit distribution of the $H_T^{\text{jets}}$ variable in the signal regions, for the fit performed with the 800 GeV signal hypothesis.	157
C.9	Pulls and constraints for the fit performed with the 200 GeV signal hypothesis.	161
C.10	Pulls and constraints for the fit performed with the 800 GeV signal hypothesis.	162
C.11	Injection tests for the 200 and 800 GeV signal hypotheses.	164
C.12	Additional tests on the fit performed with the 300 GeV signal hypothesis.	166
D.1	Performance of the classification BDT with an extended set of input variables.	169
E.1	Schematic representation of a neural network.	172
E.2	Schematic representation of an adversarial neural network.	173
E.3	Distribution of the NN discriminants for the ME generator uncertainty of the $t\bar{t} + \text{jets}$ sample, set 1.	176
E.4	Distribution of the NN discriminants for the ME generator uncertainty of the $t\bar{t} + \text{jets}$ sample, set 2.	177
E.5	Distribution of the NN discriminants for the PS and had. uncertainty of the $t\bar{t} + \text{jets}$ sample, set 1.	178

E.6	Distribution of the NN discriminants for the PS and had. uncertainty of the $t\bar{t}$ + jets sample, set 2.	179
-----	--	-----

# Introduction

The discovery of the Higgs boson announced by the ATLAS [1] and CMS [2] collaborations in 2012 [3] has been a milestone for the history of particle physics and a further confirmation of the validity of the Standard Model (SM) [4–6]. However, at the time of the discovery, both collaborations were very cautious about naming the particle as the SM Higgs boson. They only announced that they had observed a new resonance in the mass region around 125 GeV, which was compatible with the Higgs boson predicted by the Standard Model. The reason is that, despite the experimental confirmations of the theory, the SM is unable to explain several observed phenomena. There is therefore a certain expectation that new particles will be found as evidence for physics Beyond the Standard Model (BSM).

Further studies demonstrated that the newly discovered resonance had indeed the features of the SM Higgs boson. However, many open questions still stand. Among the issues left unresolved by the Standard Model, there is the fact that gravity is not included in the theory. In addition, the SM does not provide any reasonable candidate for dark matter, nor does it justify the tiny masses of the neutrinos or the matter-antimatter asymmetry in the Universe. It is therefore reasonable to think of the Standard Model as a low energy theory which is part of a larger model incorporating the yet unknown "new physics". The question is, which energy scale will this new physics appear at? Several theories, as most of the supersymmetric models [7], predict that new phenomena will appear around the TeV scale, in reach of the searches conducted at the Large Hadron Collider (LHC) [8].

A feature shared by many theoretical extensions of the SM is the inclusion of a non-minimal Higgs model, containing additional scalar states. For example, Two Higgs Doublets Models (2HDM) [9] predict five scalar particles, two neutral CP-even scalars, a neutral CP-odd pseudoscalar and two charged scalars. The observation of one of these additional bosons would clearly point to BSM physics.

In this work, a direct search for heavy charged Higgs bosons (heavier than the top quark) decaying into a top-bottom pair is presented. The primary production process is in association with top and bottom quarks. Each signal event is therefore expected to contain two bottom quarks and two top quarks. Top quarks generally decay into a  $b$ -quark and a  $W$  boson, which decays again either leptonically or hadronically. This work focuses on the case in which one of the  $W$  bosons produces a lepton (electron or muon) and a neutrino, while the other decays into light quarks. The final state therefore contains four  $b$ -quarks, one lepton, two light quarks and one neutrino.

Limits on the production of charged Higgs bosons have been previously obtained by various experiments, starting with the Aleph, Delphi, L3, and OPAL collaborations [10]. Recent

published results include a search for  $H^\pm \rightarrow tb$  in the 180 – 600 GeV mass range by CMS [11] and in the 200 – 600 GeV mass range by ATLAS [12], setting upper limits on the production cross section of 2.0 – 0.13 pb and 6 – 0.2 pb respectively.

The search presented in this thesis is performed using  $36.1 \text{ fb}^{-1}$  of proton–proton ( $pp$ ) collisions collected by the ATLAS experiment during 2015 and 2016, at a centre-of-mass energy of 13 TeV. The analysed mass range goes from 200 to 2000 GeV, including for the first time values above 600 GeV. The limits on the  $H^\pm$  cross section are set by means of a binned profile likelihood fit of the simulated signal and SM backgrounds to the collected data. The fit is performed on a discriminant obtained by combining several kinematic variables through machine learning techniques. This allows to separate signal and background, improving the sensitivity of the analysis. The largest part of background comes from  $t\bar{t} + \text{jets}$  production, whose final state can include two top quarks and two bottom quarks, similarly to the signal. The thesis is organised as follows: Chapter 1 presents the basic theoretical notions and an extended motivation for the search. The statistical methods used to set upper limits on the charged Higgs boson cross section are described in Chapter 2. Chapter 3 includes an overview of the experimental setup, while Chapter 4 describes the object reconstruction techniques. The Monte Carlo modelling used for the simulation of signal and background processes is presented in Chapter 5. Details regarding the analysis and the results are discussed in Chapter 6. The thesis follows the corresponding paper published by ATLAS [13], for which the author of this work has been a main contributor. The paper takes into account both the single-lepton and the dilepton channel of the  $H^\pm$  decay. The writer contributed to the analysis of the single-lepton channel and to the combination of the two. All the strategies discussed in this work are implemented in the paper, with the exclusion of a limited number of improvements that were developed after the start of the ATLAS publication process. These are discussed separately in Chapter 7. A summary of the work is provided in Chapter 8.

# Chapter 1

## Theoretical basis

This chapter contains the necessary theoretical notions for the understanding of the following work. It gives an overview of the most validated theory in particle physics, the Standard Model [4–6], and it continues describing possible alternative scenarios that could solve some of the issues of the current Standard Model formulation. A particular focus is given to models that include charged Higgs bosons, for which production and decay mechanisms are described. Results obtained by previous  $H^\pm$  searches are also shortly summarised. The last part of the chapter is dedicated to Monte Carlo methods for simulation of particle interactions. Natural units are used in this chapter and throughout the thesis, unless stated otherwise.

### 1.1 The Standard Model

The Standard Model (SM)<sup>1</sup> is a quantum field theory (QFT) that describes the interaction between particles. Its foundations were laid in the early sixties by Glashow, Salam, and Weinberg, who combined the electroweak theory with the Higgs mechanism. The model acquired its modern form with the theorisation of the asymptotic freedom of quarks, in 1973, and the inclusion of the strong interaction. Gravity is not incorporated in the model. Nevertheless, given that the gravitational interaction is much weaker than the other three forces, the lack of its inclusion didn't prevent the SM to become the theory of reference in particle physics, with many experimental confirmations [14].

From the mathematical point of view, the SM is a renormalisable gauge theory, described by the unitary group  $SU(3)_C \times SU(2)_L \times U(1)_Y$ .  $SU(3)_C$  is the group that describes the strong nuclear force, while  $SU(2)_L \times U(1)_Y$  is an inclusive representation of both the weak and electromagnetic (EM) forces<sup>2</sup>. The following sections, based on Ref. [15–17], introduce the particle content of the SM and provide a short summary of the most important features of the three interactions.

---

<sup>1</sup>A list of the most common abbreviations used in the thesis is provided in Appendix A.

<sup>2</sup>The subscripts C, L and Y are used to distinguish between the gauge groups of the SM and the general  $SU(N)$  and  $U(N)$  groups. The letter C stands for "colour", L for "left" and Y for "weak hypercharge" and they are related to specific features of the three interactions, as explained later in the chapter.

### 1.1.1 The particle content of the Standard Model

In the Standard Model, there are two main classes of elementary particles: fermions, and bosons. The main difference between the two is that fermions are 1/2 spin particles and they therefore obey to the Pauli exclusion principle [18], while bosons have integer spin. Figure 1.1 provides an overview of the currently known elementary particles. There are twelve fermionic particles (first three columns), four kinds of gauge bosons (fourth column) and the Higgs boson (fifth column). The first three columns are often referred to as the three "families" or "generations" of fermionic particles. Fermions are divided in leptons (electron, muon, tau, electron neutrino, muon neutrino, tau neutrino) and quarks (up, down, strange, charm, top, bottom). All fermions possess a weak charge and are subject to the weak interaction. Apart from neutrinos, the other fermions are also electrically charged and can interact via the electromagnetic force. In addition, quarks have a strong charge (colour). They can therefore interact via the strong, weak and electromagnetic forces.

Gauge bosons are the mediators of the interactions. There are eight gluons, mediators of the strong interaction, the  $W^\pm$ ,  $Z$  bosons, which mediate the weak force, and the photon, which is responsible for the electromagnetic interactions. These 11 particles have spin 1. The  $W^\pm$ ,  $Z$  bosons are massive, the others are not. The  $W$  bosons are also electrically charged. Apart from this spin-1 family, there is an additional spin-0 particle: the Higgs boson. The Higgs boson is massive and uncharged. It is not a force carrier. It is instead generated as an excitation of a scalar field, the Higgs field.

The SM predicts also antiparticles for all fermions (not depicted in Figure 1.1), with the same spin, lifetime and mass but opposite values for the other charged-like quantum numbers.

### 1.1.2 The particle interaction

The Standard Model uses the formalism of Lagrangians to describe particles and their interactions. The particles are interpreted as fields, functions of space and time. The SM is valid within the framework of special relativity, therefore all Lagrangians are Lorentz-invariant. In the simplest possible scenario of a free fermion  $\psi_f(x^\mu)$  (not interacting with other particles), the Lagrangian that describes the system can be written as:

$$\mathcal{L} = i\bar{\psi}_f\gamma^\mu\partial_\mu\psi_f - m\bar{\psi}_f\psi_f, \quad (1.1)$$

where  $m$  is the mass of the considered fermion and  $\gamma^\mu$  is a gamma matrix<sup>3</sup>. The first term of the equation is known as the kinetic term, while the second is the mass term.

---

<sup>3</sup>The gamma matrices, or Dirac matrices, are a set of 4x4 matrices with specific commutation rules that form a Clifford algebra [16].



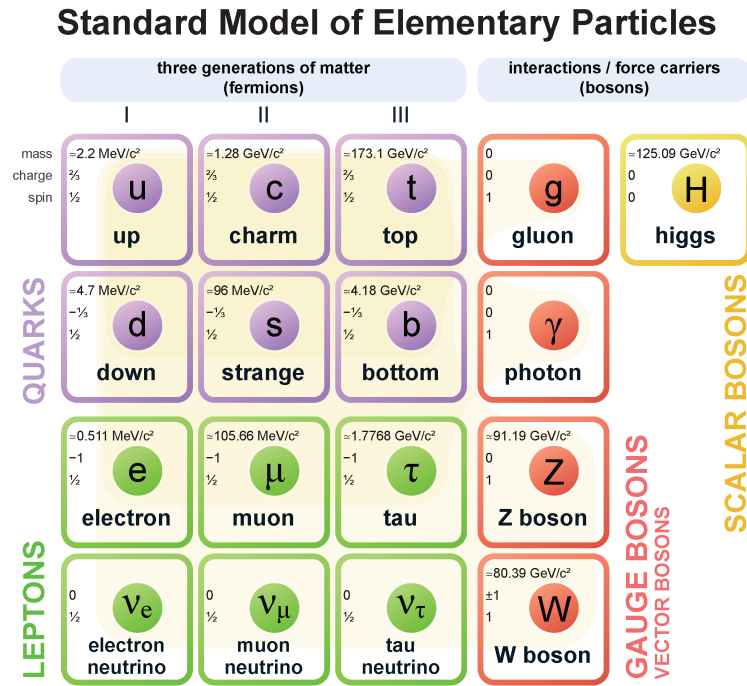


Figure 1.1: The elementary particles of the Standard Model. The first three columns contain the fermions, quarks at the top and leptons at the bottom. Gauge bosons are in the fourth column, while the Higgs boson is left alone in the fifth. Antimatter particles are not shown. Each box includes information about mass, charge and spin for the corresponding particle. No exact value is known for neutrino masses [19].

When a particle is interacting with other objects, new terms (interaction terms) appear in the Lagrangian. These terms are obtained by applying a local gauge transformation and requiring the invariance of the Lagrangian<sup>4</sup>. For example, the transformation associated to the electromagnetic interaction is:

$$\psi_f \longrightarrow e^{i\theta(x)Q}\psi_f = e^{i\theta(x)q}\psi_f, \quad (1.2)$$

where  $Q$  is the operator generating the transformation and  $q$  is the charge of the fermion. The associated gauge group is  $U(1)_Q$ , a rotation of the phase  $\theta$  of the fermionic field<sup>5</sup>. It can be easily verified that the Lagrangian of equation 1.1 is not invariant under this specific change. This can be solved by introducing a vector field  $A_\mu$  that, after the gauge transformation, becomes:

$$A_\mu(x) \longrightarrow A_\mu(x) + \partial_\mu\theta(x), \quad (1.3)$$

and by substituting the standard derivative  $\partial_\mu$  with a covariant derivative  $D_\mu = \partial_\mu - iqA_\mu(x)$ . The Lagrangian assumes the following form:

$$\mathcal{L} = i\bar{\psi}_f\gamma^\mu\partial_\mu\psi_f - m\bar{\psi}_f\psi_f + q\bar{\psi}_f\gamma^\mu\psi_f A_\mu. \quad (1.4)$$

The last term embeds the interaction between fermions and the electromagnetic field  $A_\mu$ , i.e. between fermions and photons. The addition of the kinetic term of the photon ( $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ ) completes the description of the electromagnetic interactions and brings to the Lagrangian of Quantum Electrodynamics Theory (QED):

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}_f\gamma^\mu D_\mu\psi_f - m\bar{\psi}_f\psi_f, \quad (1.5)$$

where  $F_{\mu\nu}$  is the electromagnetic field strength ( $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ ).

An analogous procedure can be used to obtain the Lagrangians of the weak and strong interactions. The main difference is that the dimensionality of the gauge group is higher:  $SU(2)_L$  is the gauge group for the weak force and  $SU(3)_C$  for the strong nuclear force. The gauge transformation associated to  $SU(N)$  is:

$$\psi \longrightarrow e^{i\theta_n(x)\tau^n}\psi \quad n = 1, 2, \dots, (N^2 - 1), \quad (1.6)$$

---

<sup>4</sup>A theory is gauge invariant when there is at least one redundant degree of freedom (gauge), which can vary without affecting the physical content of the theory.

<sup>5</sup>Similarly to C, L and Y, the letter Q is used to distinguish the gauge group of the electromagnetic interaction from other  $U(1)$  groups.

where  $\tau_n$  are the  $SU(N)$  generators, similarly to the charge  $Q$  in case of  $U(1)_Q$ .

### 1.1.3 The strong interaction

The theory describing the strong nuclear force is called the Quantum Chromodynamics (QCD) [15]. It is a gauge field theory mathematically defined by the  $SU(3)_C$  group. The corresponding Lagrangian is given by:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^A F_A^{\mu\nu} + \sum_q \bar{\psi}_{q,a} (i\gamma_\mu D_{a,b}^\mu - m_q \delta_{a,b}) \psi_{q,b}, \quad (1.7)$$

$$D_{a,b}^\mu = \partial^\mu \delta_{a,b} + ig_s t_{a,b}^C A_C^\mu.$$

The field tensor in this case is:

$$F_A^{\mu\nu} = \partial^\mu A_A^\nu - \partial^\nu A_A^\mu - g_s f_A^{BC} A_B^\mu A_C^\nu, \quad (1.8)$$

where  $f_A^{BC}$  is the "structure constant" of the gauge group and it embeds the self-interaction between the 8 carriers of the strong nuclear force  $A_\mu^C$ , called gluons. The  $\psi_{q,a}$  is the quark spinor, characterised by the colour charge  $a$ , either **red**, **green** or **blue**. The  $\gamma_\mu$  are the Dirac matrices, while  $t_{ab}^C$  is a matrix representation of the group generators (derived by the Gell-Mann matrices [20]). Colour is preserved in strong interactions, therefore gluons need to carry both colour and anti-colour.

The  $g_s$  is the strong coupling constant. It is usually redefined as  $\alpha_s = \frac{g_s^2}{4\pi}$ . Despite being called a "constant", it actually depends on the energy scale  $Q$ :

$$\alpha_s(Q^2) \approx \frac{1}{(11 - 2n_f/3) \log(Q^2/\Lambda_{QCD}^2)}. \quad (1.9)$$

In this formulation,  $n_f$  corresponds to the number of quark flavours with masses above the energy scale, while  $\Lambda_{QCD}$  is a parameter determined experimentally and known as QCD cut-off energy. At energies higher than the cut-off (ca. 200 MeV), the denominator decreases the value of  $\alpha_s$ , allowing a perturbative treatment of QCD. This feature is called "asymptotic freedom". In contrast, the strong coupling grows quickly when approaching to  $\Lambda_{QCD}$ . As a consequence, a perturbative approach is precluded at low energies and quarks are induced to combine in colour-less multiplets named mesons, formed by a quark and an anti-quark, or baryons, made by three quarks or three anti-quarks. This property is known as "quark confinement".

### 1.1.4 The electroweak model

At high energies, above  $\sim 100$  GeV, the weak and electromagnetic forces can be described by the same model: the electroweak (EW) theory. In this case, the interaction is described by the  $SU(2)_L \times U(1)_Y$  group. The weak interaction is encoded in the  $SU(2)_L$  term. The weak force acts exclusively on the left-handed particles (and right-handed antiparticles), thus the L subscript<sup>6</sup>. As a consequence, parity can be violated in weak interactions [21, 22]. The generators of the  $SU(2)_L$  transformation are defined as  $T_{1,2,3} = \frac{\tau_{1,2,3}}{2}$ , where  $\tau_{1,2,3}$  are the Pauli matrices.  $T$  is called weak-isospin<sup>7</sup>. The  $U(1)_Y$  group has only one generator, called hypercharge  $Y$ . The charge generator, mentioned in Section 1.1.2, follows as:

$$Q = T_3 + \frac{Y}{2}. \quad (1.10)$$

The electroweak Lagrangian can be written as:

$$\mathcal{L} = -\frac{1}{4}\hat{W}_{\mu\nu}\hat{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + i\sum_f \bar{\psi}_f \gamma^\mu D_\mu \psi_f, \quad (1.11)$$

where:

$$D_\mu = \partial_\mu - ig \sum_a \frac{W_\mu^a \tau_a}{2} - ig' \frac{Y}{2} B_\mu, \quad (1.12)$$

$$\hat{W}_{\mu\nu} = \partial_\mu \hat{W}_\nu - \partial_\nu \hat{W}_\mu - g \hat{W}_\mu \times \hat{W}_\nu, \quad \hat{W}_\mu = \{W_\mu^a, a = 1, 2, 3\},$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu.$$

$W_{\mu\nu}^a$  and  $B_{\mu\nu}$  are the gauge fields of  $SU(2)_L$  and  $U(1)_Y$ , which couple through the  $g$  and  $g'$  constants. The first two terms of the Lagrangian are the kinetic terms, while the third term embeds the interaction between the fermionic fields  $\psi_f$  and bosons.

The Lagrangian of Eq. 1.11 does not include bilinear (mass) terms for massive particles. Writing these terms, while preserving a complete gauge symmetry, is impossible. In fact,  $SU(2)_L \times U(1)_Y$  is not preserved in nature. At low energies, the symmetry is spontaneously broken into  $U(1)_Q$ . Mathematically, this can be achieved through the introduction of an

<sup>6</sup>Given a spinor  $\psi$ , its left and right components are defined as:  $\psi_L = \frac{1-\gamma^5}{2}\psi$ ,  $\psi_R = \frac{1+\gamma^5}{2}\psi$ .

<sup>7</sup>Left-handed fermions have a weak isospin equal to  $1/2$ . They are therefore organised in weak doublets: neutrinos and up-type quarks have the third component of the weak isospin equal to  $T_3 = 1/2$ , while down-type quarks and charged leptons have  $T_3 = -1/2$ . The right handed fermions are singlets with  $T = 0$ .

additional field, as discovered by Higgs, Englert and Brout [23, 24]. The method foresees the presence of a  $SU(2)$  scalar doublet  $\phi = (\phi^+, \phi^0)$  with positive hypercharge  $Y = 1$ , together with a corresponding scalar term in the EW Lagrangian:

$$\mathcal{L}_s = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (1.13)$$

When  $\mu^2 < 0$ , the scalar potential has a minimum at:

$$\langle \phi \rangle_0 = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad (1.14)$$

where  $v = \sqrt{-\mu^2/\lambda}$  is the vacuum expectation value (VEV) of the doublet. The VEV can't be in the charged part of the doublet, otherwise  $U(1)_Q$  is not preserved. It is possible to redefine the field  $\phi$ , expanding the neutral term around the vacuum state:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \quad (1.15)$$

where  $H$  is known as the Higgs field. Taking into account the previously defined covariant derivative and the new definition of  $\phi$ , the  $(D^\mu \phi)^\dagger (D_\mu \phi)$  term of the scalar Lagrangian can be re-written as:

$$|D_\mu \phi|^2 = \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8}g^2(v + H)^2|W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(v + H)^2|gW_\mu^3 - g'B_\mu|^2. \quad (1.16)$$

By defining new fields as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}}, \quad A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2 + g'^2}}, \quad (1.17)$$

the following bilinear terms appear in the Lagrangian:

$$M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2}M_Z^2 Z_\mu Z^\mu + \frac{1}{2}M_A^2 A_\mu A^\mu. \quad (1.18)$$

$W^\pm$ ,  $Z$  and  $A$  are the bosons of the weak and electromagnetic interactions, with masses:  $M_W = \frac{1}{2}vg$ ,  $M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}$  and  $M_A = 0$ . Fermionic masses can be obtained by introducing Yukawa couplings  $\lambda_f$  between the fermions and the Higgs field ( $m_f = \lambda_f v/\sqrt{2}$ ). The Higgs boson is also massive, with  $m_h = \sqrt{2\lambda}v$ . Ref. [25] provides a more detailed discussion.

## 1.2 Beyond the Standard Model, a supersymmetric approach

Despite the successful results of the Standard Model, a certain number of fundamental problems are still unsolved. A first example are the neutrinos, which stay massless in the SM. This is in contrast with solar and atmospheric neutrino observations, which support the theory of neutrino oscillations, only possible with massive particles [26]. Moreover, galaxy rotational curves and observations of the relic microwave background support the hypothesis of a dark matter component to the Universe [27], which again does not find explanation in the SM. The abundance of matter with respect to antimatter lacks of a justification as well. The CP violation of the Standard Model is in fact insufficient to describe the observed asymmetry, which is orders of magnitude larger than the predicted value [28, 29]. Finally, there is an important theoretical issue: the SM does not include the fourth fundamental force, gravity. While not exhaustive, these examples are emblematic of the fact that the Standard Model is unable to provide a complete description of the physical laws governing the Universe. It is therefore plausible that the SM, instead of being the "theory of everything", could be an effective theory, valid at low energies, and manifesting from a more inclusive model. This model should provide a natural solution for the problems above and contain all the features of the SM that demonstrated to be correct. In the most extreme scenario, it should appear around the Planck energy scale (ca.  $10^{19}$  GeV), when gravitational quantum effects are supposed to become relevant. However, there can be some complications if the energy scale for new physics is set to very large values. If the SM is to be considered invalid above a certain cut-off energy  $\Lambda$ , the first order corrections to the mass of the Higgs boson will include terms that are proportional to  $\Lambda^2$ . As a consequence, the larger the value of  $\Lambda$ , the larger is the fine-tuning required to match the theoretical value of the Higgs-boson mass with the one measured by the LHC experiments.

Supersymmetric (SUSY) models [7] offer a promising solution for such fine-tuning. They theorise that, above a certain energy scale ( $\sim \text{TeV}$ ), an additional symmetry will appear, producing a "supersymmetric partner" for each boson, quark and lepton, which generate corrections terms that are opposite to the ones of non-supersymmetric particles. As an example, one can consider the correction to  $m_H^2$  coming from a loop containing a fermion  $f$ , depicted in Figure 1.2 (a). This term generates a negative correction  $\sim -|\lambda_f|^2 \Lambda^2$  to the Higgs mass, where  $\lambda_f$  is the coupling between the fermion and the Higgs boson. If there existed a scalar  $\tilde{f}$

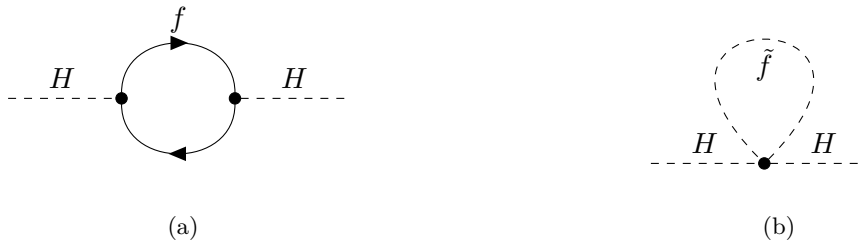


Figure 1.2: One-loop correction to the Higgs mass from (a) a fermion  $f$ , and (b) a scalar  $\tilde{f}$ .

with an analogous coupling to the Higgs, then a loop diagram as the one in Figure 1.2 (b) would generate a term  $\sim |\lambda_f|^2 \Lambda^2$ , removing the quadratic divergence<sup>8</sup>.

Apart from solving the fine-tuning problem, there are other reasons in favour of SUSY models. For example, they often include candidates for dark matter and they allow a natural unification of the coupling constants of the different interactions at high energy scales (see Figure 1.3). Ref. [7] and [31] provide a good introduction to supersymmetric models and have been widely used as guidelines for this and the next sections.

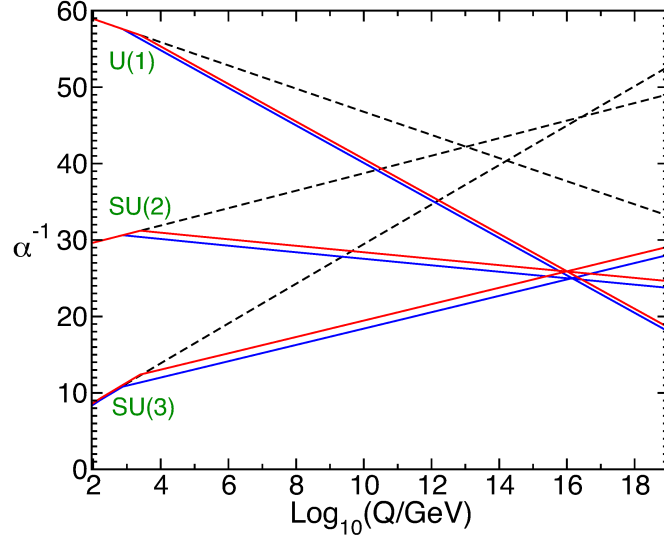


Figure 1.3: The evolution of the running constants of the three fundamental forces in the SM (dotted lines) and in the Minimal Supersymmetric Standard Model (solid lines) at two-loop level. Blue and red lines correspond to different parametrisations of the model [31].

### 1.2.1 The Minimal Supersymmetric Standard Model

The simplest SUSY model is called MSSM (Minimal Supersymmetric Standard Model), which requires the minimal amount of new degrees of freedom with respect to the Standard Model. The SM particles and their superpartners maintain the same electric charge, weak isospin and colour, while they differ in spin. The particle content of the MSSM is reported in Table 1.1. The partners of the SM gauge bosons are known as gauginos: gluons have "gluino" partners  $\tilde{g}$ , while the electroweak bosons are associated with "winos" ( $\tilde{W}^\pm, \tilde{W}^0$ ) and the "bino"  $\tilde{B}$ . The physical  $Z$  and  $\gamma$  fields are produced by mixing of the  $W^0$  and  $B$ . Similarly, the supersymmetric partner produce a zino  $\tilde{Z}^0$  and a photino  $\tilde{\gamma}$ . Superpartners of fermions are generally re-named adding an "s" in front to the SM name: quarks become "squarks", leptons are "sleptons" and so on. Finally, the theory needs two separate Higgs doublets

<sup>8</sup>It has to be noticed that the fine-tuning problem does not appear in the strict context of the Standard Model (which is renormalisable). It is only introduced when  $\Lambda$  is interpreted as a non-renormalisable cut-off that defines the range of validity of the SM. Other approaches that allow to completely reabsorb the quadratic divergences do exist [30].

$H_u = (H_u^+, H_u^0)$  and  $H_d = (H_d^-, H_d^0)$  to prevent gauge anomalies coming from the Higgs superpartners (Higgsinos):  $\tilde{H}_u = (\tilde{H}_u^+, \tilde{H}_u^0)$  and  $\tilde{H}_d = (\tilde{H}_d^-, \tilde{H}_d^0)$  [31].

Chiral fields		spin 0	spin 1/2
Squarks, quarks	$Q$	$(\tilde{u}_L, \tilde{d}_L)$	$(u_L, d_L)$
(x3 families)	$\bar{u}$	$\tilde{u}_R^*$	$u_R^\dagger$
	$\bar{d}$	$\tilde{d}_R^*$	$d_R^\dagger$
Sleptons, leptons	$L$	$(\tilde{\nu}, \tilde{e}_L)$	$(\nu, e_L)$
(x3 families)	$\bar{e}$	$\tilde{e}_R^*$	$e_R^\dagger$
Higgs, Higgsinos	$H_u$	$(H_u^+, H_u^0)$	$(\tilde{H}_u^+, \tilde{H}_u^0)$
	$H_d$	$(H_d^0, H_d^-)$	$(\tilde{H}_d^0, \tilde{H}_d^-)$

Gauge fields		spin 1/2	spin 1
gluinos, gluons		$\tilde{g}$	$g$
winos, W bosons		$\tilde{W}^\pm, \tilde{W}^0$	$W^\pm, W^0$
bino, B boson		$\tilde{B}$	B

Table 1.1: MSSM particle content. Chiral fields are reported at the top and gauge fields at the bottom.

A new quantum number is also defined in SUSY models, the  $R$ -parity.  $R$  is defined as:

$$R = (-1)^{3(B-L)+2S}, \quad (1.19)$$

where  $B$ ,  $L$  and  $S$  are the baryonic number, the leptonic number and the spin. SM particles have  $R = 1$ , while SUSY states have  $R = -1$ . In the MSSM,  $R$ -parity is conserved. As a consequence, all SUSY processes must include an even number of SUSY particles. Furthermore, the lightest SUSY particle (LSP) of the MSSM, typically associated to one of the neutralinos (mixture of Higgsinos and neutral gauginos), is stable and has a very small coupling with baryonic matter. It is therefore a very good candidate for dark matter [31].

### 1.2.2 The MSSM Lagrangian

In the MSSM, the interaction between particles is characterised by a new potential  $W$ , generally known as the "superpotential", which can be written as:

$$W = \bar{u} \mathbf{y}_u Q H_u - \bar{d} \mathbf{y}_d Q H_d - \bar{e} \mathbf{y}_e L H_d + \mu H_u \cdot H_d, \quad (1.20)$$

where  $Q$ ,  $L$ ,  $\bar{u}$ ,  $\bar{d}$  and  $\bar{e}$  are the superfields listed in Table 1.1. The  $y$ 's are Yukawa matrices



in the three-generation space. The construction of the MSSM Lagrangian goes beyond the purpose of this thesis, but its general form is:

$$\mathcal{L}^{MSSM} = \mathcal{L}_{\text{SUSY}}^{\text{MSSM}} + \mathcal{L}_{\text{BR}}^{\text{MSSM}}. \quad (1.21)$$

$\mathcal{L}_{\text{SUSY}}^{\text{MSSM}}$  is the supersymmetric invariant part, which includes chiral and gauge terms that describe the interaction between sfermions and gauginos, as well as their couplings with the superpotential.  $\mathcal{L}_{\text{BR}}^{\text{MSSM}}$  is the symmetry breaking part: to avoid the reintroduction of quadratic divergences, only some terms, known as "soft-breaking", are accepted. The supersymmetry is expected to be broken because superpartners with the same mass of SM particles have not been observed. A limited number of gauge invariant soft SUSY-breaking terms are available:

- gaugino masses:  $-M_1 \tilde{B} \tilde{B} - M_2 \tilde{W} \tilde{W} - M_3 \tilde{g} \tilde{g} + \text{h.c.}$
- sfermion masses:  $-Q^\dagger \mathbf{m}_Q^2 Q - L^\dagger \mathbf{m}_L^2 L - \bar{u} \mathbf{m}_u^2 \bar{u}^\dagger - \bar{d} \mathbf{m}_d^2 \bar{d}^\dagger - \bar{e} \mathbf{m}_e^2 \bar{e}^\dagger$
- Higgs masses:  $-m_{H_u}^2 H_u^\dagger H_u - m_{H_d}^2 H_d^\dagger H_d - (b H_u H_d + \text{h.c.})$
- trilinear couplings:  $-\bar{u} \mathbf{a}_u Q H_u + \bar{d} \mathbf{a}_d Q H_d + \bar{e} \mathbf{a}_e L H_d + \text{h.c.}$

The  $\mathbf{m}$  and  $\mathbf{a}$ 's are  $3 \times 3$  matrices in the family space [31]. A full expansion of the SUSY Lagrangian in equation 1.21, generates a large number of mass terms. The following is a focus on the Higgs sector, of particular interest for this thesis.

### 1.2.3 The MSSM Higgs sector

The scalar part of the superpotential can be expanded as:

$$\begin{aligned} V_H = & (|\mu|^2 + m_{H_u}^2)(|H_u^0|^2 + |H_u^+|^2) \\ & + (|\mu|^2 + m_{H_d}^2)(|H_d^0|^2 + |H_d^-|^2) \\ & + [b(H_u^+ H_d^- - H_u^0 H_d^0) + \text{h.c.}] \\ & + \frac{1}{8}(g^2 + g'^2)(|H_u^0|^2 + |H_u^+|^2 - |H_d^0|^2 - |H_d^-|^2)^2 \\ & + \frac{1}{2}g^2 |H_u^+ H_u^{0\dagger} + H_d^0 H_u^{-\dagger}|^2 \end{aligned} \quad (1.22)$$

The neutral components of the doublets must acquire non-zero vacuum expectation values  $v_u$  and  $v_d$  so that  $SU(2)_L \times U(1)_Y$  is broken into  $U(1)_Q$ . Re-expressing the doublets by shifting around the vacuum expectation values and expanding the potential further, mass terms for

gauge and Higgs bosons appear as:

$$\begin{aligned}
m_Z^2 &= \frac{1}{2}(g^2 + g'^2)(v_u^2 + v_d^2), \\
m_{W^\pm}^2 &= \frac{1}{2}g^2(v_u^2 + v_d^2), \\
m_{A^0}^2 &= 2|\mu|^2 + m_{H_u}^2 + m_{H_d}^2, \\
m_{H^\pm}^2 &= m_{A^0}^2 + m_W^2, \\
m_{H,h}^2 &= \frac{1}{2} \left( m_{A^0}^2 + m_Z^2 \pm \sqrt{(m_{A^0}^2 + m_Z^2)^2 - 4m_Z^2 m_{A^0}^2 \cos^2 2\beta} \right),
\end{aligned} \tag{1.23}$$

where  $\beta = \arctan(v_u/v_d)$ .  $H^\pm$  are positive and negative Higgs bosons,  $A^0$  is the CP-odd neutral state, while  $h$  and  $H$  are CP-even<sup>9</sup> neutral states ( $m_h < m_H$ ). Given the SM-like Higgs discovery, it is natural to work in the alignment limit in which either  $h$  or  $H$  couples to massive particles like the SM Higgs does. The choice generally falls on  $h$  and corresponds to the limit  $\cos(\alpha - \beta) \rightarrow 0$  [32, 33], with  $\alpha$  being the rotation angle that diagonalises the mass matrix of the CP even states and the following relation being valid between the masses of the Higgs bosons and the mixing angles  $\alpha$  and  $\beta$ :

$$\frac{\sin 2\alpha}{\sin 2\beta} = - \left( \frac{m_H^2 + m_h^2}{m_H^2 - m_h^2} \right), \quad \frac{\tan 2\alpha}{\tan 2\beta} = \left( \frac{m_{A^0}^2 + m_Z^2}{m_{A^0}^2 - m_Z^2} \right). \tag{1.24}$$

There are two final remarks to be made: the first is related to flavour-changing neutral currents (FCNC) [34]. Such processes have not yet been experimentally observed but they are allowed by models including two Higgs doublets. A possible way to suppress them is to allow fermions to only couple with one of the two doublets: for example, up-type quarks can couple to  $H_u$  and down-types and leptons with  $H_d$ . Models of this kind are known as Type-II models and they include the MSSM<sup>10</sup>.

The second remark is that 2HDMs can be incorporated in the SM without the need of SUSY. ATLAS results are however interpreted in specific MSSM scenarios, hence the decision of introducing SUSY in this chapter. The interpretation scenarios used in this analysis are  $m_h^{\text{mod-}}$  and hMSSM [35, 36]. The  $m_h^{\text{mod-}}$  model is obtained by fixing the top-squark mixing parameter [37] in order to maximise the agreement between the masses of the light CP-even Higgs boson with the one of the observed 125 GeV boson. In the hMSSM scenario, the mass of the light CP-even Higgs boson is directly fixed to 125 GeV.

A summary of the main differences between the MSSM and Two Higgs Doublets Models is provided in Ref. [38, 39]. For each 2HDM type there are 7 free parameters: the masses of the Higgs bosons,  $\tan\beta$ ,  $\cos(\alpha - \beta)$  and the diagonal term of the mass matrix

<sup>9</sup>CP stands for Charge-Parity. A CP transformation swaps all particles with their anti-particles and mirrors their spatial coordinates.

<sup>10</sup>Type-I models couple all charged fermions to the same doublet.

of the scalar doublets ( $m_{12}$ ). An MSSM-like regime would correspond to  $m_h \approx 125$  GeV,  $m_H = m_{A^0} \approx \max(m_h, m_{H^\pm})$ ,  $\cos(\alpha - \beta) \approx 0$  and  $m_{12}$  as in Eq. 2.15 of Ref. [38]. The remaining free parameters are  $m_{H^\pm}$  and  $\tan\beta$ . With this approximation, the Higgs couplings in a 2HDM of Type-II are the same as in the MSSM (at tree-level).

### 1.3 Searching for charged Higgs bosons

This thesis focuses on a search for charged Higgs bosons, targeting their decay into a top-bottom quark pair, which is one of the most probable modes for charged Higgs bosons with a mass larger than the top-quark mass. Besides SUSY models, charged Higgs bosons appear in many scenarios that include two Higgs doublets or Higgs triplets [9, 39–42]. The following provides a short summary of the phenomenology of the  $H^\pm$  decay and production.

#### 1.3.1 Charged Higgs boson decays

Charged Higgs bosons  $H^{+11}$  decay predominantly with the following modes:

$$H^+ \rightarrow c\bar{s}, \quad (1.25)$$

$$H^+ \rightarrow t\bar{b}, \quad (1.26)$$

$$H^+ \rightarrow \mu^+\nu_\mu, \quad (1.27)$$

$$H^+ \rightarrow \tau^+\nu_\tau, \quad (1.28)$$

If the  $H^+$  mass ( $m_{H^+}$ ) is smaller than the top-quark mass ( $m_t$ ), the decay of Eq. 1.26 is not possible and the decay modes of Eq. 1.28 and 1.25 would dominate. In case of  $m_{H^+} \geq m_t + m_b$ , all decays are possible<sup>12</sup>. The evolution of the branching ratios of heavy charged Higgs bosons as a function of the  $H^+$  mass and  $\tan\beta$ , in the  $m_h^{\text{mod-}}$  scenario, is shown in Figure 1.5. The dominant decay channels are  $t\bar{b}$  and  $\tau^+\nu_\tau$ . The two have a comparable branching fraction around 200 GeV, but the  $t\bar{b}$  mode prevails at higher masses, where the BR of the  $\tau^+\nu_\tau$  channel does not exceed the 15% of the total. Apart from the listed channels, possible decay modes include  $H^+ \rightarrow e^+\nu_e$ , decays through loops, as  $H^+ \rightarrow W^+\gamma$  and  $H^+ \rightarrow W^+Z$ , and additional decays into quark pairs, as  $H^+ \rightarrow c\bar{b}$  and  $H^+ \rightarrow t\bar{s}$ . They have a negligible contribution to the total branching ratio. Decays to SUSY final states can be relevant at high masses, while decays involving additional Higgs bosons ( $H^+ \rightarrow W^+h/H/A^0$ ) are only allowed in selected models. More details on the available decay modes and their dependence on the MSSM parameters can be found in Ref. [43–45].

<sup>11</sup> $H^-$  will not be explicitly mentioned from now on. It is always implied that, given the quark anti-quark symmetry, a decay like  $H^+ \rightarrow t\bar{b}$  implies the existence of  $H^- \rightarrow t\bar{b}$ .

<sup>12</sup> $m_b$  is the mass of the bottom quark. In the definition of "heavy charged Higgs bosons", heavy stands for "heavier than the sum of the masses of the top and bottom quarks". This condition will be often shortened to "heavier than the top quark" and reported as  $m_{H^+} \gtrsim m_t$ .

### 1.3.2 Charged Higgs boson production

The Yukawa coupling between charged Higgs bosons and other SM particles is proportional to their masses, therefore Higgs bosons are mainly produced in association with heavy fermions,  $\tau$ ,  $c$ ,  $b$  or  $t$ , and bosons  $W^\pm$  or  $Z$  [43]. For  $m_{H^\pm} \lesssim m_t$ , charged Higgs bosons can also be produced from a top quark decay. The decay from the top quark is the dominant production mode in the low-mass regime at the LHC, while for  $m_{H^\pm} \gtrsim m_t$ , as of interest for this work, the main production scheme is in association with a top quark and from gluon interactions. Feynman diagrams [46] for the production of a charged Higgs boson and a top quark are illustrated in Figure 1.4. This analysis makes use of the four-flavour scheme (4FS) to model the signal production, which corresponds to the Feynman diagram on the left. In this case, the  $b$ -quark is not included in the initial state as a constituent of the proton but only gluons and lighter quarks are available to start the hadronic interaction. Recent studies have shown that predictions obtained using the 4FS are in good agreement with the ones that consider five flavours (5FS). Furthermore, they provide a better modelling of the final state kinematics and they are more robust against certain types of systematic uncertainties [47]. Figure 1.6 shows the cross-section for the production of a charged Higgs boson as a function of mass and  $\tan\beta$ , in the four- and five-flavour schemes.



Figure 1.4: Leading order Feynman diagrams for the production of a charged Higgs boson with a mass larger than the top-quark mass and in association with a top quark, in (a) the 4FS and (b) the 5FS.

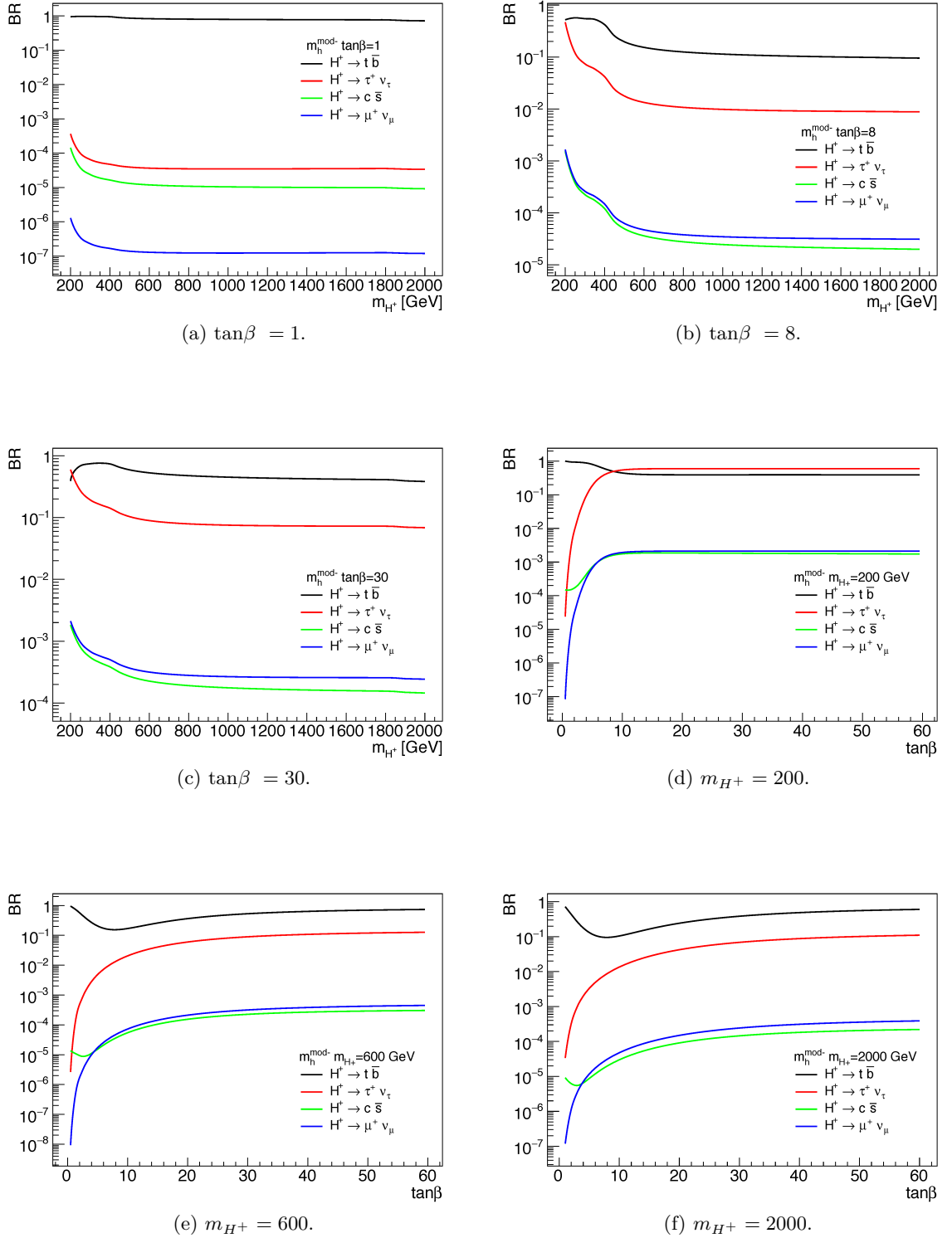


Figure 1.5: Branching ratios for a selected number of decay channels of the charged Higgs boson, shown as a function of mass, at (a)  $\tan\beta = 1$ , (b)  $\tan\beta = 8$  and (c)  $\tan\beta = 30$ , and as a function of the  $\tan\beta$  parameter, at (d)  $m_{H^+} = 200$  GeV, (e)  $m_{H^+} = 600$  GeV and (f)  $m_{H^+} = 2000$  GeV. The  $H^+ \rightarrow t\bar{b}$  decay is drawn in black. The values do not sum to one because small SM decay channels and modes involving BSM final states are not shown but they are considered in the BR evaluation. The plots have been produced with the hplumssm package [48].

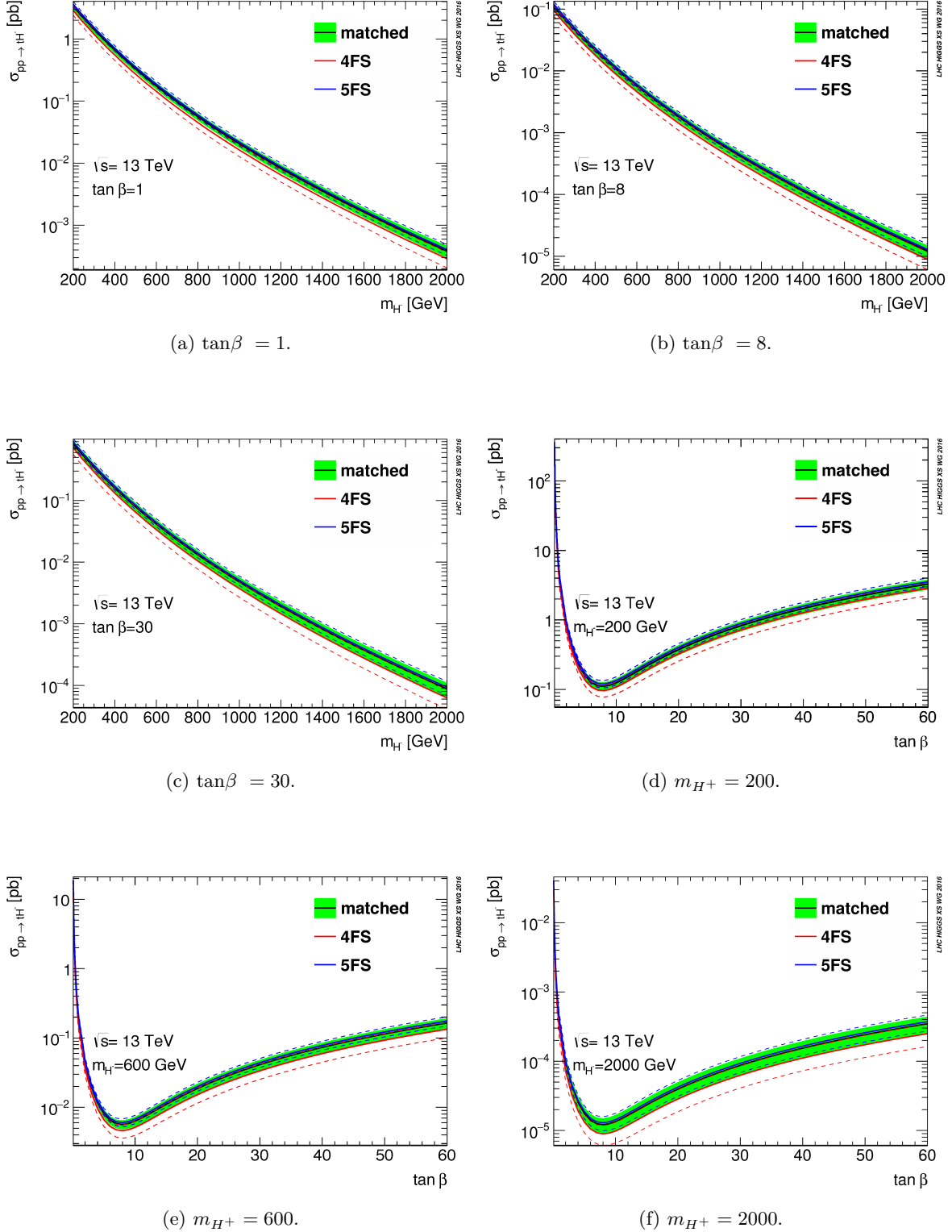


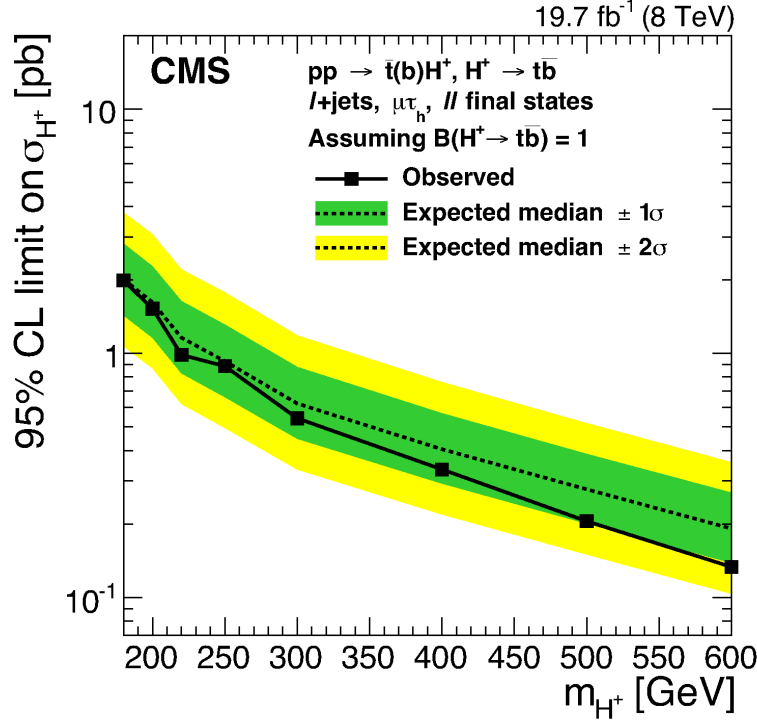
Figure 1.6: Cross-section for the production of charged Higgs bosons as a function of mass, at (a)  $\tan\beta = 1$ , (b)  $\tan\beta = 8$  and (c)  $\tan\beta = 30$ , and as a function of the  $\tan\beta$  parameter, at (d)  $m_{H^+} = 200$  GeV, (e)  $m_{H^+} = 600$  GeV and (f)  $m_{H^+} = 2000$  GeV. The predictions for the four- and five-flavour schemes are shown as red and blue solid lines. The corresponding error bands at 95% CL are drawn as dashed lines. A combination of the 4FS and 5FS predictions is shown in black with the corresponding uncertainty as a green band [45].

### 1.3.3 Experimental constraints

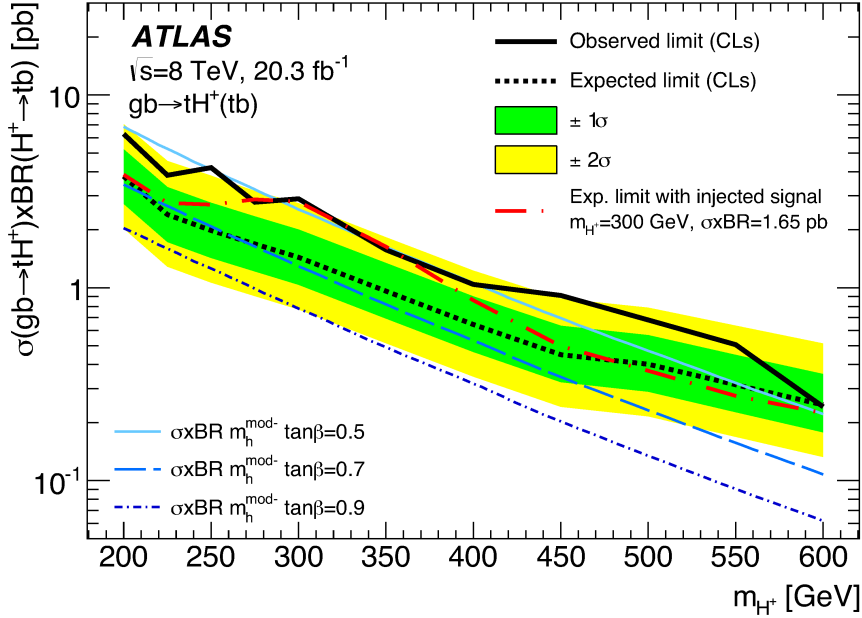
First limits on the production of charged Higgs bosons were put by the LEP collaborations [10], studying the 40–100 GeV mass range. Similarly, the CDF and D0 collaborations at the Tevatron investigated masses between 80 and 150 GeV [49, 50]. In addition, the D0 collaboration performed a dedicated study for the  $H^+ \rightarrow t\bar{b}$  decay, in the 180–300 GeV mass range [51]. The top-bottom final state has been investigated at the LHC as well. The CMS collaboration studied the 180–600 GeV mass range at a centre-of-mass energy of 8 TeV, producing upper limits on the  $H^+$  cross-section between 2.0 and 0.13 pb [11]. With the same centre-of-mass energy, the ATLAS collaboration investigated masses ranging from 200 to 600 GeV, setting limits between 6 and 0.2 pb [12]. The two results are reported in Figure 1.7. Recent searches from ATLAS and CMS also include the  $H^+ \rightarrow W^+Z$ ,  $H^+ \rightarrow c\bar{s}$  and  $H^+ \rightarrow \tau^+\nu_\tau$  channels [52–61].

Constraints are also available on the  $(m_{H^+}, \tan\beta)$  parameter space, in various interpretation scenarios. Figure 1.8 shows the constraints obtained by the ATLAS collaboration, investigating the  $H^+ \rightarrow t\bar{b}$  and  $H^+ \rightarrow \tau^+\nu_\tau$  channels at 8 TeV. The results are reported for masses above 200 GeV and in the  $m_h^{\text{mod-}}$  scenario of the MSSM. The two analysis are sensitive to different areas of the parameter space: the  $t\bar{b}$  analysis has a stronger reach at low  $\tan\beta$  values, while the  $\tau^+\nu_\tau$  analysis dominates at high values. The corresponding limits at 13 TeV [13, 54] will be shown at the conclusion of the thesis, in Chapter 6.

Indirect constraints on  $m_{H^+}$  and  $\tan\beta$  are also available from flavour-physics observables, as the relative branching ratio of  $B$ -meson decays [62–68]. In this regard, a combination of the most recent flavour-physics results excluded masses below 600 GeV, in a 2HDM scenario of Type-II (see Figure 1.9). Such results have a small dependence on the 2HDM parameters and can be compared to constraints obtained in most MSSM scenarios [38].



(a) CMS.



(b) ATLAS.

Figure 1.7: Expected and observed limits on the cross-section of the  $H^+ \rightarrow t\bar{b}$  production, as well as error bands for the 68% (in green) and 95% (in yellow) confidence intervals. The limits correspond to the results from (a) the CMS collaboration [11] and (b) the ATLAS collaboration [12], using  $pp$  collisions at 8 TeV. Final states including electrons or muons have been investigated by ATLAS. The CMS collaboration included the  $\mu\tau$  channel.



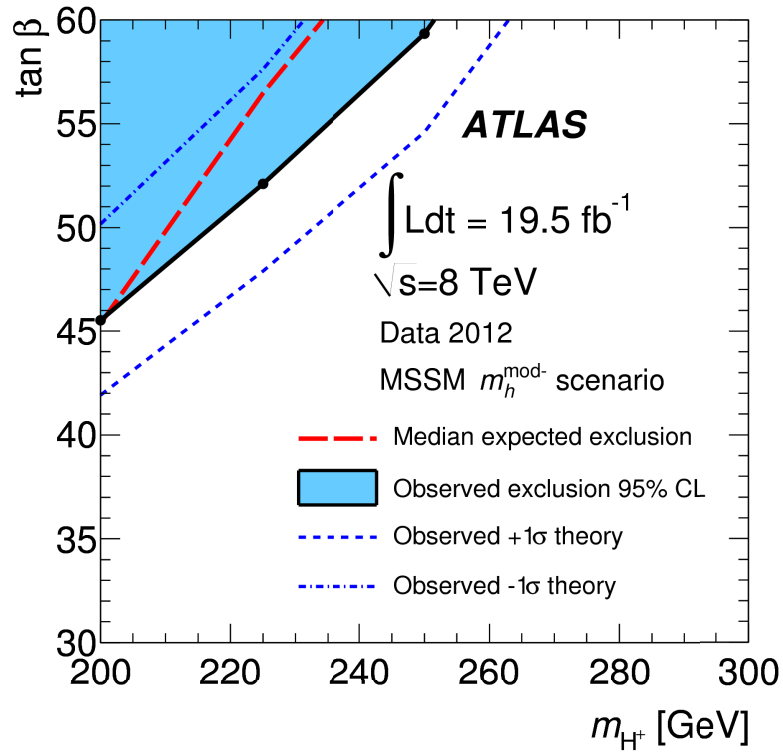
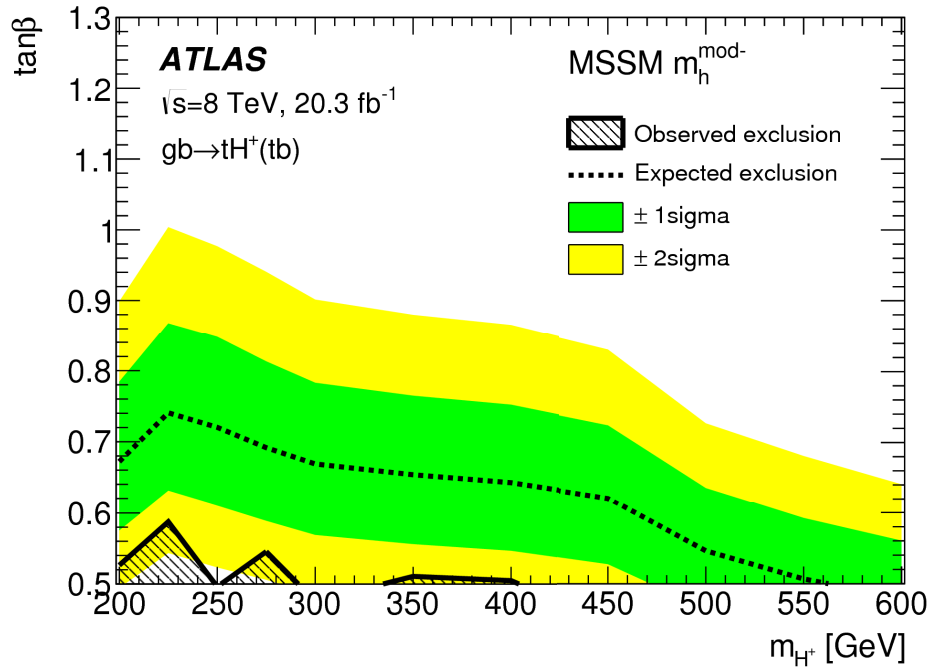

 (a)  $H^+ \rightarrow \tau^+ \nu_\tau$ .

 (b)  $H^+ \rightarrow t\bar{b}$ .

Figure 1.8: Expected and observed limits on  $\tan\beta$  as a function of the  $H^+$  mass for (a) the  $H^+ \rightarrow \tau^+ \nu_\tau$  [53] and (b) the  $H^+ \rightarrow t\bar{b}$  [12] decay channels, obtained by the ATLAS collaboration at 8 TeV. Error bands for the 68% (in green) and 95% (in yellow) confidence intervals are provided for the  $t\bar{b}$  decay mode. Errors at 68% CL are shown for the  $\tau^+ \nu_\tau$  channel (dashed lines). The exclusion limits are valid in the  $m_h^{\text{mod-}}$  scenario of the MSSM.

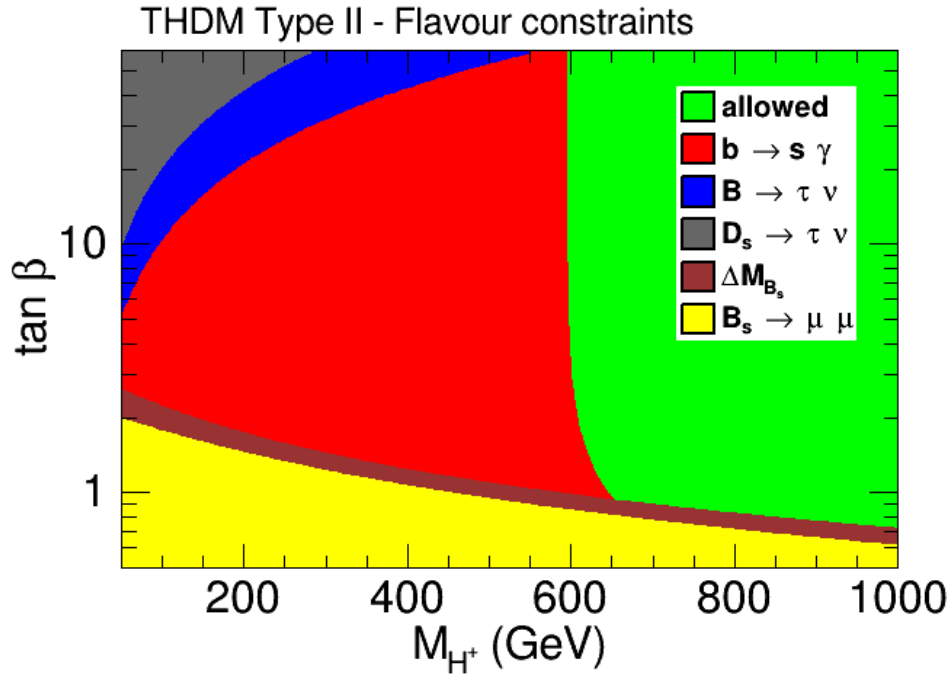


Figure 1.9: Region of the  $(m_{H^+}, \tan \beta)$  parameter space excluded at 95% CL by flavour physics constraints, in a 2HDM scenario of Type-II. The strongest constraints at low-masses come from the measurement of the branching ratio of  $B \rightarrow X_s \gamma$  (in red). The constraints at high masses are provided by the  $B_s \rightarrow \mu^+ \mu^-$  decay and the mixing of  $B$  mesons ( $\Delta M_{B_s}$ ). The green area (above 600 GeV) is allowed [38].

### 1.3.4 Targeted channel

This thesis focuses on a search for a charged Higgs boson  $H^+$ , decaying into a top-bottom final state, analysing  $pp$  collisions collected at 13 TeV. The investigated mass interval ranges between 200 and 2000 GeV. The considered production mode is in association with a top quark in the 4FS. Only the semi-leptonic decay channel of the  $H^+$  decay is considered. In this case, one of the  $W$  bosons (produced via  $t \rightarrow Wb$ ) decays hadronically, while the other one decays leptonically<sup>13</sup>. The corresponding Feynman diagram is depicted in Figure 1.10.

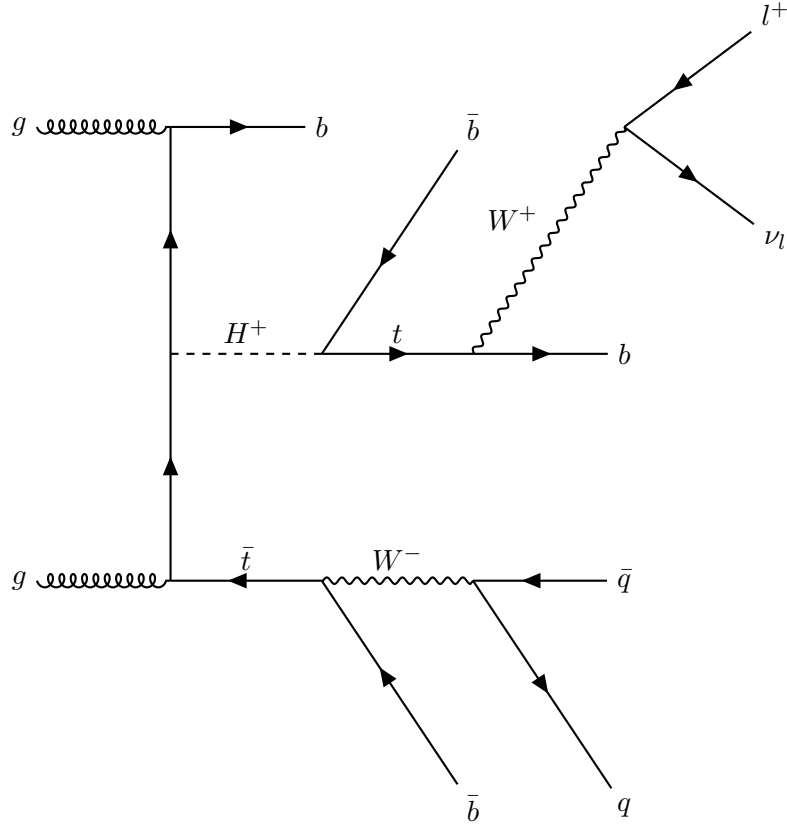


Figure 1.10: Leading order Feynman diagram for the production of a charged Higgs boson with a mass larger than the top-quark mass, in association with a top quark and the consequent decay in the semi-leptonic channel. The diagram shows the  $W^+$  decaying into leptons and the  $W^-$  decaying into hadrons. The opposite is also possible.

<sup>13</sup>The branching ratio of the  $W$  boson decay is dominated by the hadronic channel ( $ud$  and  $cs$  final states):  $\text{BR}_{\text{had}} \approx 67.6\%$ . Each of the leptonic decays has a contribution of  $\text{BR}_{e,\mu,\tau} \approx 1/9$ . Despite being the most probable, a charged Higgs boson decay with a full hadronic final state would be hidden by a large multi-jet background and it is therefore ignored. The dilepton decay is instead very rare and has a small contribution to the overall sensitivity.

## 1.4 Monte Carlo simulations

Monte Carlo (MC) generators are powerful software libraries that can simulate the particle interactions, allowing to compare the predictions of a specific model with the data recorded by the experiments. They are used in this work to provide predictions for all signal ( $H^+$ ) and background (SM) final states, with the exclusion of background events characterised by non-prompt leptons and misidentified jets (see Section 4.3.2), for which a data-driven procedure has been used. A schematic representation of a typical hadronic interaction produced in a  $pp$  particle collider is shown in Figure 1.11. The simulation can be divided into three steps: simulation of the hard scattering (innermost red lines), simulation of the parton showering and hadronisation (the blue, green and most external red lines), and simulation of the particle interaction with the detector (not shown). The parton shower takes also into account soft interactions between secondary partons (in violet).

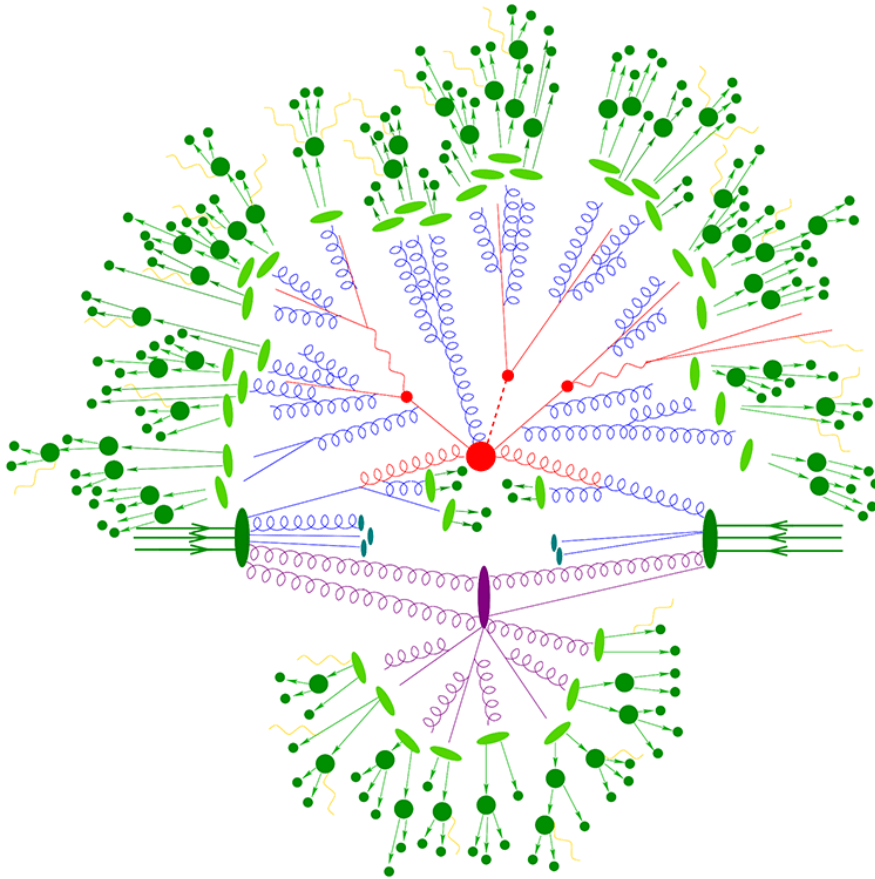


Figure 1.11: Schematic representation of a hadronic interaction produced in a  $pp$  particle collider. Two partons contained in the protons generate the hard scattering, in red, together with some softer interaction, in violet. Gluons (in blue) and quarks (in red) are subject to the hadronisation, which produces the visible final state, in green [69].

### 1.4.1 Event generation

#### The hard scattering

When two bunches of protons collide at the LHC, most of the particles continue their travel untouched, while only a small number interacts. Even between these, only few will start a hard scattering process, meaning an interaction characterized by a large momentum transfer ( $\gtrsim 1$  GeV). In first approximation, one could consider that only two partons (either quarks or gluons) inside the protons are originating such an interaction. In this case, the cross-section of their transition into a final state containing  $m$  partons can be approximated by the following expression:

$$\sigma(pp \rightarrow m) = \sum_{i,j} \int dx_1 dx_2 f_{i,p}(x_1, \mu_F^2) f_{j,p}(x_2, \mu_F^2) \hat{\sigma}_{i,j \rightarrow m}(x_1 x_2 s, \mu_R^2, \mu_F^2). \quad (1.29)$$

The  $f_{i,p}$  are the parton density functions (PDFs), which correspond to the probability density for finding the parton  $i$  inside the proton, carrying a fraction  $x$  of the proton momentum. The scattering itself is regulated by the cross-section  $\hat{\sigma}_{i,j \rightarrow m}$ . This cross-section is proportional to the square of the matrix element ( $M_{\text{XO}}$ ) of the transition, i.e. the sum over all transition amplitudes corresponding to processes with initial state  $i, j$  and final state  $m$ , at a centre-of-mass energy  $\sqrt{s}$ . The XO subscript characterises the order of perturbation theory used for the matrix element computation. The Leading Order calculation (LO) would correspond to the case where only  $m$  partons are considered in the final state. If additional emissions are taken into account, the calculation will be at the next-to-leading order (NLO) or higher. In this case, the cross-section can be written as:

$$\sigma \propto |M_{\text{NLO}}|^2 = |M_{\text{LO}}|^2 + 2\text{Re}(M_{\text{LO}} M_{\text{Virtual}}) + |M_{m+1}|^2, \quad (1.30)$$

where  $M_{m+1}$  is the matrix element corresponding to the final state with an additional parton, while the second term contains the interference between the LO diagram and the virtual Feynman diagrams, which include closed loops of virtual particles. Integrals involving loops can diverge when the momentum of the particles goes to very large values. The calculation is therefore truncated at a finite energy scale, known as the renormalisation scale  $\mu_R$  [70]. The  $\mu_F$  of Eq. 1.29 is instead the factorisation scale, an arbitrary parameter that defines the energy scale of the processes that are taken into account by the ME computation. Partons with energies below  $\mu_F$  are considered as part of the proton and accounted for within the parton distribution functions.

The generators responsible for the computation of the matrix element are known as fixed-order Monte Carlo generators. Examples are MG5\_aMC@NLO [71] and POWHEG [72]. The parton distribution functions are extracted from data, through measurements of deep

inelastic scatterings (DIS) and from collider experiments. A large number of PDF sets have been produced by different collaborations. The most used in ATLAS are PDF4LHC [73], NNPDF [74], CT14 [75] and MSTW [76].

### The parton shower

To take into account the branching of secondary objects, a parton shower (PS) algorithm is applied on top of the matrix element calculation. The PS approximates particle emissions at higher order of perturbation theory, neglecting virtual corrections. Parton showers can be seen as a sequence of  $p \rightarrow l, m$  partonic splittings, where the momentum of the parton  $p$  is shared between the partons  $l$  and  $m$ . These processes are controlled by the Altarelli-Parisi splitting functions  $P(z)$  [77], which define the probability of a parent parton to produce a daughter parton with a fraction  $z$  of its energy. The emissions generated by the PS are named initial state radiation (ISR) and final state radiation (FSR) accordingly to the origin of the particles, i.e. if they are produced by the initial state or the final state of the hard scattering process. Parton shower algorithms are commonly part of multi-purpose MC generators, which can be used as stand-alone generators or interfaced with fixed-order generators. Two common examples are PYTHIA [78] and HERWIG [79].

The perturbative approach employed by the PS works as long as the energy of the particles is larger than approximately 1 GeV. At lower energies, the strong coupling constant becomes very large, forcing quarks to combine into baryons and mesons. The transition of quarks and gluons to non-coloured particles is called hadronisation and it is approximated by phenomenological models. A common way to imagine the low-energy quark interactions is the Lund model<sup>14</sup> [80], which describes the strong interaction as an elastic string. In this case, the potential of the strong interaction between two quarks can be approximated by:

$$V(r) = \frac{4\pi}{3r} + kr, \quad (1.31)$$

where  $r$  is the distance between the particles and  $k$  is the elastic constant. By moving apart, quarks pull the elastic string and increase the potential. At a certain point, breaking the string (and producing a new pair of quarks) becomes energetically convenient. The process continues until the energy of the particles falls below a certain threshold. Resulting hadrons for which the decay length is larger than 10 mm are considered stable. Decay processes are taken into account for the others.

The hadronisation involves not only the partons generated by the primary hard scattering, but also other partons produced by secondary interactions between the incoming protons. The final state contribution from these processes is generally known as the underlying event (UE) and constitutes a non-irrelevant background.

---

<sup>14</sup>Another method used by many parton shower algorithms to simulate low energy interactions between partons is the Cluster model [80].

### 1.4.2 Detector simulation

While details regarding the ATLAS experiment are given in Chapter 3, this section would not be complete without introducing the method used to simulate the particle interaction with the detector. The most complete approach is the so called "Full-Simulation" (FullSim), used for the most important SM processes. The method exploits a package called GEANT [81], which is used to simulate the trajectory of the particles (and their decay products) through ATLAS and their interaction with the various elements of the detector. A geometry model is used to specify all materials, volumes and the intensity of the magnetic fields. The most time consuming part of this process is the showering of particles inside the calorimeters, which requires a large computational power. The smallest backgrounds and BSM processes are therefore simulated using the ATLAS-Fast II Simulation (AFII) [82]. In this case, the shape of the electromagnetic and hadronic showers is approximated using a parametrised simulation of the detector response, reducing the computational time of a factor 20 or more with respect to the FullSim. The two methods have a good agreement in terms of energy scale and resolution of the simulated showers, but some discrepancies are observed at low energies and need to be taken into account by using dedicated systematic uncertainties.





## Chapter 2

### Statistical methods

In order to test the predictions of a given model, experimental data and Monte Carlo simulations have to be compared using statistical methods. This chapter describes the tools used to set upper limits on the charged Higgs boson cross-section, defining an appropriate test statistic and introducing the multivariate technique adopted to enhance the separation between signal and SM processes. More details on the statistical methods used by the ATLAS collaboration can be found in Ref. [83].

#### 2.1 Hypothesis testing

Searching for new phenomena generally involves the comparison of distributions extracted from real data with distributions obtained from an idealised model. It is therefore necessary to define a figure of merit to quantify the agreement between the two. The approach followed in particle physics is called hypothesis testing. The method relies on the definition of a certain number of statistical hypotheses: a null-hypothesis  $H_0$ , which is generally associated to the absence of new physics, and one or more alternate hypotheses  $H_\mu$  that take into account BSM interactions, like the charged Higgs boson production. A single alternate hypothesis can be used when the BSM process is fully defined by the model under investigation. However, this analysis makes no assumption regarding the cross-section of the charged Higgs boson production. It is therefore necessary to test a variety of signal hypotheses, for all possible cross-sections. This is achieved by defining the signal strength:

$$\mu = \frac{\sigma_x}{\sigma_{\text{ref}}}, \quad (2.1)$$

where  $\sigma_x$  can be any cross-section and  $\sigma_{\text{ref}}$  is a reference value, typically 1 pb. With this definition,  $H_\mu$  can be imagined as a continuous spectrum of signal hypotheses that will approach to the background-only hypothesis  $H_0$  when  $\sigma_x$  is close to 0.

The level of agreement between data and predictions for a given signal strength is quantified by computing the  $p$ -value,  $p_\mu$ . The  $p$ -value corresponds to the probability of observing a deviation from  $H_\mu$  that is more extreme than the one measured with data, assuming that  $H_\mu$

is true. When  $p_\mu$  is lower than a certain threshold, the hypothesis should be rejected. The  $p$ -value is defined as:

$$p_\mu = \int_{t_{\text{obs}}}^{\infty} f(t|H_\mu) dt, \quad (2.2)$$

where  $t$  is a test statistic, i.e. a function related to a given observable of the analysed sample, and  $f(t|H_\mu)$  is the distribution predicted for such function.  $t_{\text{obs}}$  is the value of the test statistic measured on data. The significance  $Z = \Phi^{-1}(1 - p_\mu)$  is often preferred to the  $p$ -value to quantify the level of disagreement between data and predictions ( $\Phi$  is the cumulative distribution of a Standard Gaussian). The significance corresponds to the quantile of the distribution, i.e. the number of standard deviations at which  $\Phi$  is equal to  $1 - p_\mu$ . In particle physics, the common approach is to require an observed significance of at least  $Z = 5$  to reject the background-only hypothesis and  $Z = 1.64$  to reject an alternate hypothesis. These correspond to  $p_0 = 2.87 \cdot 10^{-7}$  and  $p_\mu = 0.05$  respectively. A schematic representation of the relation between the  $p$ -value and the test statistic is given in Figure 2.1.

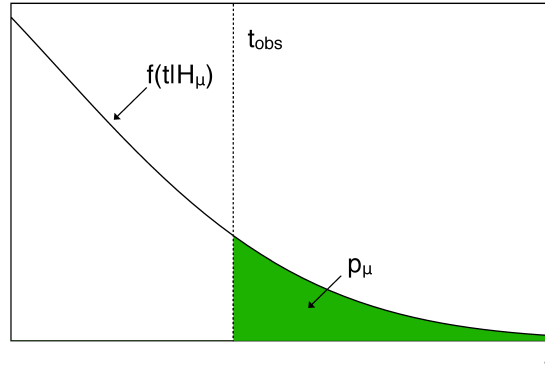


Figure 2.1: Graphical representation of the  $p$ -value obtained by the measurement of a test statistic  $t$  for a certain hypothesis  $H_\mu$ .

A powerful test statistic can be defined by using likelihood functions. Given a binned distribution of a kinematic variable for which both measured and expected results are available, the likelihood  $L$  is defined as:

$$L(\mu, \boldsymbol{\theta}) = \prod_{i=1}^N \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))} \prod_{\theta_j \in \boldsymbol{\theta}} P(\theta_j). \quad (2.3)$$

The first term is the product of the Poisson probability density functions for all bins  $i$ , where  $N$  is the total number of bins, each with  $n_i$  measured entries. The expected number of entries is given by the sum of the expected number of background entries  $b_i$  and signal entries  $s_i$  (normalised to 1 pb). The signal is scaled by the signal strength  $\mu$ .  $s_i$  and  $b_i$  depend on  $\boldsymbol{\theta}$ , the array of the nuisance parameters (NPs). Such parameters quantify the

impact of the systematic uncertainties related to the signal and background modelling, as well as experimental uncertainties coming from the detector response and resolution. The second term is the product of the probability density functions  $P(\theta_j)$  of all the NPs, called penalty terms. Gaussian PDFs are generally used for systematic uncertainties that can assume both positive and negative values, Log-Normal distributions for cross-section uncertainties, which must be positive, and Gamma PDFs for the statistical uncertainties associated to the bins of the kinematic variable distribution [84]. As their name suggests, they penalise large deviations of the uncertainties from their nominal values.

The optimal  $\mu$  and  $\theta$ 's, as well as their errors, are not known a-priori and need to be extracted from a fit of the predicted distributions to the observed data, maximising the likelihood. The agreement between data and predictions for a specific signal strength can be obtained from the likelihood ratio<sup>1</sup>:

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})}, \quad (2.4)$$

where  $\hat{\theta}$  is the array of NPs that maximises the likelihood for the considered value of  $\mu$ , while  $\hat{\mu}$  and  $\hat{\theta}$  are unconditional values that maximise  $L$ . Although  $\lambda(\mu)$  is already a suitable test statistic, it is commonly rearranged as:

$$t_\mu = -2\ln\lambda(\mu). \quad (2.5)$$

With this formulation, a good agreement between data and  $H_\mu$  corresponds to values of  $t_\mu$  close to 0. Depending on the purpose of the measurement, the test statistic can be subject to further modifications. For example, the version used in this analysis to set upper limits on the cross-section of the  $H^+$  production is:

$$\tilde{t}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} < \mu \\ 0 & \hat{\mu} \geq \mu \end{cases} = \begin{cases} -2\ln\frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0 \\ -2\ln\frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & 0 < \hat{\mu} < \mu \\ 0 & \hat{\mu} \geq \mu \end{cases}. \quad (2.6)$$

With respect to Eq. 2.5,  $\tilde{t}_\mu$  assumes a positive signal and therefore negative values of  $\hat{\mu}$  are treated as  $\hat{\mu} = 0$ . Furthermore, the test statistic is set to 0 for  $\hat{\mu} \geq \mu$  because there is no interest in considering signal hypotheses for which the signal strength is lower than the observed value.

As already mentioned, a signal hypothesis would be conventionally excluded when the corresponding  $p$ -value is reasonably small ( $< 0.05$ ). This approach is however not appropriate

<sup>1</sup>The choice is justified by the Neyman-Pearson lemma [85].

when signal and background are poorly separated, because negative fluctuations of the background could lead to the rejection of valid signal hypotheses. This is particularly true if the expected number of signal events is small. A different quantity is therefore used for this purpose:

$$\text{CL}_s = \frac{p_\mu}{p_0}. \quad (2.7)$$

A signal hypothesis is rejected when  $\text{CL}_s < 0.05$ , avoiding the exclusion of areas of the phase space in which both  $p_\mu$  and  $p_0$  are small. Signal strength values for which  $\text{CL}_s = 0.05$  define the 95% confidence level (CL) limits [86].

Fitting the likelihood function to data leads to the observed limit. It is however important to be able to predict the expected limit either in a background-only hypothesis, or for a given signal model, before looking at the observed data. When the size of the sample under analysis is large enough, both the likelihood ratio and the distribution of the test statistic can be approximated to some analytical form. A single dataset, named Asimov, can then be used to extrapolate the median sensitivity of the measurement. This dataset is created in such a way that, when used to extrapolate the parameter estimators, their "true" value is obtained, suppressing statistical fluctuations [83]. This corresponds to setting the content of each bin to the expected amount  $\mu s_i + b_i$ , the nuisance parameters to 0 and the signal strength to the value of interest (0 for the background-only case). The general procedure to set upper limits on a given signal hypothesis is to first compute the expected limit under the background-only hypothesis, and then compare such result with the observed limit computed on data, eventually quantifying the deviation between the two.

## 2.2 Boosted decision trees

The choice of the distribution used to perform the likelihood fit can have a strong impact on the signal strength sensitivity. It is therefore important to select the variable that shows the best separation between signal and background. However, a single kinematic variable is often not suited for this purpose. In fact, the features of the signal can be embedded in correlations between multiple variables. It is partly possible to overcome this issue by combining multiple features into engineered variables. As an example, one could sum the transverse momenta of multiple particles, so as to include kinematic information for all the objects in a given event. Unfortunately, it is difficult to create the optimal variable "by hand", therefore this analysis makes use of a more refined technique. A set of variables, which encode some of the most important kinematic features of the  $H^+$  decay, is chosen, and then a boosted decision tree (BDT) [87] is trained on the entire set, using MC events. The BDT learns how to exploit the correlation between variables, in order to distinguish between signal and background events. The BDT is then applied on data and MC events, generating an output proportional to the

probability that each event has of being generated by signal. Such output is used to perform the profile likelihood fit.

The fundamental component of a BDT is a single decision tree. A tree is a construct that performs a certain number of cuts on the input variables so as to separate between the classes of interest. For example, the decision tree of Figure 2.2 is trained to distinguish between two classes, signal and background. The tree can take its decision based on three variables: the  $p_T$ ,  $\eta$  and  $\phi$  of a lepton. A metric called Gini-Index is used to select the variable and the position at which each cut is performed. The Gini-Index is defined as  $G = P_S(1 - P_S)$ , where  $P_S$  is the signal purity of the fraction of events passing a cut. The optimal cut is defined as the one that maximises the separation gain ( $SG$ ):

$$SG = G(\text{branch}) - G(\text{leaf 1}) - G(\text{leaf 2}), \quad (2.8)$$

where  $G(\text{branch})$  is the Gini-Index computed before performing the cut while  $G(\text{leaf 1})$  and  $G(\text{leaf 2})$  are the ones computed on the two subsets of events divided by the cut. The maximum number of subsequent cuts is known as the depth of the tree (the decision tree of Figure 2.2 has a depth of 3).

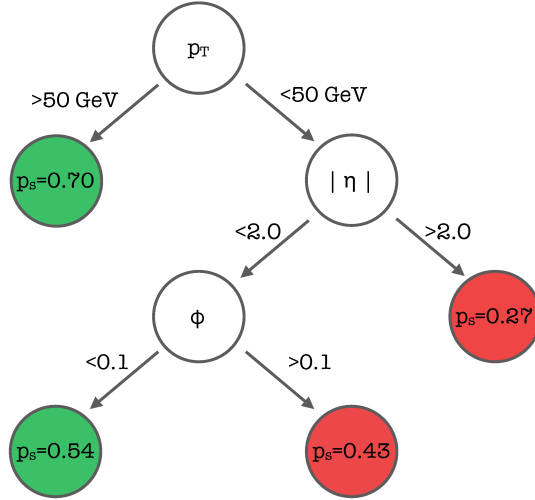


Figure 2.2: Schematic representation of a decision tree trained on a dataset composed of signal and background events. The training variables are the  $p_T$ ,  $\eta$  and  $\phi$  of a lepton. The output of the tree is the probability that each event has of being generated by signal. The probabilities do not correspond to the result of a real training but have been randomly generated. Values below 0.5 (in red) correspond to background-like events, values above 0.5 (in green) are associated to signal-like events.

Multiple decision trees can be combined into ensembles, known as "forests", to mitigate statistical fluctuations and improve the performances. In the case of BDTs, the trees are trained one after each other, using events that are weighted to give larger importance to the ones that were misclassified by the previous tree. The final probability is produced as

a weighted average of the scores of all trees. While the output of a single tree is a discrete distribution, BDTs produce pseudo-continuous distributions, spanning values from  $-1$  to  $1$  (or  $0$  to  $1$ ). The smaller the BDT output, the more background-like is the input event.

To maximise the performance of the BDT, one could increase the size of the forest or the depth of each tree. However, there is a limiting factor given by the number of training events. If the forest is too large or the trees are too deep, the BDT could overtrain, i.e. pick up statistical fluctuations characteristic of the training sample only. This would result in a sub-optimal classifier. To mitigate these effects, a  $k$ -fold training is performed:  $k$  BDTs are trained on different sub-samples that exclude  $1/k$  of the events, which are then used for the validation ( $k \geq 2$ ). The optimal BDT parameters are chosen as the ones that minimise the differences between the BDT output distributions for the training and validation events, while maximising the separation between signal and background. The separation power of the BDT is estimated using the Receiver Operating Characteristic (ROC) curve. The ROC curve is a graphical representation of the background rejection efficiency vs the signal acceptance efficiency of a given distribution. Larger values of the area under the ROC curve (AUC) correspond to better probabilities of assigning the correct class to the events. The AUC must be computed on the validation set to avoid biases coming from the training. An optimal classifier would have  $\text{AUC}=1$ , while no separation power corresponds to  $\text{AUC}=0.5$ . An example of ROC curve is provided in Figure 2.3.

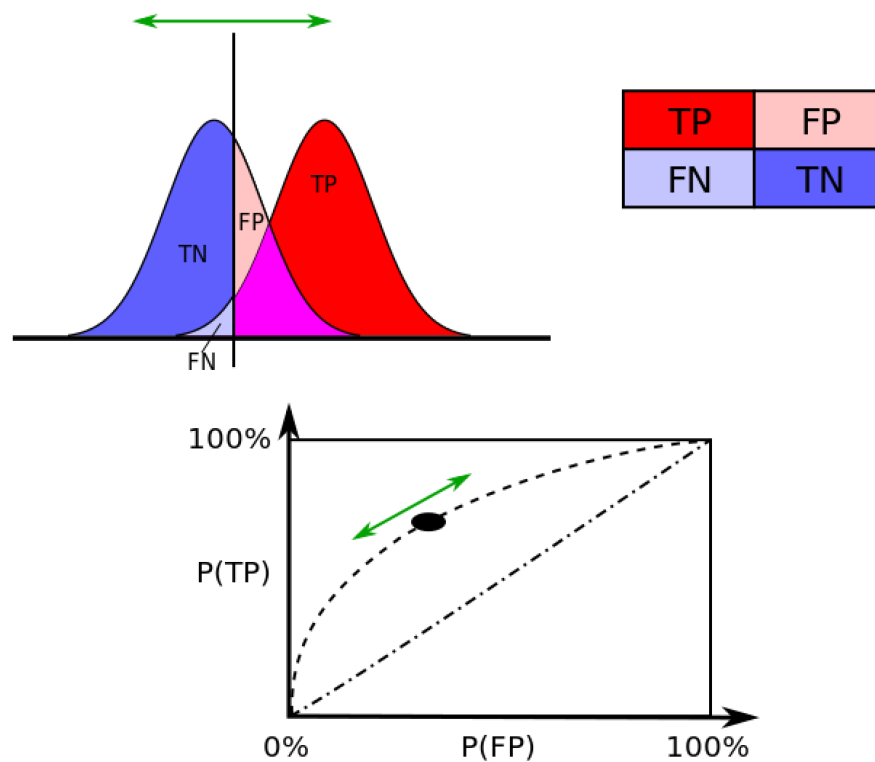


Figure 2.3: Example of a Receiver Operating Characteristic curve. The ROC curve, at the bottom, is a graphical representation of the separation between the Gaussian distributions drawn in the top left corner. The red distribution can be interpreted as the signal distribution and the blue as the background distribution. The ROC curve is built by drawing the rate of true positive (signal) events versus the rate of false positive (background) events above a moving threshold (the black vertical line between the Gaussian distributions) [88].





## Chapter 3

### The LHC and the ATLAS experiment

The Large Hadron Collider (LHC) [8] is a particle accelerator built by the European Organisation for Nuclear Research (CERN, Conseil Européen pour la Recherche Nucléaire), in close proximity to Geneva, across the Swiss-French border. It was constructed to verify the validity of the Standard Model and to find answers to the open questions that the current formulation of the SM is not able to explain. With this goal, various detectors have been built along the LHC to record the interactions of high energy particles. Among these detectors there is ATLAS (A Toroidal LHC ApparatuS) [1], responsible for the acquisition of the data used in this thesis. The first part of this chapter describes the Large Hadron Collider and its pre-accelerators, while the second part is dedicated to the ATLAS experiment.

#### 3.1 The Large Hadron Collider

The Large Hadron Collider is the largest particle accelerator ever built in human history. It is a synchrotron, an accelerator where the electric and magnetic fields are synchronised in order to maintain the particles along a closed circular path. The machine is located underground, in a tunnel with a circumference of 27 km. It consists of two parallel pipes, where two beams of particles travel at the same time in opposite directions. The particles are accelerated by radio-frequency (RF) cavities and kept on their trajectory by a large number of superconducting magnets. The two beam-lines cross in four intersection points, where the main experiments are located. The LHC is able to generate both protons and heavy-ions collisions, but, for the purpose of this thesis, the following will focus on  $pp$  collisions only.

The flux of particles travelling inside the LHC pipes is not continuous. Protons are grouped in bunches, each containing hundreds of billions of particles. When fully operational, the LHC can accommodate 2800 bunches, colliding with a frequency of 40 MHz. The design centre-of-mass energy of the collisions is  $\sqrt{s} = 14$  TeV, but the first collisions were operated at 7 TeV. The energy moved up to  $\sqrt{s} = 13$  TeV only recently, in 2015.

A set of pre-accelerators is used to gradually accelerate the particles before feeding them into the LHC. At first, the LINAC 2 seeds protons from a hydrogen-gas source and accelerates them to 50 MeV. The particles are then injected into the Proton Synchrotron Booster (PSB),

followed by the Proton Synchrotron (PS), where they reach 26 GeV. The protons are then passed to the Super Proton Synchrotron (SPS), which accelerates them to 450 GeV. Finally, the protons are delivered to the Large Hadron Collider and brought to 6.5 TeV.

An important quantity to be defined when dealing with particle accelerators is the Luminosity  $L$ , which relates the cross-section of a physical process to the event rate:

$$\frac{dN}{dt} = L\sigma. \quad (3.1)$$

$L$  depends on the specifications of the accelerator and it can be computed as:

$$L = \frac{nN_{p,1}N_{p,2}f}{A}, \quad (3.2)$$

where  $n$  is the number of particle bunches in each of the two beams,  $N_{p,i}$  is the number of protons per bunch and  $f$  is the collider revolution frequency ( $\sim 11$  kHz for the LHC).  $A$  is the area of the overlap region between the two beams at the collision point. The time integrated value of  $L$ , called integrated Luminosity, is used to measure the total amount of interactions that took place in a considered time interval.

### 3.1.1 The detectors

There are seven experiments that are currently using data coming from LHC collisions, plus many others in the close vicinity, profiting from beam-lines exiting the pre-accelerators. Two of these seven experiments are general-purpose detectors, the ATLAS experiment and the Compact Muon Solenoid (CMS) experiment [2]. They can detect a wide range of phenomena, aiming to perform both precise measurements of Standard Model processes and searches of BSM physics. The other five detectors are: ALICE [89], LHCb [90], TOTEM [91], MoEDAL [92] and LHCf [93]. The accelerator chain of LHC and the detector complex are shown in Figure 3.1.

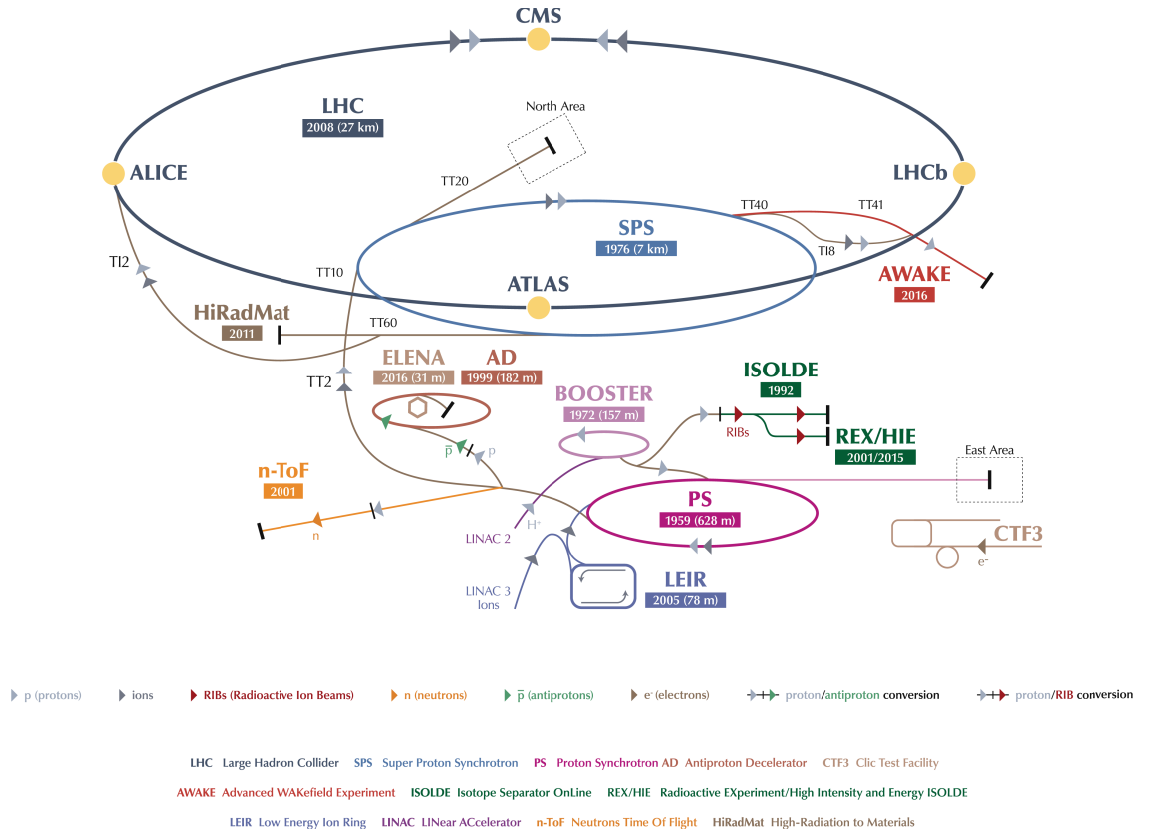


Figure 3.1: The LHC pre-accelerators and the detector complex. The ring of LHC is drawn in blue, at the top, together with the main experiments located at the four interaction points [94].

### 3.2 The ATLAS experiment

As already mentioned, there are four interaction points along the LHC ring. The closest to the CERN laboratories is called Point-1 (P1) and it is the one assigned to the ATLAS experiment. The facilities at Point-1 include the ATLAS control room and the detector itself, located underground in a cavern. The detector is of cylindrical shape, 44 m long, with a lateral diameter of 25 m and a weight of 7000 tons. Being a general-purpose experiment, ATLAS was built to measure all possible particle properties, as their mass, momentum, energy, charge and lifetime. Figure 3.2 shows a cut-away image of the detector. Going from the inner-most layer to the outside, there are the tracking system, the calorimeters and the muon detector. The tracking system detects the momentum and direction of the particles, the calorimeters measure their energy, while the muon spectrometer gathers further information about muons, which are highly penetrating particles and can easily pass through the previous stages of the detector. There are also two independent magnet systems, outside the tracking system and in the muon spectrometer, which contribute to the measurement of the momentum of the particles. All together, the different parts cover approximately the entire solid angle, guaranteeing an almost hermetic coverage. This is important to estimate the "missing energy", i.e. the amount of energy carried away by neutrinos or other undetected particles. The following pages contain a description of the different sub-detectors, for which additional details are available in Ref. [1].

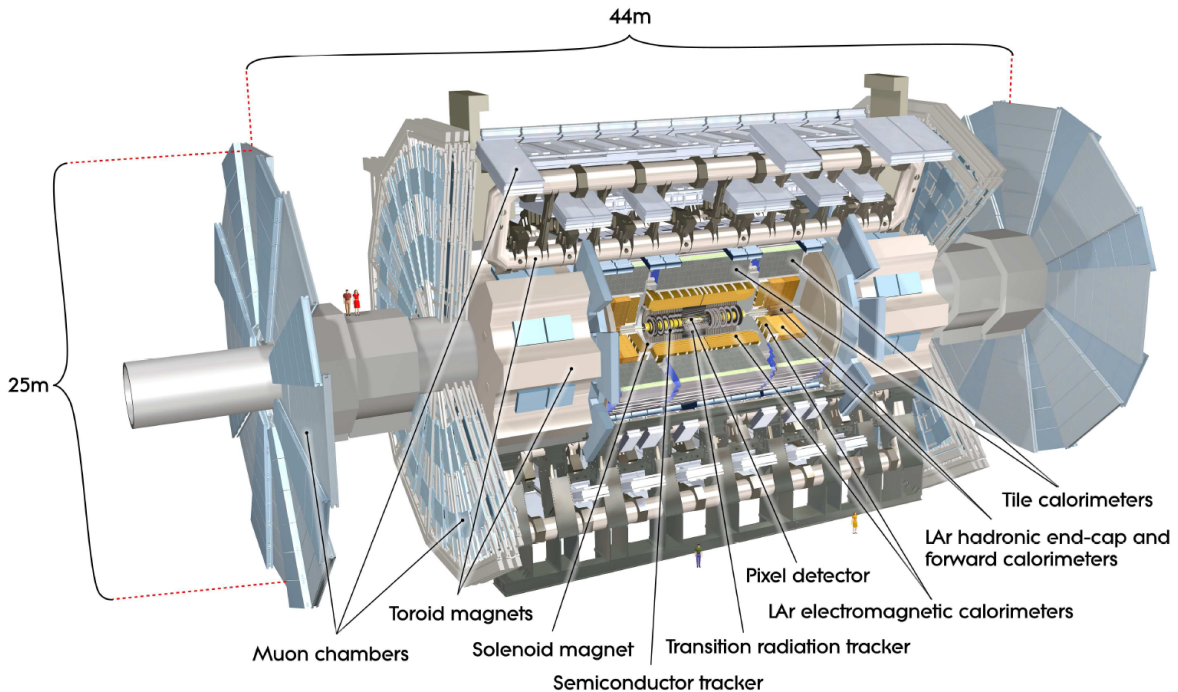


Figure 3.2: The ATLAS experiment and its sub-detectors. Going from the inner-most layer to the outside, there are the the Pixel detector, the semiconductor tracker, the transition radiation tracker, the solenoid magnet, the calorimeters (electromagnetic and hadronic), the toroid magnets and the muon spectrometer [95].

### 3.2.1 The ATLAS coordinate system

The ATLAS experiment uses a 3D Cartesian coordinate system, with the nominal interaction point defined as the origin. The  $x$ -axis points to the centre of the LHC ring, the  $y$ -axis is perpendicular to the ground and pointing up to the sky, while the  $z$ -axis is along the beam-line, pointing in the anticlockwise direction when looking at the LHC from above.

Cylindrical coordinates  $(\theta, \phi, z)$  are also used:  $\theta$  has a  $[0, \pi]$  range around the  $x$ -axis, on the  $z$ - $y$  plane, while  $\phi$  moves around the  $z$ -axis from  $-\pi$  to  $\pi$ . The angle  $\theta$  is rarely employed, in favour of a variable called pseudorapidity:

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

The advantage of using  $\eta$  over  $\theta$  is that the separation in pseudorapidity between two massless particles is an invariant quantity under a boost in the  $z$  direction. The angular distance between objects (in the  $\eta$ - $\phi$  space) is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (3.4)$$

Finally, variables such as energy and momentum, even if used, are generally less common than their transverse counterparts  $E_T = E \sin \theta$  and  $p_T = p \sin \theta$ , which are conserved in the interaction<sup>1</sup>. Figure 3.3 shows a schematic representation of the coordinate system.

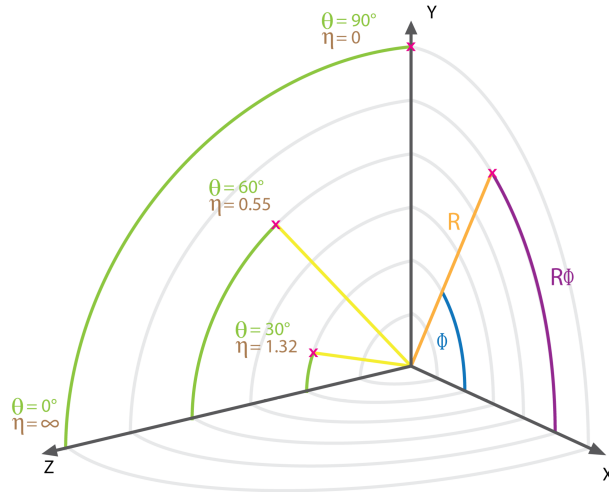


Figure 3.3: The ATLAS coordinate system [96].

<sup>1</sup>The longitudinal component of the parton momentum is unknown before the interaction, while the transverse component is approximately 0.

### 3.2.2 The inner detectors

The innermost part of the ATLAS detector is responsible for tracking charged particles generated in the proton–proton collision. Three sub-detectors contribute to the track reconstruction: the Pixel tracker (PIX), the semiconductor tracker (SCT) and the transition radiation tracker (TRT). Their layout is illustrated in Figure 3.4. They extend radially from 3.3 cm to about 1.1 m (the beam pipe radius is approximately 2.7 cm). The Pixel and SCT cover a solid angle up to  $\eta = \pm 2.5$ . The TRT is limited to  $|\eta| \leq 2$ . A precise tracking information not only allows to reconstruct the particle decays, it also gives important insights about the particle species and their momentum. In this regard, a 2 Tesla magnetic field permeates the inner detectors (ID), bending the trajectory of charged particles and allowing to reconstruct their momentum and charge. The magnetic field is generated by a superconducting solenoidal magnet that surrounds the tracking system. The transverse momentum resolution of the inner detectors is approximately:

$$\frac{\sigma p_T}{p_T} \approx \frac{0.05\%}{\text{GeV}} \cdot p_T \oplus 1\%. \quad (3.5)$$

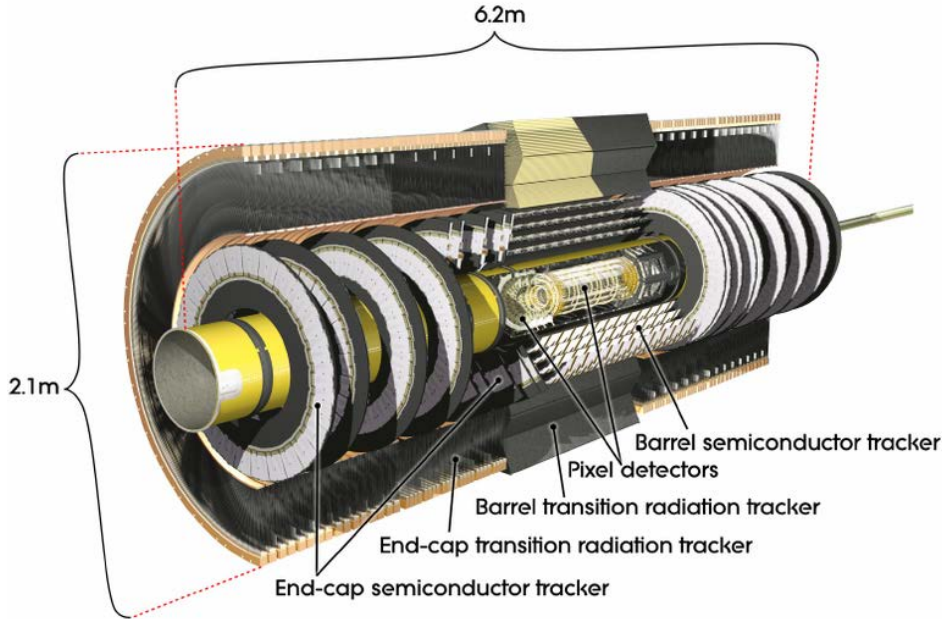


Figure 3.4: The ATLAS inner detectors: the Pixel tracker, the semiconductor tracker and the transition radiation tracker [1].

The Pixel tracker is a silicon detector, composed of four concentric layers of modules in the central part (barrel) and three disks at the two ATLAS extremities (end-caps). The modules consist of a layer of oxygen enriched n-in-n doped silicon [97], on which a certain depletion voltage (up to 600 V) is applied in order to maximise the collection efficiency for the charge deposited by particles passing through. The smallest sensible unit of the modules is a "pixel". Pixels have a nominal size of  $50 \times 250 \mu\text{m}^2$  in the innermost barrel layer (called Insertable

b-Layer, IBL [98]) and  $50 \times 400 \mu\text{m}^2$  in the other three layers. Such a small dimension is a necessary requirement to guarantee a low hit occupancy, allowing to reconstruct the trajectory of the many particles escaping from the interaction point and the displaced vertices produced by the decays of long-lived particles such as the  $B$ -mesons.

The SCT is the second silicon detector inside ATLAS. It is located outside the Pixel tracker, with four layers in the barrel and nine in each end-cap. The main difference with respect to the previous sub-detector is that the smallest sensible unit is not a pixel but a "strip". Strips are 6 cm long. They are disposed in parallel in the barrel region, with a pitch of  $80 \mu\text{m}$ , and they fan out radially in the end-caps, with a pitch ranging between 70 and  $90 \mu\text{m}$ . Each of the 4000 modules composing the system is constructed with two back-to-back layers of strips, slightly tilted ( $40 \text{ mrad}$ ) with respect to the each other. This allows to compensate the dimensional asymmetry of the sensors, guaranteeing a good 3D reconstruction of the particle hits.

Finally, the TRT is a gaseous detector. It is composed of drift tubes that alternate with a transition radiation detector. There are 73 tube layers in the barrel and 160 in the end-caps. The tubes have a diameter of 4 mm and they contain a  $30 \mu\text{m}$  diameter tungsten wire used as anode. Each tube is filled with an Argon-based gas mixture. The space between the tubes is filled with polypropylene. A charged particle passing through the tubes ionises the gas. Given the negative potential applied to the straws, the electrons are driven to the anode creating a signal. Furthermore, when the particle moves from the gas to the polypropylene, it emits transition radiation, with an intensity proportional to its Lorentz factor.

### 3.2.3 The calorimeters

Three calorimeters are placed outside the solenoidal magnet: the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the forward calorimeter (FCAL). They are instruments dedicated to the measurement of the particle energy. They cover a pseudorapidity range of  $|\eta| < 4.9$  and they extend up to 4.25 m of radius. They are sampling calorimeters, i.e. they are composed of alternate layers of absorbers and active media. Absorbers absorb the energy from the original particle, converting it into radiation and secondary particles, whose energy is then measured in the active medium. A cut-away image of the calorimeter system is presented in Figure 3.5.

The electromagnetic calorimeter measures the energy of photons, electrons and positrons. The electromagnetic interaction manifests as a combination of two processes: the emission of Bremsstrahlung radiation by electrons and positrons, and the conversion of photons into  $e^+/e^-$  pairs. These two effects combine together generating a cascade of particles (shower). Depending on the interacting material and the energy of the primary particle, the particle shower will be larger (and deeper) or smaller. The ability of the calorimeters of containing such showers is measured in radiation lengths  $X_0$ , where  $X_0$  corresponds to the length after

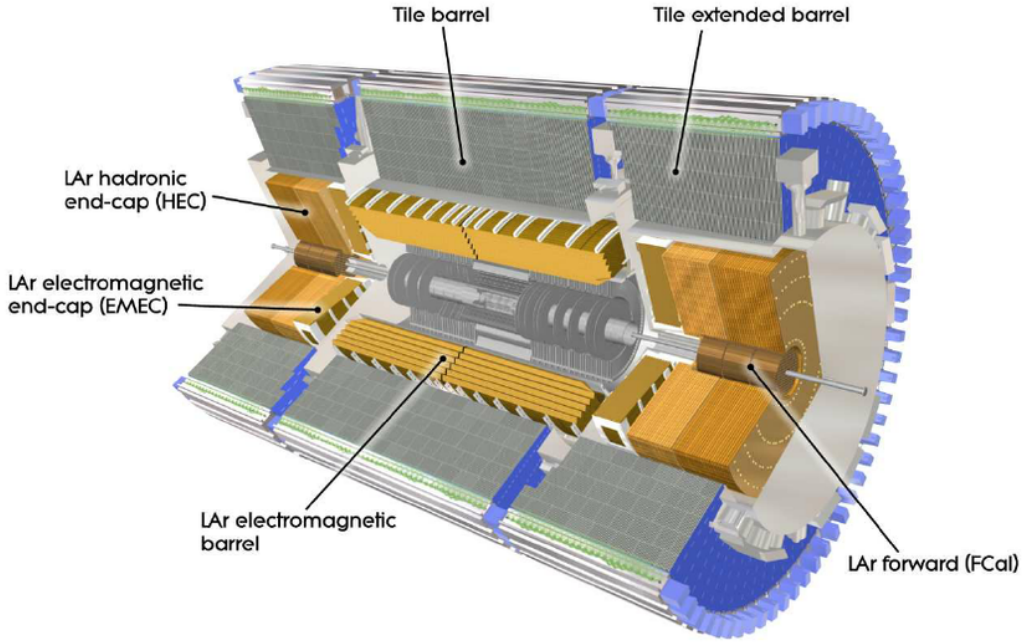


Figure 3.5: The electromagnetic, hadronic and forward calorimeters of ATLAS [1].

which the energy of a relativistic electron would be reduced of a factor  $1/e$  due to the exclusive emission of Bremsstrahlung radiation. The ATLAS electromagnetic calorimeter uses a combination of stainless steel and lead as absorber, and liquid Argon (LAr) as sampling material. The thickness of the calorimeter is approximately  $22 X_0$  in the barrel and  $24 X_0$  in the end-caps. The calorimeter is divided into different sections, with different granularities, depending on the radial depth. It covers a region up to  $\eta = \pm 3.9$ , but the area between the barrel and the end-caps ( $1.37 < |\eta| < 1.52$ ) is used for detector services. The energy of the particles passing in such region is therefore not measurable. The design energy resolution of ECAL is:

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus \frac{17\%}{E} \oplus 0.7\%. \quad (3.6)$$

The processes governing hadronic showers are different. The loss of energy is mainly caused by interactions with nuclei, production of secondary hadrons and hadronic decays. The extension of the showers is in this case characterized by the nuclear interaction length  $\lambda$ , i.e. the mean free path of a particle before undergoing an inelastic scattering. The hadronic calorimeter can be divided into 2 regions: the Tile calorimeter, used in the barrel, and the liquid Argon calorimeter, for the end-caps. The first uses steel as absorber and plastic scintillating tiles as the active medium, while the second alternates copper with liquid Argon. These two parts cover an  $\eta$  range of  $\pm 3.2$  and have an approximate thickness of  $7.4$  and  $12 \lambda$  respectively. The overall energy resolution for the hadronic system is  $50\%/\sqrt{E}$  in the Tile calorimeter and  $70\%/\sqrt{E}$  in the LAr calorimeter.



Given the necessity of an hermetic measurement, there is an additional calorimeter that covers a range of  $3.1 < |\eta| < 4.9$ . Such a forward calorimeter consists of three disks in each end-cap: the first one is optimised for electromagnetic showers and it is made of copper, while the other two disks are made of tungsten and they are mostly dedicated to hadronic processes. The active medium is liquid Argon. The total thickness is  $9.5 \lambda$ . The three sub-detectors overlap with each other, guaranteeing a good control over the transition regions.

### 3.2.4 The muon spectrometer

The muon system (MS) is the outermost part of the ATLAS detector. It covers a range up to  $|\eta| < 2.7$  and a radius between 4.25 and 11 meters. It is the biggest part of the detector and it is dedicated to the identification of muons. Similarly to the inner detectors, a strong magnetic field is responsible for bending muon trajectories, to allow a precise measurement of the particle momenta. The field is created by a toroid embedded in the muon detector, which has a bending power ranging between 2 and 8 T-m. The system is composed of four sub-detectors (see Figure 3.6): monitored drift tubes (MDTs), cathode strip chambers (CSCs), resistive plate chambers (RPCs), and thin-gap chambers (TGCs).

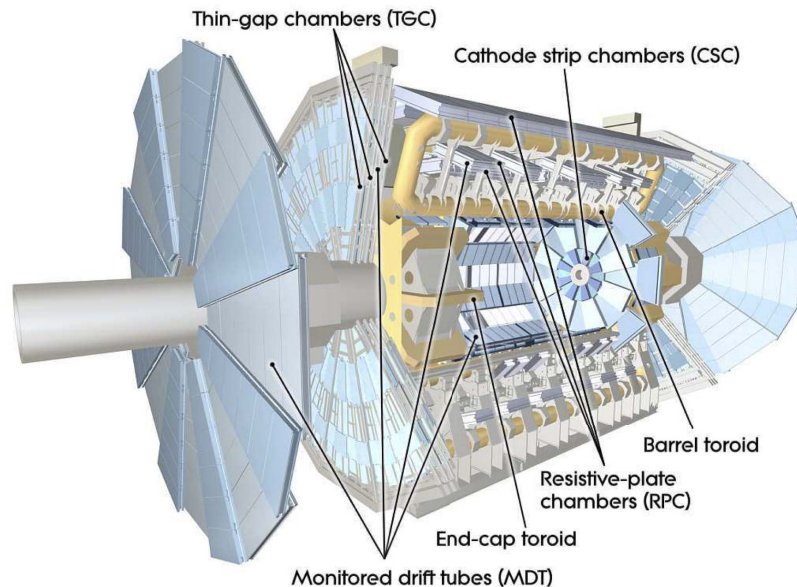


Figure 3.6: The muon system of ATLAS [1].

The MDTs and CSCs provide precise tracking information for muons. They have a comparable spatial resolution, but the cathode strip chambers have a higher granularity and a better radiation tolerance. The MDTs are drift chambers, composed of aluminium tubes filled with a mixture of Argon and  $\text{CO}_2$ . Particles moving in the MDTs create electrons that travel in the chambers to reach the tungsten-Rhenium anode wire at the centre and produce a signal. Three to eight layers of tubes surround the barrel and the end-caps, with a spatial resolution

of  $80\text{ }\mu\text{m}$  per layer, in the bending direction. The CSCs are multi-wire proportional chambers, also filled with Argon and  $\text{CO}_2$ , but with a resolution of  $60\text{ }\mu\text{m}$ . They are used exclusively in the end-caps, positioned in the innermost layer of the spectrometer, covering the forward region ( $2 < |\eta| < 2.7$ ). The last two sub-detectors are mainly used for triggering purposes. The RPCs are gaseous (tetrafluoroethane) detectors with a fast response (below  $2\text{ ns}$ ), and a spatial resolution of  $1\text{ cm}$ . The TGCs are multi-wire proportional chambers that use  $\text{CO}_2$  and n-pentan, with a response of about  $5\text{ ns}$ . The transverse momentum resolution of the muon spectrometer is  $\sigma_{p_T}/p_T \sim 10\%$  for  $1\text{ TeV}$  muon tracks.

### 3.2.5 Additional detectors

The performances of ATLAS and the LHC are monitored by two small detectors: LUCID (LUminosity measurement using a Cerenkov Integrating Detector) [99] and ALFA (Absolute Luminosity For ATLAS) [100]. They are positioned along to the beam pipe, in the forward region. They are dedicated to the measurement of the beam luminosity and to the estimation of the total cross-section of the  $pp$  collisions. It is also worth mentioning the ATLAS Beam Conditions Monitor (BCM), which consists of two groups of diamond sensors, installed at the two sides of the Pixel tracker [101]. The BCM is designed to detect unexpected proton hits due to unstable beam conditions, which could damage the ATLAS systems. It also provides additional information regarding the beam luminosity.

### 3.2.6 The data acquisition system

The  $40\text{ MHz}$  collision rate translates into a very large amount of data to read and save from the hundred of millions of channels of the various sub-detectors. A Trigger and Data Acquisition System (TDAQ) [102] was therefore designed to perform a quick selection of the events, saving exclusively the most interesting. Such events generally include high- $p_T$  objects or large amounts of missing energy. The TDAQ system is schematically represented in Figure 3.7. It is composed by two triggers: the level 1 trigger (L1) and the level 2 trigger (L2) or high level trigger (HLT). The L1 analyses the raw data from a limited part of the detector, namely the calorimeters and the trigger chambers of the muon spectrometer. It is very fast and it reduces the event rate from  $40\text{ MHz}$  to  $100\text{ kHz}$  in about  $2.5\text{ }\mu\text{s}$ . When the L1 finds an interesting event, it tags the region of interest (RoI) of the detector and it passes the event to the L2. The L2 analyses the RoI with the full available granularity and takes a final decision about the event, reducing the rate to  $1\text{ kHz}$  in  $40\text{ ms}$ . Events surviving the trigger selection are permanently saved in a distributed computer grid [103]. A preliminary object reconstruction is performed during the triggering process. It is therefore possible to filter the saved events and select the most relevant subset for each analysis. With the goal of investigating the charged Higgs boson decay to a single-lepton final state, this work requires each event to contain either one electron or one muon. Further details are given in Section 6.1.

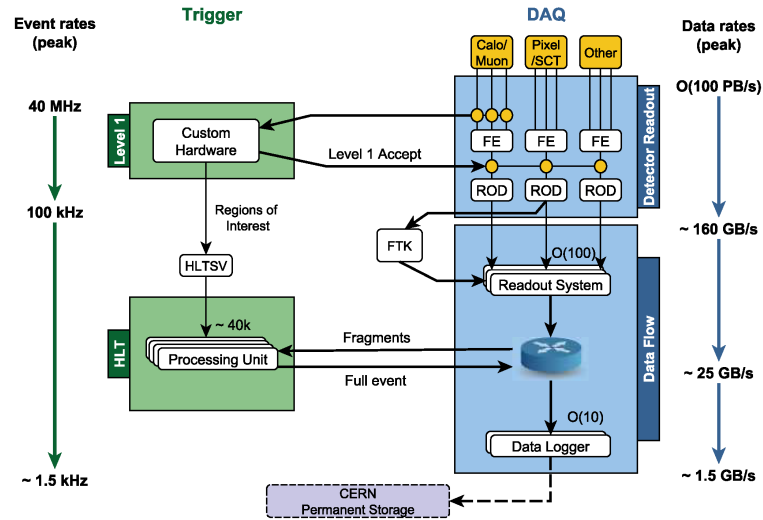
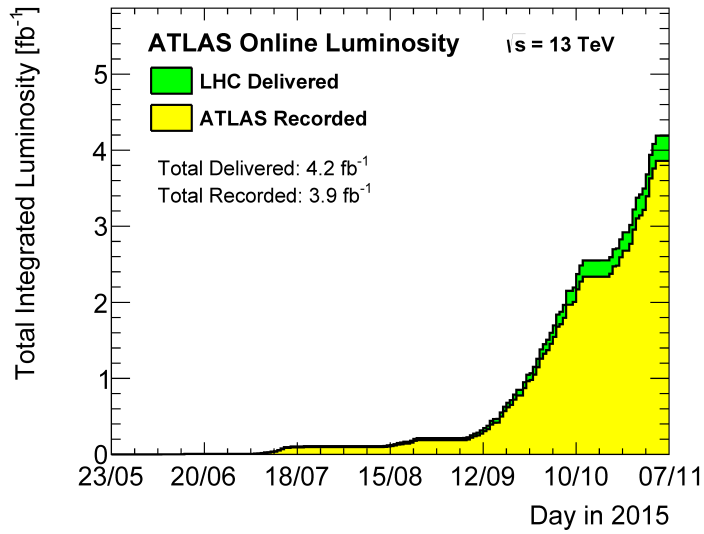


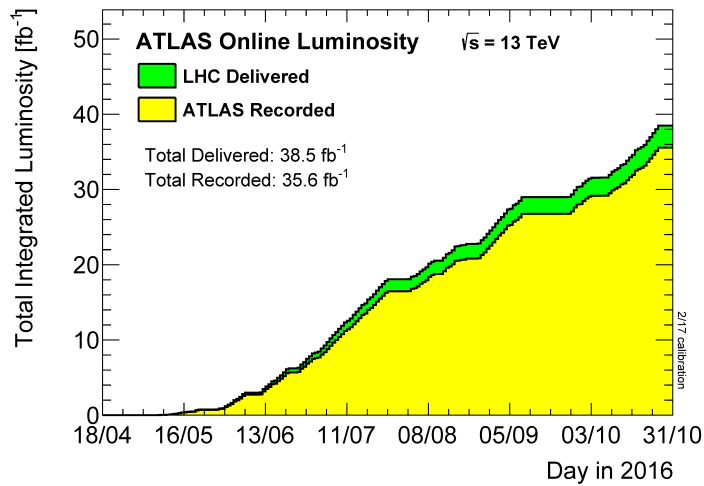
Figure 3.7: The data acquisition system of ATLAS. The information from the calorimeters and the muon spectrometer is processed by the L1 trigger, which reduces the event rate from 40 MHz to 100 kHz. The output of the L1 trigger is further analysed by the HLT which finalises the event selection reducing the event rate to 1 kHz. The events passing the HLT selection are saved permanently [104].

### 3.2.7 The data recorded by ATLAS

The integrated luminosity provided by the LHC during 2015 and 2016 was  $42.7 \text{ fb}^{-1}$ , while the dataset used for the thesis corresponds to  $36.1 \text{ fb}^{-1}$ . The difference is due to ATLAS quality criteria that discard data taking periods in which the experiment was not working properly. This can be due to wrong detector configurations or misbehaviours of single sub-detectors. Figure 3.8 shows the delivered and recorded integrated luminosities for 2015 and 2016. The peak instantaneous luminosity achieved by the LHC during such period is  $1.38 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . The period going from 2015 to 2019 is generally referred to as the Run-2 of the LHC data taking. Run-1 started in 2010 and finished in 2012, Run-3 will start after 2021.



(a) Integrated luminosity 2015.



(b) Integrated luminosity 2016.

Figure 3.8: The integrated luminosity delivered by the LHC and recorded by ATLAS in (a) 2015 and (b) 2016 [105].

## Chapter 4

### Object reconstruction in ATLAS

Before being passed on to the analysis teams, data (and Monte Carlo) events need to be processed by a certain number of off-line algorithms so as to reconstruct and calibrate the physical objects produced by the interactions. This chapter provides a short summary of the reconstruction procedures and the quality requirements applied to the various objects. Particular focus is given to leptons (electrons and muons), jets and missing energy. Tau leptons, photons and forward objects ( $|\eta| > 2.5$ ) are not used in the analysis; details regarding their reconstruction are therefore excluded from the discussion.

#### 4.1 Base-object reconstruction: tracks and vertices

The first step in the object reconstruction is to identify the trajectory of the particles inside the inner detectors (described in Section 3.2.2). This is done by converting the charge deposited in the ID modules into space-point locations, called hits. Combinations of three hits in the silicon detectors are used as track seeds for a Kalman Filter algorithm [106], which extrapolates the tracks up to the last layer of the SCT. They are subsequently extended into the TRT and re-fit with the ATLAS Global  $\chi^2$  Track Fitter [107]. Tracks are discarded if they don't satisfy the following quality criteria:

- the track includes at least seven hits in the silicon detectors;
- the track has a maximum of one hole in the Pixel detector and two holes in the combined Pixel and SCT detectors, where a hole is defined as a layer with a missing hit between two layers with recorded hits.

The collection of tracks is then extrapolated in the direction of the beam, in order to identify the starting points (vertices) of the hadronic interactions. This is performed by the Iterative Vertex Finding (IVF) algorithm [108]. Vertices along the beam line that are generated by the scattering of partons in the proton–proton collisions are known as primary vertices (PVs). Vertices originated by particle decays are secondary vertices. They are distinguishable from PVs only if they are generated by particles with a long lifetime, which produce a certain number of hits before decaying. The distance between the beam-line and a particle track

on the transverse plane is known as the transverse impact parameter  $d_0$ , while the difference between the  $z$  coordinate of a primary vertex and the  $z$  coordinate of the point of closest approach of the track to the same PV is known as the longitudinal impact parameter  $z_0$  (see Figure 4.1).

All primary vertices are required to have at least two associated tracks with  $p_T \geq 400$  MeV. The primary vertex with the largest sum of the transverse momenta of its associated tracks is the hard scattering vertex (HSV). The other identified vertices along the beam-line correspond to secondary soft interactions, known as pile-up. For the data collected during 2015 and 2016, an average of 24 pile-up interactions per bunch crossing were recorded [105]. Pile-up can be caused by interactions between protons of the primary bunches ("in-time" pile-up) or by proton-proton interactions from previous bunch crossings ("out-of-time" pile-up).

The tracks reconstructed in the inner detectors can be combined with the track segments reconstructed in the muon spectrometer and with the energy clusters of the calorimeters (known as topological cell clusters, or topo-clusters [109]) to form the physical objects defined in the following sections. Additional information regarding the reconstruction of tracks can be found in Ref. [110].

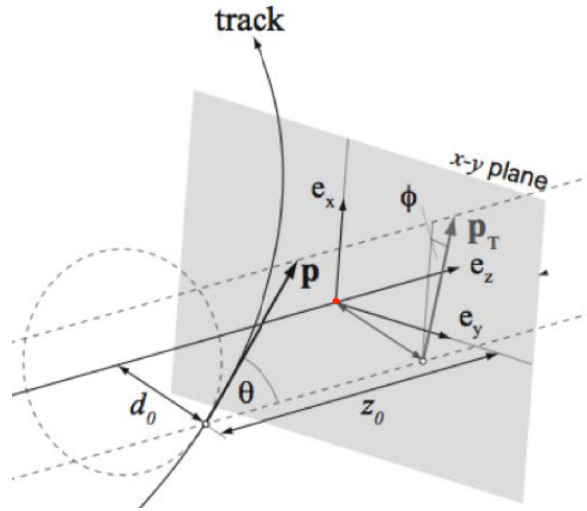


Figure 4.1: Schematic representation of the transverse ( $d_0$ ) and longitudinal ( $z_0$ ) impact parameters of the tracks. The red dot corresponds to the primary vertex [111].

## 4.2 Electrons

### 4.2.1 Reconstruction

The electron reconstruction starts with the identification of energy clusters in the electromagnetic calorimeter. The clusters are selected by sliding a window of size  $\Delta\eta \times \Delta\phi = 0.075 \times 0.125$  over the calorimeter and by tagging the regions in which the deposited energy is larger than 2.5 GeV. Such regions are matched with the tracks reconstructed in the ID, which are re-fit using a Gaussian Sum Filter (GSF) that takes into account the deposit of Bremsstrahlung radiation in the calorimeter [112]. Quality criteria are imposed to select only one track per cluster. When the matching fails, the cluster is identified as an uncovered photon. Electron candidates found in the calorimeter transition region  $1.37 < \eta < 1.57$  are excluded. The energy of the particle is then reconstructed using multivariate techniques, exploiting track features and properties of the clusters. The efficiency of the reconstruction is measured using  $Z \rightarrow ee$  and  $J/\psi \rightarrow ee$  events, selected using the Tag-and-Probe method<sup>1</sup> [113]. The electron reconstruction efficiency for  $Z \rightarrow ee$  events is shown in Figure 4.2, comparing data and MC predictions.

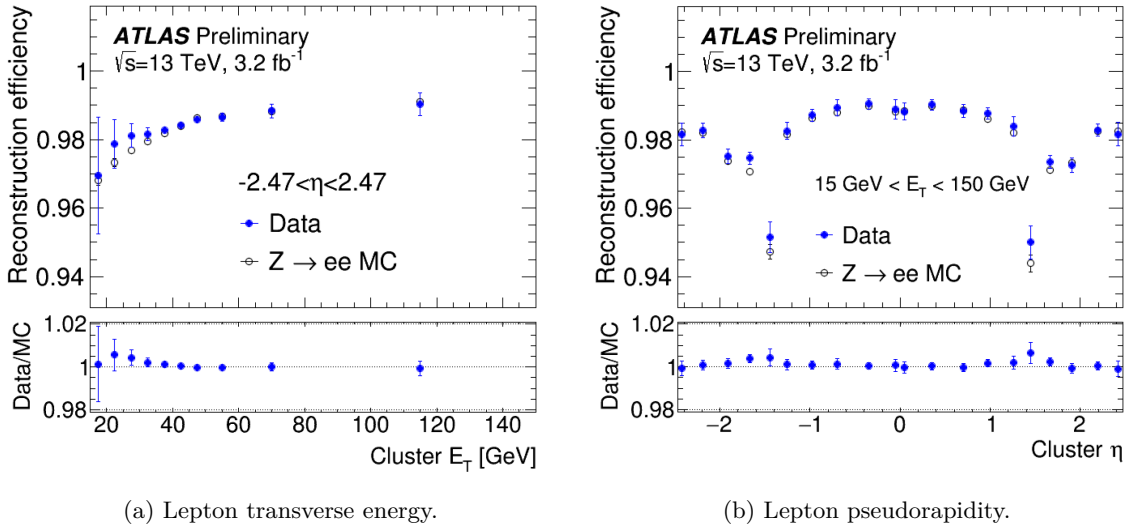


Figure 4.2: Electron reconstruction efficiency in simulated (grey) and measured (blue) events using  $Z \rightarrow ee$  decays, as a function of (a) the transverse energy and (b) the pseudorapidity of the electrons. The error bars include statistical and systematic uncertainties [113].

<sup>1</sup>The Tag-and-Probe method is a procedure often used to estimate the reconstruction or selection efficiency of physical objects on real data. The following focuses on the  $Z \rightarrow ee$  decay but it is generalisable to other processes. Given a certain number of events containing two electrons, strict criteria are applied to one of the two electrons (Tag), while very loose requirements are applied to the second (Probe). The two electrons are also requested to pass a requirement on their combined invariant mass to guarantee the presence of a  $Z$  boson in the event. The efficiency is then estimated as the fraction of combinations in which the Probe electron passes the strict criteria versus the total number of combinations.

### 4.2.2 Identification

Most of the electron candidates defined during the reconstruction process are background-like objects originating from hadronic showers or converted photons. An identification procedure was developed to distinguish between these objects and signal-like electrons (generated in the hard scattering). A total of 20 variables, including parameters related to the shower shape, information from the TRT, track properties and others, are fed into a likelihood discriminant. The output of the discriminant is the probability, for each electron candidate, of being a signal-like object. Three working points (WP), Loose, Medium and Tight, are defined by cutting at increasing values of the likelihood output [113]. Figure 4.3 shows the identification efficiency of the three WPs, as a function of the transverse energy and the pseudorapidity of the electron candidates. The tighter working point is the one used in this thesis<sup>2</sup> and it is always assumed when mentioning electrons.

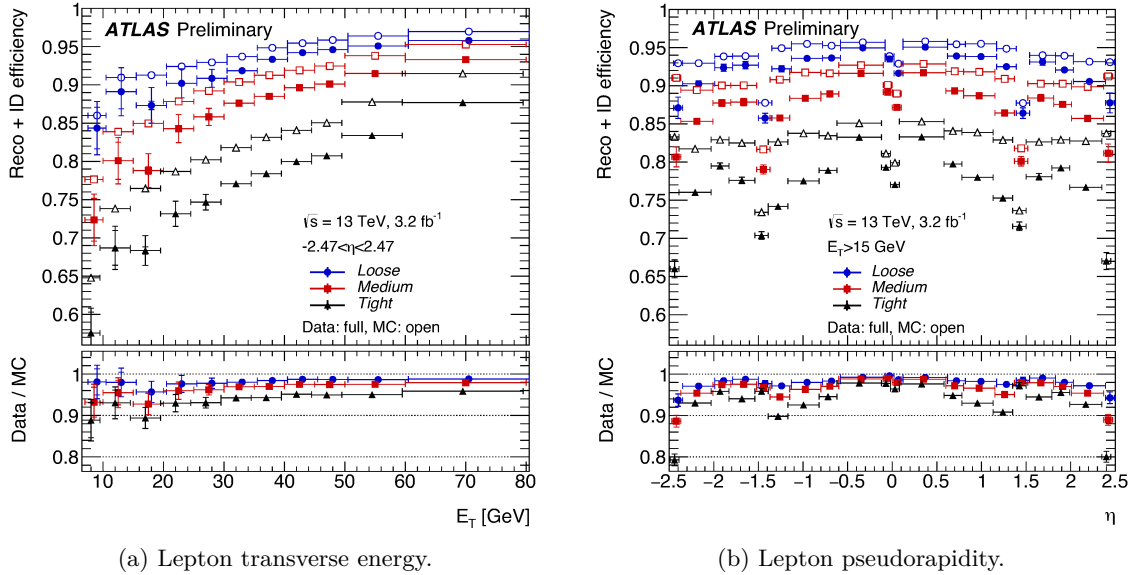


Figure 4.3: Electron identification efficiency using  $Z \rightarrow ee$  events for different operating points, as a function of (a) the transverse energy and (b) the pseudorapidity of the electrons. The error bars include statistical and systematic uncertainties [113].

### 4.2.3 Calibration

With the goal of performing highly precise measurements, having properly calibrated objects is a mandatory requirement. Numerous effects are taken into account. The response of the calorimeter is tuned to consider energy leaks and passive materials. Gravity-induced modifications to the structure of the detector, non-optimal voltages during data taking and other instrumental issues are adjusted as well. Lastly, differences in the energy scale of data

<sup>2</sup>A looser selection is used to estimate the background produced by non-prompt leptons and misidentified jets, as described in Appendix B



and MC events are corrected by applying a scale factor (SF) to all simulated events [114]. The nominal scale factor is provided together with a certain number of alternative weights, used to evaluate the systematic uncertainties (see Section 6.3).

## 4.3 Muons

### 4.3.1 Reconstruction

The first step of the muon reconstruction is performed using the output of the muon spectrometer. Hits in the muon chambers are combined to identify track segments that are then fit together to identify the muon trajectory. If possible, the hits are matched to the tracks reconstructed in the inner detectors and/or to the energy clusters of the calorimeters [115]. Four muon types are defined: if a successful combination between the ID and the MS tracks is found, the muons are called combined (CB). If the MS trajectory doesn't find a match in the inner detectors, the muons are stand-alone (SA). Segment-tagged (ST) muons correspond to ID tracks matched to partial segments in the MS, while calorimeter-tagged (CT) muons are ID tracks matched with the deposits in the calorimeters that are compatible with minimum ionising particles but do not find a match in the MS. This last kind typically corresponds to particles travelling in the region  $|\eta| < 0.1$ , where service lines create a fracture in the MDT chambers. Tracks of the CB, CT and ST categories are re-fit to exploit the additional information provided by the other sub-detectors. Only combined muons are used in this work.

### 4.3.2 Identification

Similarly to the electron case, a large part of reconstructed muons comes from secondary decays (pions and kaons mostly). The identification procedure allows to identify the muons generated in the hard scattering. Different quality criteria are applied for each of the muon categories, defining Loose, Medium and Tight working points with increasing background rejection [116]. The most important variables used for the muon identification include the  $\chi^2$  of the combined track, the  $q/p$  track parameter<sup>3</sup> and the difference between the transverse momentum of the ID track and the one of the MS segment. Figure 4.4 shows the combined reconstruction and identification efficiency for the Medium selection on CB muons, evaluated with the Tag-and-Probe method on  $Z \rightarrow \mu\mu$  events. The Medium selection is used as default in this thesis.

---

<sup>3</sup>The values  $q$  and  $p$  correspond to the charge and momentum of the track.

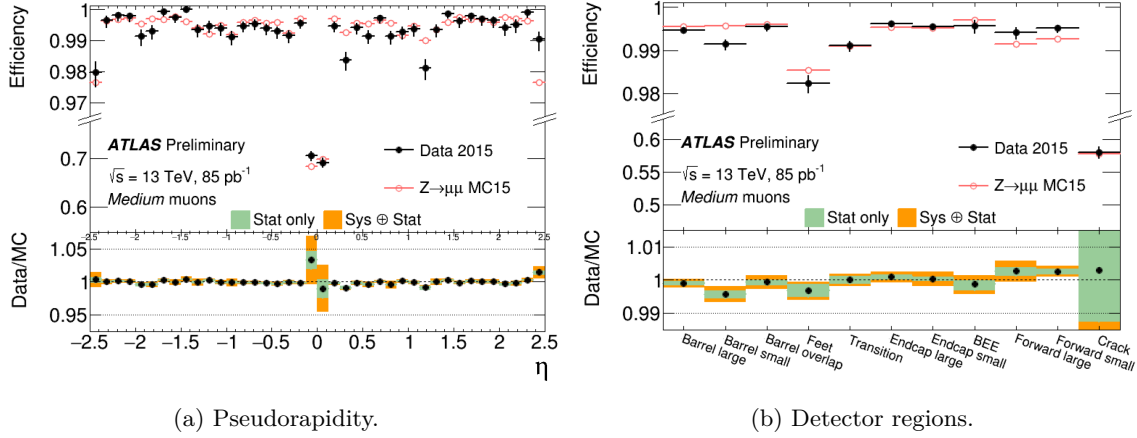


Figure 4.4: Muon reconstruction and identification efficiency using  $Z \rightarrow \mu\mu$  events as a function of (a) the pseudorapidity and (b) for various regions of the detector. Real data corresponds to the black dots, while the red circles correspond to Monte Carlo simulations. The ratio between the expected and observed efficiency is shown in the bottom panel. Green bands correspond to the statistical uncertainty only. Orange bands include statistical and systematic uncertainties [115].

### 4.3.3 Calibration

Muon candidates after the identification phase already describe real physical objects with a good accuracy. Only minor adjustments are needed, mainly related to the material profile of the detector and to the modelling of the magnetic field. As for the electrons, a set of scale factors is used to correct residual differences between the energy scales of data and Monte Carlo events [116].

## 4.4 Additional requirements on leptons

Electrons and muons produced in the hadronic collisions can be classified in two broad categories: prompt and non-prompt leptons. Prompt leptons are produced from the hard scattering, while non-prompt leptons originate from photon conversions or from the decays of other particles, such as tau leptons or heavy-hadrons. Leptons can be also reconstructed from wrongly identified hadrons. Non-prompt leptons and badly identified hadrons can be put together in the same category of "fake" leptons and represent an important background for the analysis.

Two features are particularly useful to distinguish between prompt and fake leptons: their impact parameters with respect to the vertex that generated the hard scattering and the amount of secondary tracks in close proximity of the considered lepton. In fact, most prompt leptons are emitted close to the HSV and they are very well isolated. The first feature is exploited by imposing the selection of Table 4.1. In addition, an isolation cut is implemented

by requiring the deposited energy and the sum of transverse momenta of the tracks within a certain cone around the lepton to be lower than a given threshold. The specific isolation working point used in the analysis is called Gradient isolation and it employs a threshold that increases with the transverse momentum of the lepton, making the selection tighter for low  $p_T$  objects [113]. The efficiency of such selection is approximately  $\epsilon = 0.1143 \cdot p_T / (\text{GeV}) + 0.92$ .

	Electron	Muon
$ d_0 /\sigma(d_0)$	$< 5$	$< 3$
$ z_0  \sin \theta$	$< 0.5 \text{ mm}$	$< 0.5 \text{ mm}$

Table 4.1: Requirements on the track impact parameters of lepton candidates.  $d_0$  and  $z_0$  are the transverse and longitudinal impact parameters of their tracks with respect to the HSV.  $\theta$  is the polar angle of the track while  $\sigma(d_0)$  is the error on  $d_0$  extracted from the track fit [113, 116].

## 4.5 Jets

### 4.5.1 Reconstruction

The hadronisation of quarks and gluons produces localised bundles of hadrons, known as jets. Jets are reconstructed by exploiting the energy deposited in the calorimeters, merging together topo-clusters that are close to each other. The procedure relies on the anti- $k_t$  algorithm [117, 118]. The algorithm makes use of an effective distance  $d_{i,j}$  between clusters, computed as:

$$d_{i,j} = \min\left(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}\right) \frac{\Delta R_{i,j}^2}{r^2}. \quad (4.1)$$

The indexes  $i, j$  identify two clusters, with transverse momenta  $p_{T_i}$  and  $p_{T_j}$ .  $\Delta R_{i,j}$  is their distance in the  $\eta$ - $\phi$  plane, while  $r = 0.4$  is a parameter that defines the "radius" of the jets. An effective distance between each cluster and the axis of the beam is also defined as:

$$d_{C,i} = \frac{1}{p_{T_i}^2}. \quad (4.2)$$

The algorithm iterates over all possible cluster combinations, computing the distance  $d_{i,j}$  for each pair. When the minimum  $d_{i,j}$  of the collection is larger than the corresponding  $d_{C,i}$  ( $d_{C,j}$ ), the cluster  $i$  ( $j$ ) is considered as a fully-formed jet. It is therefore removed from the collection and it is not used any more in the following iterative steps. On the other hand, if

the minimum  $d_{i,j}$  is smaller than the  $d_{C,i}$ 's, the  $i,j$  clusters are merged together and a new cluster is defined as their combination, recomputing the associated distances. The anti- $k_t$  algorithm is also applied on clusters produced by hadronic decays of tau leptons, which are treated as jets [119].

The anti- $k_t$  algorithm has two important features: it is collinear safe and infrared safe. Collinear safety means that the reconstructed jets do not depend on the transverse momentum carried by each parton, i.e. a jet does not change if two particles are merged together forming one particle with double momentum, or vice versa if a single particle is split in two. Infrared safety means that the reconstruction is driven by high- $p_T$  objects and that low- $p_T$  particles have a negligible impact on the jet momentum.

### 4.5.2 Calibration

The jet calibration consists of multiple steps, combining MC-based and "in-situ" techniques. The aim of the MC-based calibration is to adjust the energy scale of simulated jets that are reconstructed using the calorimeter topo-clusters (reco-jets), to the one of simulated jets that are reconstructed using stable final-state particles at generator-level (truth-jets). The procedure starts by correcting the reco-jet trajectory to point to the primary vertex and by removing the excess energy due to pile-up. The four-momentum of the resulting jet is then adjusted to match the generator-level energy scale. Additional corrections are applied to remove effects due to energy leaks in the calorimeters and dependencies from the jet flavour in the response of the detector. The last step is the in-situ calibration, which is dedicated to matching the energy scale of jets in Monte Carlo simulations and data. It consists of a series of corrections that are computed by balancing the transverse momentum of the jets against reference objects, as  $Z$  bosons or photons [120]. Residual discrepancies between data and MC are covered by a dedicated set of systematic uncertainties (see Section 6.3).

### 4.5.3 Pile-up rejection

A large part of reconstructed jets, in particular the ones with a low transverse momentum, are produced from pile-up interactions. In order to distinguish between such objects and the jets originating from the hard scattering, the analysis applies a dedicated discriminant, known as Jet Vertex Tagger (JVT). The JVT assigns a score to each jet, proportional to the probability that the jet has originated by the HSV [121]. It is required that all jets with a transverse momentum lower than 60 GeV have a JVT score of at least 0.59. Further criteria are applied to discard jets produced by detector noise or non-collision sources [122].

#### 4.5.4 Flavour tagging

The flavour tagging step can be interpreted as the jet "identification". Depending on the parton that originated the parton shower, the jet manifests with different features. In particular, jets including  $b$ -hadrons are distinguishable from lighter objects due to a distinct secondary vertex produced by the decay of the  $b$ -quark. Four independent algorithms are exploited in order to classify the jets: IP2D, IP3D, SV, and the JetFitter [123]. IP2D and IP3D focus on the tracks that compose the jets and on their impact parameters, SV estimates various properties related to the secondary vertex, while the JetFitter aims to reconstruct the decay of the  $b$ -hadron inside the jet. The outputs of the four algorithms are passed to a Boosted Decision Tree, known as MV2, which is used as the final discriminant to identify  $b$ -jets. The specific BDT used in this analysis is the MV2c10, which is designed to reject both light-jets (originating from gluons and  $u$ -,  $d$ -,  $s$ -quarks) and  $c$ -jets (from  $c$ -quarks). This is achieved by training the BDT with a background composed of a mixture of the two kinds. Simulated  $t\bar{t}$  + jets events are used for the training. Four calibrated working points are available: 60%, 70%, 77% and 85%, with increasing  $b$ -tagging efficiencies and decreasing background rejection. The efficiencies are obtained cutting at different values of the BDT output distribution (shown in Figure 4.5).

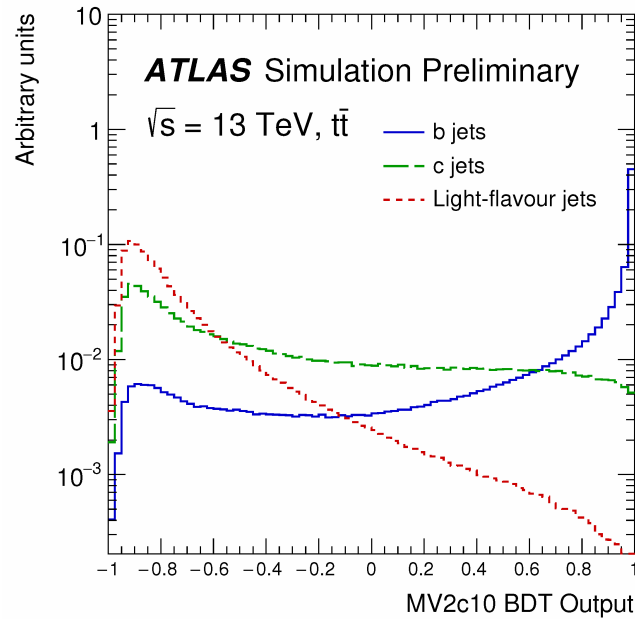


Figure 4.5: Distribution of the MV2c10 output for different jet flavours [123].

The higher the value at which the cut is performed the smaller is the probability for light and  $c$ -jets of being wrongly identified as  $b$ -jets<sup>4</sup>. A set of weights is provided to account for the differences between the  $b$ -tagging efficiency in data and Monte Carlo events. These correction factors are derived separately for light,  $c$ - and  $b$ -jets. The default  $b$ -tagging working point

<sup>4</sup>Light and  $c$ -jets tagged as  $b$ -jets are generally referred to as mistag jets.

used in this analysis is 70%. From now on, whenever an object is referred to as  $b$ -tagged, the 70% efficiency is assumed.

## 4.6 Overlap removal

Leptons and jets are reconstructed at the same time. It is therefore possible for different objects to contain the same particle. This issue is taken care of by applying an overlap removal procedure, which identifies double counted objects and subtracts their contribution from reconstructed jets or leptons. At first, all jets (and tau leptons decaying hadronically) reconstructed within a  $\Delta R < 0.2$  from an electron are discarded. Electrons are removed if they overlap with muons ( $\Delta R < 0.01$ ) or when they lie within  $0.2 < \Delta R < 0.4$  from a jet. Muons are removed when their distance from a jet is  $\Delta R < 0.4$  and the jet has at least three associated tracks. On the other hand, if the jet contains less than three tracks, the muon is kept and the jet is removed<sup>5</sup>.

## 4.7 Missing transverse momentum

The missing transverse momentum ( $\mathbf{E}_T^{\text{miss}}$ ) is a quantity defined as the momentum imbalance in the plane perpendicular to the direction of the beam. It is computed as shown in Eq. 4.3. The first term corresponds to the negative vector sum of the transverse momentum  $\mathbf{p}_i \sin \theta_i$  of all calibrated jets, leptons and photons produced by the hard scattering. The second is a correction term that accounts for reconstructed tracks that have not been matched to calibrated objects, but that are still associated to the HSV [124]. The transverse momentum of the initial state particles is zero. Therefore, a positive contribution to  $\mathbf{E}_T^{\text{miss}}$  can only come from objects that have been badly reconstructed or from non-detectable particles, as neutrinos. The magnitude of  $\mathbf{E}_T^{\text{miss}}$  is called missing transverse energy.

$$\mathbf{E}_T^{\text{miss}} = - \sum_{i \in \text{objects}} \mathbf{p}_i \sin \theta_i + \mathbf{E}_T^{\text{soft}} \quad (4.3)$$

---

<sup>5</sup>This condition allows to maintain a good efficiency in the selection of muons that lose a large amount of energy in the calorimeters.

## Chapter 5

### Signal and background modelling

This chapter describes the Monte Carlo modelling of signal and background samples, as well as the method used to estimate the background of fake leptons. As mentioned in Section 1.3.4, each signal event is expected to include one lepton, one neutrino, four  $b$ -jets and two light jets. Such a final state can be fully or partly reproduced by a large number of background processes. The main background for this analysis is  $t\bar{t}$ +jets, i.e. the associated production of top quarks and jets. Additional contributions to the background come from the production of  $W$  and  $Z$  bosons with jets, single-top-quark production, diboson ( $WW$ ,  $WZ$ ,  $ZZ$ ) processes and the associated production of bosons and top quarks ( $t\bar{t}W$ ,  $t\bar{t}Z$ ,  $t\bar{t}H$ )<sup>1</sup>. Non-prompt leptons and misidentified jets also contribute as a multi-jet background. A Monte Carlo simulation is used for all these backgrounds, except for the fake leptons background, which is estimated using a data-driven technique. Various ME and PS generators have been used for the production of the MC samples. The showering of partons has been simulated with PYTHIA [78] for all nominal<sup>2</sup> samples, except for those created using the SHERPA generator [125]. All nominal background samples have been interfaced with GEANT for the simulation of the ATLAS detector, while the signal samples and some alternative samples, used for the evaluation of the systematic uncertainties, have been simulated using the ATLAS-Fast II technique (the FullSim and AFII techniques have been introduced in Section 1.4.2). The top-quark mass is assumed to be  $m_t = 172.5$  GeV for all samples, while the SM Higgs-boson mass is set to  $m_H = 125$  GeV. Data and MC events are reconstructed using the same techniques, of which an overview has been given in the previous chapter. All simulated samples are also reweighted to match the average number of pile-up interactions per bunch crossing observed in data [126]. Figure 5.1 shows examples of Feynman diagrams for some of the most important background processes. Table 5.1 provides an overview of the setup used for the event generation of the nominal samples, which is detailed in the following sections.

---

<sup>1</sup>The three categories together will be often reported as  $t\bar{t}X$ , while the  $t\bar{t}W$  and  $t\bar{t}Z$  backgrounds are collectively named  $t\bar{t}V$ .

<sup>2</sup>Nominal refers to samples that are not used for the systematic evaluation but are the default in the MC modelling.

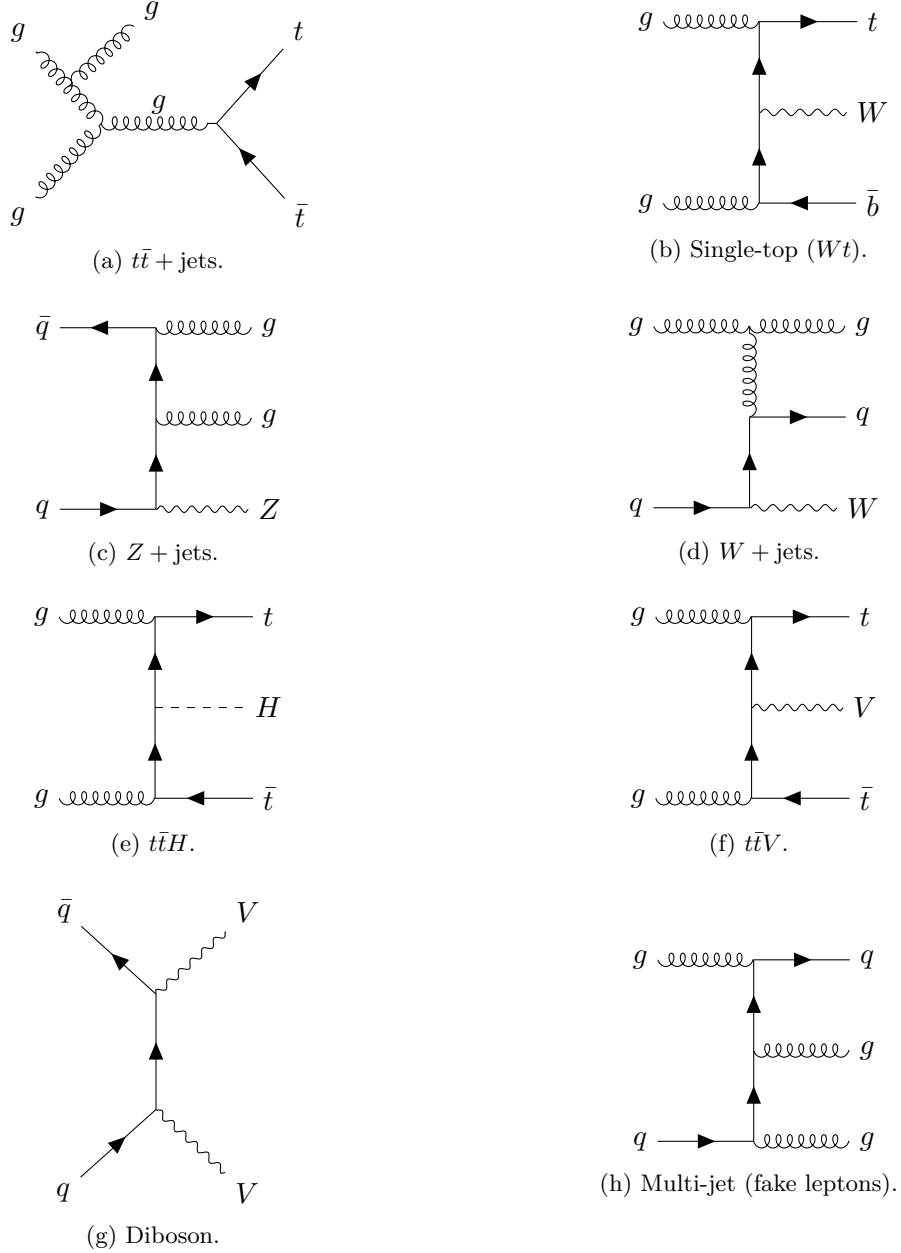


Figure 5.1: Examples of Feynman diagrams for the most important backgrounds: (a)  $t\bar{t}$  + jets, (b) single-top  $Wt$ , (c)  $Z$  + jets, (d)  $W$  + jets, (e)  $t\bar{t}H$ , (f)  $t\bar{t}V$ , (g) diboson, (h) fake leptons.



Sample	ME Generator	ME PDF	PS Generator	Cross-section Normalisation	Tune
$\bar{t}bH^+$	MG5_aMC@NLO	NNPDF2.3NLO	PYTHIA 8.1	NLO	A14
$t\bar{t} + \text{jets}$	POWHEG v2	NNPDF3.0NLO	PYTHIA 8.2	NNLO + NNLL	A14
$t\bar{t}X$	MG5_aMC@NLO	NNPDF3.0NLO	PYTHIA 8.2	NLO	A14
$V + \text{jets}$	SHERPA 2.2.1	NNPDF3.0NNLO	SHERPA 2.2.1	NNLO	SHERPA default
diboson	SHERPA 2.2.1	CT10	SHERPA 2.2.1	NNLO	SHERPA default
single-top ( $s$ -channel, $Wt$ )	POWHEG v2	CT10	PYTHIA 6.4	aNNLO	Perugia 2012
single-top ( $t$ -channel)	POWHEG v1	CT10f4	PYTHIA 6.4	aNNLO	Perugia 2012

Table 5.1: Setup used for the event generation of the nominal signal samples and main background samples. References for the ME and PS generators, and for the tunes, can be found in the main text.

## 5.1 Signal modelling

The  $\bar{t}bH^+$  production is simulated at the next-to-leading order in QCD using MG5\_aMC@NLO [71]. The parton distribution functions are modelled with the NNPDF2.3NLO set [127], using the four-flavour scheme (see Section 1.3.2). PYTHIA 8.1 (A14 tune) [128] is used to simulate the parton shower and the hadronisation. The tune embeds the choice made for some parameters of the MC simulation, which are "tuned" to obtain the best agreement with data (see Ref. [129]). The detector interaction is modelled using the ATLAS-Fast II simulation. A total of 18 mass points have been generated, ranging between 200 and 2000 GeV. The narrow-width approximation has been used, neglecting the charged Higgs boson natural width, which is significantly smaller than the experimental resolution in most of the  $\tan\beta$  parameter space<sup>3</sup>. Interference effects with the SM backgrounds have not been considered. Their impact is under discussion in the scientific community but they are absent in the narrow-width approximation. The expected cross-sections for the  $\bar{t}bH^+$  production are listed in Table 5.2, for two values of the  $\tan\beta$  parameter, together with the total number of generated events.

Four additional samples (at 250, 600, 1200 and 2000 GeV) have been generated to compute the systematic uncertainties related to the choice of the renormalisation and factorisation scales. They include a set of eight weights, corresponding to all possible combinations of  $\mu_R, \mu_F$  where  $\mu_{R,F} = S \cdot \bar{\mu}_{R,F}$  and  $S = 0.5, 1.0, 2.0$ . The hyphenated energy scales correspond to the default values of the nominal sample ( $\bar{\mu}_R = \bar{\mu}_F = m_{H^+}/3$ ). Such variations can strongly affect the normalisation of the samples, in particular for the high-mass hypotheses. Details

<sup>3</sup>The experimental resolution can be inferred by performing injection tests, as shown in Appendix C. It is approximately 18 GeV for a 200 GeV mass hypothesis and 260 GeV for a 800 GeV mass hypothesis. The width predicted by the  $m_h^{\text{mod-}}$  scenario of the MSSM for a 200 GeV mass hypothesis is lower than 4.2 in the  $0.5 - 60 \tan\beta$  range and lower than 1.1 for  $\tan\beta \geq 1$ . The width predicted for  $m_{H^+} = 800$  GeV stays below 150 in the  $0.5 - 60 \tan\beta$  range and below 45 at  $\tan\beta \geq 1$ .

regarding the computation of the related systematic uncertainties are given in Section 6.3.

$H^+$ mass (GeV)	Size (nominal)	$\sigma(\tan\beta = 1)$ [pb]	$\sigma(\tan\beta = 60)$ [pb]
200	5.0M	3.3642	3.1218
225	1.5M	2.6823	2.4761
250	1.5M	2.4642	1.9838
275	1.0M	1.7517	1.5993
300	1.0M	1.4224	1.2931
350	0.8M	0.9626	0.8697
400	0.8M	0.6626	0.5915
500	0.7M	0.3300	0.2927
600	0.6M	0.1749	0.1534
700	0.6M	0.0969	0.0844
800	0.6M	0.0559	0.0482
900	0.6M	0.0333	0.0286
1000	0.7M	0.0204	0.0175
1200	0.9M	0.0082	0.0069
1400	1.2M	0.0036	0.0030
1600	1.2M	0.0016	0.0014
1800	2.0M	0.0008	0.0006
2000	2.0M	0.0004	0.0003

Table 5.2: Generated events for various mass hypotheses of the  $t\bar{t}bH^+$  production. The last two columns report the expected cross-sections for two values of the  $\tan\beta$  parameter, in the  $m_h^{\text{mod-}}$  scenario of the MSSM.

## 5.2 Background modelling

### 5.2.1 $t\bar{t} + \text{jets}$ background

The  $t\bar{t} + \text{jets}$  process is the dominant background in the search for charged Higgs bosons. In particular, the subset of  $t\bar{t} + \text{jets}$  events where the additional jets are produced by  $b$ -quarks can have the same final state of the signal ( $t\bar{t} + b\bar{b}$ ) and very similar kinematics. The  $t\bar{t} + \text{jets}$  nominal sample is generated with the POWHEG v2 NLO generator [72] and the NNPDF3.0NLO set [74], using the five-flavour scheme. The parton shower of choice is PYTHIA 8.2 [78], with the A14 tune. A filtered  $t\bar{t} + \geq 1b$  sample is also generated, requiring that at least one of the additional jets originated from a  $b$ -quark. This strongly reduces the statistical uncertainty

associated to such subset of background events. The combination of the two samples is normalised to the  $t\bar{t}+2.0$  [130] theoretical cross-section of  $832^{+46}_{-51}$  pb, computed at the NNLO in QCD and including next-to-next-to-leading logarithmic (NNLL) soft gluon terms [131]. The factorisation and renormalisation scales are set to the top-quark mass, while the  $h_{\text{damp}}$  parameter is set to 1.5 times the top-quark mass<sup>4</sup>. The choice of the ME generator, PS and related parameters are justified by dedicated studies performed by the ATLAS collaboration on the  $t\bar{t}+\text{jets}$  modelling [132]. The set of provided recommendations includes recipes for the computation of systematic uncertainties, some of which require a comparison with alternative samples. Samples produced using SHERPA 2.2.1 [125] and POWHEG + Herwig 7 [79, 133] are used to evaluate the uncertainties related to the choice of the ME and PS generators, while two samples produced with POWHEG + PYTHIA 8.2 are used to estimate the error associated to the amount of radiation in the events. One of such samples is obtained by scaling the  $h_{\text{damp}}$  parameter to twice the default value and dividing by two the renormalisation and factorisation energy scales. The other corresponds to  $\mu_R$  and  $\mu_F$  scaled to twice the default value and the  $h_{\text{damp}}$  parameter left unchanged. Both configurations also include different settings for the Var3c parameter of the A14 tune, which control the amount of ISR/FSR in the events [134]. The distributions of the mass and  $p_T$  of the  $t\bar{t}$  system for the alternative  $t\bar{t}+\text{jets}$  samples are compared in Figure 5.2.

A series of reweightings are applied to the  $t\bar{t}+\text{jets}$  nominal and alternative samples in order to improve their modelling. The technique is similar to the one described in Ref. [135]. For the purpose of the reweightings,  $t\bar{t}+\text{jets}$  events have been categorised as follows: events where at least one jet is matched<sup>5</sup> to a  $b$ -hadron (not produced by a top quark decay), with a transverse momentum of at least 5 GeV, are called  $t\bar{t}+\geq 1b$  events. When the jet is matched to a  $c$ -hadron (not produced by  $W$ -boson decays), the event is named  $t\bar{t}+\geq 1c$ . The  $t\bar{t}+\geq 1c$  and  $t\bar{t}+\geq 1b$  categories are collectively referred to as the  $t\bar{t}+\text{HF}$  (heavy flavours) category. Remaining events are called  $t\bar{t}+\text{light}$ . A finer categorisation, described in Table 5.3, is applied to  $t\bar{t}+\geq 1b$  events.

Category	Description
$t\bar{t}+\geq 1b$ (MPI/FSR)	One or more jets matched to $b$ -hadrons produced by MPI or FSR
$t\bar{t}+b$	One jet matched to a $b$ -hadron not by MPI or FSR
$t\bar{t}+b\bar{b}$	Two jets each matched to a single $b$ -hadron not by MPI or FSR
$t\bar{t}+B$	One jet matched to two $b$ -hadrons not by MPI or FSR
$t\bar{t}+\geq 3b$	At least 3 $b$ -hadrons (not by MPI or FSR) matched to any number of jets

Table 5.3: The  $t\bar{t}+\text{HF}$  subcategories. MPI stands for multiple parton-parton interactions, not produced by the hard scattering.

A first reweighting is applied to the  $t\bar{t}+\geq 1b$  events. The relative normalisations of the  $t\bar{t}+\geq 1b$  subcategories of the nominal sample are matched to the normalisations predicted by

<sup>4</sup>The  $h_{\text{damp}}$  parameter is related to the ME to PS matching. It controls the transverse momentum of the first jet emitted beyond the tree-level.

<sup>5</sup>An hadron is matched to a jet when it is found within  $\Delta R < 0.3$  to the jet axis.

a  $t\bar{t} + b\bar{b}$  sample produced with the SHERPA+OPENLOOPS generator at the NLO [136]. The events are generated with the CT10 PDF set [137], using the four-flavour scheme. This sample should provide a better modelling than the nominal POWHEG + PYTHIA 8.2 because the NLO calculation is done on the final state of four partons<sup>6</sup>. Once the relative normalisations have been fixed, a 2D reweighting is performed to correct the shape of the  $p_T$  distribution of the top quarks and the  $t\bar{t}$  system. For  $t\bar{t} + \text{jets}$  categories that include more than one additional  $b$ -quark, this is followed by a 2D reweighting of the  $\Delta R$  between the leading  $b$ -jets and the  $p_T$  of the leading  $b$ -jet. For the other categories, the  $b$ -jet transverse momentum and  $\eta$  are used. The relative fractions predicted by the nominal sample and the SHERPA+OPENLOOPS sample, for the various categories, are shown in Figure 5.3. The correction factors are applied to the alternative samples used for the systematic evaluation as well. The  $t\bar{t}+b$  (MPI/FSR) category is not simulated by SHERPA+OPENLOOPS and it is therefore not reweighted. Table 5.4 provides an overview of the setup used for the generation of the  $t\bar{t} + \text{jets}$  alternative samples used in the analysis.

Application	Generator	PDF	Shower	Additional information
Nominal	POWHEG	NNPDF3.0NLO	PYTHIA 8.2	-
PS variation	POWHEG	CT10	Herwig 7	-
ME variation	SHERPA 2.2.1	NNPDF3.0NNLO	SHERPA	-
Radiation Down	POWHEG	NNPDF3.0NLO	PYTHIA 8.2	$2.0 \cdot \bar{\mu}_r, 2.0 \cdot \bar{\mu}_f$
Radiation Up	POWHEG	NNPDF3.0NLO	PYTHIA 8.2	$2.0 \cdot \bar{h}_{\text{damp}}, 0.5 \cdot \bar{\mu}_r, 0.5 \cdot \bar{\mu}_f$
$t\bar{t} + \geq 1b$ reweighting	SHERPA 2.1.1	CT10	SHERPA	4FS, $\mu_r = \mu_f = \sum_i^{\text{jets}} p_{T,i}/2$

Table 5.4: Overview of the Monte Carlo samples used for the estimation of the systematic uncertainties of the  $t\bar{t} + \text{jets}$  modelling. The hyphenated variables under "additional information" refer to the default values of the nominal sample.

Finally, a data-driven reweighting is performed on the inclusive  $t\bar{t} + \text{jets}$  sample, correcting the shape of the distribution of the leading-jet  $p_T$  of POWHEG + PYTHIA 8.2 to the one predicted by data. The reweighting factors are computed in a region of the phase space ( $\geq 4$  jets,  $2b$  jets) that is dominated by the SM background (the expected amount of signal is  $S/B = 0.4\%$  at 200 GeV and  $S/B = 0.01\%$  at 2000 GeV). A second-order polynomial function is fit on the ratio between the two distributions and applied to all  $t\bar{t} + \text{jets}$  samples in the analysis. The function is visible in the bottom panel of Figure 5.4.

<sup>6</sup>It must be added that, in the 5FS, as used in the nominal POWHEG sample, the  $b$ -quarks emitted as initial state radiation are approximated as massless particles. This is a constraint coming from the fact that  $b$ -quarks are heavier than protons. In the SHERPA 4FS sample, they are instead produced as massive FSR.

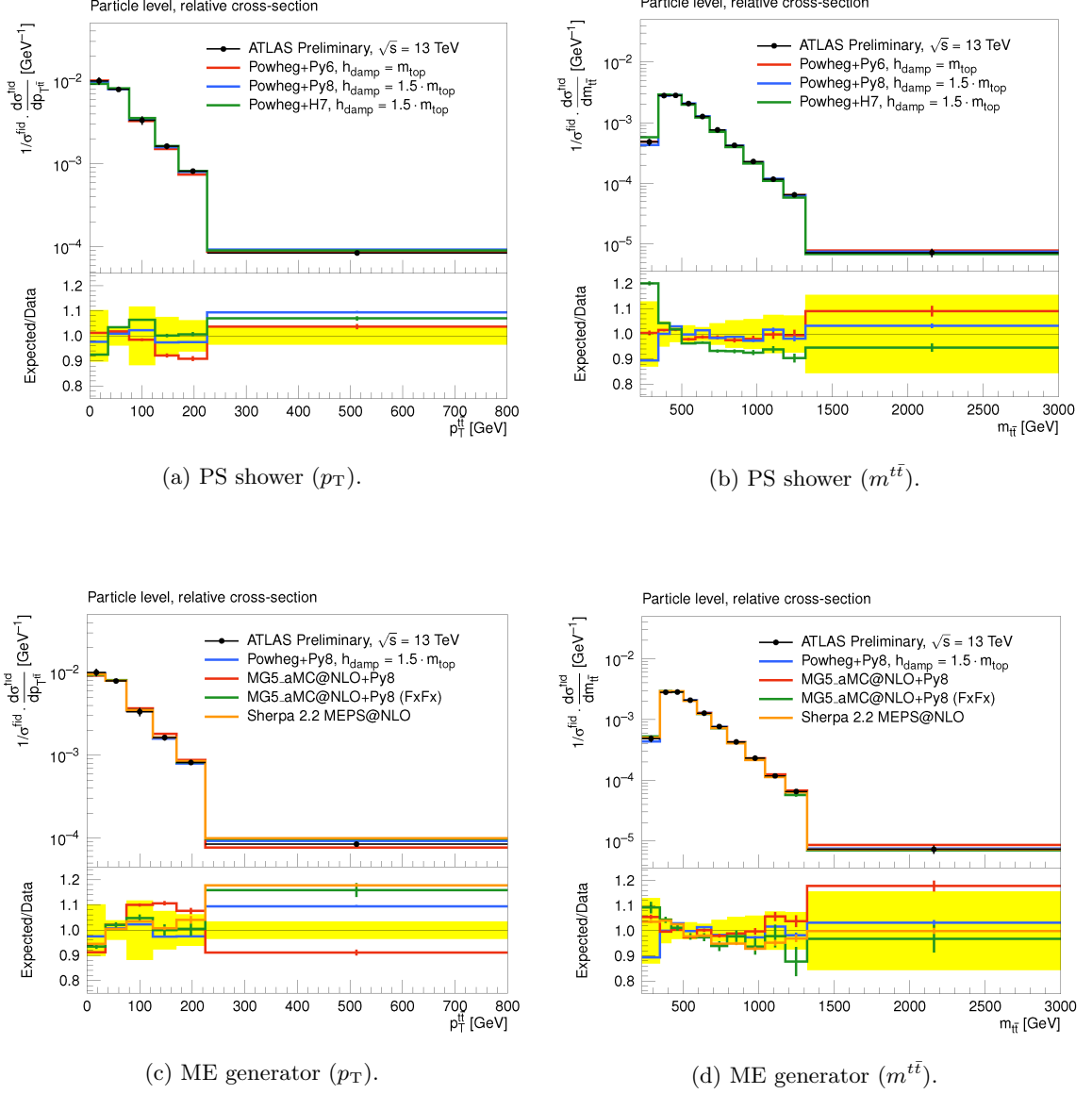


Figure 5.2: Comparison between data and  $t\bar{t}$  + jets events produced with (a,b) different showering algorithms and (c,d) different ME generators. The contribution of non- $t\bar{t}$  and  $t\bar{t}X$  backgrounds is subtracted from data. The comparison is performed for the distributions of (a,c) the transverse momentum and (b,d) the mass of the  $t\bar{t}$  system [132].

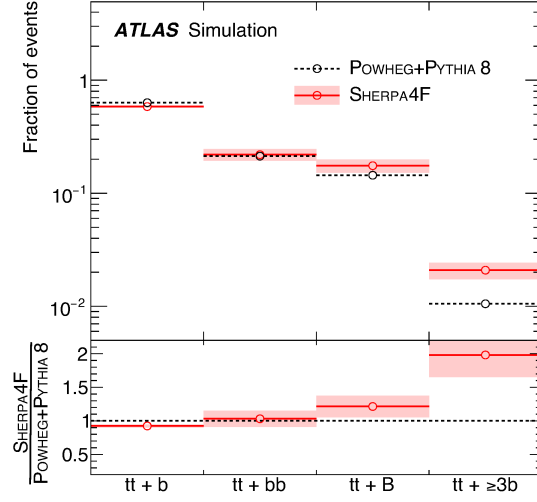


Figure 5.3: Comparison between the relative fractions of the  $t\bar{t} + \geq 1b$  subcategories predicted by the nominal  $t\bar{t} + \text{jets}$  sample (in black) and the alternative SHERPA+OPENLOOPS sample (in red). The error bands include statistical and systematic uncertainties [135].

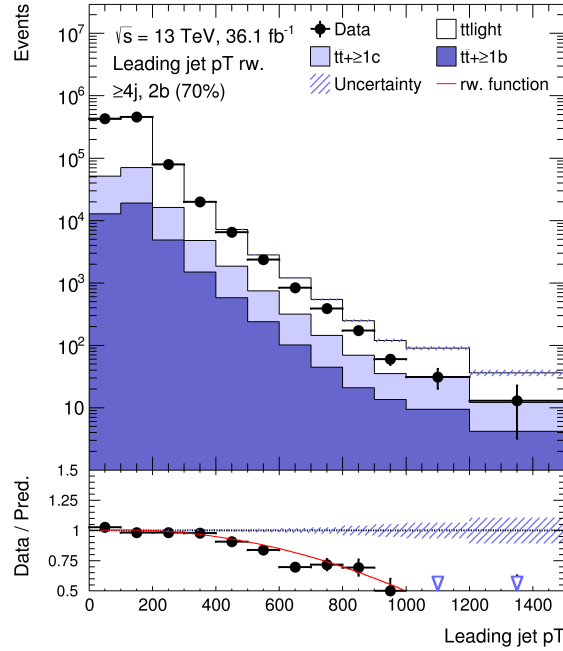


Figure 5.4: Distributions of data and simulated  $t\bar{t} + \text{jets}$  events as a function of the leading-jet  $p_T$  in the ( $\geq 4$  jets,  $2b$  jets) region. The contribution of non- $t\bar{t}$  and  $t\bar{t}X$  backgrounds is subtracted from data. The uncertainty band corresponds to the statistical uncertainty of the MC sample. The reweighting function is drawn in red in the ratio panel.

### 5.2.2 Other backgrounds

$W/Z$ +jets samples are generated using SHERPA 2.2.1, with the NNPDF3.0NNLO set and a tune developed by the SHERPA authors. The samples are normalised to the NNLO cross-section [138]. A correction of +30% is applied to the normalisation of  $Z$ +jets events that include  $c$ - or  $b$ -hadrons<sup>7</sup>. Diboson samples are generated with SHERPA 2.2.1 and using the CT10 PDF. Samples for the  $Wt$  process and the single-top  $s$ -channel are generated with POWHEG v2 + PYTHIA 6.4 [139], using the CT10 PDF and the Perugia 2012 tune [140]. POWHEG v1 and the CT10f4 set [137] are used for the  $t$ -channel. All single-top processes are normalised to the aNNLO (approximate NNLO) cross-section [141–143]. Samples for  $t\bar{t}X$  processes are generated using MG5\_aMC@NLO + PYTHIA 8.2 with the NNPDF3.0NNLO set. Alternative samples for  $t\bar{t}X$  and single-top processes are used for the computation of ME and PS uncertainties. An additional sample is also produced for the  $Wt$  process, in order to compare the diagram-removal and diagram-subtraction schemes<sup>8</sup>. Minor backgrounds as  $tH$ ,  $tWZ$ , 4-top and  $t\bar{t}WW$  are considered as well, but they produce less than 1% of the total background. Contributions to the  $tH$  background come from the production of a top quark in association with jets ( $tHjb$ ) or with a  $W$  boson ( $tHW$ ). Both processes are simulated using MG5\_aMC@NLO with the CT10 PDF. The parton shower of choice is PYTHIA 8.2 for  $tHjb$  events and Herwig++ [79] for  $tHW$  events. The  $tWZ$ , 4-top and  $t\bar{t}WW$  processes are all simulated with MG5\_aMC@NLO + PYTHIA 8.2.

### 5.2.3 Fake leptons

Fake leptons events are strongly suppressed by the isolation and identification cuts applied to leptons (see Section 4.4). However, they still constitute a non-negligible background, in particular in the low- $p_T$  region. Their contribution is estimated via a data-driven method, called Matrix-Method (MM) [145]. The MM compares two kinds of reconstructed leptons: the ones passing the nominal ID and isolation requirements (named "tight" leptons), and a larger group with looser identification and isolation criteria ("loose" leptons). The selection criteria for tight leptons have been detailed in the previous chapter: they correspond to a Tight (Medium) ID for electrons (muons), a Gradient isolation and cuts on the impact parameters of both leptons. The selection employed for loose leptons has lower requirements: Medium ID for both leptons and no isolation cuts. The method allows to extrapolate the number of events with leptons passing the tight criteria by using events with leptons that satisfy the loose criteria, which are a larger sample and therefore have a smaller statistical error.

<sup>7</sup>The correction has been extrapolated from dedicated data control regions in which the invariant mass of the dilepton system was compatible with a  $Z$  resonance:  $|m_{ll} - m_Z| < 8$  GeV.

<sup>8</sup>The  $Wt$  process at NLO and  $t\bar{t}$  process at LO can have the same final state. A non-negligible interference effect must be taken into account. Two methods are available: the diagram-removal and the diagram-subtraction schemes. The default in the analysis is the diagram-removal scheme, while the diagram-subtraction scheme is used as systematic uncertainty [144].

Both for the tight and the loose selections, a certain number of leptons is expected to be real and a certain number fake. By definition, the events with at least one real (fake) tight lepton are a subset of the events with at least one real (fake) loose lepton. It is therefore possible to compute the total number of events of the first kind  $N_{\text{real}}^{\text{tight}}$  ( $N_{\text{fake}}^{\text{tight}}$ ) by multiplying the number of events of the second kind  $N_{\text{real}}^{\text{loose}}$  ( $N_{\text{fake}}^{\text{loose}}$ ) by a certain efficiency  $\epsilon_{\text{real}}$  ( $\epsilon_{\text{fake}}$ ):

$$\begin{aligned} N^{\text{loose}} &= N_{\text{real}}^{\text{loose}} + N_{\text{fake}}^{\text{loose}}, \\ N^{\text{tight}} &= N_{\text{real}}^{\text{tight}} + N_{\text{fake}}^{\text{tight}} = \epsilon_{\text{real}} N_{\text{real}}^{\text{loose}} + \epsilon_{\text{fake}} N_{\text{fake}}^{\text{loose}}, \end{aligned} \quad (5.1)$$

where  $N^{\text{loose}}$  and  $N^{\text{tight}}$  correspond to the total number of events passing the loose and tight selections. The equations can be rearranged to obtain:

$$N_{\text{fake}}^{\text{tight}} = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} (\epsilon_{\text{real}} N^{\text{loose}} - N^{\text{tight}}). \quad (5.2)$$

Eq. 5.2 gives the total number of events with at least one tight fake lepton. However, the interest is more often for the distribution of the fake leptons events as a function of some kinematic variable. Such distribution can be extrapolated by re-expressing the result of Eq. 5.2 as a weighting factor  $w_i$ , to be applied to all loose events:

$$w_i = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{real}} - \epsilon_{\text{fake}}} (\epsilon_{\text{real}} - t_i), \quad (5.3)$$

where  $t_i$  is equal to 1 for events passing both selections and equal to 0 for events passing the loose selection only. The  $\epsilon_{\text{real}}$  and  $\epsilon_{\text{fake}}$  efficiencies are computed in dedicated control regions, enriched of real and fake leptons respectively. The fake region corresponds to events including at least one loose lepton and one jet, with a low amount of missing transverse energy  $E_T^{\text{miss}} < 20$  GeV and low values of the reconstructed  $W$  boson transverse mass<sup>9</sup>. The fake efficiency  $\epsilon_{\text{fake}}$  is computed by taking the ratio between the number of tight events and loose events in the fake control region, subtracting the contribution from the other SM backgrounds:

$$\epsilon_{\text{fake}} = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}} = \frac{N_{\text{data}}^{\text{tight}} - N_{\text{MC}}^{\text{tight}}}{N_{\text{data}}^{\text{loose}} - N_{\text{MC}}^{\text{loose}}}. \quad (5.4)$$

The real efficiency  $\epsilon_{\text{real}}$  is measured on  $Z \rightarrow ll$  events using the Tag-and-Probe method

---

<sup>9</sup>The  $W$  boson transverse mass is defined as:

$$m_T^W = \sqrt{2p_T^{\text{lep}} E_T^{\text{miss}} (1 - \cos \Delta\phi)},$$

where  $\Delta\phi$  is the azimuthal angle difference between the lepton and the transverse missing energy.



(tagging the tight leptons and probing on the loose leptons). The efficiencies are computed separately for the electron and muon channels and they are parametrised as functions of the following kinematic variables: number of jets, leading jet  $p_T$ , lepton  $\eta$ , angular distance between the lepton and the leading jet.



## Chapter 6

### Search for $\bar{t}bH^+(H^+ \rightarrow t\bar{b})$

This chapter presents the search for a charged Higgs boson produced in association with a top quark, where the Higgs boson decays in a top-bottom pair with a single-lepton final state. The analysis strategy is described in the first part, starting with the event selection and categorisation. Events are classified into different regions in order to enhance the sensitivity to the signal and to improve the control over the SM backgrounds. The regions are designed to be enriched in different  $t\bar{t}$  + jets categories: a low number of  $b$ -jets provides control on the  $t\bar{t}$  + light category, while a request of three or more  $b$ -jets isolates the  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \geq 1b$  backgrounds. Regions with at least three  $b$ -jets are also heavily enriched in signal and are therefore named signal regions. A boosted decision tree is used to improve the separation between signal and background processes in the signal regions. The signal strength is extracted with a binned profile likelihood fit, performed on the distribution of the BDT output. The fit runs simultaneously on all regions and separately for each of the  $H^+$  mass hypotheses. A large set of nuisance parameters is used to cover all systematic uncertainties. They are described in the second part of the chapter, together with the results. A 95% CL limit is provided for the cross-section of the signal production (times the BR of the decay) and for the  $\tan\beta$  parameter of the MSSM, as a function of the signal mass. The full set of techniques used in this chapter has been implemented in the search for charged Higgs bosons published by ATLAS and available at [13]. The publication includes a combination with the dilepton channel and a superposition of the resulting limits with the ones obtained by the analysis of the  $\tau^+\nu_\tau$  decay channel. While this work does not provide a detailed description of the dilepton and  $\tau^+\nu_\tau$  analyses, the outcome of their combination with the single-lepton channel is shortly described at the end of the chapter.

The analysis follows on from the previous result obtained by the ATLAS collaboration at 8 TeV, which set 95% CL upper limits on the  $H^+$  cross-section ranging between 6 and 0.2 pb for masses between 200 and 600 GeV (see Section 1.3.3).

## 6.1 Event selection and categorisation

As mentioned in Section 3.2.6, the analysis is performed exclusively on the events passing the trigger selection, which requires each event to include at least one electron or one muon. Various triggers are combined together with a logical OR, depending on the transverse momentum of the lepton candidates. Different requirements are also applied for 2015 and 2016 datasets, due to different data taking conditions (see Table 6.1)<sup>1</sup>.

Trigger name (2015)	Object	minimum $p_T$ [GeV]
e24_lhmedium_L1EM20VH	electron	24
e60_lhmedium	electron	60
e120_lhloose	electron	120
mu_20_iloose_L1MU15	muon	20
mu_50	muon	50

Trigger name (2016)	Object	minimum $p_T$ [GeV]
e26_lhtight_nod0_ivarloose	electron	26
e60_lhmedium_nod0	electron	60
e120_lhloose_nod0	electron	120
mu_26_ivarmedium	muon	26
mu_50	muon	50

Table 6.1: Triggers used for the event selection. The first column corresponds to the trigger name. The name includes information about the identification and isolation criteria applied on the objects. The ID working points used for electrons are identified by the strings "lhtight", "lhmedium" and "lhloose". Muons always require a medium identification. The tag "nod0" means that the cuts on the impact parameters have been removed, while "ivarloose" and "ivarmedium" refer to the isolation working points, described in Ref. [115].

The number of events passing the trigger selection is reduced further by imposing the quality criteria described in Chapter 4. These include a Tight identification working point for electrons and a Medium identification WP for muons. They also require both kind of leptons to pass the Gradient isolation and to satisfy a certain number of cuts on their impact parameters, to reduce the amount of fake leptons.

Finally, a set of cuts is applied at analysis level. Given the large number of expected jets (and  $b$ -jets) in the signal final state, the analysis is performed solely on events that contain at least five jets, of which at least two are  $b$ -tagged, with a minimum transverse momentum of

<sup>1</sup>A different set of low- $p_T$  triggers is used for the estimate of the fake background. Details are given in Appendix B.

25 GeV. A cut on the transverse momentum is also applied on leptons, which are required to have a  $p_T$  of at least 27 GeV. Both leptons and jets are required to be within  $|\eta| \leq 2.5$ . More forward objects are neglected. Events containing more than one electron or muon are vetoed, preventing possible overlaps with the dilepton analysis. Events that include tau leptons contribute to the analysis in the rare occasion in which the leptons produced by their decays survive to the quality requirements or when another lepton is found in the events. Events are then categorised into regions, using the multiplicity of the jets and  $b$ -jets. A total of six regions is defined: two control regions (CRs), dominated by  $t\bar{t}$  + light events, and four signal regions (SRs), enriched in signal and dominated by  $t\bar{t}$  + HF events. The region definitions are given in Table 6.2: the control regions correspond to events with either five jets and two  $b$ -jets (5j2b) or at least six jets and two  $b$ -jets ( $\geq 6j2b$ ). The four signal regions require either three or more than three  $b$ -jets (5j3b, 5j $\geq 4b$ ,  $\geq 6j3b$ ,  $\geq 6j\geq 4b$ ).

	2 $b$ -jets	3 $b$ -jets	$\geq 4b$ -jets
5 jets	CR	SR	SR
$\geq 6$ jets	CR	SR	SR

Table 6.2: Control and signal regions used in the analysis.

The signal over background ratio (S/B) of the event yields in the six regions is presented in Figure 6.1, for the 18 signal mass hypotheses (the signal events are normalised to 1 pb). The sensitivity to the  $H^+$  production is expected to come from the regions with the highest (relative) amount of signal, i.e. the 5j $\geq 4b$  and  $\geq 6j\geq 4b$  regions in the low-mass range, and the  $\geq 6j3b$  and  $\geq 6j\geq 4b$  regions in the high mass range.

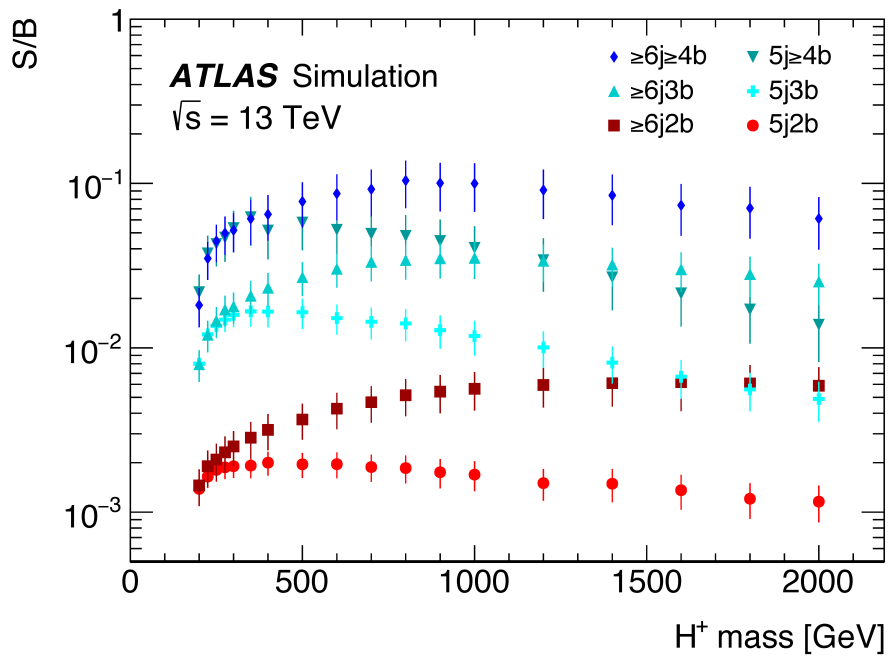


Figure 6.1: Signal over background ratio (S/B) as a function of the  $H^+$  mass in the control and signal regions. The signal is normalised to 1 pb for all mass points, while the backgrounds are normalised to the SM cross-sections [13].

The expected background composition for the six regions is shown in Figure 6.2. The yields of the  $t\bar{t}$  + jets background are split into the three flavour subcategories:  $t\bar{t}$  + light,  $t\bar{t}$  +  $\geq 1c$  and  $t\bar{t}$  +  $\geq 1b$ . The  $t\bar{t}W$ ,  $t\bar{t}Z$  and  $t\bar{t}H$  processes are included in the  $t\bar{t}X$  category, while all other backgrounds are classified as non- $t\bar{t}$ . The dominant background in the control regions is produced by  $t\bar{t}$  + light events, while the  $5j \geq 4b$  and  $\geq 6j \geq 4b$  signal regions are dominated by  $t\bar{t}$  +  $\geq 1b$  events. The  $5j3b$  and  $\geq 6j3b$  signal regions consist of a mixture of the three categories: 50% of  $5j3b$  is made of  $t\bar{t}$  + light events, with equal amounts of  $t\bar{t}$  +  $\geq 1c$  and  $t\bar{t}$  +  $\geq 1b$ . The percentage of  $t\bar{t}$  + HF events slightly increases in the  $\geq 6j3b$  region, which is almost evenly divided between the three flavour categories. The analysis is missing a region dominated by  $t\bar{t}$  +  $\geq 1c$ , which is the most difficult  $t\bar{t}$  + jets category to control. Observed and expected event yields for the six regions are presented in Figure 6.3. The signal event yields are shown for two mass hypotheses, at 200 and 800 GeV, normalised to 1 pb and superimposed to the distribution of the SM backgrounds. Data and predictions agree within the systematic uncertainties, however the SM background is slightly underestimated in the signal regions.

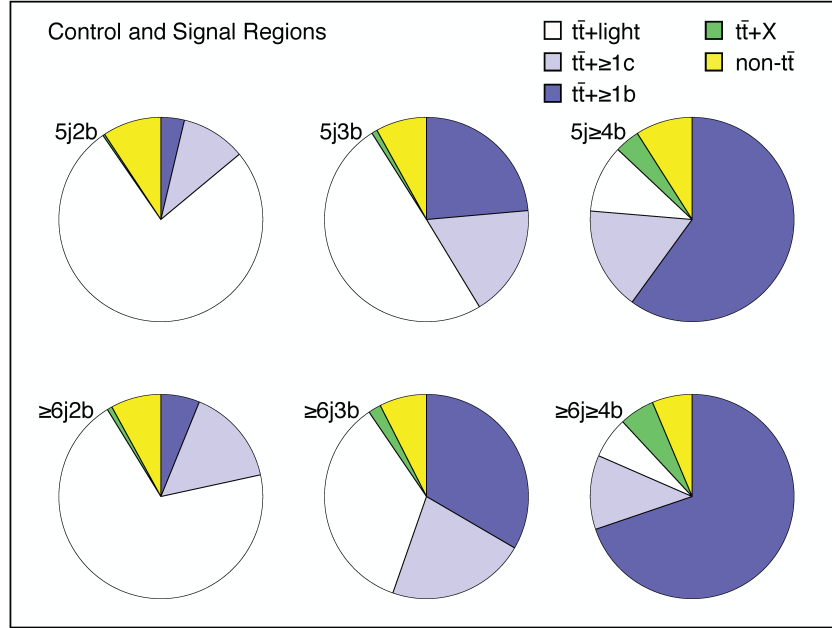


Figure 6.2: Background composition of control and signal regions. The  $t\bar{t}$  + jets background is divided in the three flavour categories  $t\bar{t}$  + light,  $t\bar{t}$  +  $\geq 1c$  and  $t\bar{t}$  +  $\geq 1b$ . The  $t\bar{t}X$  category includes the  $t\bar{t}Z$ ,  $t\bar{t}W$  and  $t\bar{t}H$  processes. All other backgrounds are included in the non- $t\bar{t}$  category.

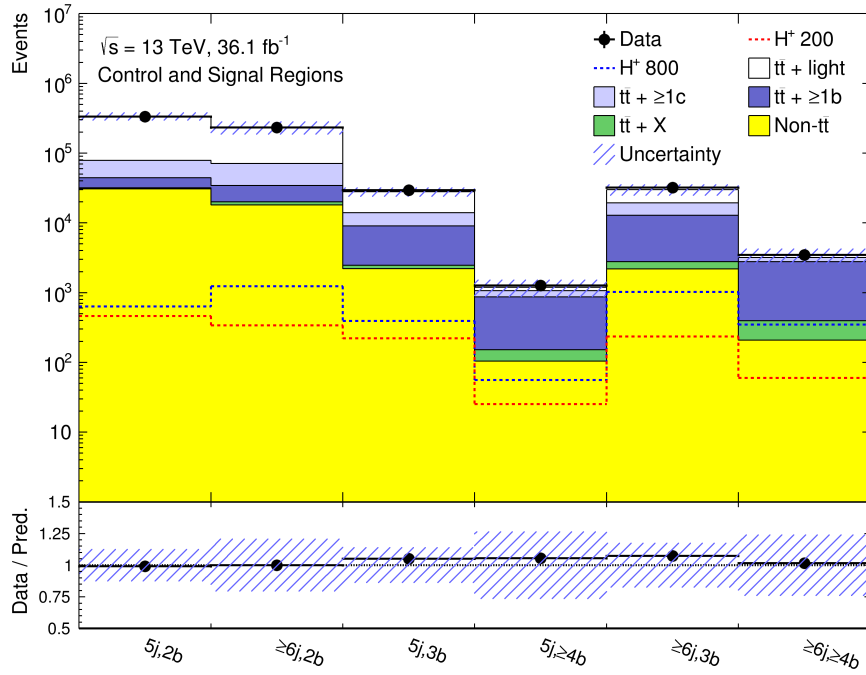


Figure 6.3: Pre-fit comparison of the predicted and observed event yields in control and signal regions. Each background is normalised to the predicted SM cross-section. Two signal hypotheses ( $m_{H^+} = 200$  GeV and  $m_{H^+} = 800$  GeV) are shown as dashed lines, normalised to 1 pb and superimposed to the background distribution. The lower panel shows the ratio between the observed yields and predicted yields for the SM backgrounds. The error bands include all systematic and statistical uncertainties.



## 6.2 Multivariate methods

The usage of multiple regions contributes to improve the sensitivity of the analysis to the signal. However, signal and  $t\bar{t} + \geq 1b$  events have very similar kinematics and the signal regions are enriched in both components. Multivariate techniques are therefore used to further enhance the separation between signal and background. A set of BDTs is trained with the Toolkit for Multivariate Data Analysis (TMVA, [87]), separately for each of the 18 signal hypotheses. The training is split into three regions:  $5j \geq 3b$ ,  $\geq 6j3b$  and  $\geq 6j \geq 4b$ . The 5-jets signal regions are trained together to minimise the overtraining in  $5j \geq 4b$ , which includes only few thousands events. The samples used in the training include the full set of SM backgrounds, properly normalised to the expected cross-section. For  $m_{H^+} \geq 350$  GeV, BDTs are trained on 12 kinematic variables, while the training at low masses includes an additional variable, named "likelihood discriminant" (LHD), built to enhance the separation between  $H^+$  and  $t\bar{t} + \text{jets}$  events. Boosted decision trees were introduced in Chapter 2; the following focuses on the performance of the specific setup used in the analysis.

### 6.2.1 Input variables

The BDT is trained using several types of input variables. The list includes:

- $p_T^{\text{jet}_0}$ : the transverse momentum of the leading-jet, i.e. the jet with the highest transverse momentum in the event. This variable is particularly useful to identify charged Higgs bosons with large invariant masses, because they produce very boosted decay products;
- $H_T^{\text{jets}}$ : the sum of the transverse momenta of all jets in the event. Similarly to  $p_T^{\text{jet}_0}$ , this variable is sensible to charged Higgs bosons with large invariant masses;
- $M_{bb}^{\text{Min}(\Delta R)}$ : the invariant mass of the system composed of the two  $b$ -jets with the smallest separation in the  $\eta$ - $\phi$  plane. It aims to partially reconstruct the top quark (and  $H^+$ ) invariant mass, as all other mass variables below;
- $M_{bb}^{\text{Max}(p_T)}$ : the invariant mass of the system composed of the two  $b$ -jets with the largest vector sum  $p_T$ ;
- $M_{bb}^{\text{Max}(M)}$ : the maximum invariant mass of all pairs of  $b$ -jets;
- $M_{jjj}^{\text{Max}(p_T)}$ : the invariant mass of the system composed of the three jets with the largest vector sum  $p_T$ ;
- $p_T^{\text{jet}_5}$ : the transverse momentum of the fifth jet (ordered from the highest to the lowest  $p_T$ ). It characterises the low energy component of the event;
- $H_1^{\text{all}}$ : the second Fox-Wolfram moment calculated using all leptons and jets in the event [146];

- $\Delta R_{bb}^{\text{avg}}$ : the average of the angular distance between all pairs of  $b$ -jets in the event;
- $M_{uu}^{\text{Min}(\Delta R)}$ : the invariant mass of the untagged pair of jets with the smallest separation in the  $\eta$ - $\phi$  plane. It is a proxy for the mass of the  $W$  boson that decayed hadronically;
- $\Delta R_{\text{lep},bb}^{\text{Min}(\Delta R)}$ : the minimum angular distance between the lepton and all pairs of  $b$ -jets. It is used to distinguish between charged Higgs bosons that decay into leptons and charged Higgs bosons that decay into hadrons;
- $C_{\text{all}}$ : the centrality of the event. It is an event shape variable, computed as the scalar sum of the  $p_T$  of all jets and leptons, divided by the sum of their energies.

As mentioned before, the list includes an additional variable when training BDTs in the low-mass region. The likelihood discriminant is computed using a set of probability density functions that exploit the truth-kinematics of  $H^+$  and  $t\bar{t} + \text{jets}$  events. The LHD calculation starts by performing the matching between reconstructed jets and quarks, identified at generator level. A quark is matched to a jet when they are found within  $\Delta R \leq 0.3$ . When the full set of partons produced by the interaction can be matched to reconstructed jets, the event is named Fully-Matched (FM). If some of the partons cannot be matched, the event is called Partially-Matched (PM). Partially-Matched events are mostly constituted of combinations in which some of the partons were produced outside the detector acceptance region or the associated jet didn't pass the selection requirements. Most of the times the missing parton is one of the quarks produced by the  $W$  boson decaying hadronically.

The second step is to create a probability density function for each of the following invariant masses, using exclusively the jets that were matched during the previous step:

- $M_{qq}$ : the mass of the hadronic  $W$  boson (hadronic stands for "decaying into hadrons");
- $M_{bl\nu}$ : the mass of the leptonic top quark<sup>2</sup> (leptonic stands for "decaying into leptons");
- $X_T^h = M_{bqq} - M_{qq}$ : the difference between the masses of the hadronic top quark and the hadronic  $W$  boson;
- $X_H^h = M_{bbqq} - M_{bqq}$ : the difference between the mass of the charged Higgs boson and the mass of the top quark, in the case of an hadronic decay of the  $H^+$  (and the associated top quark);
- $X_H^l = M_{bb\nu} - M_{bl\nu}$ : the difference between the mass of the charged Higgs boson and the mass of the top quark, in the case of a leptonic decay of the  $H^+$  (and the associated

---

<sup>2</sup>The  $z$  component of the neutrino is recovered by reconstructing the  $W \rightarrow l\nu$  decay in the hypothesis that all the missing energy in the event is produced by a single  $W$  boson decay. This results in a quadratic equation with solutions:

$$p_{\nu,z}^{\pm} = \frac{p_{l,z}\beta \pm \sqrt{\Delta}}{2(E_l^2 - p_{l,z}^2)}, \quad (6.1)$$

where  $\beta = M_W^2 - M_l^2 + 2p_{l,x}p_{\nu,x} + 2p_{l,y}p_{\nu,y}$  and  $\Delta = E_l^2[\beta^2 + (2p_{l,z}p_{\nu,t})^2 + (2E_l p_{\nu,t})^2]$ . The solution with the smallest  $p_{\nu,z}$  is chosen when  $\Delta > 0$ . When  $\Delta < 0$ ,  $p_{\nu,z}$  is estimated by reducing the missing transverse energy to the maximum amount that returns a positive  $\Delta$ .

top quark).

The PDFs are produced separately for FM and PM events and for each of the signal (and  $t\bar{t}$  + jets) nominal samples. A charged Higgs boson is not included in background processes, therefore the PDFs for the  $X_H^l$  and  $X_H^h$  variables in  $t\bar{t}$  + jets events are constructed with a different approach. For regions with five jets, the two variables are built by combining the reconstructed top quarks with the jet that is unmatched to their decay products. For regions with at least six jets, the variables are substituted with the invariant mass of the two jets with the highest transverse momentum (between those that are not matched to the decay products of the top quarks).

The following step is to iterate over all possible jet combinations, in data and MC events, and to estimate the probability that each of these combinations has of being generated by signal or background events. This is achieved by combining the aforementioned PDFs in a single "kinematic" probability defined as:

$$P_{\text{sig/bkg}}^{\text{kin}} = P_{\text{sig/bkg}}(X_H) \cdot P_{\text{sig/bkg}}(X_T^h) \cdot P_{\text{sig/bkg}}(M_{b\nu}) \cdot P_{\text{sig/bkg}}(M_{qq}), \quad (6.2)$$

where each term corresponds to the weighted average of the FM and PM cases:

$$P_{\text{sig/bkg}}(x) = \omega_{\text{sig/bkg}}^{FM} \cdot P_{\text{sig/bkg}}^{FM}(x) + \omega_{\text{sig/bkg}}^{PM} \cdot P_{\text{sig/bkg}}^{PM}(x). \quad (6.3)$$

The  $\omega$ 's are extrapolated from the corresponding fractions of simulated events, i.e.  $\omega_{PM}$  is the fraction of Partially-Matched events, while  $\omega_{FM}$  corresponds to the fraction of Fully-Matched events. The  $P_{\text{sig/bkg}}(X_H)$  of Eq. 6.2 is obtained by averaging between the hadronic and leptonic cases:

$$\begin{aligned} P_{\text{sig/bkg}}^{PM}(X_H) &= \omega_{\text{sig/bkg}}^{l,PM} \cdot P_{\text{sig/bkg}}^{PM}(X_H^l) + \omega_{\text{sig/bkg}}^{h,PM} \cdot P_{\text{sig/bkg}}^{PM}(X_H^h), \\ P_{\text{sig/bkg}}^{FM}(X_H) &= \omega_{\text{sig/bkg}}^{l,FM} \cdot P_{\text{sig/bkg}}^{FM}(X_H^l) + \omega_{\text{sig/bkg}}^{h,FM} \cdot P_{\text{sig/bkg}}^{FM}(X_H^h), \end{aligned} \quad (6.4)$$

where  $\omega^l$  is the fraction of events in which the  $H^+$  decays into leptons and  $\omega^h$  corresponds to the fraction of events in which the  $H^+$  decays into hadrons (computed separately for FM and PM events).

Finally, the probability that each event has of being signal or background is computed by averaging over the kinematic probabilities of all possible jet combinations (using only the eight jets with the largest  $p_T$ ):

$$P_{\text{sig/bkg}} = \frac{\sum_{k=1}^{n_{\text{comb}}} P_{\text{sig/bkg}}^{k,\text{btag}} P_{\text{sig/bkg}}^{k,\text{kin}}}{\sum_{k=1}^{n_{\text{comb}}} P_{\text{sig/bkg}}^{k,\text{btag}}}. \quad (6.5)$$

$P^{\text{kin}}$  is the kinematic PDF defined by Eq. 6.2, while  $P^{\text{btag}}$  is a PDF associated to the  $b$ -tagging

probability of the jets. It is defined differently for signal and background classes:

$$P_{\text{sig}}^{\text{btag}} = P_b^{\text{jet}_1} P_l^{\text{jet}_2} P_l^{\text{jet}_3} P_b^{\text{jet}_4} P_b^{\text{jet}_5} \left\{ \begin{array}{ll} 1, & \text{for } 5j \\ P_l^{\text{jet}_6}, & \text{for } 6j3b \\ P_b^{\text{jet}_6}, & \text{for } 6j4b \end{array} \right\} \cdot \prod_{i=7}^{n_{\text{jets}}} P_l^{\text{jet}_i}, \quad (6.6)$$

$$P_{\text{bkg}}^{\text{btag}} = P_b^{\text{jet}_1} P_l^{\text{jet}_2} P_l^{\text{jet}_3} P_b^{\text{jet}_4} \left\{ \begin{array}{ll} P_l^{\text{jet}_5}, & \text{for } 5j \\ P_l^{\text{jet}_5} P_l^{\text{jet}_6}, & \text{for } 6j3b \\ P_b^{\text{jet}_5} P_b^{\text{jet}_6}, & \text{for } 6j4b \end{array} \right\} \cdot \prod_{i=7}^{n_{\text{jets}}} P_l^{\text{jet}_i}. \quad (6.7)$$

The use of the  $b$ -tagging information allows to reduce the impact of wrong combinations, given that each signal event is expected to contain at least three  $b$ -tagged jets.  $P_l$  and  $P_b$  are the probabilities that each jet has of including a light or  $b$ -quark. They are binned probabilities, obtained through the integration of the normalised MV2c10 output distribution for simulated light and  $b$ -tagged jets (shown in Section 4.5.4). A jet that is  $b$ -tagged at 60% efficiency will have the highest  $P_b$ , a jet that is  $b$ -tagged at 70% but not at 60% efficiency will have a smaller  $P_b$  but a higher  $P_l$  and so on. Table 6.3 provides the values of the binned distributions of  $P_b$  and  $P_l$ .

Efficiency	1 – 0.85	0.85 – 0.77	0.77 – 0.70	0.70 – 0.60	0.60 – 0.0
$P_b$	0.15	0.08	0.07	0.1	0.60
$P_l$	0.970	0.022	0.005	0.002	0.001

Table 6.3: Pseudo-continuous  $b$ -tagging probabilities for light and  $b$ -jets.

The approach of using multiple  $b$ -tagging efficiencies is known as the Pseudo-Continuous  $b$ -tagging (PCBT) and it requires a different set of systematic uncertainties with respect to the case in which only one (fixed) working-point is used throughout the analysis (the systematic uncertainties are described later, in Section 6.3.1). With respect to the fixed WP case, the PCBT systematic uncertainties are not calibrated for jets with transverse momenta larger than 300 GeV. Less than 5% of the jets have a  $p_T$  larger than 300 GeV for  $m_{H^+} < 350$  GeV, while the percentage grows significantly at higher masses. Hence the decision of using the likelihood discriminant only for the training at low masses.

The LHD variable passed to the classification BDT is finally defined as:

$$\text{LHD} = \frac{P_{\text{sig}}(x)}{P_{\text{sig}}(x) + P_{\text{bkg}}(x)}, \quad (6.8)$$

where the  $P_{\text{sig}}(x)$  and  $P_{\text{bkg}}(x)$  are obtained from Eq. 6.5 under the signal or background

hypotheses, i.e. using signal or background PDFs. The distribution of the likelihood discriminant for the 200 GeV mass point is shown in Figure 6.4, comparing signal and  $t\bar{t} + \text{jets}$  events in the BDT training regions. The LHD performs at its best in the  $5j \geq 3b$  region, while the kinematic information is strongly diluted by the larger number of combinations in the  $\geq 6$ -jets regions.

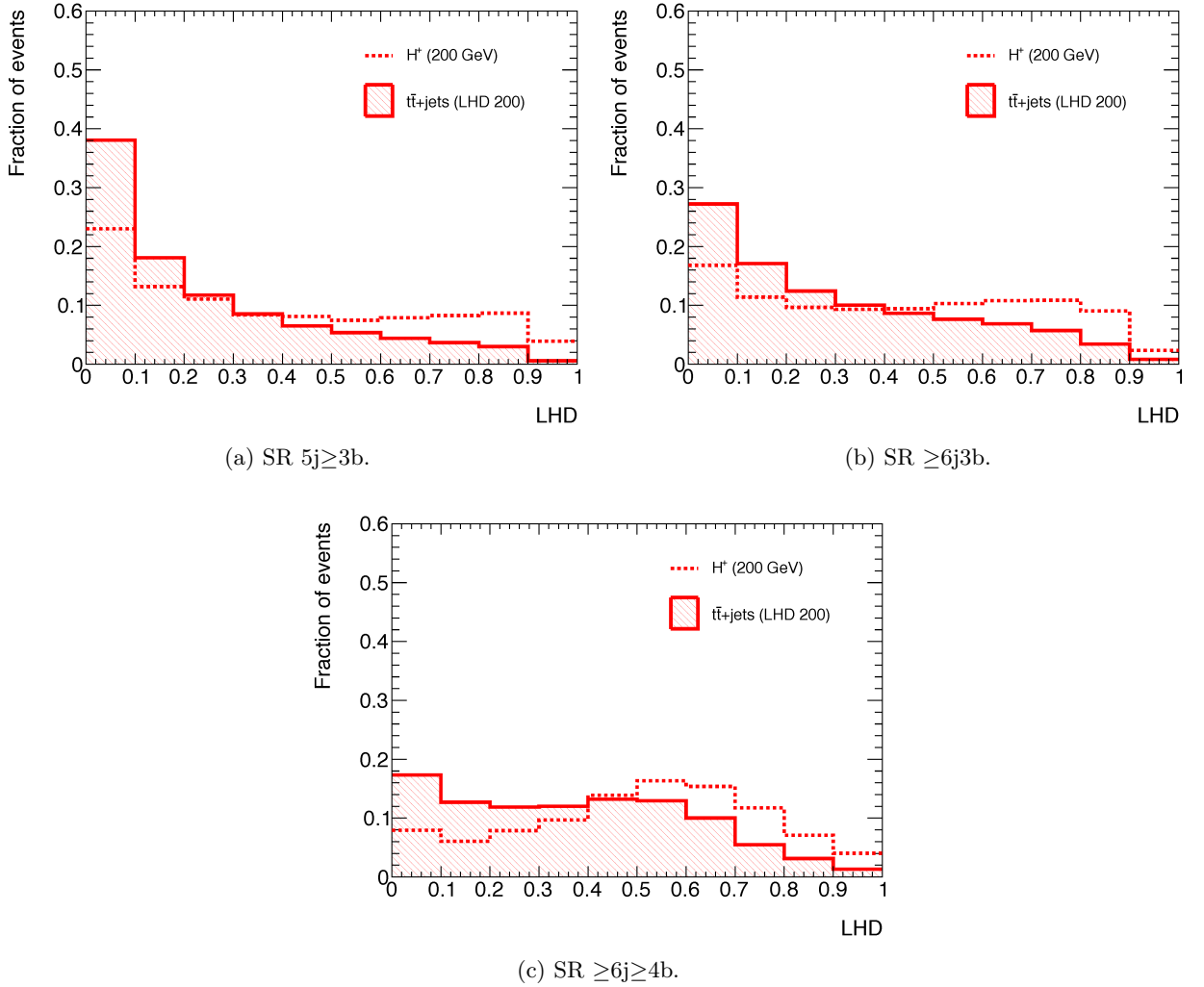


Figure 6.4: The output distributions (normalised to unity) of the LHD variable for the 200 GeV signal hypothesis. The  $t\bar{t} + \text{jets}$  and  $H^+$  distributions are compared in the three SRs used in the BDT training: (a)  $5j \geq 3b$ , (b)  $\geq 6j3b$  and (c)  $\geq 6j \geq 4b$ .

### 6.2.2 Classification BDT: training and performance

The classification BDTs are created using a two-fold training process in which one half of the events is used for the training and another half is used for validation and evaluation in the fit. All MC events are assigned an event-number when they are produced: events with an odd event-number are evaluated using the BDTs trained on even events and even events are evaluated using the BDTs trained on odd events. This method prevents overtraining biases while exploiting the full statistics of the MC samples. The BDTs are trained to distinguish the signal from the full set of SM backgrounds. The disproportion between the number of training events of the signal and background classes is taken care by weighting each signal event by the ratio between the total number of background events and the total number of signal events.

The setup used for the hyper-parameters of the training is shown in Table 6.4. The parameters are optimised to avoid overtraining and maximise the discrimination power, i.e. maximising the AUCs of the Receiver Operating Characteristic curves. The performance of the BDT increases with the mass, ranging from  $\text{AUC} \approx 0.67$  at 250 GeV, to  $\text{AUC} \approx 0.99$  at 2000 GeV<sup>3</sup>.

Option	Value	Description
NTrees	160 ( $5j \geq 3b, \geq 6j3b$ ) 120 ( $\geq 6j \geq 4b$ )	Number of trees
MaxDepth	5	Maximum depth of each tree
MinNodeSize	0.1%	Minimum percentage of events in each node
nCuts	50	Number of cuts used to find the optimal splitting point for each variable
BoostType	Grad	Boosting type
Shrinkage	0.13	Learning rate for the gradient descent algorithm
SeparationType	GiniIndex	Criterion used for the node splitting

Table 6.4: Hyper-parameters used in the BDT training.

The kinematic of the charged Higgs boson decay is very different across the mass range. Consequently, BDTs trained with different signal hypotheses learn to exploit different variables. Figure 6.5 presents the ranking of the input variables based on their "importance", which is a quantity estimated by counting how often a given variable is used to split a node of the BDT [87]. The likelihood discriminant is the most important variable below 350 GeV, while the leading-jet  $p_T$  and  $H_T^{\text{jets}}$  are the most relevant at high masses. The output distributions of the BDTs trained with the 200 and 800 GeV signal hypotheses are shown in Figure 6.6.

<sup>3</sup>The AUC values are computed inclusively in the  $\geq 5j \geq 3b$  region. The minimum is at 250 GeV, not at 200 GeV: the 250 GeV mass point seems to correspond to the signal hypothesis for which the  $t\bar{t}$  + jets background is mimicking the  $H^+$  decay at its best. For example, the distribution of the  $M_{b\bar{b}q\bar{q}}$  mass used in the likelihood discriminant shows the smallest separation between  $t\bar{t}$  + jets and  $H^+$  at 250 GeV.

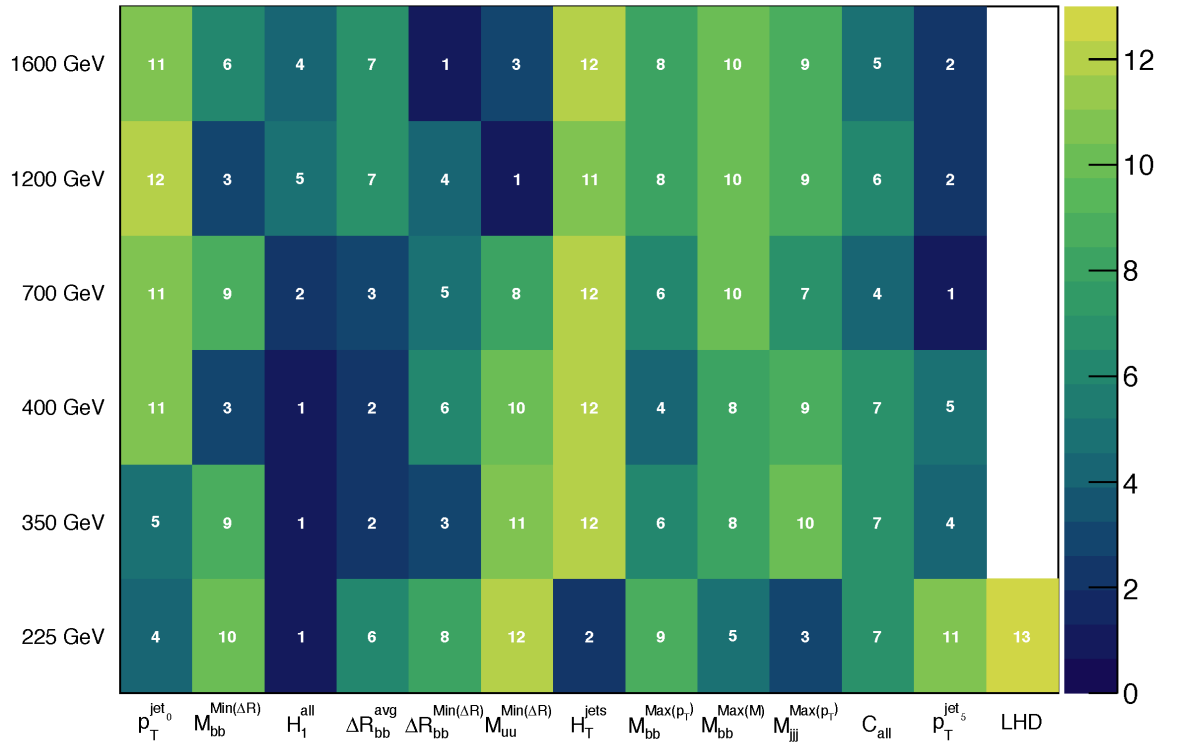


Figure 6.5: Ranking of the BDT input variables for a selection of masses. The higher the number, the higher the variable importance, i.e. the number of nodes created using that variable. The LHD variable is only shown for the 225 GeV mass point, because it is not used above 300 GeV.

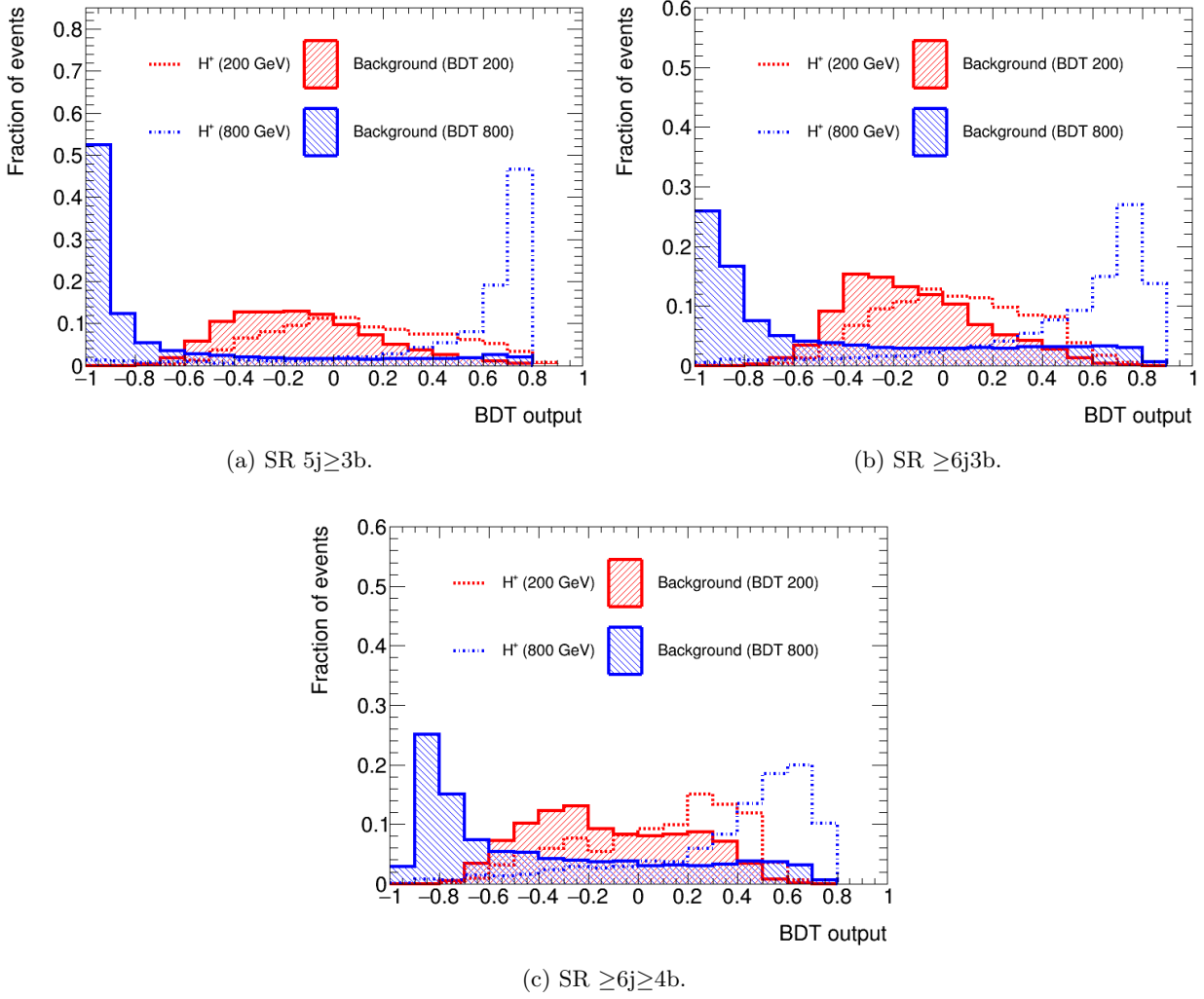


Figure 6.6: The output distributions (normalised to unity) of the classification BDTs for the 200 GeV and 800 GeV signal hypotheses. The signal distributions are drawn as dashed lines while the total background distributions are drawn as solid lines. The distributions are shown in the three SRs used in the BDT training: (a)  $5j \geq 3b$ , (b)  $\geq 6j3b$  and (c)  $\geq 6j \geq 4b$ . The binning does not correspond to the one used in the fit.



### 6.3 Systematic uncertainties

The analysis takes into account a large number of systematic uncertainties, shortly described in this section. It is possible to define two categories of systematic uncertainties: experimental uncertainties, associated to the reconstruction of physics objects with the ATLAS detector, and theoretical uncertainties, which arise from the modelling of the background and signal processes. In practical terms, all systematic uncertainties are estimated by varying the distribution of the variables passed to the profile likelihood fit. They can be obtained by scaling the total yields of the nominal distribution by a constant factor (normalisation-systematic), applying a weight to the events (weight-systematic), or comparing the nominal distribution to the one produced by an alternative sample (sample-systematic). Normalisation-systematics are mostly used to describe uncertainties on the cross-sections, while weight-systematics are used to model variations of the PDF sets or uncertainties on the efficiencies of the  $b$ -tagging and lepton reconstruction. Finally, sample-systematics are used to cover some of the uncertainties related to the signal and background modelling, as the ones related to the choice of the ME generator or the parton shower. Weight- and sample-systematics change both the shape and normalisation of the samples. Table 6.5 offers a complete list of the systematic uncertainties used in this work. Each systematic uncertainty is kept 100% correlated with itself across samples and regions, unless stated otherwise. The uncertainties are derived according to the recommendations of specialised ATLAS working groups.

To reduce the computational time required by the fits, uncertainties that show a difference of less than 1% (in shape and/or normalisation) with respect to the nominal distribution are dropped. It is possible to only exclude the shape component of a given uncertainty and preserve the normalisation, or vice-versa. This procedure is known as "pruning" and it is performed separately in each region and for each sample<sup>4</sup>. In addition, all uncertainties are symmetrised, providing the fit perfectly symmetric  $+1\sigma$  and  $-1\sigma$  variations. Some systematic uncertainties are provided as a single variation with respect to the nominal distribution (one-sided), while others have both a positive and a negative variation (two-sided). One-sided systematic uncertainties are symmetrised mirroring the variation around the nominal distribution, while for two-sided systematic uncertainties the difference between the two variations is centred around the nominal distribution, bin-by-bin. Finally, all uncertainties are "smoothed", i.e. they are re-binned until the slope of the systematic variation distribution shows less than a maximum number of direction changes with respect to the nominal distribution (four by default). They are then processed with the "TH1:Smooth()" function of ROOT [147]. This prevents the fit from being sensitive to statistical fluctuations in the distribution of the systematic uncertainties.

<sup>4</sup>The shape pruning is performed checking if any bin has a deviation larger than 1% after renormalising the distribution of the systematic uncertainty to the integral of the nominal sample. Dedicated studies were performed to demonstrate that a pruning of 1% has a negligible impact on the final results.

Systematic uncertainty	Type	Number of components
Luminosity	N	1
Pile-up reweighting	SN	1
<b>Reconstructed objects</b>		
Electron trigger+reco+ID+isolation	SN	4
Electron energy scale+resolution	SN	2
Muon trigger+reco+ID+isolation+TTVA	SN	10
Muon momentum scale+resolution	SN	3
Jet vertex tagger (JVT)	SN	1
Jet energy scale (JES)	SN	23
Jet energy resolution (JER)	SN	1
$E_T^{\text{miss}}$ resolution and scale	SN	3
$b$ -tagging efficiency, fixed WP (*)	SN	6
$b$ -tagging efficiency, PCBT (*)	SN	30
$c$ mistag rate, fixed WP (*)	SN	3
$c$ mistag rate, PCBT (*)	SN	15
Light mistag rate, fixed WP (*)	SN	16
Light mistag rate, PCBT (*)	SN	80
Mistag extrapolation $c \rightarrow \tau$ , fixed WP (*)	SN	1
Mistag extrapolation $c \rightarrow \tau$ , PCBT (*)	SN	1
High- $p_T$ $b$ -tagging extrapolation, fixed WP (*)	SN	1
<b>Signal model</b>		
Renormalisation and factorisation scales	N	1
Parton distribution function (PDF)	SN	30
<b>Background models</b>		
$t\bar{t}$ + jets cross-section	N	1
$t\bar{t} + \geq 1c$ normalisation	N (floating in fit)	1
$t\bar{t} + \geq 1b$ normalisation	N (floating in fit)	1
$t\bar{t} + \geq 3b$ normalisation	N	1
$t\bar{t} + \geq 1b$ modelling (SHERPA+OPENLOOPS rw.)	SN	8
$t\bar{t} + \geq 1c$ modelling (FS)	SN	1
$t\bar{t}$ + jets modelling: IS and FS radiation	SN	3
$t\bar{t}$ + jets modelling: ME generator	SN	3
$t\bar{t}$ + jets modelling: PS and hadronisation	SN	3
$t\bar{t}$ + jets modelling: jet $p_T$ reweighting	SN	6
Single-top $Wt$ cross-section	N	1
Single-top $Wt$ modelling	SN	3
Single-top $s$ -, $t$ -channels cross-section	N	2
Single-top $t$ -channel radiation and parton shower	SN	2
$t\bar{t}V$ cross-section (QCD and PDF)	N	4
$t\bar{t}V$ ME generator	SN	2
$t\bar{t}H$ cross-section (QCD and PDF)	N	2
$t\bar{t}H$ model	SN	1
$tH$ ( $tHjb$ and $WtH$ ) cross-section (QCD and PDF)	N	2
$tZ$ cross-section (QCD and PDF)	N	2
$t\bar{t}WW$ cross-section (QCD and PDF)	N	2
$tWZ$ cross-section	N	1
$4t$ cross-section	N	1
$W$ +jets normalisation	N	5
$Z$ +jets normalisation	N	1
Diboson normalisation	N	1
Fake leptons normalisation	N	6

Table 6.5: List of the considered systematic uncertainties. "N" indicates normalisation-systematics, while "SN" means that the uncertainty affects both the shape and the normalisation of the sample. The number of independent components considered for each uncertainty is indicated in the last column on the right. (\*) The fixed-point uncertainties are used at high masses ( $m_{H^+} > 300$  GeV) and the PCBT uncertainties at low masses ( $m_{H^+} \leq 300$  GeV).

### 6.3.1 Experimental uncertainties

Experimental uncertainties cover for possible mismodelling in the simulation of the ATLAS detector as well as for differences in the reconstruction of objects obtained from data and simulated events. They are applied to all Monte Carlo samples. The category also includes an uncertainty on the integrated luminosity and one on the pile-up modelling. The systematic uncertainty on the luminosity is computed with the methodology described in Ref. [148] and it corresponds to a normalisation-uncertainty of  $\pm 2.1\%$ . Differences in the pile-up distribution of data and simulated events are covered by a two-sided systematic uncertainty obtained through a reweighting procedure [126]. The two sides correspond to positive and negative variations of the ratio between the measured and predicted number of visible  $pp$  interactions per bunch crossing. Details about the other experimental uncertainties are given below.

#### Charged leptons

Charged leptons are affected by systematic uncertainties related to the trigger selection, the reconstruction and identification efficiencies, and the isolation requirements. In addition, uncertainties on the lepton energy scale and resolution are considered. The uncertainty on the muon energy resolution is separated between the inner detector and the muon spectrometer contributions. Dedicated low- $p_T$  systematic uncertainties and track-to-vertex association (TTVA) uncertainties are also included for the muons [113, 116]. The majority of the charged leptons uncertainties is very small and thrown away by the pruning procedure.

#### Jets and missing transverse momentum

There are three main sources of uncertainty to consider for the jet reconstruction: the Jet Energy Scale (JES), the Jet Energy Resolution (JER) and the efficiency of the Jet Vertex Tagger (JVT) [120, 149]. The latter is a weight-systematic with a negligible impact on the signal strength. The JES and JER uncertainties are instead sample-systematics. The JES uncertainty is split into 23 independent nuisance parameters:

- eight NPs generated by in-situ corrections, related to the event topology, the MC statistics and to effects propagating from the calibration of other objects;
- four NPs associated to pile-up corrections;
- four NPs associated to the simulation of jets with different flavours;
- three NPs from the  $\eta$  inter-calibration, to ensure equal calorimeter responses in all pseudorapidity regions;
- two NPs for the jet punch-through into the Muon Spectrometer, to correct the energy scale of the showers leaking outside the calorimeters (one for AFII samples and one for

FullSim samples);

- one NP to cover for the differences between the AFII vs. FullSim methodologies (only applied to AFII samples);
- one NP for the calibration of high- $p_T$  jets;

The JER systematic uncertainty is included as a single nuisance parameter and it is obtained by smearing the energy of simulated jets. The JES and JER systematic uncertainties can produce variations larger than 10%, in particular in the low- $p_T$  region.

The systematic uncertainties on the energy scale of all reconstructed objects are propagated to the missing transverse energy with three additional nuisance parameters. They have a negligible effect on the analysis.

### Jet flavour tagging

The uncertainties on the flavour tagging arise from differences in the performance of the  $b$ -tagging algorithm when applied to data and Monte Carlo events [150]. They are divided in three categories: uncertainties on the tagging efficiencies of  $b$ -jets and uncertainties on the mistag efficiencies of light and  $c$ -jets. The number of components to consider for each category depends on the number of  $b$ -tagging working-points simultaneously used in the analysis<sup>5</sup>. If only one (fixed) working-point is used, then 6, 3 and 16 independent components are used for the three categories respectively. In the low-mass range ( $m_{H^+} \leq 300$  GeV), where the Likelihood Discriminant is included, a set of Pseudo Continuous  $b$ -tagging uncertainties must be used. The PCBT systematic uncertainties take into account the correlations across multiple working points: they include 30, 15 and 80 components for the three categories respectively (5 times the number of variations used for the fixed working-point  $b$ -tagging). In addition, both sets include an uncertainty that accounts for the fact that the mistag rates of  $c$ -jets are also used for the decays of tau leptons into hadrons.

It is to be noted that a precise  $b$ -tagging calibration is only possible for jets with transverse momenta below 300 GeV. Only few data events are available at higher energies, therefore the calibration has to be extrapolated. The set of uncertainties used for the fixed working-points include an additional systematic uncertainty that covers for the minor precision of the extrapolation procedure. This systematic uncertainty is missing in the PCBT set. As mentioned in Section 6.2.1, this forbids the use of the PCBT calibration, and the likelihood discriminant, for the analysis of signal hypotheses with large masses, for which the decay products are very boosted. Given the large number of  $b$ -tagged jets required by the analysis selection, the uncertainties on the jet flavour tagging can have a strong impact on the signal strength.

---

<sup>5</sup>Multiple components are needed because the correction factors are computed as functions of various kinematic variables, as the jet  $p_T$  and  $\eta$ .

### 6.3.2 Theoretical uncertainties

Theoretical uncertainties arise from the simulation of signal and background processes as well as the method used to evaluate the data-driven backgrounds. In this analysis, the largest uncertainties are related to the  $t\bar{t}$  + jets modelling and in particular to the modelling of the  $t\bar{t}$  + HF category.

#### Signal modelling

Two sources of uncertainty are considered on the signal modelling. The first is associated to the choice of the energy scales used to generate the Monte Carlo samples. As described in Section 5.1, four mass hypotheses (250, 600, 1200 and 2000 GeV) have been generated with variations of the renormalisation and factorisation scales. These alternative samples include a set of weights that allows to change the  $\mu_R$  and  $\mu_F$  energy scales from 0.5 to 2.0 of the nominal value. The variation corresponding to the maximum difference in acceptance is taken as a normalisation uncertainty on the cross-section. The uncertainty is extrapolated to the full set of signal hypotheses by fitting a second order polynomial on the four available masses, as shown by Figure 6.7. The uncertainty is 7% at 200 GeV and it reaches a maximum of 15% at 2000 GeV. In addition, a PDF uncertainty is estimated using the Hessian reduction method [73], considering 30 independent nuisance parameters. Differences between the 4FS and 5FS of the  $H^+$  production have been neglected.

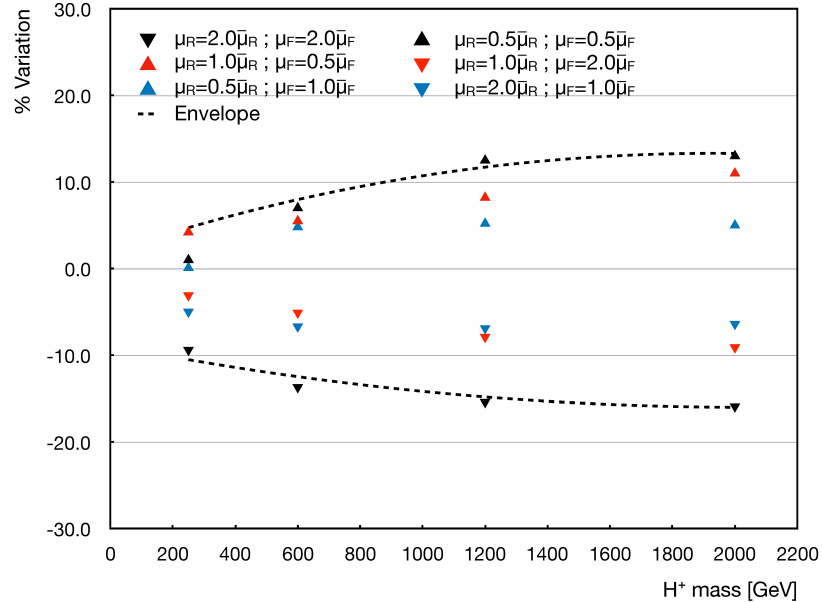


Figure 6.7: Impact of the energy scale variations on the acceptance of  $H^+$  events. The envelope is obtained by fitting a second order polynomial on the maximum difference in acceptance, separately for the positive and negative variations.

**$t\bar{t}$  + jets modelling**

A large number of systematic uncertainties is considered for the  $t\bar{t}$ +jets background, to which the analysis is very sensitive [132]. A flat uncertainty of 6% is applied to the cross-section of  $t\bar{t}$ +jets events to cover for the choice of the energy scales and the PDF set. Furthermore, the cross-sections of the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  categories are poorly modelled in theory and they are therefore included in the fit as free floating parameters, similarly to the signal strength. In addition, an uncertainty of 50% is applied to the cross-section of the  $t\bar{t} + \geq 3b$  category, to cover for the large differences between the 4FS and the 5FS predictions.

An uncertainty on the choice of the ME generator is estimated by comparing the nominal POWHEG + PYTHIA 8 sample to an alternative sample generated with SHERPA, while a comparison with a POWHEG + Herwig 7 sample allows to extrapolate the uncertainty associated to the choice of the parton shower generator. Two POWHEG + PYTHIA 8 samples, with different tuning options, are used to evaluate the uncertainty on the modelling of the initial and final state radiation. These uncertainties are referred to as the ME generator, PS generator and radiation systematic uncertainties and they are de-correlated across the three  $t\bar{t}$ +jets categories:  $t\bar{t}$  + light,  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \geq 1b$  (details regarding the Monte Carlo production of the alternative samples have been given in Section 5.2.1). Two additional uncertainties are related to the choice of the flavour scheme: one is applied to the  $t\bar{t} + \geq 1b$  category, comparing the nominal 5FS sample to one produced with SHERPA+OPENLOOPS in the 4FS. For the  $t\bar{t} + \geq 1c$  category, an uncertainty is derived by comparing the nominal sample to a MG5\_aMC@NLO + Herwig++ sample generated using the 3FS<sup>6</sup> and a CT10f3 PDF set [151]. Another set of eight systematic uncertainties arises from the SHERPA+OPENLOOPS reweighting of the  $t\bar{t} + \geq 1b$  category. These uncertainties are obtained by reweighting the nominal events to different configurations of the SHERPA+OPENLOOPS sample. Three systematic uncertainties are related to the choice of the energy scales, another two are evaluated using alternative sets of parton distribution functions (MSTW2008NLO [76] and NNPDF2.3NLO [127]), two accounts for the modelling of the underlying event and the last is computed using a different shower recoil scheme in the PS generator [132]. Finally, there are two systematic uncertainties associated to the leading-jet  $p_T$  reweighting. The first is obtained by comparing the distributions of the  $t\bar{t}$  + jets categories before and after the reweighting, while the second is a 15% normalisation-systematic applied to events where  $p_T^{\text{jet0}} \geq 400$  GeV, which are the most affected by the procedure<sup>7</sup>. Plots showing the differences between the prediction of the nominal sample and the ones of the systematic variations have been presented in Section 5.2.1.

<sup>6</sup>In the four- and five-flavour schemes the  $c$ -quarks produced as ISR are considered massless particles. In the three-flavour scheme, only  $u$ -,  $d$ - and  $s$ -quarks are available as constituents of the proton in the initial state, while  $c$ -quarks are produced as massive FSR.

<sup>7</sup>The value of 15% is derived by considering the variation in normalisation before and after the reweighting of the distribution of the leading jet  $p_T$ , for  $p_T^{\text{jet0}} \geq 400$  GeV, in the ( $\geq 4$  jets,  $2b$  jets) region.

### Modelling of the other top-quark backgrounds

A 5% uncertainty is applied to the total cross-section of the single-top-quark production, de-correlated across channels [141–143]. Similarly to the  $t\bar{t}$ +jets modelling, alternative samples are used to extract the radiation and PS systematic uncertainties. Additionally, the model includes an uncertainty on the interference between the  $t\bar{t}$ +jets and  $Wt$  processes, obtained by comparing the default "diagram removal" scheme to the "diagram subtraction" scheme [144]. The  $t\bar{t}X$  and  $tH$  processes have cross-section uncertainties that are de-correlated between QCD and PDF effects, with contributions of approximately 10% and 4-6% respectively [45, 152–155]. Higher-order QCD effects are estimated by varying the  $\mu_R$  and  $\mu_F$  energy scales, while the PDF uncertainty is associated to the choice of the parton distribution function. A ME generator uncertainty is also considered for the  $t\bar{t}W$  and  $t\bar{t}Z$  processes, comparing samples generated with MG5\_aMC@NLO and SHERPA. A PS systematic uncertainty is considered for the  $t\bar{t}H$  background, comparing samples produced with PYTHIA and Herwig++. The uncertainties on the  $tZ$  and  $t\bar{t}WW$  cross-sections are de-correlated between QCD (8-10%) and PDF (1-2%) effects, while a 50% systematic uncertainty is applied to the cross-section of the  $4t$  and  $tWZ$  processes, accounting for both PDF and QCD contributions.

### Modelling of the remaining backgrounds

A flat uncertainty of 30% is assumed for the cross-section of  $Z$  + jets samples. A 40% uncertainty is used for the  $W$  + jets process, de-correlated among the number of jets. An additional systematic uncertainty of 30% is applied to  $W$  + jets events that include at least one  $c$ - or  $b$ -hadron matched to a jet, de-correlated among the number of matched jets. These uncertainties are extracted by varying the renormalisation and factorisation energy scales, as well as other parameters related to the ME to PS matching. A 50% uncertainty is used for the cross-section of diboson processes [156]. Finally, fake leptons have a cross-section uncertainty of 50%, de-correlated among the number of  $b$ -jets and the electron and muon channels. Such uncertainty covers effects related to the choice of the parametrisation scheme and to the lepton identification efficiencies. An overview of the theoretical uncertainties applied to the cross-section of the SM processes is provided in Table 6.6.

Process	Normalisation uncertainty	
$t\bar{t} + \text{jets}$	$\pm 6\%$	
$t\bar{t} + \geq 3b$	$\pm 50\%$	
$Z + \text{jets}$	$\pm 30\%$	
$W + \text{jets}$	$\pm 40\%$	
$W + \text{HF}$	$\pm 30\%$	
Single-top	$\pm 5\%$	
$t\bar{t}W$	QCD: $+12.9\%$ $-11.5\%$	PDF: $\pm 3.4\%$
$t\bar{t}Z$	QCD: $+9.6\%$ $-11.3\%$	PDF: $\pm 4.0\%$
$t\bar{t}H$	QCD: $+5.8\%$ $-9.2\%$	PDF: $\pm 3.6\%$
$tH$	QCD: $+10.0\%$ $-10.0\%$	PDF: $\pm 6.0\%$
$t\bar{t}WW$	QCD: $+10.9\%$ $-11.8\%$	PDF: $\pm 2.1\%$
$tZ$	QCD: $+7.8\%$ $-7.8\%$	PDF: $\pm 0.9\%$
$tWZ$	$\pm 50\%$	
$4t$	$\pm 50\%$	
Diboson	$\pm 50\%$	
Fake leptons	$\pm 50\%$	

Table 6.6: Theoretical uncertainties applied to the cross-section of the SM processes. QCD uncertainties correspond to variations of the energy scales, while the PDF uncertainty is associated to the choice of the parton distribution function.

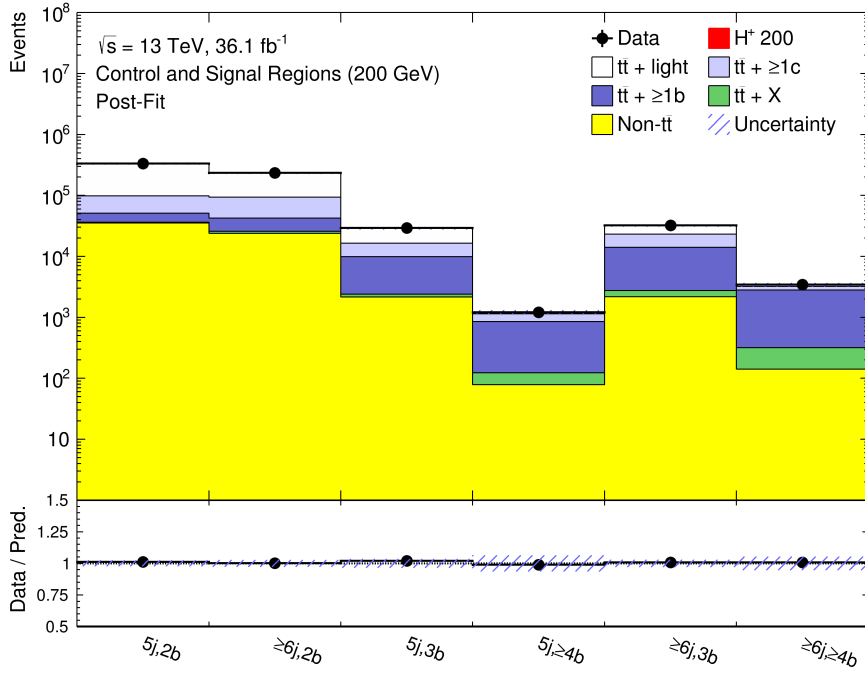


## 6.4 Statistical analysis

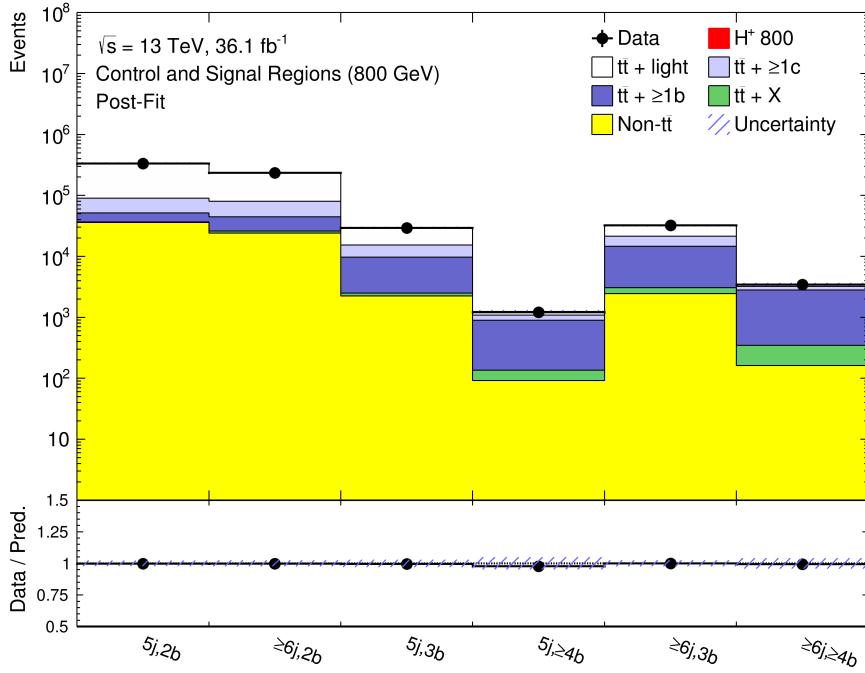
The fit used to extract the signal strength is performed with the RooStats framework [157], integrated in a package called TRexFitter [158]. The likelihood function includes one nuisance parameter for each source of systematic uncertainty and one NP for each bin of the discriminating variable entering the fit, to account for the statistical uncertainties of the MC samples. As mentioned in Chapter 2, the penalty terms associated to the systematic uncertainties are initialised with Gaussian or Log-Normal PDFs, while Gamma functions are used for the statistical errors. Uniform PDFs are used for the signal strength  $\mu$  and the floating parameters related to the normalisations of  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \geq 1b$  events, named  $k_{t\bar{t} + \geq 1c}$  and  $k_{t\bar{t} + \geq 1b}$ . The fit is performed simultaneously on all regions. The control regions are passed to the fit as single-bin distributions, i.e. simply providing the total number of events in each region. A binned distribution of the BDT output is used in the signal regions. The number of bins depends on the specific region and on the signal hypothesis<sup>8</sup>. The agreement between the observed and expected event yields in all regions, after performing the fit, is shown in Figure 6.8. The total uncertainty (systematic and statistical) is shown as a light-blue shaded-band. The results are reported for the 200 and 800 GeV signal hypotheses. With respect to the pre-fit comparison (Figure 6.3), data and MC show an improved compatibility. The signal does not appear because the signal strength is fitted to slightly negative values.

Pre-fit and post-fit distributions of the BDT output in the signal regions are shown from Figure 6.9 to Figure 6.12. The agreement between the observed and expected yields is already good at pre-fit level, but an improvement can be observed after the fit. The nuisance parameters are in fact moved from their original values, accommodating for normalisation and shape differences between the observed and predicted distributions. The largest discrepancies between data and MC are visible in the 5j3b and  $\geq 6j3b$  signal regions, where data overshoots the Monte Carlo predictions in the last bins of the distributions of the BDT output. This is particularly true for the 200 GeV mass hypothesis. However, the agreement is improved at post-fit level and, as shown in few pages, the resulting signal strength is compatible with the absence of signal. The post-fit agreement is mostly due to the pull of the nuisance parameters related to the  $t\bar{t} + \text{jets}$  background and to the pull of the  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \geq 1b$  normalisation factors. The reduced uncertainty band in the post-fit plots is due to the ability of the fit to constrain the systematic uncertainties by exploiting their correlations. The scale of the  $x$ -axis of the figures has been modified in order to regularise the width of the bins. The result is that the content of each bin is clearly visible, independently from its width (which can be very small for the high-masses BDTs).

<sup>8</sup>The binning in the SRs has been chosen as the one that minimises the error on the signal strength, while preserving small statistical errors. The optimisation has been performed at Asimov level, separately for each region and mass point. The number of bins is larger for the  $3b$  regions, which have a larger number of events, and for the low-mass signal hypotheses. The improved separation provided by the classification BDT at high masses can result in a low amount of background events for the last bins of the BDT output. It is therefore necessary to enlarge their width in order to prevent the fit from being dominated by statistical fluctuations.



(a) 200 GeV.



(b) 800 GeV.

Figure 6.8: Post-fit comparison of the predicted and observed event yields in control and signal regions. The results correspond to the fits performed with (a) the 200 GeV and (b) the 800 GeV signal hypotheses. The lower panel corresponds to the ratio between the observed and expected yields. The error bands include all systematic and statistical uncertainties.

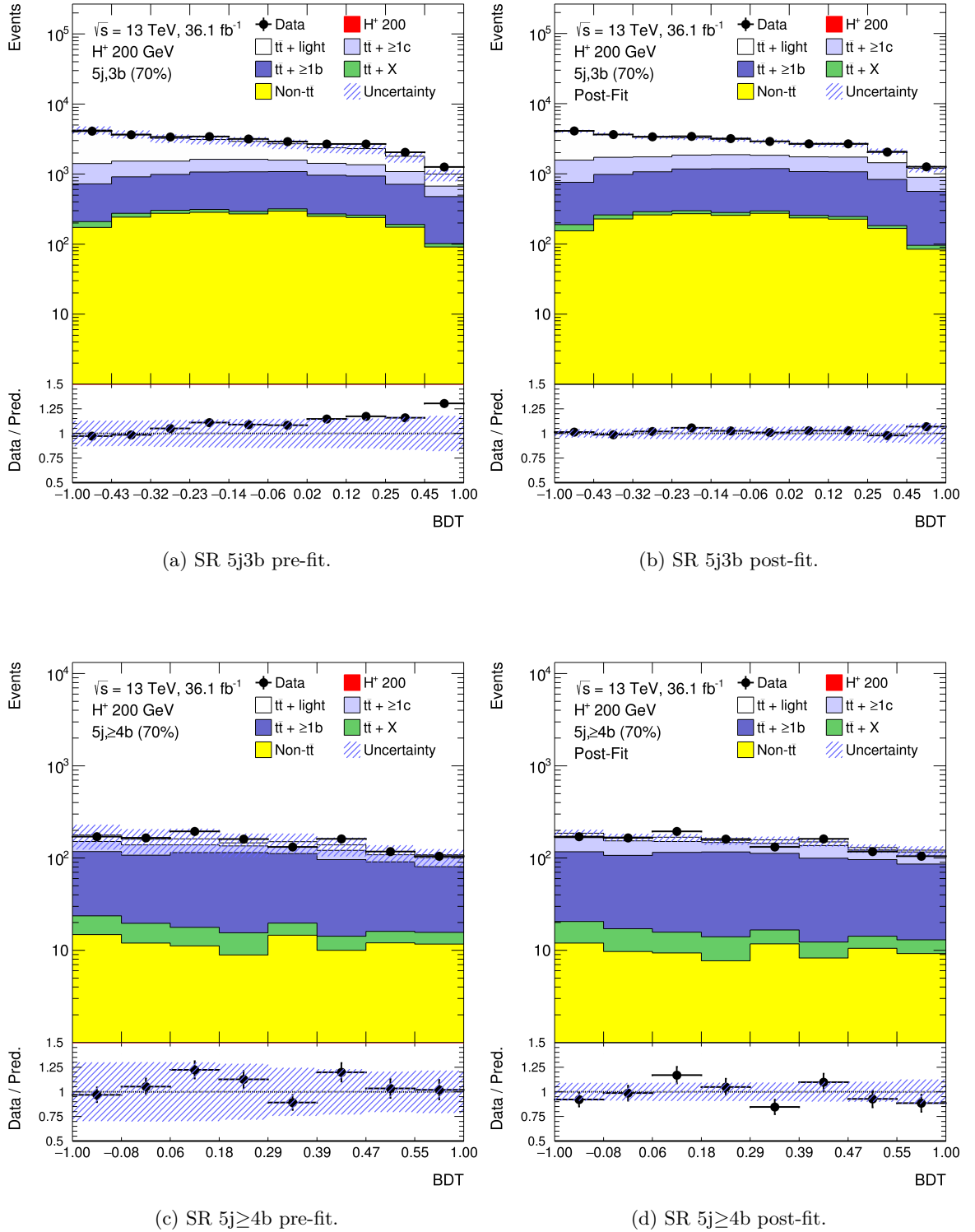


Figure 6.9: (left) Pre-fit and (right) post-fit distributions of the classification BDT output in the 5-jets signal regions, for the fit performed with the 200 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and expected yields. The error bands include all systematic and statistical uncertainties. The scale of the  $x$ -axis has been modified in order to regularise the width of the bins.

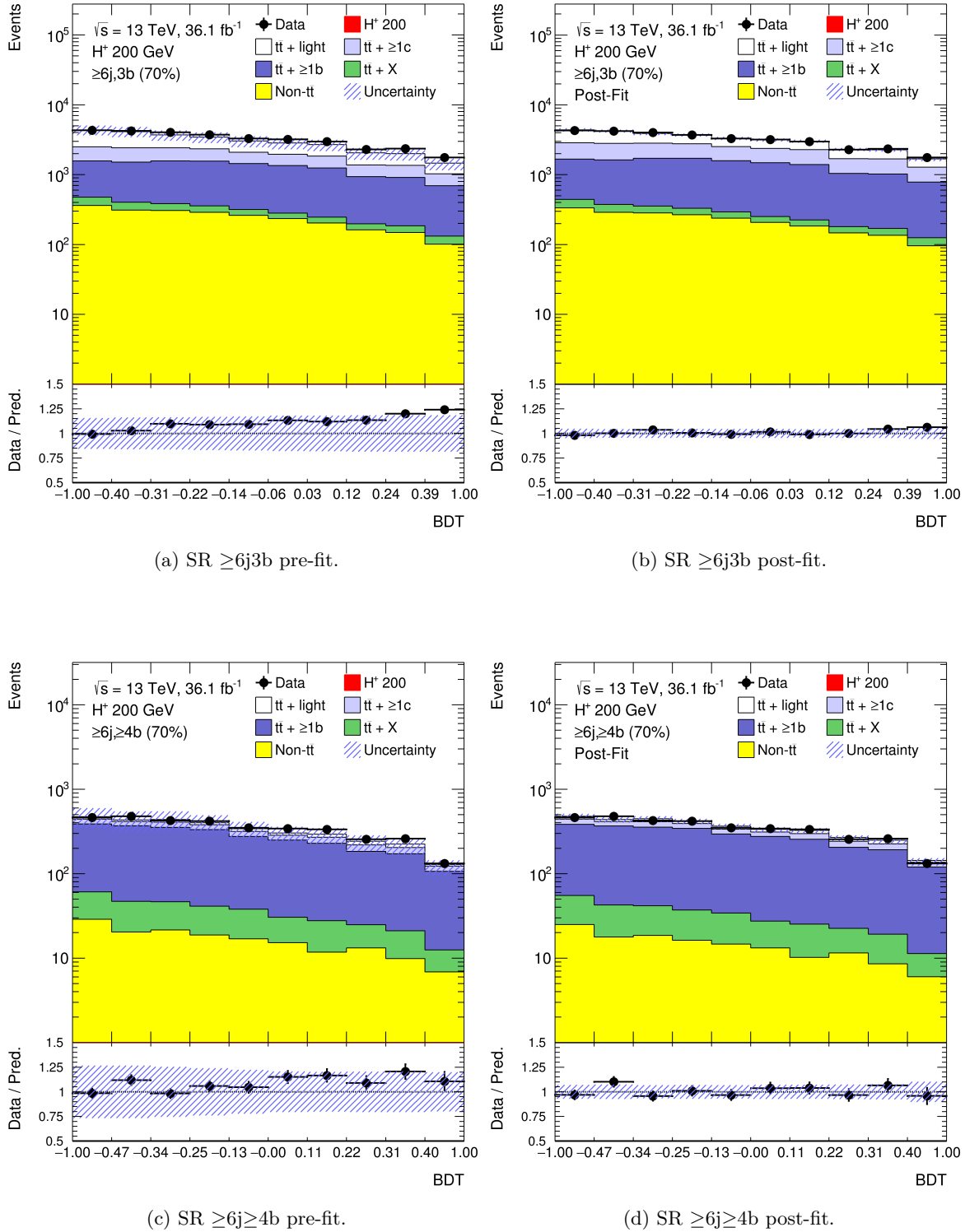


Figure 6.10: (left) Pre-fit and (right) post-fit distributions of the classification BDT output in the  $\geq 6$ -jets signal regions, for the fit performed with the 200 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and expected yields. The error bands include all systematic and statistical uncertainties. The scale of the  $x$ -axis has been modified in order to regularise the width of the bins.

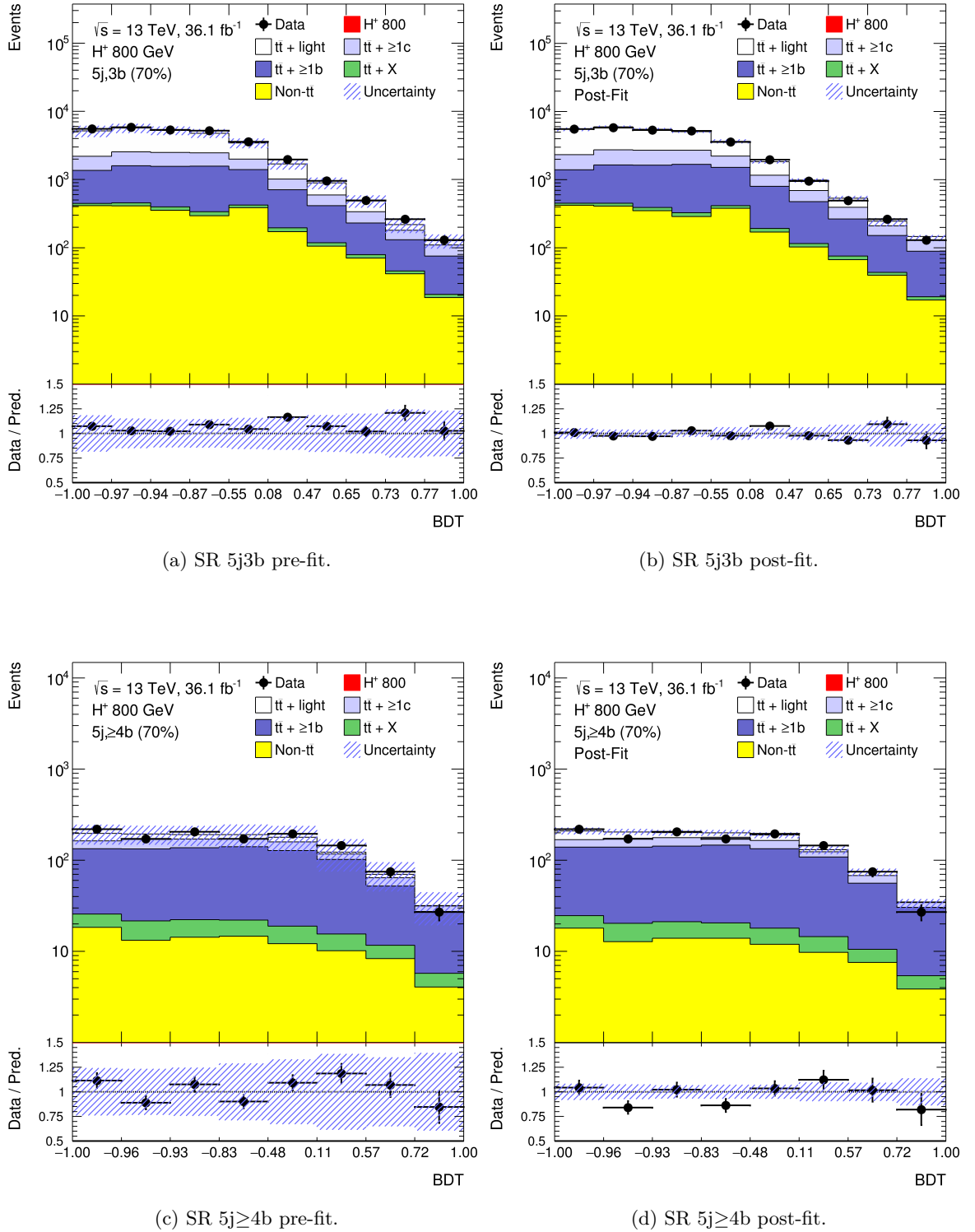


Figure 6.11: (left) Pre-fit and (right) post-fit distributions of the classification BDT output in the 5-jets signal regions, for the fit performed with the 800 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and expected yields. The error bands include all systematic and statistical uncertainties. The scale of the  $x$ -axis has been modified in order to regularise the width of the bins.

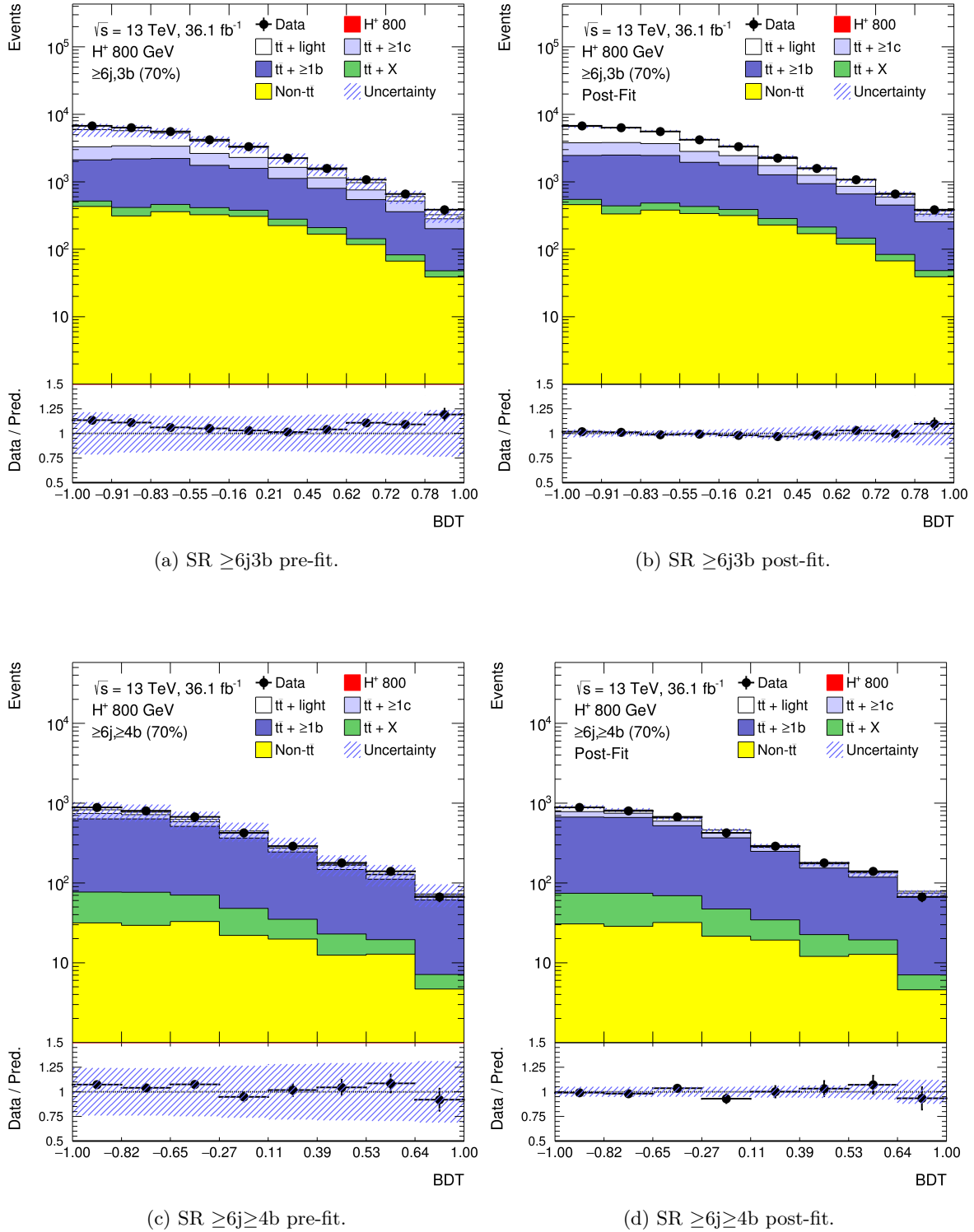


Figure 6.12: (left) Pre-fit and (right) post-fit distributions of the classification BDT output in the  $\geq 6$ -jets signal regions, for the fit performed with the 800 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and expected yields. The error bands include all systematic and statistical uncertainties. The scale of the  $x$ -axis has been modified in order to regularise the width of the bins.

The effect of the systematic uncertainties on the analysis is estimated by evaluating their impact on the signal strength. The impact  $\Delta\mu$  of a systematic uncertainty is defined as the difference between the signal strength obtained with the nominal fit and the one obtained by performing a fit in which the corresponding NP is fixed to its best fit value shifted by its pre-fit or post-fit uncertainty.

Figure 6.13 and Figure 6.14 list the 20 most important nuisance parameters for the fits performed with the 200 and 800 GeV signal hypotheses, ranked from the largest to the smallest impact on  $\mu$ . The three largest contributions to the signal strength uncertainty at 200 GeV are all related to the  $t\bar{t} + \geq 1b$  category. They are the normalisation of the  $t\bar{t} + \geq 1b$  background ( $k_{t\bar{t} + \geq 1b}$ ), the ME generator systematic uncertainty and the 4vs5 flavour scheme systematic uncertainty. At 800 GeV, the dominant systematic uncertainties are related to the modelling of the parton distribution functions and to the leading-jet  $p_T$  reweighting of the  $t\bar{t} + \geq 1b$  category. The difference in ranking of the uncertainties between mass points is justified by the fact that, at low masses, the  $t\bar{t} + \geq 1b$  and  $H^+$  processes produce very similar BDT distributions, therefore the fit is mostly affected by NPs that vary the overall normalisation. At high masses, most signal events are in the tails of the  $t\bar{t} + \geq 1b$  distribution, therefore systematic uncertainties with large shape effects become more relevant. Each NP is shown with its pull (shift) with respect to the pre-fit value (black dots) and its pre-fit and post-fit impact on the signal strength (empty and filled rectangles). The error bars on the pulls correspond to the ratio between the post-fit and pre-fit errors of the NPs. Small ("constrained") errors typically correspond to uncertainties that have high correlation with other nuisance parameters, i.e. uncertainties with similar (or opposite) shape across multiple regions<sup>9</sup>. The fit is able to exploit such correlation to improve its understanding of the model and reduce the pre-fit uncertainty. The largest pulls and the strongest constraints are observed for the systematic uncertainties of the  $t\bar{t} + \text{jets}$  background<sup>10</sup>.

Table 6.7 provides a summary of the impact of the systematic uncertainties by grouping similar sources together. In this case, the impact of each category is obtained by fixing all NPs of the category of interest to their best fit value, repeating the fit, and computing the root square difference between the error on the signal strengths obtained with this fit and the nominal. The statistical uncertainty includes the contribution of the floating normalisation parameters for the  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \geq 1b$  backgrounds. Statistical uncertainties can have a non-negligible impact on  $\mu$ , in particular at high masses, where the most signal-like bins include only few background events. The analysis is however dominated by systematic uncertainties, in particular by the ones related to the  $t\bar{t} + \text{HF}$  modelling and the  $b$ -tagging calibration.

Multiple studies were performed to verify the validity of the fits. Possible biases in the signal

<sup>9</sup>Systematic uncertainties for which the shape is very different across regions or for which the  $+1\sigma$  and  $-1\sigma$  variations are not consistently increasing or decreasing the normalisation of the sample (for example, the  $+1\sigma$  variation correspond to a positive difference in normalisation in one region and a negative variation in another) can also show constrained error bars.

<sup>10</sup>With an ideal MC modelling, the pulls of  $k_{t\bar{t} + \geq 1b}$  and  $k_{t\bar{t} + \geq 1c}$  should be 1, while the pulls of the other nuisance parameters should be 0.

modelling were checked by performing background-only fits and removing the most signal-like bins, obtaining compatible pulls. Asimov fits were performed to validate the constraints on the error bars of the nuisance parameters. Furthermore, a good agreement between data and predicted yields is observed in the post-fit distributions of the BDT input variables (see Appendix C).

Uncertainty Source	$\Delta\mu(H_{200}^+)$	$\Delta\mu(H_{800}^+)$
$t\bar{t} + \geq 1b$ modelling	0.77	0.09
Jet flavour tagging	0.72	0.05
$H^+$ modelling	0.05	0.01
MC statistics	0.47	0.05
$t\bar{t} + \geq 1c$ modelling	0.49	0.03
Jet energy scale and resolution	0.54	0.03
$t\bar{t}$ + light modelling	0.61	0.02
Other background modelling	0.42	0.04
Jet-vertex assoc., pile-up modelling	0.22	0.01
Luminosity	0.27	0.01
Lepton, $E_T^{\text{miss}}$ , ID, isol., trigger	<0.001	<0.001
Total systematic uncertainty	1.81	0.13
$t\bar{t} + \geq 1b$ normalisation	0.65	0.03
$t\bar{t} + \geq 1c$ normalisation	0.30	0.01
Total statistical uncertainty	0.75	0.05
Total uncertainty	1.96	0.14

Table 6.7: Post-fit impact of the systematic uncertainties on the signal strength parameter, for the 200 and 800 GeV mass hypotheses. For simplicity, similar sources of systematic uncertainties are grouped together under the same category. The statistical component includes the contributions produced by the  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  normalisation factors.



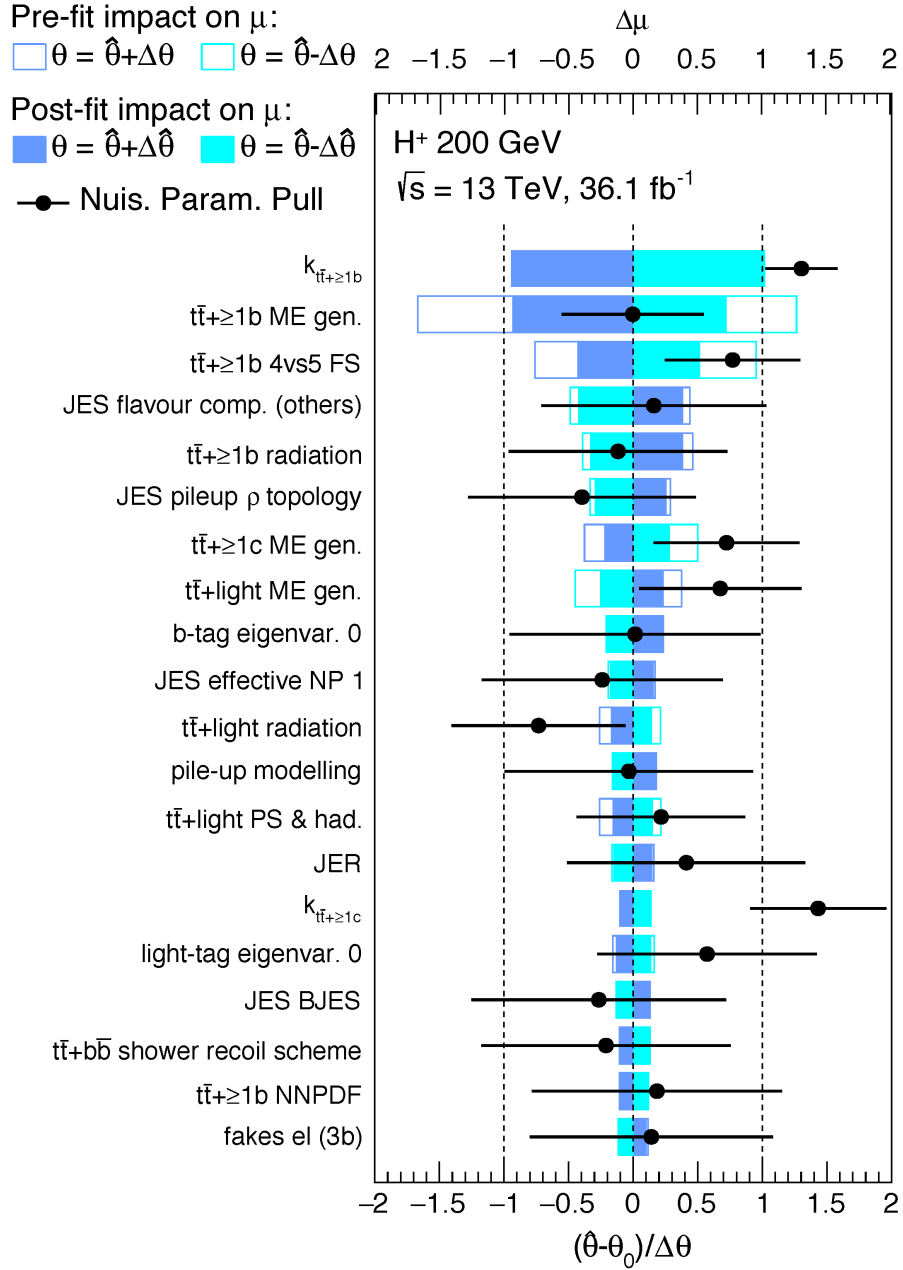


Figure 6.13: Ranking of the 20 systematic uncertainties with the largest impact on the signal strength for the fit performed with the 200 GeV signal hypothesis. The NPs are sorted by their post-fit impact  $\Delta\mu$ , from the highest to the lowest. The empty rectangles correspond to the pre-fit impact on  $\mu$ , while the filled rectangles correspond to the post-fit impact on  $\mu$ . The black dots represent the post-fit pulls of the systematic uncertainties, computed as the difference between the best-fit value and the pre-fit value divided by the pre-fit error  $((\hat{\theta} - \theta_0)/\Delta\theta)$ . The axis at the top corresponds to the scale of the impact, while the scale of the pulls is shown at the bottom. The dashed vertical lines represent the  $\pm 1\sigma$  variations for the NPs.

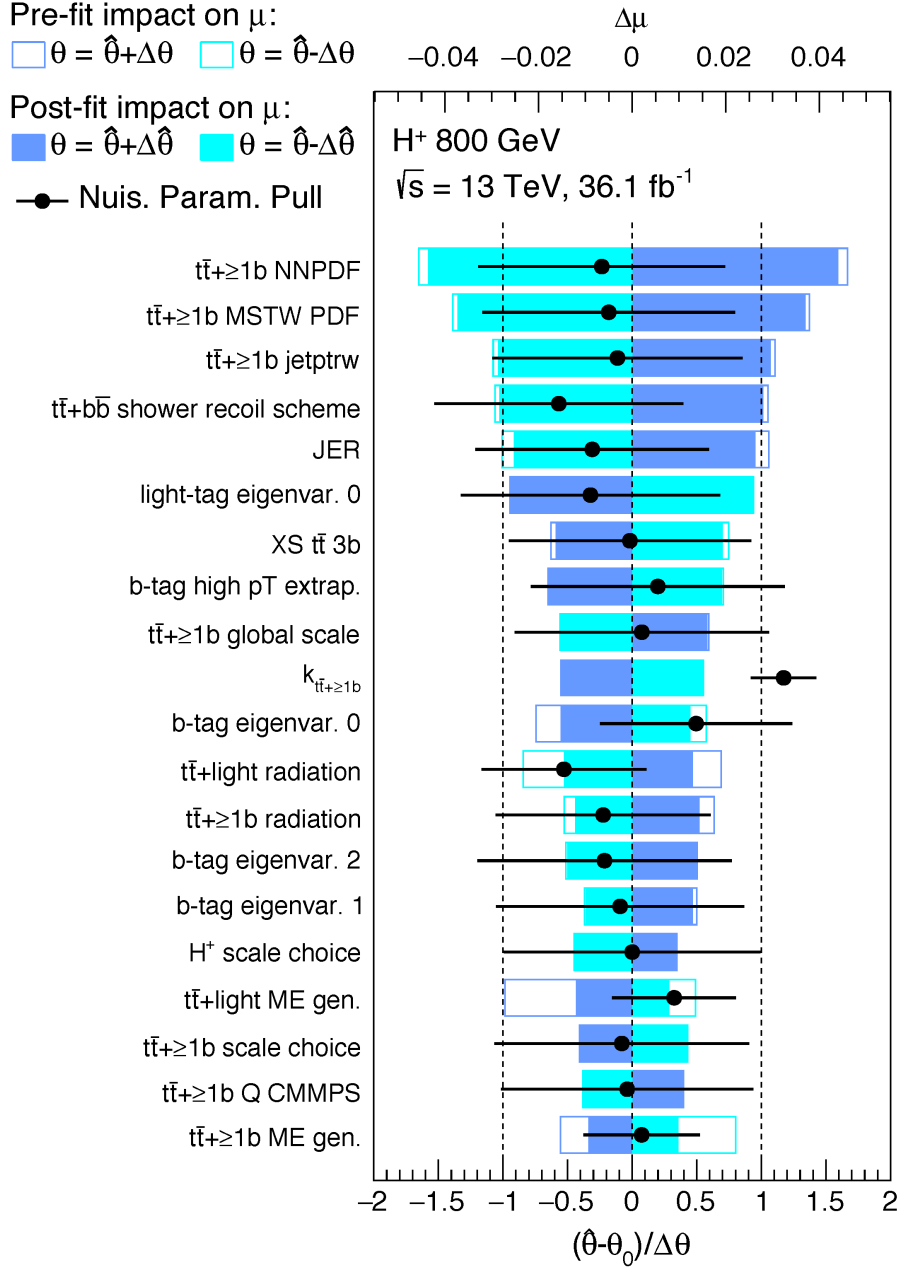
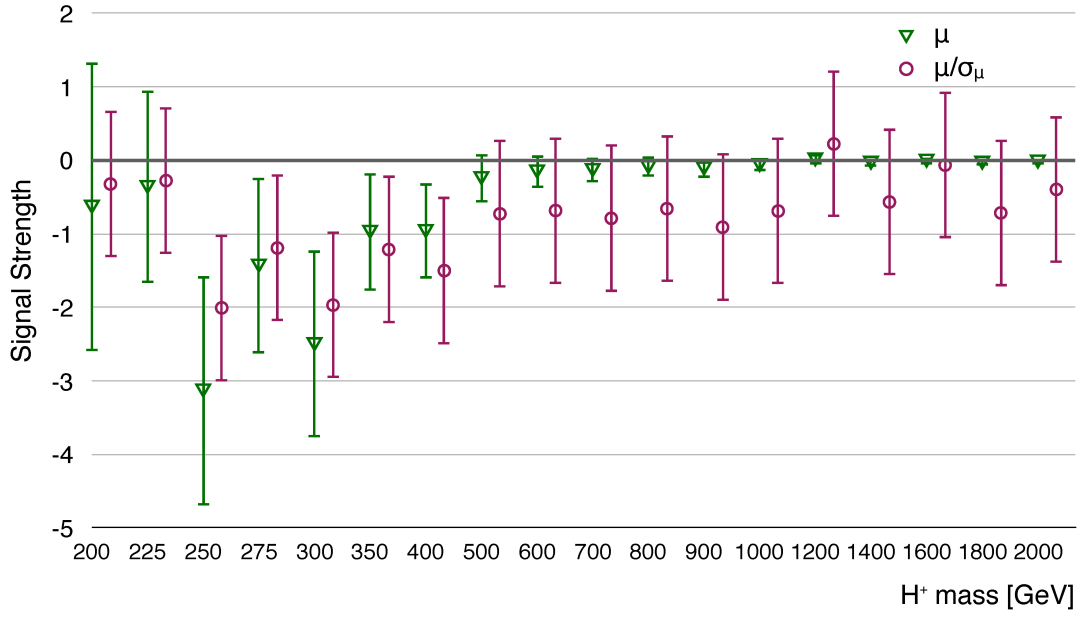
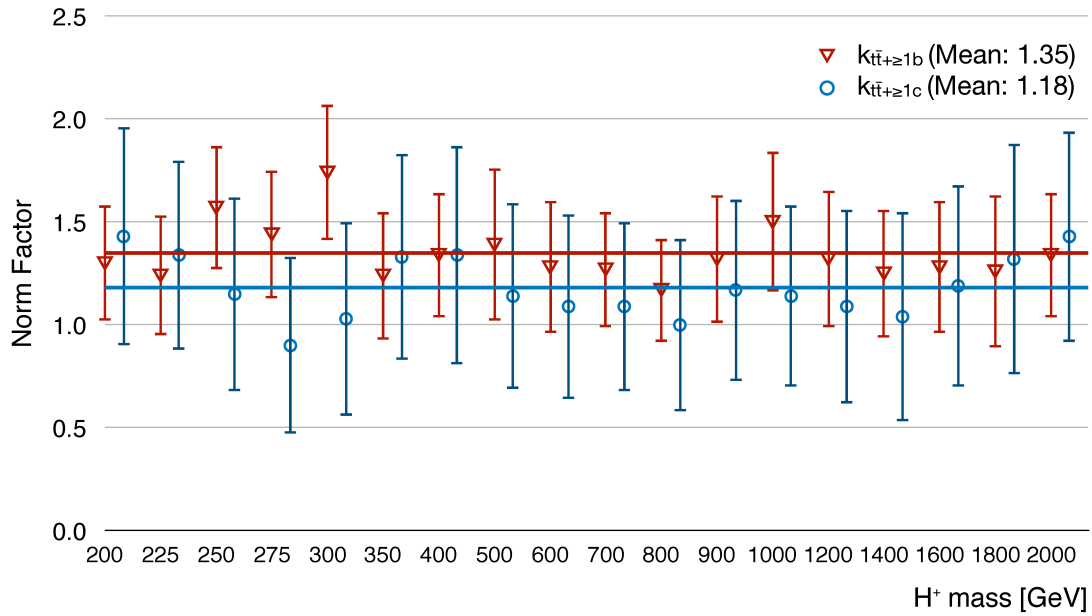


Figure 6.14: Ranking of the 20 systematic uncertainties with the largest impact on the signal strength for the fit performed with the 800 GeV signal hypothesis. The NPs are sorted by their post-fit impact  $\Delta\mu$ , from the highest to the lowest. The empty rectangles correspond to the pre-fit impact on  $\mu$ , while the filled rectangles correspond to the post-fit impact on  $\mu$ . The black dots represent the post-fit pulls of the systematic uncertainties, computed as the difference between the best-fit value and the pre-fit value divided by the pre-fit error  $(\hat{\theta} - \theta_0)/\Delta\theta$ . The axis at the top corresponds to the scale of the impact, while the scale of the pulls is shown at the bottom. The dashed vertical lines represent the  $\pm 1\sigma$  variations for the NPs.

Finally, the evolution of the  $t\bar{t} + \geq 1c$  and  $t\bar{t} + \geq 1b$  normalisation factors as well as the one of the signal strength, as a function of the  $H^+$  mass, are presented in Figure 6.15. The signal strength is reported twice, normalised to 1 pb and divided by its own error. The largest deviation with respect to the pure SM hypothesis is observed in the low-mass region, around 250 – 300 GeV. The deviation is however negative and does not represent evidence of a physical signal. It is more likely related to the high anti-correlation between  $k_{t\bar{t} + \geq 1b}$  and the signal strength (approximately -62% at 250 GeV and -66% at 300 GeV). The local  $p$ -value of the deviation is  $p = 2.28\%$  at 250 GeV and  $p = 2.51\%$  at 300 GeV. The  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  normalisation factors are fairly stable on the entire mass range, with average values of 1.35 and 1.18 respectively. The error on  $k_{t\bar{t} + \geq 1c}$  is generally larger than the error on  $k_{t\bar{t} + \geq 1b}$  due to the absence of control regions for the  $t\bar{t} + \geq 1c$  category.



(a) Signal strength.



(b)  $t\bar{t} + \geq 1b$  and  $t\bar{t} + \geq 1c$  normalisation factors.

Figure 6.15: Evolution of (a) the signal strength and (b) the  $t\bar{t} + \text{HF}$  normalisation factors as a function of the  $H^+$  mass. The horizontal lines indicate the average values of the normalisation factors and the expected value of the signal strength. The signal strength is reported normalised to 1 pb (in green) and divided by its own error (in violet).

## 6.5 Results

As the data do not show any significant excess with respect to the expected Standard Model background, the fit is used to set exclusion limits on the cross-section of the signal production and on the  $\tan\beta$  parameter of the MSSM. The limits are set at 95% CL, using the test-statistic of Eq. 2.6 and the  $\text{CL}_s$  method. The observed and expected upper limits on  $\sigma(pp \rightarrow H^+) \times \text{BR}(H^+ \rightarrow t\bar{b})$  are presented in Figure 6.16. The limits range from 3.4 pb at  $m_{H^+} = 200$  GeV to 0.072 pb at  $m_{H^+} = 2000$  GeV. The largest (negative) deviations with respect to the pure SM hypothesis are observed in the low-mass range, at 250 and 300 GeV. The observed limit is superimposed with the theoretical cross-sections predicted by the  $m_h^{\text{mod-}}$  model, for three different values of the  $\tan\beta$  parameter (0.5, 1 and 60). The limits between simulated masses are obtained by linear interpolation. This approach is justified by the fact that, despite being optimised for a single signal hypothesis, each signal-plus-background model is also sensitive to the adjacent mass points. This was proven by performing signal injection tests, as shown in Appendix C.

Limits are also provided for the  $\tan\beta$  parameter as a function of the  $H^+$  mass, in the  $m_h^{\text{mod-}}$  and hMSSM scenarios (presented in Figure 6.17). The strongest limits are set in the low-mass region, where both high ( $\geq 60$ ) and low ( $\leq 1$ ) values of  $\tan\beta$  can be excluded at 95% CL. Values of  $\tan\beta$  below 1.84 and above 39 are excluded at 250 GeV, while values below 1.76 and above 40 are excluded at 300 GeV (in the  $m_h^{\text{mod-}}$  scenario).

In comparison to the results obtained by ATLAS at 8 TeV (see Section 1.3.3), the analysis at 13 TeV includes an extended mass range and stricter constraints on the  $(m_{H^+}, \tan\beta)$  parameter space. In particular,  $\tan\beta \leq 1$  or larger can be excluded in the mass range between 200 and 600 GeV, while previous limits were below  $\tan\beta = 0.6$ . Limits at high values of  $\tan\beta$  are set at 13 TeV for the first time. Values of  $\tan\beta$  lower than 0.5 are also excluded for masses between 600 and 1000 GeV. This is particularly relevant in comparison to the constraints produced by flavour physics (Figure 1.9, Chapter 1), for which the only allowed region is at high masses. While this work confirms those indirect constraints, the high-mass region is the most promising for future discoveries.

### 6.5.1 Combination with the dilepton channel

In addition to searching for the  $H^+$  production in the single-lepton final state, the dilepton channel was also investigated by the ATLAS collaboration [13]. The signal and background modelling used by the two searches is the same, but the classification BDT was optimised independently for the kinematics of the dilepton decay. The dilepton analysis includes three additional signal regions (3j3b,  $\geq 4j3b$ ,  $\geq 4j\geq 4b$ ) and two control regions (3j2b,  $\geq 4j2b$ ). The combination between the two analyses allows the extraction of a more precise value for the signal strength. This is achieved by using a likelihood function built as the product of the ones of the two analyses. A large part of the nuisance parameters are therefore correlated between

the two channels. The 95% CL upper limits on the cross-section of the  $H^+$  production times the BR of the decay, obtained by combining the two searches, are shown in Figure 6.18. The inclusion of the dilepton channel allows to improve the expected limit of approximately 5%. The observed limits range from 2.9 pb at  $m_{H^+} = 200$  GeV to 0.070 pb at  $m_{H^+} = 2000$  GeV. The largest deviation with respect to the SM background hypothesis is still found in the low-mass region, at 300 GeV, with a  $p$ -value of 1.13%. The  $m_h^{\text{mod-}}$  and hMSSM interpretation plots are shown in Figure 6.19. Similarly to the single-lepton channel, values of  $\tan\beta$  above 60 and below 1 are excluded in the mass range 200 – 600 GeV. At 300 GeV, the  $m_h^{\text{mod-}}$  scenario excludes  $\tan\beta$  values below 1.9 and above 36.

### 6.5.2 Combination of the $t\bar{b}$ and $\tau^+\nu_\tau$ decays

A dedicated search for the  $H^+ \rightarrow \tau^+\nu_\tau$  decay at 13 TeV was also performed by the ATLAS collaboration [54]. As mentioned in Chapter 1, the  $\tau^+\nu_\tau$  and  $t\bar{b}$  channels have the largest branching ratios above 200 GeV. They are however sensitive to different areas of the MSSM parameter phase space. Figure 6.20 shows the expected and observed limits on  $\tan\beta$  obtained by superimposing the  $t\bar{b}$  and  $\tau^+\nu_\tau$  results. The limits obtained with the  $t\bar{b}$  final state are shown in black, while the ones obtained with the  $\tau^+\nu_\tau$  final state are drawn in red. Masses lower than 180 GeV were considered in the  $\tau^+\nu_\tau$  analysis, for which the full  $\tan\beta$  range is excluded at 95% CL. For masses between 200 and 600 GeV,  $\tan\beta$  below one is excluded by the  $t\bar{b}$  analysis, while values above 40 are excluded by the  $\tau^+\nu_\tau$  final state. At 300 GeV, the  $m_h^{\text{mod-}}$  scenario excludes  $\tan\beta$  values below 1.9 and above 21.

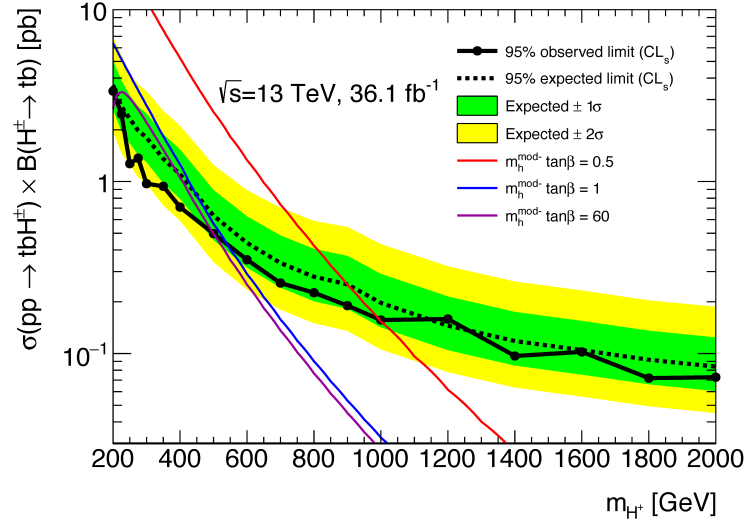


Figure 6.16: Expected and observed limits on the cross-section of the  $H^+ \rightarrow t\bar{b}$  production, as well as error bands for the 68% (in green) and 95% (in yellow) confidence intervals. The limits correspond to the single-lepton decay only. The theoretical predictions of the  $m_h^{\text{mod-}}$  benchmark scenario are shown for three values of  $\tan\beta$ .

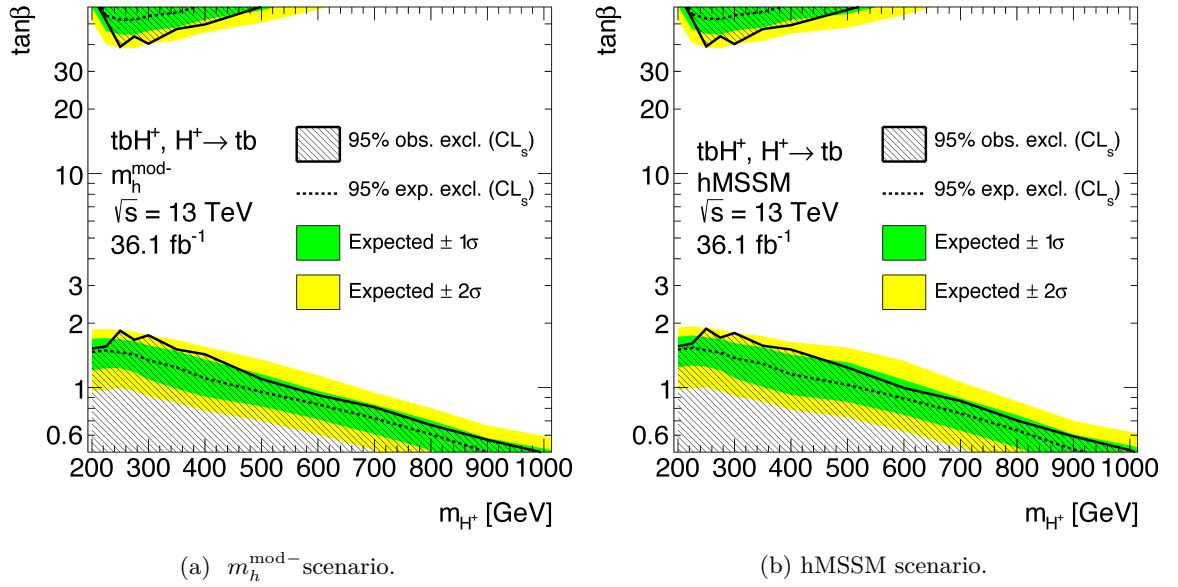


Figure 6.17: Expected and observed limits on  $\tan\beta$ , as well as error bands for the 68% (in green) and 95% (in yellow) confidence intervals, in the  $m_h^{\text{mod-}}$  scenario of (a) the MSSM and (b) the hMSSM model. The limits correspond to the single-lepton decay only. All values covered by the shaded area are excluded at 95% CL.

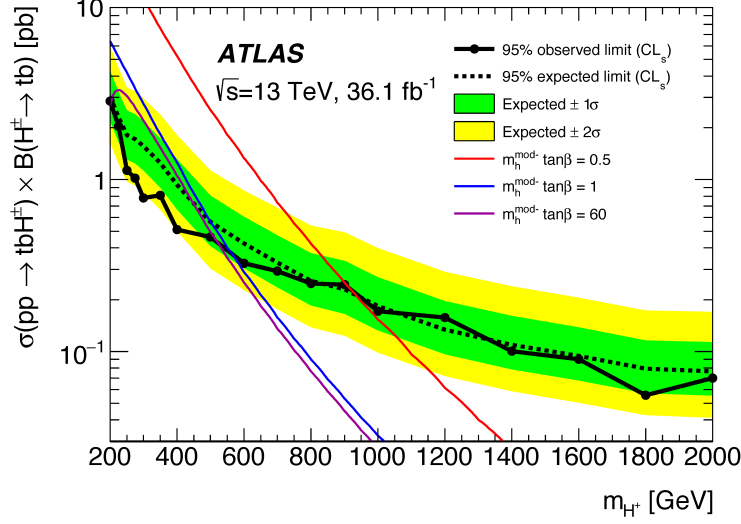


Figure 6.18: Expected and observed limits on the cross-section of the  $H^+ \rightarrow t\bar{b}$  production, as well as error bands for the 68% (in green) and 95% (in yellow) confidence intervals. The limits correspond to the combination between the single-lepton and dilepton decays. The theoretical predictions of the  $m_h^{\text{mod-}}$  benchmark scenario are shown for three values of  $\tan\beta$  [13].

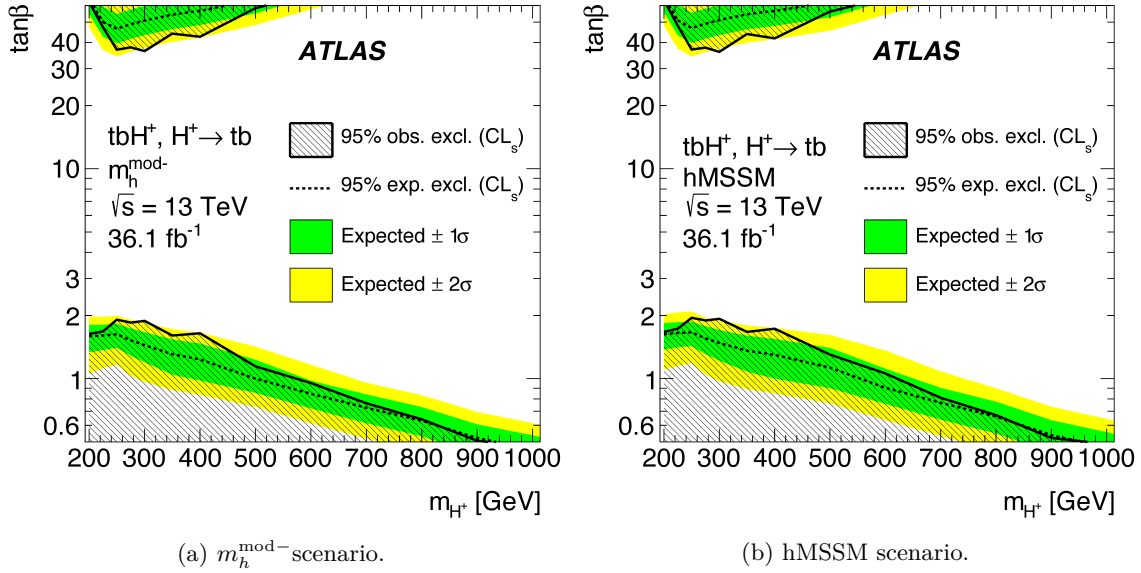


Figure 6.19: Expected and observed limits on  $\tan\beta$ , as well as error bands for the 68% (in green) and 95% (in yellow) confidence intervals, in the  $m_h^{\text{mod-}}$  scenario of (a) the MSSM and (b) the hMSSM model. The limits correspond to the combination between the single-lepton and dilepton decays [13]. All values covered by the shaded area are excluded at 95% CL.



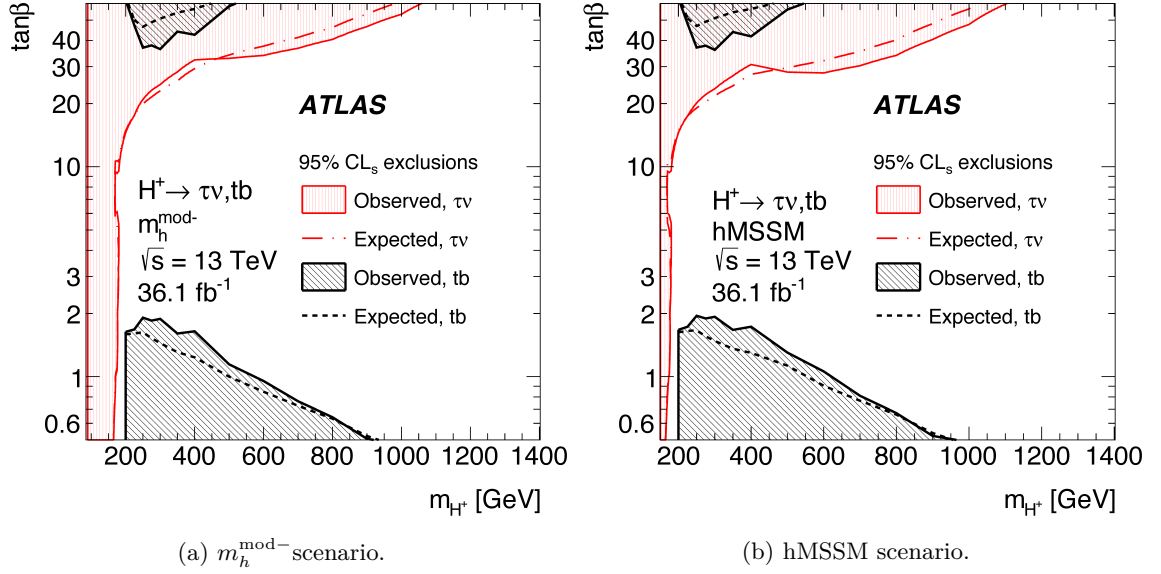


Figure 6.20: Expected and observed limits on  $\tan\beta$ , in the  $m_h^{\text{mod-}}$  scenario of (a) the MSSM and (b) the hMSSM model. The limits are a superposition of the results obtained with the  $t\bar{b}$  final state (combining the single-lepton and dilepton channels), and the results derived using the  $H^+ \rightarrow \tau^+\nu_\tau$  decay [13]. The limits from the  $t\bar{b}$  final state are shown in black, while the limits for the  $\tau^+\nu_\tau$  final state are shown in red. All values covered by the shaded area are excluded at 95% CL.



## Chapter 7

### Future prospects

The analysis strategy presented in the previous chapter was fixed before the start of the publication process and could not change afterwards, in order to proceed with the publication itself. The analysis development has however continued in view of the  $100 \text{ fb}^{-1}$  of  $pp$  collisions that the LHC is expected to deliver by the end of Run-2. This chapter proposes a set of refinements that could be implemented to improve the sensitivity and robustness of the analysis. The first part is dedicated to the low-mass region, with the aim of reducing the correlation between the signal and the  $t\bar{t} + \geq 1b$  background, while the second part focuses on discrimination techniques. The conclusion provides the expected limit obtained with the proposed changes and the projection of such a limit to the integrated luminosity of  $100 \text{ fb}^{-1}$ . All results are to be intended as preliminary.

#### 7.1 Low-mass range and correlation

As mentioned in Section 6.4, a large (anti)correlation is observed between  $\mu$  and  $k_{t\bar{t} + \geq 1b}$  (the  $t\bar{t} + \geq 1b$  normalisation), for the fits performed in the  $250 - 300 \text{ GeV}$  mass range. This correlation is imputable to the very similar kinematics of the two processes, which are almost indistinguishable by the BDT. The parameters  $\mu$  and  $k_{t\bar{t} + \geq 1b}$  are unconstrained in the likelihood function, therefore their correlation can drive the signal strength to large pulls, reducing the meaningfulness of the result. Possible ways of addressing the issue are illustrated below.

A first method to reduce the correlation between  $\mu$  and  $k_{t\bar{t} + \geq 1b}$  is to relax some of the constraints on the most relevant  $t\bar{t} + \text{jets}$  systematic uncertainties, making it preferable for the fit to use such uncertainties rather than the floating parameters. In fact, the pre-fit uncertainties are large enough to cover for the differences between data and MC in all regions, but the correlation between uncertainties, and the correlation of each uncertainty with itself across regions, lead to constrained post-fit errors. The choice of keeping each nuisance parameter 100% correlated across regions is generally justified by the assumption that their behaviour should be consistent in various areas of the phase space. However, uncertainties like those related to the choice of the ME generator, or to the choice of the PS, embed multiple systematic effects and can be of difficult physical interpretation. A dedicated decorrelation scheme

could be more effective for such nuisance parameters.

To better understand, one can consider the systematic variations depicted in Figure 7.1. The distributions correspond to the output of the BDT trained with the 300 GeV signal hypothesis, for the nominal  $t\bar{t} + \geq 1b$  background (in black) and the associated PS uncertainty (in red and blue), in two signal regions  $5j \geq 4b$  (a) and  $\geq 6j \geq 4b$  (b). As explained in Section 6.3.2, the uncertainty on the PS is associated to the choice of the generator employed for the simulation of the showering and hadronisation of partons, and it is obtained by comparing the default POWHEG + PYTHIA 8 sample to one produced with POWHEG + Herwig 7. Such uncertainty affects both the shape and normalisation of  $t\bar{t} + \geq 1b$  events, however the  $+1\sigma$  variation in  $5j \geq 4b$  corresponds to an overall positive (and small) variation in normalisation, while in the  $\geq 6j \geq 4b$  region the difference in normalisation is much larger and negative. The impact of the uncertainty on the shape of the  $t\bar{t} + \geq 1b$  distribution is visible from the ratio plot. It is almost negligible in the  $5j \geq 4b$  region, where the variation is flat within statistical uncertainties, and stronger in the  $\geq 6j \geq 4b$  region, where the ratio rises with the BDT output.

Similar differences have the potential of producing unexpected constraints. This is particularly true when the agreement between data and MC is improved by the pull of a NP in some regions, but not in others, or if the change in shape predicted by the systematic variation is in good agreement with data, but not the change in normalisation (and vice-versa). A possible solution is to decorrelate the corresponding systematic uncertainties between regions, and/or between shape and normalisation, allowing the fit to only exploit the "useful" part of the uncertainty and leave the rest unchanged. Promising results were obtained by decorrelating the PS, ME generator, radiation and 4vs5 flavour-scheme systematic uncertainties of the  $t\bar{t} + \text{jets}$  sample among the number of jets, i.e. between 5-jets and  $\geq 6$ -jets regions, and the 4vs5 FS, PS and ME systematic uncertainties of the  $t\bar{t} + \geq 1b$  category between shape and normalisation. The net result is a considerably lower correlation between  $\mu$  and  $k_{t\bar{t} + \geq 1b}$  with only a slightly decreased sensitivity. Table 7.1 provides a comparison between the values of the signal strength and the ( $\mu$  vs  $k_{t\bar{t} + \geq 1b}$ ) correlations, obtained by performing Asimov fits with the default model or implementing the new decorrelation scheme. The tests were performed using BDTs trained over the set of 12 variables described in Section 6.2.1, without exploiting the likelihood discriminant. The results are presented for  $m_{H^+} \leq 350$  GeV, which is the most sensitive region to such correlations. The fact that the result is more conservative, together with the reduced risk of an unwanted pull of the signal strength, make this decorrelation scheme preferable to the default model.

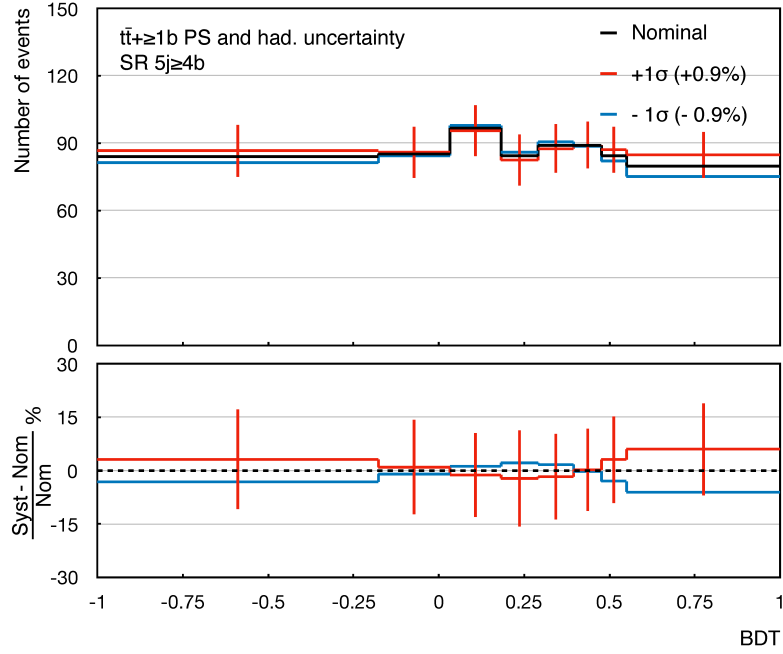
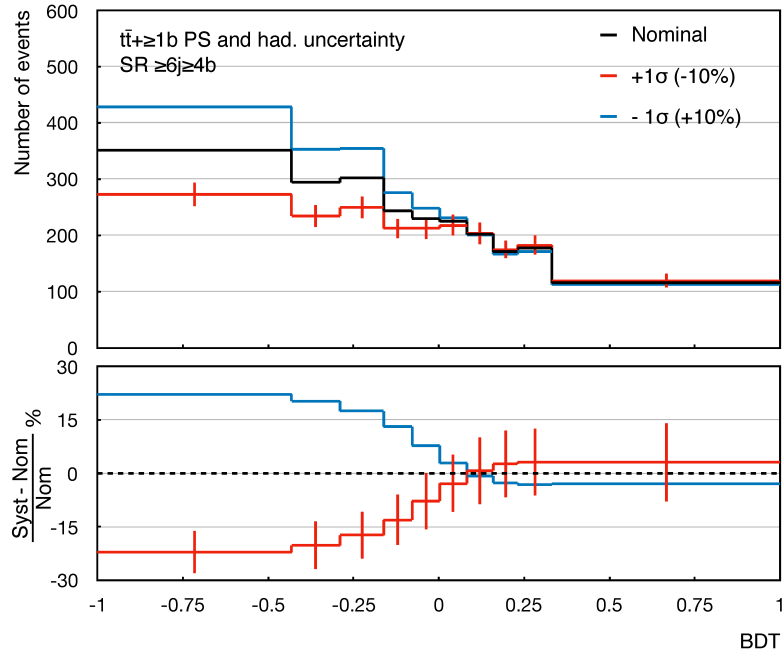

 (a) SR  $5j \geq 4b$ .

 (b) SR  $\geq 6j \geq 4b$ .

Figure 7.1: The  $t\bar{t} + \geq 1b$  PS and hadronisation systematic uncertainty for the distribution of the BDT trained with the 200 GeV signal hypothesis, in (a) the  $5j \geq 4b$  region and (b) the  $\geq 6j \geq 4b$  region. The red and blue lines represent the  $+1\sigma$  and  $-1\sigma$  systematic variations (after smoothing and symmetrisation). The bottom panel corresponds to the ratio between the systematic variations and the nominal distribution. Given the symmetry between the two sides of the systematic uncertainty, the statistical error is only reported for the  $+1\sigma$  variation.

default model			new model	
mass	signal strength	$\mu$ vs $k_{t\bar{t}+\geq 1b}$ correlation	signal strength	$\mu$ vs $k_{t\bar{t}+\geq 1b}$ correlation
200 GeV	$0 \pm 2.10$	-48	$0 \pm 2.16$	-19
225 GeV	$0 \pm 1.91$	-56	$0 \pm 1.96$	-20
250 GeV	$0 \pm 1.36$	-43	$0 \pm 1.39$	-25
275 GeV	$0 \pm 1.22$	-73	$0 \pm 1.29$	-49
300 GeV	$0 \pm 1.12$	-68	$0 \pm 1.20$	-49
350 GeV	$0 \pm 0.74$	-47	$0 \pm 0.81$	-47

Table 7.1: Signal strengths and ( $\mu$  vs  $k_{t\bar{t}+\geq 1b}$ ) correlations obtained with the default  $t\bar{t}$  + jets modelling and with a new model in which the PS, ME generator, 4vs5 flavour scheme and radiation  $t\bar{t}$  + jets systematics are decorrelated among the number of jets and the 4vs5 FS, PS and ME systematic uncertainties for the  $t\bar{t} + \geq 1b$  category are also decorrelated between shape and normalisation.

A second way of improving the ability of the fit to distinguish between the two processes is to modify the analysis selection in order to create regions that are heavily enriched in  $t\bar{t} + \geq 1b$  events. The  $H^+$  and  $t\bar{t} + \geq 1b$  distributions are in fact similar but not identical, therefore a high purity can enhance the control over both components. A large part of the  $5j \geq 4b$  and  $\geq 6j \geq 4b$  regions is already composed by  $t\bar{t} + \geq 1b$  events, however better results can be achieved by modifying the  $b$ -tagging efficiency used for the region definition. Figure 7.2 shows the background composition of the six regions used in the analysis when they are tagged using the 60%, 70% or 77%  $b$ -tagging working point.

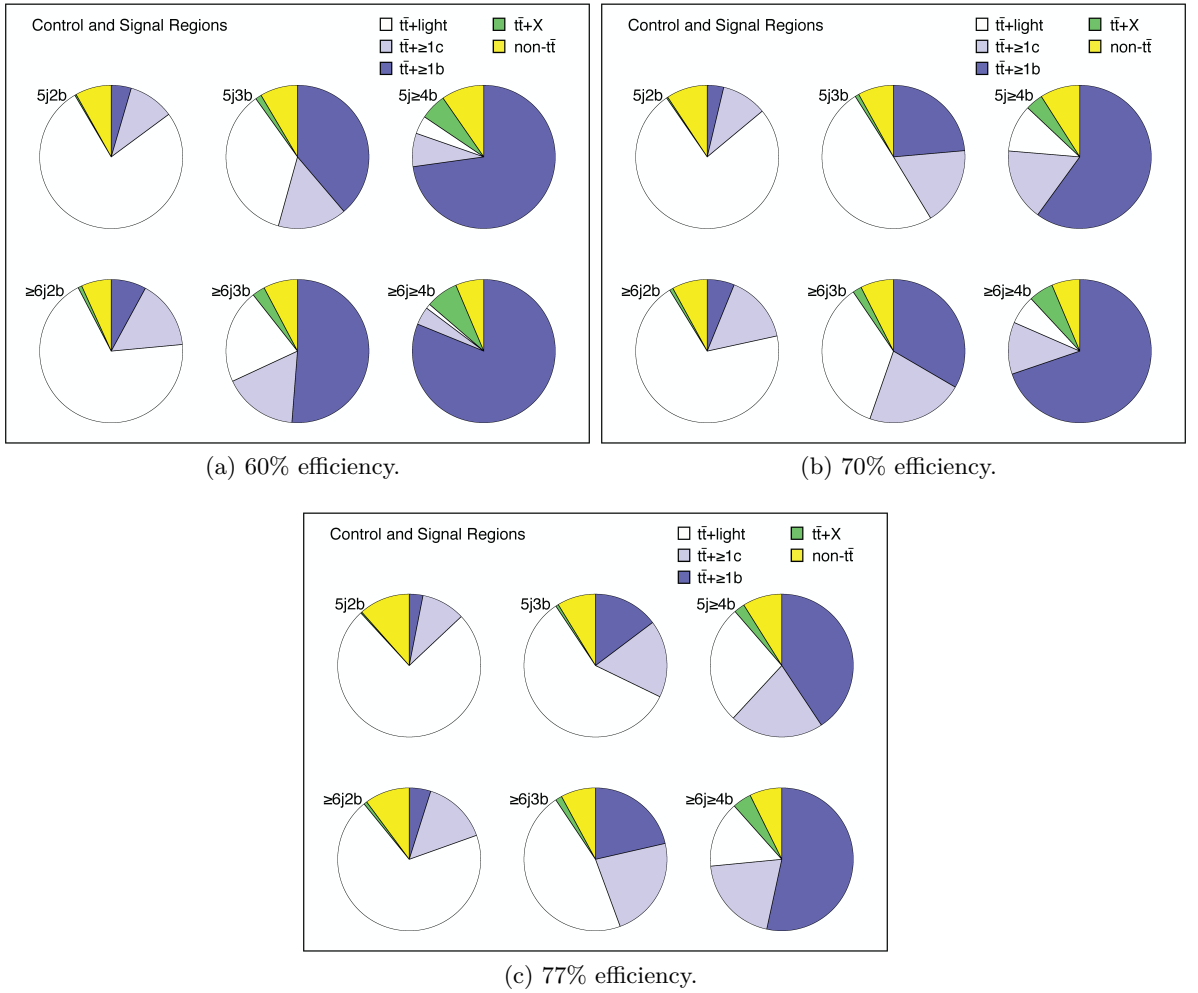


Figure 7.2: Background composition of control and signal regions tagged with different efficiencies: (a) 60%, (b) 70% and (c) 77%.

The larger the  $b$ -tagging efficiency, the larger is the mistag probability. At 77% efficiency, there is a large amount of  $t\bar{t}$  + light and  $t\bar{t} + \geq 1c$  events in the  $\geq 4b$  regions, while they are expected to be mainly composed by  $t\bar{t} + \geq 1b$  events. The  $t\bar{t}$  + light and  $t\bar{t} + \geq 1c$  categories are consistently reduced at 70% efficiency and they almost disappear at 60%. Table 7.2 provides a comparison between the values of the signal strength and the ( $\mu$  vs  $k_{t\bar{t} + \geq 1b}$ ) correlation obtained with Asimov fits performed on regions tagged at 60%, 70% or 77% efficiency (dedicated BDTs were trained for each of the different  $b$ -tagging working points). The results show how the increased purity of the 60% regions allows to further reduce the correlation between  $\mu$  and  $k_{t\bar{t} + \geq 1b}$ , with a positive impact on the sensitivity of the analysis. Using a stricter efficiency at higher masses is disadvantageous because of the reduced statistics. At 77% efficiency, the total amount of  $t\bar{t} + \geq 1b$  events in the  $\geq 5j \geq 3b$  region is approximately 27000, but only 12000 events pass the selection at 60%. This has a small impact at low masses, where the uncertainty on the signal strength is dominated by systematic errors, but it has a strong effect at high masses, where the most signal-like bins of the BDT distributions are populated by few background events and the statistical error plays a more relevant role.

60% efficiency			70% efficiency		77% efficiency	
mass	signal strength	$\mu$ vs $k_{t\bar{t} + \geq 1b}$ correlation	signal strength	$\mu$ vs $k_{t\bar{t} + \geq 1b}$ correlation	signal strength	$\mu$ vs $k_{t\bar{t} + \geq 1b}$ correlation
200 GeV	$0 \pm 1.74$	-14	$0 \pm 2.16$	-19	$0 \pm 2.25$	-29
225 GeV	$0 \pm 1.37$	-21	$0 \pm 1.96$	-20	$0 \pm 1.85$	-34
250 GeV	$0 \pm 1.15$	-27	$0 \pm 1.39$	-25	$0 \pm 1.80$	-40
275 GeV	$0 \pm 1.01$	-40	$0 \pm 1.29$	-49	$0 \pm 1.43$	-45
300 GeV	$0 \pm 1.02$	-39	$0 \pm 1.20$	-49	$0 \pm 1.40$	-60
350 GeV	$0 \pm 0.70$	-37	$0 \pm 0.81$	-47	$0 \pm 0.90$	-50

Table 7.2: Signal strengths and ( $\mu$  vs  $k_{t\bar{t} + \geq 1b}$ ) correlations obtained with control and signal regions tagged at 60%, 70% or 77% efficiency.

## 7.2 Improving the discrimination power

This section is dedicated to improving the signal-background separation. With respect to the BDT of Section 6.2.2, which was trained separately in the  $5j \geq 3b$ ,  $\geq 6j \geq 3b$  and  $\geq 6j \geq 4b$  regions, the following tests were performed using BDTs trained inclusively in the  $\geq 5j \geq 3b$  region. Training in a single region reduces the risk of overtraining while quickening the analysis process. The improvement of each version is estimated by comparing the corresponding AUC with the AUC obtained with a reference version (named v0), which was trained on the 12 input variables described in Section 6.2.1 (with the exclusion of the LHD and with the addition of the number of jets and the number of  $b$ -jets variables). The AUC values are computed inclusively in the  $\geq 5j \geq 3b$  region. The training is performed with TMVA, using



the full set of SM backgrounds. The BDTs are built using 300 trees, with a maximum depth of five.

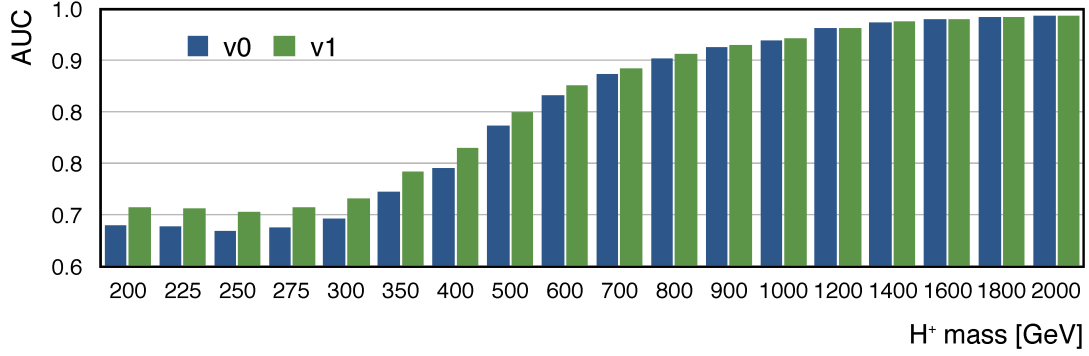
### 7.2.1 The likelihood discriminant

As already mentioned, the BDTs trained in the low-mass region can profit from the inclusion of the likelihood discriminant. Unfortunately, the Pseudo-Continuous  $b$ -tagging uncertainties are not calibrated for high- $p_T$  jets. Therefore, the version of the LHD described in Section 6.2.1 cannot be used for signal hypotheses with  $m_{H^+} > 300$  GeV, which include a large number of high- $p_T$  objects. The PCBT uncertainties are however only necessary because the  $P_b$  and  $P_l$  probabilities of the  $P^{\text{btag}}$  PDF (Equations 6.6 and 6.7 of Section 6.2.1) can take on five different values depending on the output of the  $b$ -tagging BDT. Reducing the number of bins from five to two allows to use the Fixed-WP uncertainties, and the LHD, over the entire mass range. This means that each jet is either  $b$ -tagged or not and the  $P_b$  and  $P_l$  probabilities can be parametrised as step-functions (at 70% efficiency, the distribution of  $P_b$  and  $P_l$  would be the one of Table 7.3). Such a change will become unnecessary as soon as the systematic uncertainty on the high- $p_T$  extrapolation will be included in the PCBT set by the responsible ATLAS working group, however the change in performance with respect to the PCBT version of the likelihood discriminant is negligible. Figure 7.3 (a) provides a comparison between the performances of the default BDT (v0) and the one trained with the LHD (v1). The relative improvement of v1 with respect to v0, estimated as the difference between the AUCs of the two versions divided by  $1 - \text{AUC}_{v0}$ , is shown in Figure 7.3 (b)<sup>1</sup>. The performance of the BDT is considerably improved by the addition of the likelihood discriminant, in particular at low and intermediate masses.

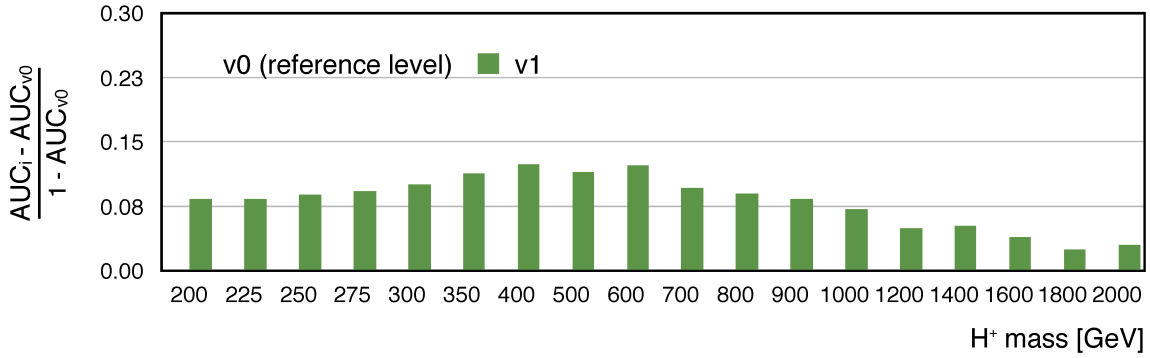
Efficiency	1 – 0.70	0.70 – 0.0
$P_b$	0.3	0.7
$P_l$	0.997	0.003

Table 7.3: The flavour-tagging probabilities for light and  $b$ -jets at 70% efficiency.

<sup>1</sup>Dividing by  $1 - \text{AUC}_{v0}$  instead than by  $\text{AUC}_{v0}$  allows to compare the improvement to the residual number of misclassified events. This provides a sensible estimate of the change in performance even for mass points where the AUC values are very close to 1.



(a) AUCs.



(b) Relative improvement.

Figure 7.3: Performance of the classification BDTs. Sub-figure (a) present the AUC values obtained with versions v0 and v1 of the BDT, while sub-figure (b) shows the relative improvement provided by version v1 with respect to v0. Version v0 was trained on the 12 input variables described in Section 6.2.1 with the addition of the number of jets and the number of  $b$ -jets variables. The LHD is also included in version v1. The AUCs are computed comparing the BDT distribution of the signal with the one produced by the full set of SM backgrounds, in the  $\geq 5j \geq 3b$  region.

### 7.2.2 Reconstruction BDTs

The separation power of the classification BDT can be further improved by extending the set of input variables. The following presents the results obtained by performing a two-layers BDT training, using a reconstruction BDT as input variable to the classification BDT. Reconstruction (Reco) BDTs aim to use the jets, leptons and missing energy of the events to reconstruct the original particles produced by the interaction, as the charged Higgs boson. They have been recently used by other ATLAS analyses, with promising results [135]. Reconstruction BDTs are not trained on event variables but on properties of object combinations, in order to match the reconstructed jets to the partons of the decay. While the method could appear to provide similar information to the likelihood discriminant, the reconstruction BDT has the advantage of being able to exploit the correlation between input variables.

The final state of the  $\bar{t}bH^+$  decay is composed of four  $b$ -quarks (two from the  $H^+$  decay, one from the anti-top quark and one directly produced by the hard scattering), two light quarks produced by one of the  $W$  bosons and one lepton and neutrino produced by the other  $W$  boson (the neutrino is reconstructed with the same technique used for the likelihood discriminant, see Section 6.2.1). The Reco-BDT is trained to distinguish between two classes: the "signal" class, made of all jet combinations produced by  $\bar{t}bH^+$  events in which at least four partons out of the three  $b$ -quarks (produced by the  $H^+$  decay and the anti-top decay) and two light quarks (produced by the  $W$  boson decay) are matched to reconstructed jets, and the "background" class, which corresponds to all possible jet combinations produced by  $t\bar{t} + \geq 1b$  events.

For each combination, regardless of the correctness of the assignment, the jets are used to build a set of kinematic variables<sup>2</sup>:

- the masses of the hadronic  $W$  boson, hadronic and leptonic top quarks, hadronic and leptonic charged Higgs bosons (hadronic and leptonic stand for "hadronically decaying" and "leptonically decaying". For example, the hadronic  $H^+$  is reconstructed by assembling two light-jets and two  $b$ -jets, while the leptonic  $H^+$  is reconstructed using the lepton, the neutrino and two  $b$ -jets);
- the  $p_T$  of the hadronic and leptonic top quarks and the  $p_T$  of the leptonic  $H^+$ ;
- the sum of the transverse momenta of all jets used to reconstruct the hadronic and leptonic charged Higgs bosons;
- the angular separation between the lepton and the leptonic  $W$  boson, the angular separation between the  $b$ -jet (directly) produced by the  $H^+$  decay and the leptonic  $W$  boson, the angular separation between the  $b$ -jet (directly) produced by the  $H^+$  decay and the  $b$ -jet produced by the top decay and the angular separation between the  $b$ -jet

<sup>2</sup>The 14 variables were chosen in a previous step, selecting the ones with the highest separation power from a larger set.

(directly) produced by the  $H^+$  decay and the  $b$ -jet produced by the anti-top decay.

The XGBoost package [159] is then used to train one BDT for each of the signal hypotheses, with the aforementioned variables as inputs. Once the discriminants are trained, they are applied to all data and MC events in order to identify the "correct" jet combination, i.e. the one with the highest BDT output. This BDT output is passed as an input variable to the classification BDT, together with the difference between the target  $H^+$  mass and the mass of the  $H^+$  built with the jet combination associated to the highest score (the variable is later referred to as  $MD_{H^+}$ ). Given that both the hadronic and the leptonic masses are defined for each combination, the "correct"  $H^+$  mass is selected as the one closer to the target mass.

The efficiency of the reconstruction is shown in Table 7.4. It is calculated as the number of  $H^+$  events in which the best combination corresponds to at least four properly matched jets, divided by the total number of events in which at least four matchable jets are available. Similarly to the LHD, the worst performance is observed for signal of masses around 250 – 275 GeV. The performance also decreases at very high masses due to the increasing number of possible combinations.

mass [GeV]	200	225	250	275	300	350	400	500	600
efficiency	0.60	0.49	0.36	0.36	0.40	0.46	0.50	0.55	0.59

mass [GeV]	700	800	900	1000	1200	1400	1600	1800	2000
efficiency	0.61	0.63	0.63	0.64	0.65	0.65	0.64	0.63	0.62

Table 7.4: Efficiency of the reconstruction BDT across the 200 – 2000 GeV mass range. The error on the efficiency is not reported because it is lower than 0.5% at all mass points.

Figure 7.4 provides a comparison between the performances of the default BDT (v0), the one trained including the LHD (v1), and a version trained with the addition of the LHD, the Reco-BDT output and  $MD_{H^+}$  (v2). The output of the reconstruction BDT is always ranked between the top three variables for importance by TMVA, independently from the mass point. With respect to version v1, slightly larger AUC values are observed over the entire mass range, with the largest (relative) improvement at high masses. It is however clear that the performance of the classification BDT is reaching a plateau and that the addition of more input variables has little impact on the overall separation power. Tests performed with larger sets of input variables are available in Appendix D.

Figure 7.5 and Figure 7.6 show the distributions of the Reco-BDT output,  $MD_{H^+}$  and some of the input variables, created by selecting the object combination with the highest score, for the training performed with the 200 and 800 GeV signal hypotheses. Figure 7.5 (a) and (b) correspond to the masses of the hadronic and leptonic top quarks, which, as expected, show a negligible separation between signal and background. A larger separation is visible in the distribution of the Reco-BDT output,  $MD_{H^+}$  and the masses of the hadronic and leptonic

charged Higgs bosons (Figure 7.5 (c) and Figure 7.6 (a,b,c)). The separation is also larger for higher mass hypotheses.

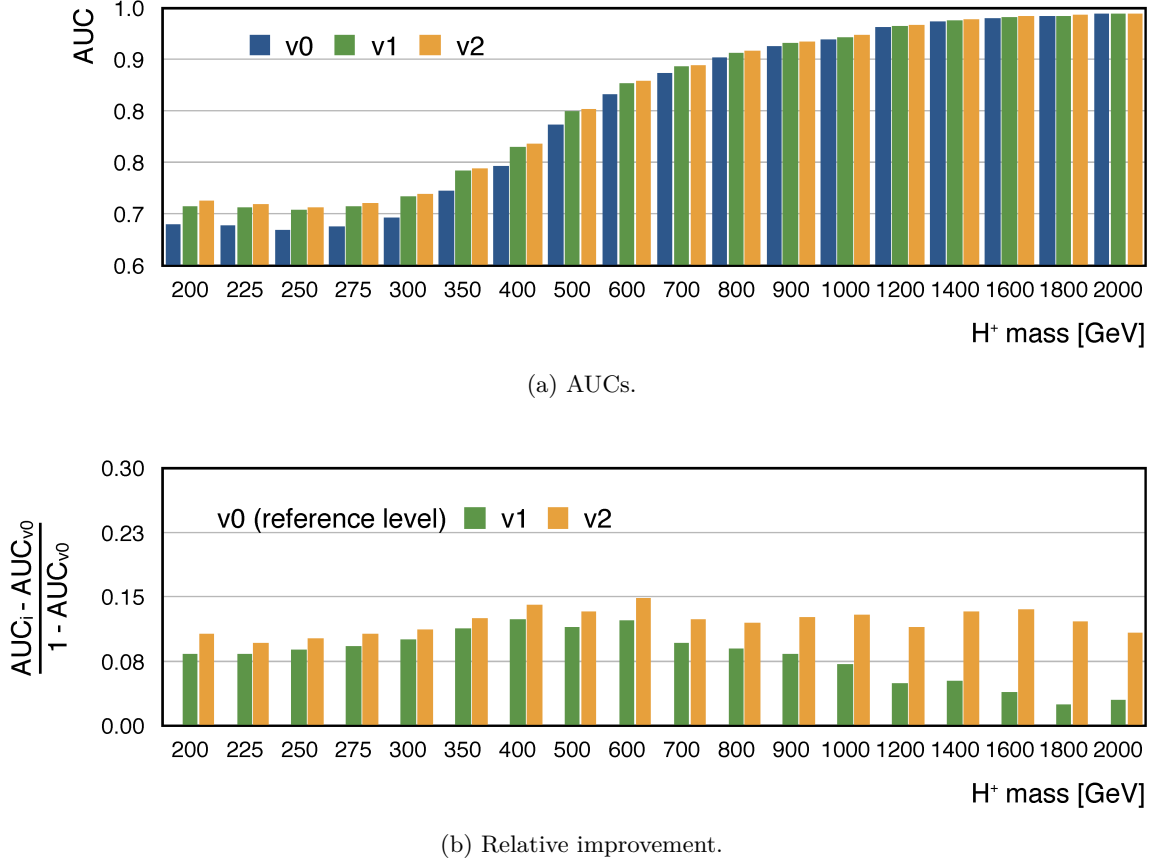


Figure 7.4: Performance of the classification BDTs. Sub-figure (a) presents the AUC values obtained with versions v0, v1 and v2 of the BDT, while sub-figure (b) shows the relative improvement provided by versions v1 and v2 with respect to v0. Version v0 was trained on the 12 input variables described in Section 6.2.1 with the addition of the number of jets and the number of  $b$ -jets, v1 includes the LHD, v2 includes the LHD discriminant and the variables associated to the reconstruction BDT. The AUCs are computed comparing the BDT distribution of the signal with the one produced by the full set of SM backgrounds, in the  $\geq 5j \geq 3b$  region.

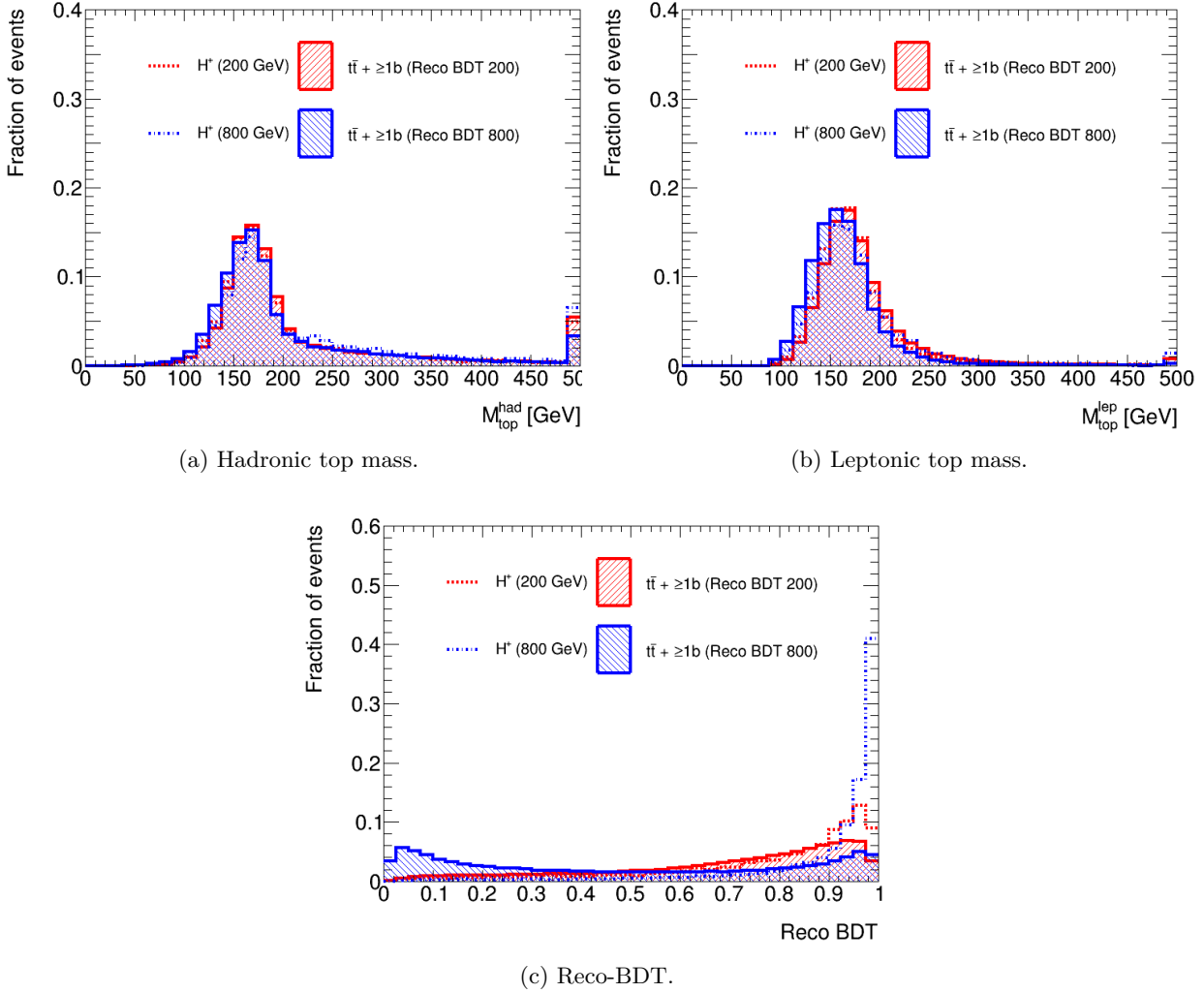


Figure 7.5: Distributions of (a) the mass of the hadronic top quark, (b) the mass of the leptonic top quark and (c) the Reco-BDT output for the object combination with the highest Reco-BDT output. The signal and background distributions are reported as dotted and solid lines respectively. The red distributions correspond to the training performed at 200 GeV, the blue distributions to the one at 800 GeV. The distributions are normalised to unity.

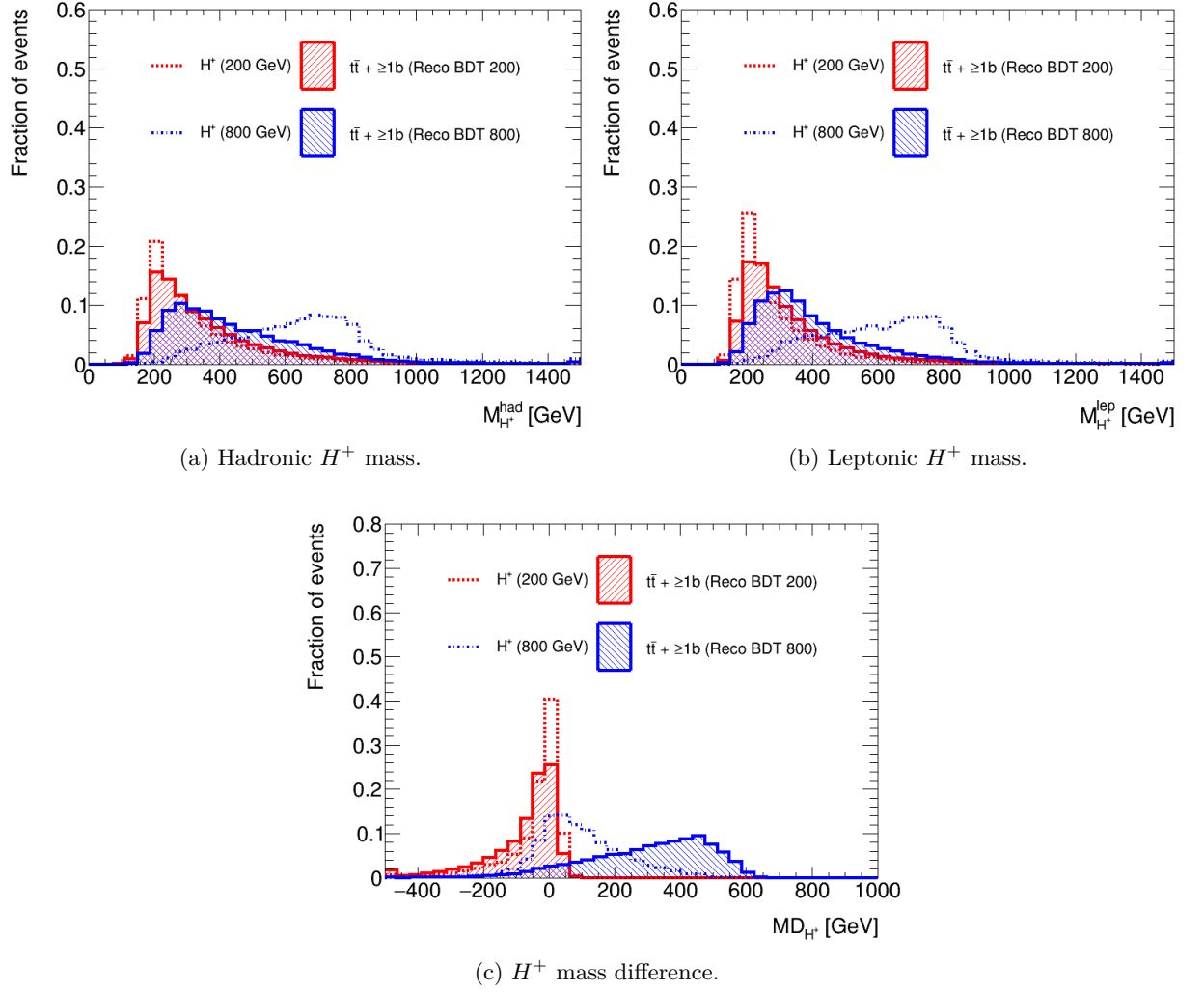


Figure 7.6: Distributions of (a) the mass of the hadronic  $H^+$ , (b) the mass of the leptonic  $H^+$  and (c)  $MD_{H^+}$  for the object combination with the highest Reco-BDT output. The signal and background distributions are reported as dotted and solid lines respectively. The red distributions correspond to the training performed at 200 GeV, the blue distributions to the one at 800 GeV. The distributions are normalised to unity.

### 7.2.3 Expected limits

Figure 7.7 offers a comparison between the expected (Asimov) limit obtained with the default setup, described in the previous chapter, and the limit obtained with version v2 of the classification BDT<sup>3</sup>. The latter also includes the new decorrelation scheme and the usage of 60%  $b$ -tagged regions in the 200 – 350 GeV mass range. The default setup already included the likelihood discriminant below 350 GeV, therefore the differences in that region are limited to the addition of the reconstruction BDT and to the usage of a different  $b$ -tagging working point. The largest improvement is observed in the mid-mass range, between 350 and 600 GeV, where the new expected limit is approximately 20% lower than the default version. The improved limit comes together with a lower error on the signal strength. For an hypothetical signal with a cross section set to the upper value of the default expected limit (3.2 pb at 200 GeV and 0.26 pb at 800 GeV), the relative error on  $\mu$  would decrease from 54% to 48% at 200 GeV and from 49% to 42% at 800 GeV. Observed limits are not shown because the new strategy must be further validated and optimised before being officially approved by ATLAS. A projection of the "revised" expected limit to the integrated luminosity of 100 fb<sup>-1</sup> is also provided, as a dashed purple line. The increased luminosity is expected to improve the limit of approximately 15 – 20%. This corresponds to a relative error on the signal strength of 41% at 200 GeV and 35% at 800 GeV (with the same assumption on the cross section as above). The reduced correlation between the signal strength and the  $t\bar{t} + \geq 1b$  normalisation is also expected to have a positive impact on the analysis. Further benefits could come from discriminants built to minimise the impact of the systematic uncertainties (see Appendix E) or from advancements in the theoretical modelling of the backgrounds.

---

<sup>3</sup>The "default" limit has been recreated as well, so as to guarantee that the exact same code was used for all versions.



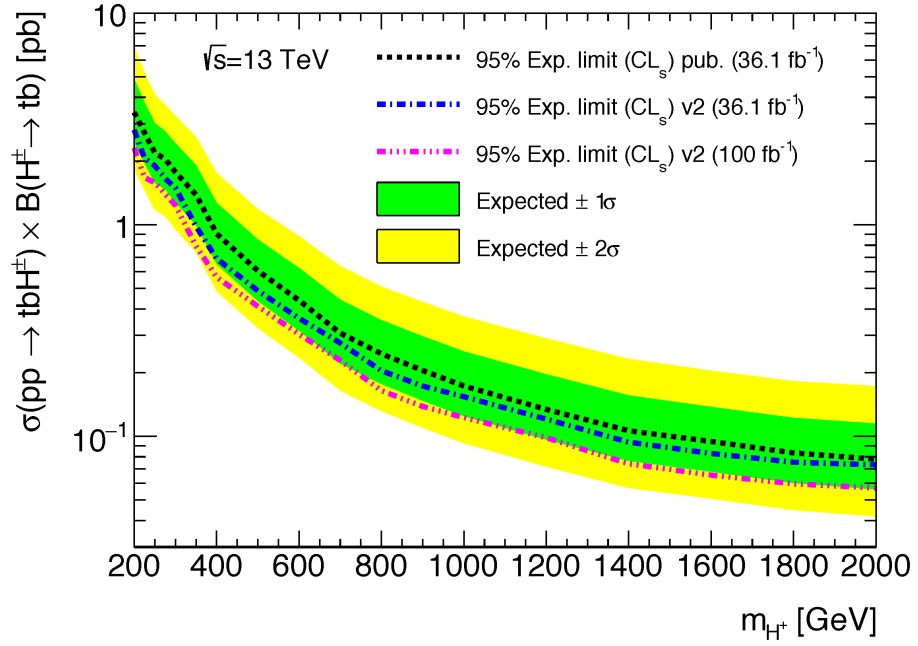


Figure 7.7: Expected limits for the production of  $H^+ \rightarrow t\bar{b}$  in association with a top quark and a bottom quark (single-lepton channel), as well as error bands for the 68% (in green) and 95% (in yellow) confidence intervals. The black dashed line corresponds to the limit for the single-lepton channel obtained with the setup of Chapter 6. The limit obtained with the techniques described in this chapter is reported twice: as a blue dashed line for an integrated luminosity of  $36.1 \text{ fb}^{-1}$  and as a purple dashed line for  $100 \text{ fb}^{-1}$ .



## Chapter 8

### Summary and conclusions

In this thesis, the search for charged Higgs bosons produced in association with top and bottom quarks has been presented. The search was performed using  $36.1 \text{ fb}^{-1}$  of  $pp$  collisions collected by ATLAS during 2015 and 2016, at a centre-of-mass energy of 13 TeV. Signal hypotheses with masses ranging from 200 to 2000 GeV were investigated, targeting the  $H^+$  decay into a top-bottom pair.

The thesis focused on the single-lepton decay, categorising the events into regions by using the multiplicity of jets and  $b$ -jets. The signal strength associated to the  $H^+$  cross-section was measured by performing a profile likelihood fit, separately for each of the signal hypotheses. The sensitivity of the analysis was enhanced by fitting the distribution of a boosted decision tree trained on multiple kinematic variables and optimised independently for each mass point. No relevant excess was found with respect to the SM background, therefore 95% CL upper limits were put on the cross-section of the production times the branching ratio of the decay ( $H^+ \rightarrow t\bar{b}$ ). The limits range from  $\sigma \times \text{BR} = 3.4 \text{ pb}$  at 200 GeV to  $\sigma \times \text{BR} = 0.073 \text{ pb}$  at 2000 GeV. Limits were also set on the  $\tan\beta$  parameter in the context of the  $m_h^{\text{mod-}}$  and hMSSM scenarios of the MSSM. The most stringent results were obtained at low masses, where both high and low  $\tan\beta$  values can be excluded, e.g.  $\tan\beta < 1.76$  and  $\tan\beta > 40$  at 300 GeV in the  $m_h^{\text{mod-}}$  scenario.

Stronger limits were obtained by combining the single-lepton and dilepton channels, and by superimposing the resulting limits with the ones produced by the  $H^+ \rightarrow \tau^+ \nu_\tau$  analysis. Values of  $\tan\beta$  below 1 and above 40 are excluded in the 200 – 600 GeV mass range, with the strongest constraints at 300 GeV, where values below 1.9 and above 21 are excluded.

With respect to previous ATLAS searches for  $H^+ \rightarrow t\bar{b}$ , the results of this thesis provide more stringent limits on the cross-section of the  $H^+$  production and stricter constraints on the  $(m_{H^+}, \tan\beta)$  parameter space. Moreover, limits were set for the first time on high values of  $\tan\beta$  and in the mass range above 600 GeV.

Moving forward, the LHC will provide approximately  $100 \text{ fb}^{-1}$  of  $pp$  collisions by the end of Run-2. The increased integrated luminosity, together with a refined analysis strategy, are expected to improve the reach of the search. Stronger limits are foreseeable, together with the increased chance of discovering BSM physics. Future iterations of the search are expected to focus on the high-mass region, where the results from flavour physics set the loosest constraints, and on reducing the impact of the systematic uncertainties, which are

indeed a limiting factor of the current analysis.

## Bibliography

- [1] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [2] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [3] ATLAS and CMS Collaboration, *Combined measurement of the Higgs boson mass in pp collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS experiments*, Phys. Rev. Lett. **114** (2015) p. 191803, arXiv: 1503.07589 [hep-ex].
- [4] Weinberg, Steven, *A model of leptons*, Phys. Rev. Lett. **19** (1967) p. 1264.
- [5] Salam, Abdus, *Weak and electromagnetic interactions*, Conf. Proc. **C680519** (1968) p. 367.
- [6] Glashow, S. L., *Partial symmetries of weak interactions*, Nucl. Phys. **22** (1961) p. 579.
- [7] I. Aitchison, *Supersymmetry in particle physics: an elementary introduction*, Cambridge University Press, 2007.
- [8] L. Evans and P. Bryant, *LHC machine*, JINST **3** (2008) S08001.
- [9] Branco, G. C. and others, *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) p. 1, arXiv: 1106.0034 [hep-ph].
- [10] Abbiendi, G. and others, *Search for charged Higgs bosons: combined results using LEP data*, Eur. Phys. J. **C73** (2013) p. 2463, arXiv: 1301.6065 [hep-ex].
- [11] CMS Collaboration, *Search for a charged Higgs boson in pp collisions at  $\sqrt{s} = 8$  TeV*, JHEP **11** (2015) p. 018, arXiv: 1508.07774 [hep-ex].
- [12] ATLAS Collaboration, *Search for charged Higgs bosons in the  $H^\pm \rightarrow tb$  decay channel in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, JHEP **03** (2016) p. 127, arXiv: 1512.03704 [hep-ex].
- [13] ATLAS Collaboration, *Search for charged Higgs bosons decaying into top and bottom quarks at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, JHEP **11** (2018) p. 085, arXiv: 1808.03599 [hep-ex].
- [14] Weinberg, Steven, *The making of the Standard Model*, Eur. Phys. J. **34** (2004) p. 5, arXiv: hep-ph/0401010 [hep-ph].
- [15] F. Halzen and A. D. Martin, *Quarks and leptons: an introductory course in modern particle physics*, 1984.

- [16] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*, Westview, 1995.
- [17] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013.
- [18] A. A. Klyachko, *The Pauli exclusion principle and beyond*, ArXiv e-prints (2009), arXiv: 0904.2009 [quant-ph].
- [19] Wikipedia, *Standard Model particles*, (2006), URL: [http://commons.wikimedia.org/wiki/File:Standard\\_Model\\_of\\_Elementary\\_Particles.svg](http://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg).
- [20] Gell-Mann, Murray, *Symmetries of baryons and mesons*, Phys. Rev. **125** (1962) p. 1067.
- [21] Lee, T. D. and Yang, Chen-Ning, *Question of Parity Conservation in Weak Interactions*, Phys. Rev. **104** (1956) p. 254.
- [22] Wu, C. S. and others, *Experimental test of parity conservation in beta decay*, Phys. Rev. **105** (1957) p. 1413.
- [23] Higgs, Peter W., *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) p. 508.
- [24] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) p. 321.
- [25] Djouadi, Abdelhak, *The anatomy of electro-weak symmetry breaking. I: the Higgs boson in the Standard Model*, Phys. Rept. **457** (2008) p. 1, arXiv: hep-ph/0503172 [hep-ph].
- [26] T. Kajita, *Discovery of neutrino oscillations*, Reports on Progress in Physics **69** (2006) p. 1607.
- [27] J. Silk and others, *Particle dark matter: observations, models and searches*, ed. by G. Bertone, Cambridge Univ. Press, 2010.
- [28] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) p. 32, [Usp. Fiz. Nauk161, no.5, 61 (1991)].
- [29] M. B. Gavela et al., *Standard model CP violation and baryon asymmetry*, Mod. Phys. Lett. **A9** (1994) p. 795, arXiv: hep-ph/9312215 [hep-ph].
- [30] Fujikawa, Kazuo, *Dimensional regularization is generic*, Int. J. Mod. Phys. **31** (2016) p. 1630042, arXiv: 1605.05813 [hep-th].
- [31] Martin, Stephen P., *A Supersymmetry primer*, (1997), arXiv: hep-ph/9709356 [hep-ph].
- [32] P. S. Bhupal Dev and A. Pilaftsis, *Natural alignment in the two Higgs doublet model*, J. Phys. Conf. Ser. **873** (2017) p. 012008, arXiv: 1703.05730 [hep-ph].
- [33] J. F. Gunion and H. E. Haber, *The CP conserving two Higgs doublet model: the approach to the decoupling limit*, Phys. Rev. **D67** (2003) p. 075019, arXiv: hep-ph/0207010 [hep-ph].

- 
- [34] S. L. Glashow and S. Weinberg, *Natural Conservation Laws for Neutral Currents*, Phys. Rev. **D15** (1977) p. 1958.
- [35] Carena, M. and others, *MSSM Higgs boson searches at the LHC: benchmark scenarios after the discovery of a Higgs-like particle*, Eur. Phys. J. **C73** (2013) p. 2552, arXiv: 1302.7033 [hep-ph].
- [36] Djouadi, A. and others, *The post-Higgs MSSM scenario: Habemus MSSM?*, Eur. Phys. J. **C73** (2013) p. 2650, arXiv: 1307.5205 [hep-ph].
- [37] F. Brummer, S. Kraml, and S. Kulkarni, *Anatomy of maximal stop mixing in the MSSM*, JHEP **8** (2012) p. 89, arXiv: 1204.5977 [hep-ph].
- [38] Arbey, A. and others, *Status of the charged Higgs boson in two Higgs doublet models*, Eur. Phys. J. **C78** (2018) p. 182, arXiv: 1706.07414 [hep-ph].
- [39] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model*, Phys. Rept. **459** (2008) p. 1, arXiv: hep-ph/0503173 [hep-ph].
- [40] J. E. Cieza Montalvo et al., *Searching for doubly charged Higgs bosons at the LHC in a 3-3-1 model*, Nucl. Phys. **B756** (2006) p. 1, [Erratum: Nucl. Phys.B796,422(2008)], arXiv: hep-ph/0606243 [hep-ph].
- [41] J. F. Gunion, R. Vega, and J. Wudka, *Higgs triplets in the standard model*, Phys. Rev. **D42** (1990) p. 1673.
- [42] H. Georgi and M. Machacek, *Doubly charged Higgs bosons*, Nuclear Physics B **262** (1985) p. 463, ISSN: 0550-3213.
- [43] Akeroyd, A. G. and others, *Prospects for charged Higgs searches at the LHC*, Eur. Phys. J. **C77** (2017) p. 276, arXiv: 1607.01320 [hep-ph].
- [44] J. R. Andersen et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, (2013), ed. by S Heinemeyer et al., arXiv: 1307.1347 [hep-ph].
- [45] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, (2016), arXiv: 1610.07922 [hep-ph].
- [46] K. Kumericki, *Feynman diagrams for beginners*, (2016), arXiv: 1602.04182.
- [47] Degrande, C. and others, *Heavy charged Higgs boson production at the LHC*, JHEP **10** (2015) p. 145, arXiv: 1507.02549 [hep-ph].
- [48] Schaarschmidt, J., *hplusmssm*, (2018), URL: [url:https://github.com/janascha/hplusmssm](https://github.com/janascha/hplusmssm).
- [49] Aaltonen, T. and others, *Search for charged Higgs bosons in decays of top quarks in p anti-p collisions at  $\sqrt{s} = 1.96$  TeV*, Phys. Rev. Lett. **103** (2009) p. 101803, arXiv: 0907.1269 [hep-ex].
- [50] Abazov, V. M. and others, *Search for charged Higgs bosons in top-quark decays*, Phys. Lett. **B682** (2009) p. 278, arXiv: 0908.1811 [hep-ex].

- [51] Abazov, V. M. and others, *Search for charged Higgs bosons decaying to top and bottom quarks in  $p\bar{p}$  collisions*, Phys. Rev. Lett. **102** (2009) p. 191802, arXiv: 0807.0859 [hep-ex].
- [52] ATLAS Collaboration, *Search for charged Higgs bosons decaying via  $H^+ \rightarrow \tau\nu$  in top quark pair events using  $pp$  collision data at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, JHEP **06** (2012) p. 039, arXiv: 1204.2760 [hep-ex].
- [53] ATLAS Collaboration, *Search for charged Higgs bosons decaying via  $H^\pm \rightarrow \tau^\pm\nu$  in fully hadronic final states using  $pp$  collision data at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, JHEP **03** (2015) p. 088, arXiv: 1412.6663 [hep-ex].
- [54] ATLAS Collaboration, *Search for charged Higgs bosons decaying via  $H^\pm \rightarrow \tau^\pm\nu_\tau$  in the  $\tau$ +jets and  $\tau$ +lepton final states with  $36\text{ fb}^{-1}$  of  $pp$  collision data recorded at  $\sqrt{s} = 13$  TeV with the ATLAS experiment*, JHEP **09** (2018) p. 139, arXiv: 1807.07915 [hep-ex].
- [55] CMS Collaboration, *Search for a light charged Higgs boson in top quark decays in  $pp$  collisions at  $\sqrt{s} = 7$  TeV*, JHEP **07** (2012) p. 143, arXiv: 1205.5736 [hep-ex].
- [56] ATLAS Collaboration, *Search for a light charged Higgs boson in the decay channel  $H^+ \rightarrow c\bar{s}$  in  $t\bar{t}$  events using  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, Eur. Phys. J. **C73** (2013) p. 2465, arXiv: 1302.3694 [hep-ex].
- [57] CMS Collaboration, *Search for a light charged Higgs boson decaying to  $c\bar{s}$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV*, JHEP **12** (2015) p. 178, arXiv: 1510.04252 [hep-ex].
- [58] ATLAS Collaboration, *Search for a charged Higgs boson produced in the vector-boson fusion mode with decay  $H^\pm \rightarrow W^\pm Z$  using  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS experiment*, Phys. Rev. Lett. **114** (2015) p. 231801, arXiv: 1503.04233 [hep-ex].
- [59] CMS Collaboration, *Search for charged Higgs bosons produced via vector-boson fusion and decaying into a pair of  $W$  and  $Z$  bosons using  $pp$  collisions at  $\sqrt{s} = 13$  TeV*, Phys. Rev. Lett. **119** (2017) p. 141802, arXiv: 1705.02942 [hep-ex].
- [60] CMS Collaboration, *Search for a charged Higgs boson decaying to charm and bottom quarks in proton-proton collisions at  $\sqrt{s} = 8$  TeV*, JHEP **11** (2018) p. 115, arXiv: 1808.06575 [hep-ex].
- [61] CMS Collaboration, *Search for charged Higgs bosons with the  $H^\pm \rightarrow \tau^\pm\nu_\tau$  decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV*, CMS-PAS-HIG-18-014, 2018, URL: <https://cds.cern.ch/record/2640359>.
- [62] Babar Collaboration, *Measurement of an excess of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  decays and implications for charged Higgs bosons*, Phys. Rev. **D88** (2013) p. 072012, arXiv: 1303.0571 [hep-ex].
- [63] Belle Collaboration, *Measurement of the  $\tau$  lepton polarization and  $R(D^*)$  in the decay  $\bar{B} \rightarrow D^*\tau^-\bar{\nu}_\tau$  with one-prong hadronic  $\tau$  decays at Belle*, Phys. Rev. **D97** (2018) p. 012004, arXiv: 1709.00129 [hep-ex].



- 
- [64] Belle Collaboration, *Measurement of the branching ratio of  $\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau$  relative to  $\bar{B}^0 \rightarrow D^{*+}\ell^-\bar{\nu}_\ell$  decays with a semileptonic tagging method*, Phys. Rev. **D94** (2016) p. 072007, arXiv: 1607.07923 [hep-ex].
  - [65] Belle Collaboration, *Measurement of the branching ratio of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  relative to  $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$  decays with hadronic tagging at Belle*, Phys. Rev. **D92** (2015) p. 072014, arXiv: 1507.03233 [hep-ex].
  - [66] M. Misiak and M. Steinhauser, *Weak radiative decays of the B meson and bounds on  $M_{H^\pm}$  in the Two-Higgs-Doublet Model*, Eur. Phys. J. **C77** (2017) p. 201, arXiv: 1702.04571 [hep-ph].
  - [67] A. G. Akeroyd, F. Mahmoudi, and D. M. Santos, *The decay  $B_s \rightarrow \mu^+ \mu^-$ : updated SUSY constraints and prospects*, JHEP **12** (2011) p. 088, arXiv: 1108.3018 [hep-ph].
  - [68] A. Arbey et al., *Supersymmetry confronts  $B_s^{+-}$  : Present and future status*, Phys. Rev. **D87** (2013) p. 035026, arXiv: 1212.4887 [hep-ph].
  - [69] Stanford University, *Hard scattering*, (2016), URL: [https://theory.slac.stanford.edu/sites/theory.slac.stanford.edu/files/styles/lightbox\\_image\\_popup/public/event\\_800px.png?itok=qjhwv\\_hy](https://theory.slac.stanford.edu/sites/theory.slac.stanford.edu/files/styles/lightbox_image_popup/public/event_800px.png?itok=qjhwv_hy).
  - [70] G. Hooft and M. J. G. Veltman, *Regularization and renormalization of gauge fields*, Nucl. Phys. **B44** (1972) p. 189.
  - [71] Alwall, J. and others, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) p. 079, arXiv: 1405.0301 [hep-ph].
  - [72] Frixione, S. and others, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007) p. 070.
  - [73] Butterworth, J. and others, *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) p. 023001, arXiv: 1510.03865 [hep-ph].
  - [74] Ball, Richard D. and others, *Parton distributions for the LHC Run II*, JHEP **04** (2015) p. 040, arXiv: 1410.8849 [hep-ph].
  - [75] Dulat, S. and others, *New parton distribution functions from a global analysis of quantum chromodynamics*, Phys. Rev. **D93** (2016) p. 033006, arXiv: 1506.07443 [hep-ph].
  - [76] Martin, A. D. and others, *Heavy-quark mass dependence in global PDF analyses and 3- and 4-flavour parton distributions*, Eur. Phys. J. **C70** (2010) p. 51, arXiv: 1007.2624 [hep-ph].
  - [77] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nucl. Phys. **B126** (1977) p. 298.
  - [78] Sjostrand, Torbjorn and others, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) p. 159, arXiv: 1410.3012 [hep-ph].

- [79] Bahr, M. and others, *Herwig++ physics and manual*, Eur. Phys. J. **C58** (2008) p. 639, arXiv: 0803.0883 [hep-ph].
- [80] Buckley, Andy and others, *General-purpose event generators for LHC physics*, Phys. Rept. **504** (2011) p. 145, arXiv: 1101.2599 [hep-ph].
- [81] Agostinelli, S. and others, *GEANT4: a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) p. 250.
- [82] ATLAS Collaboration, *The ATLAS simulation infrastructure*, Eur. Phys. J. C **70** (2010) p. 823, arXiv: 1005.4568 [physics.ins-det].
- [83] Cowan, Glen and others, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) p. 1554, arXiv: 1007.1727 [physics.data-an].
- [84] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011, URL: <https://cds.cern.ch/record/1379837>.
- [85] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, ed. by S. Kotz and N. L. Johnson, Springer New York, 1992.
- [86] A. L. Read, *Presentation of search results: the CL<sub>s</sub> technique*, J. Phys. G **28** (2002).
- [87] Hoecker, A. and others, *TMVA - Toolkit for Multivariate Data Analysis*, ArXiv Physics e-prints (2007), eprint: physics/0703039.
- [88] Wikipedia, *Receiver operating characteristic*, (2015), URL: [https://en.wikipedia.org/wiki/Receiver\\_operating\\_characteristic](https://en.wikipedia.org/wiki/Receiver_operating_characteristic).
- [89] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.
- [90] LHCb Collaboration, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.
- [91] TOTEM Collaboration, *The TOTEM experiment at the LHC*, JINST **3** (2008) S08007.
- [92] MoEDAL Collaboration, *The MoEDAL experiment at the LHC*, PoS **LHCP2016** (2016) p. 099.
- [93] LHCf Collaboration, *Measurements of very forward particles production spectra at LHC: the LHCf experiment*, PoS **LHCP2016** (2017) p. 195.
- [94] Wikipedia, *CERN acceleration complex*, (2008), URL: [https://upload.wikimedia.org/wikipedia/commons/7/72/CAC-acc\\_45.jpg](https://upload.wikimedia.org/wikipedia/commons/7/72/CAC-acc_45.jpg).
- [95] ATLAS Collaboration, *Studies of the performance of the ATLAS detector using cosmic-ray muons*, Eur. Phys. J. **C71** (2011) p. 1593, arXiv: 1011.6665 [physics.ins-det].
- [96] Lenzi, Thomas, *Development and study of different muon track reconstruction algorithms for the Level-1 trigger for the CMS muon upgrade with GEM detectors*, (2013), arXiv: 1306.0858 [physics.ins-det].

- 
- [97] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.
  - [98] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19, 2010, URL: <https://cds.cern.ch/record/1291633>.
  - [99] Bruschi, Marco, *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.
  - [100] Abdel Khalek, S. and others, *The ALFA Roman Pot detectors of ATLAS*, JINST **11** (2016) P11013.
  - [101] Cindro, V. and others, *The ATLAS beam conditions monitor*, JINST **3** (2008) P02004.
  - [102] S. Ballestrero, W. Vandelli, and G. Avolio, *ATLAS TDAQ system: current status and performance*, Phys. Procedia **37** (2012) p. 1819.
  - [103] Bird, I. and others, *LHC computing Grid. Technical design report*, (2005).
  - [104] Panduro Vazquez, William, *The ATLAS data acquisition system: from Run 1 to Run 2*, Nucl. Part. Phys. Proc. **273-275** (2016) p. 939.
  - [105] ATLAS Collaboration, *Luminosity public results*, (2016), URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
  - [106] Fruhwirth, R., *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) p. 444.
  - [107] Cornelissen, T. G. and others, *The global  $\chi^2$  track fitter in ATLAS*, Journal of Physics: Conference Series **119** (2008) p. 032013, URL: <http://stacks.iop.org/1742-6596/119/i=3/a=032013>.
  - [108] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at  $\sqrt{s} = 13$  TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: <https://cds.cern.ch/record/2037717>.
  - [109] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. **C77** (2017) p. 490, arXiv: 1603.02934 [hep-ex].
  - [110] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. **C77** (2017) p. 673, arXiv: 1704.07983 [hep-ex].
  - [111] N. A. Asbah and J. Katzy, *Search for the production of a Standard Model Higgs boson in association with top-quarks and decaying into a pair of bottom-quarks with 13 TeV ATLAS data*, 2018.
  - [112] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for Bremsstrahlung*, ATLAS-CONF-2012-047, 2012, URL: <https://cds.cern.ch/record/1449796>.

- [113] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024, 2016, URL: <https://cds.cern.ch/record/2157687>.
- [114] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74** (2014) p. 3071, arXiv: 1407.5063 [hep-ex].
- [115] ATLAS Collaboration, *Muon reconstruction performance in early  $\sqrt{s} = 13$  TeV data*, ATL-PHYS-PUB-2015-037, 2015, URL: <http://cds.cern.ch/record/2047831>.
- [116] 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) p. 292, arXiv: 1603.05598 [hep-ex].
- [117] Cacciari, M. and others, *FastJet user manual*, Eur. Phys. J. **C72** (2012) p. 1896, arXiv: 1111.6097 [hep-ph].
- [118] Cacciari, M. and others, *The Anti- $k(t)$  jet clustering algorithm*, JHEP **04** (2008) p. 063, arXiv: 0802.1189 [hep-ph].
- [119] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, URL: <http://cds.cern.ch/record/2064383>.
- [120] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Phys. Rev. **D96** (2017) p. 072002, arXiv: 1703.09665 [hep-ex].
- [121] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, Eur. Phys. J. **C76** (2016) p. 581, arXiv: 1510.03823 [hep-ex].
- [122] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: <http://cds.cern.ch/record/2037702>.
- [123] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: <https://cds.cern.ch/record/2160731>.
- [124] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at  $\sqrt{s} = 13$  TeV*, ATL-PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904>.
- [125] Gleisberg, T. and others, *Event generation with SHERPA 1.1*, JHEP **02** (2009) p. 007, arXiv: 0811.4622 [hep-ph].
- [126] Buttinger, William, *Using event weights to account for differences in instantaneous Luminosity and trigger prescale in Monte Carlo and data*, ATL-COM-SOFT-2015-119, 2015, URL: <https://cds.cern.ch/record/2014726>.

- 
- [127] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) p. 244, arXiv: 1207.1303 [hep-ph].
- [128] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) p. 852, arXiv: 0710.3820 [hep-ph].
- [129] ATLAS Collaboration, *ATLAS Run 1 Pythia 8 tunes*, ATL-PHYS-PUB-2014-021, 2014, URL: <http://cds.cern.ch/record/1966419>.
- [130] M. Czakon and A. Mitov, *Top++: a program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. **185** (2014) p. 2930, arXiv: 1112.5675 [hep-ph].
- [131] Cacciari, M. and others, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) p. 612, arXiv: 1111.5869 [hep-ph].
- [132] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, URL: <https://cds.cern.ch/record/2216168>.
- [133] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. **C76** (2016) p. 196, arXiv: 1512.01178 [hep-ph].
- [134] ATLAS Collaboration, *Simulation of top quark production for the ATLAS experiment at  $\sqrt{s} = 13$  TeV*, ATL-PHYS-PUB-2016-004, 2016, URL: <https://cds.cern.ch/record/2120417>.
- [135] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a  $b\bar{b}$  pair in pp collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*, Phys. Rev. **D97** (2018) p. 072016, arXiv: 1712.08895 [hep-ex].
- [136] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering amplitudes with open loops*, Phys. Rev. Lett. **108** (2012) p. 111601, arXiv: 1111.5206 [hep-ph].
- [137] Lai, H. and others, *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) p. 074024, arXiv: 1007.2241 [hep-ph].
- [138] Butterworth, J. and others, *Single boson and diboson production cross sections in pp collisions at  $\sqrt{s} = 7$  TeV*, ATL-COM-PHYS-2010-695, 2010, URL: <https://cds.cern.ch/record/1287902>.
- [139] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) p. 026, arXiv: hep-ph/0603175 [hep-ph].
- [140] Skands, Peter Zeiler, *Tuning Monte Carlo generators: the Perugia tunes*, Phys. Rev. **D82** (2010) p. 074018, arXiv: 1005.3457 [hep-ph].
- [141] Kidonakis, Nikolaos, *Two-loop soft anomalous dimensions for single-top-quark associated production with a  $W^-$  or  $H^-$* , Phys. Rev. **D82** (2010) p. 054018, arXiv: 1005.4451 [hep-ph].

- [142] Kidonakis, Nikolaos, *NNLL resummation for s-channel single-top-quark production*, Phys. Rev. **D81** (2010) p. 054028, arXiv: 1001.5034 [hep-ph].
- [143] Kidonakis, Nikolaos, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single-top-quark production*, Phys. Rev. **D83** (2011) p. 091503, arXiv: 1103.2792 [hep-ph].
- [144] Frixione, S. and others, *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) p. 029, arXiv: 0805.3067 [hep-ph].
- [145] Derue, Frederic, *Estimation of fake lepton background for top analyses using the Matrix Method with the 2015 dataset at  $\sqrt{s}=13$  TeV with AnalysisTop-2.3.41*, ATL-COM-PHYS-2016-198, 2016, URL: <https://cds.cern.ch/record/2135116>.
- [146] Bernaciak, C. and others, *Fox-Wolfgram moments in Higgs physics*, Phys. Rev. **D87** (2013) p. 073014, arXiv: 1212.4436 [hep-ph].
- [147] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) p. 81.
- [148] ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) p. 653, arXiv: 1608.03953 [hep-ex].
- [149] 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, URL: <https://cds.cern.ch/record/2037613>.
- [150] ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex].
- [151] ATLAS Collaboration, *Studies of tt+cc production with MG5\_aMC@NLO and Herwig++ for the ATLAS experiment*, ATL-PHYS-PUB-2016-011, 2016, URL: <https://cds.cern.ch/record/2153876>.
- [152] A. Lazopoulos et al., *Next-to-leading order QCD corrections to  $t\bar{t}Z$  production at the LHC*, Phys. Lett. **B666** (2008) p. 62, arXiv: 0804.2220 [hep-ph].
- [153] J. M. Campbell and R. K. Ellis,  *$t\bar{t}W^{+-}$  production and decay at NLO*, JHEP **07** (2012) p. 052, arXiv: 1204.5678 [hep-ph].
- [154] Beenakker, W. and others, *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl. Phys. **B653** (2003) p. 151, arXiv: hep-ph/0211352 [hep-ph].
- [155] Y. Zhang et al., *QCD NLO and EW NLO corrections to  $t\bar{t}H$  production with top quark decays at hadron collider*, Phys. Lett. **B738** (2014) p. 1, arXiv: 1407.1110 [hep-ph].
- [156] ATLAS Collaboration, *Multi-boson simulation for 13 TeV ATLAS analyses*, ATL-PHYS-PUB-2016-002, 2016, URL: <https://cds.cern.ch/record/2119986>.
- [157] Moneta, L. and others, *The RooStats project*, (2010) p. 57, arXiv: 1009.1003 [physics.data-an].

- [158] Pinamonti, M. and others, *TRexFitter*, (2008), URL: [url:https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TtHFitter](https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TtHFitter).
- [159] T. Chen and C. Guestrin, *XGBoost: a scalable tree boosting system*, ArXiv e-prints (2016), arXiv: 1603.02754 [cs.LG].
- [160] Nachman, B. and others, *Jets from Jets: Re-clustering as a tool for large radius jet reconstruction and grooming at the LHC*, JHEP **02** (2015) p. 075, arXiv: 1407.2922 [hep-ph].
- [161] ATLAS Collaboration, *Jet re-clustering in ATLAS*, ATL-COM-PHYS-2014-1117, 2014, URL: <https://cds.cern.ch/record/1753401>.
- [162] G. Louppe, M. Kagan, and K. Cranmer, *Learning to pivot with adversarial networks*, (2016), arXiv: 1611.01046 [stat.ME].
- [163] R. D. Reed and R. J. Marks, *Neural smithing : supervised learning in feedforward artificial neural networks*, MIT press, 1999.
- [164] Brilliant.org, *Backpropagation*, (2018), URL: <https://brilliant.org/wiki/backpropagation/>.
- [165] Wikipedia, *Artificial neural network*, (2013), URL: [https://en.wikipedia.org/wiki/Artificial\\_neural\\_network](https://en.wikipedia.org/wiki/Artificial_neural_network).
- [166] Chollet, François and others, *Keras*, (2015), URL: <https://keras.io>.





## **Additional material**



## Appendix A

### Glossary of abbreviations

Abbreviation	Meaning
2HDM	Two Higgs Doublets Models
AFII	AtlasFast II
ANN	Adversarial Neural Network
AUC	Area Under the Curve
BDT	Boosted Decision Tree
BR	Branching Ratio
BSM	Beyond the Standard Model
CB	ComBined (muons)
CERN	Conseil Européen pour la Recherche Nucléaire
CL	Confidence Level
CP	Charge Parity
CR	Control Region
CSC	Cathode Strip Chambers
CT	Calorimeter Tagged (muons)
DAQ	Data Acquisition System
ECAL	Electromagnetic CALorimeter
EM	ElectroMagnetic
EW	ElectroWeak
FCAL	Forward CALorimeter
FM	Fully Matched (events)
FS	Flavour Scheme
FSR	Final State Radiation
FullSim	Full Simulation
GSF	Gaussian Sum Filter
HCAL	Hadronic CALorimeter
HF	Heavy Flavour
HLT	High Level Trigger
HSV	Hard Scattering Vertex
IBL	Insertable b-Layer
ID	IDentification

Abbreviation	Meaning
IP	Impact Parameter
ISR	Initial State Radiation
JER	Jet Energy Resolution
JES	Jet Energy Scale
JVT	Jet Vertex Tagger
LAr	Liquid Argon
LHC	Large Hadron Collider
LHD	LikeliHood Discriminant
LSP	Lightest SUSY Particle
LO	Leading Order
MC	Monte Carlo
MDT	Monitored Drift Tubes
ME	Matrix Element
MM	Matrix Method
MS	Muon System or Muon Spectrometer
MSSM	Minimal Supersymmetric Standard Model
NLO	Next to Leading Order
NN	Neural Network
NNLL	Next to Next Leading Logarithmic
NNLO	Next to Next Leading Order
NP	Nuisance Parameter
PCBT	Pseudo Continuous $b$ -tagging
PDF	Parton Distribution Function or Probability Density Function
PIX	Pixel detector
PM	Partially Matched (events)
PS	Parton Shower or Proton Synchrotron
PV	Primary Vertex
QCD	Quantum Chromo Dynamics
QED	Quantum Electrodynamics Theory
QFT	Quantum Field Theory
RF	Radio-Frequency
ROC	Receiver Operating Characteristic
RPC	Resistive Plate Chambers
SA	Stand Alone (muons)
SCT	Semiconductor Tracker
SG	Separation Gain
SM	Standard Model
SPS	Super Proton Synchrotron
SR	Signal Region
ST	Segment Tagged (muons)

---

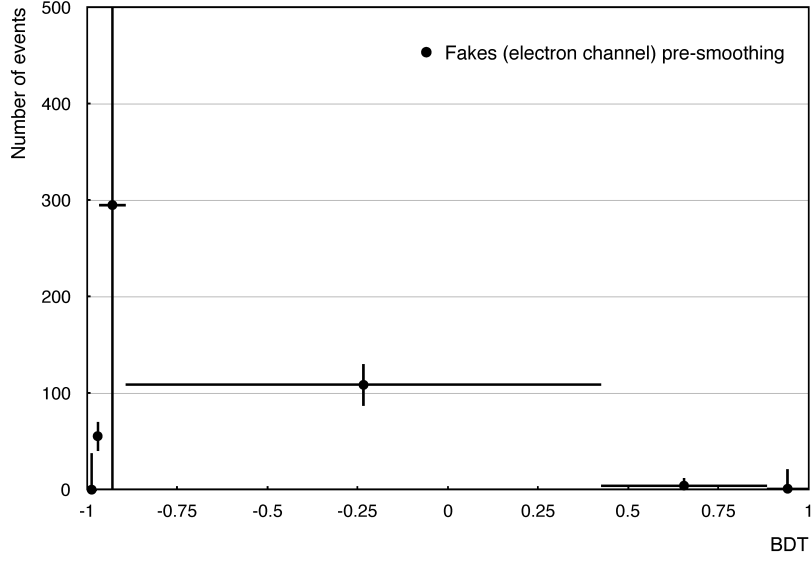
Abbreviation	Meaning
SUSY	SUperSYmmetric theory
TDAQ	Trigger and Data Acquisition System
TGC	Thin-Gap Chambers
TMVA	Toolkit for Multivariate Data Analysis
TRT	Transition Radiation Tracker
UE	Underlying Event
UV	UltraViolet
VEV	Vacuum Expectation Value
WP	Working Point



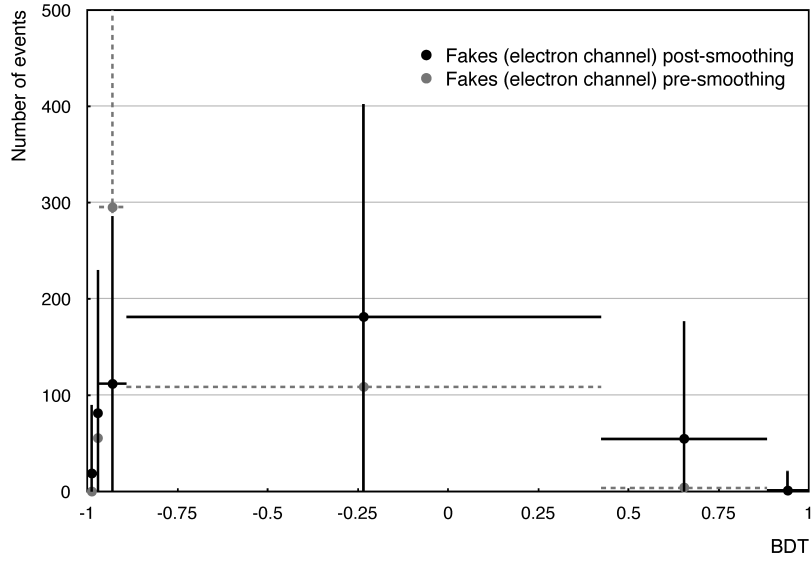
## Appendix B

### Fake leptons

As mentioned in Chapter 6, different triggers are used for the nominal analysis and the fake extrapolation. The reason behind this choice is that the low- $p_T$  triggers of the 2016 dataset (*e26\_lhtight\_nod0\_ivarloose* and *mu\_26\_ivarmedium*) include an isolation cut, which cannot be removed at analysis level. The loose and tight selections used for the fake estimation would therefore be very similar, limiting the amount of available events for the computation of the fake efficiencies. They were therefore replaced with the triggers *e26\_lhvloose\_nod0\_L1EM20VH* and *mu\_26*, which have looser ID requirements and do not include any cut on the isolation (nor on the electron impact parameters). Such triggers are prescale triggers, i.e. they only save a fraction of the events passing the selection criteria. This is necessary because the looseness of the triggers can significantly increase the event rate and overload the data acquisition system of the experiment. The nominal trigger efficiency is recovered by applying an additional weight, known as prescale weight, to the events. The weight can vary from a minimum of 20 to a maximum of 1000, accordingly to the trigger rate which in turn depends on the instantaneous luminosity. Large weights are problematic for regions with limited statistics. The runs with prescale weights larger than 200 were therefore removed, rescaling the remaining fake events to the proper integrated luminosity. However, a prescale weight of 200 would still produce statistical fluctuations. For this reason, the fake distributions were smoothed before being passed to the likelihood fit. The smoothing is applied by using the corresponding "Smooth" option of the TRexFitter [158]. The result of this procedure is shown below: Figure B.1 (a) and Figure B.1 (b) correspond to the output of the BDT trained on the 1600 GeV signal hypothesis, for the electron channel in the  $\geq 6j3b$  region, before and after the smoothing. The smoothing removes the statistical fluctuation of the third bin, but it preserves the overall integral. The statistical error is also distributed over all bins, proportionally to their content. Large prescale weights are not observed in the muon channel, for which the trigger rate is kept constant. The method has a small impact on the results produced at the low and intermediate masses, but it has a larger effect at high masses, where the sensitivity of the analysis is strongly dependent on the statistical error of the most signal-like bins. However, the sensitivity is generally reduced, leading to more conservative results. There is therefore no risk of biasing the analysis towards excessively tight limits. A better fake modelling is under study for future iterations of the analysis.



(a) Before smoothing.



(b) After smoothing.

Figure B.1: Effect of the smoothing function applied to the fake electrons. The distributions correspond to the output of the BDT trained with the 1600 GeV signal hypothesis, in the  $\geq 6j3b$  region, (a) before and (b) after the smoothing.



## Appendix C

### Fit validation

This chapter includes a selection of plots used to validate the fit results of Chapter 6.

Figures C.1 to C.8 show the post-fit distributions of some of the BDT input variables, for the fits performed with the 200 and 800 GeV signal hypotheses:  $p_T^{\text{jet}_0}$ ,  $M_{bb}^{\text{Max}(p_T)}$ ,  $H_1^{\text{all}}$  and  $H_T^{\text{jets}}$ .

Table C.1 and Table C.2 present the post-fit event yields determined from the two fits.

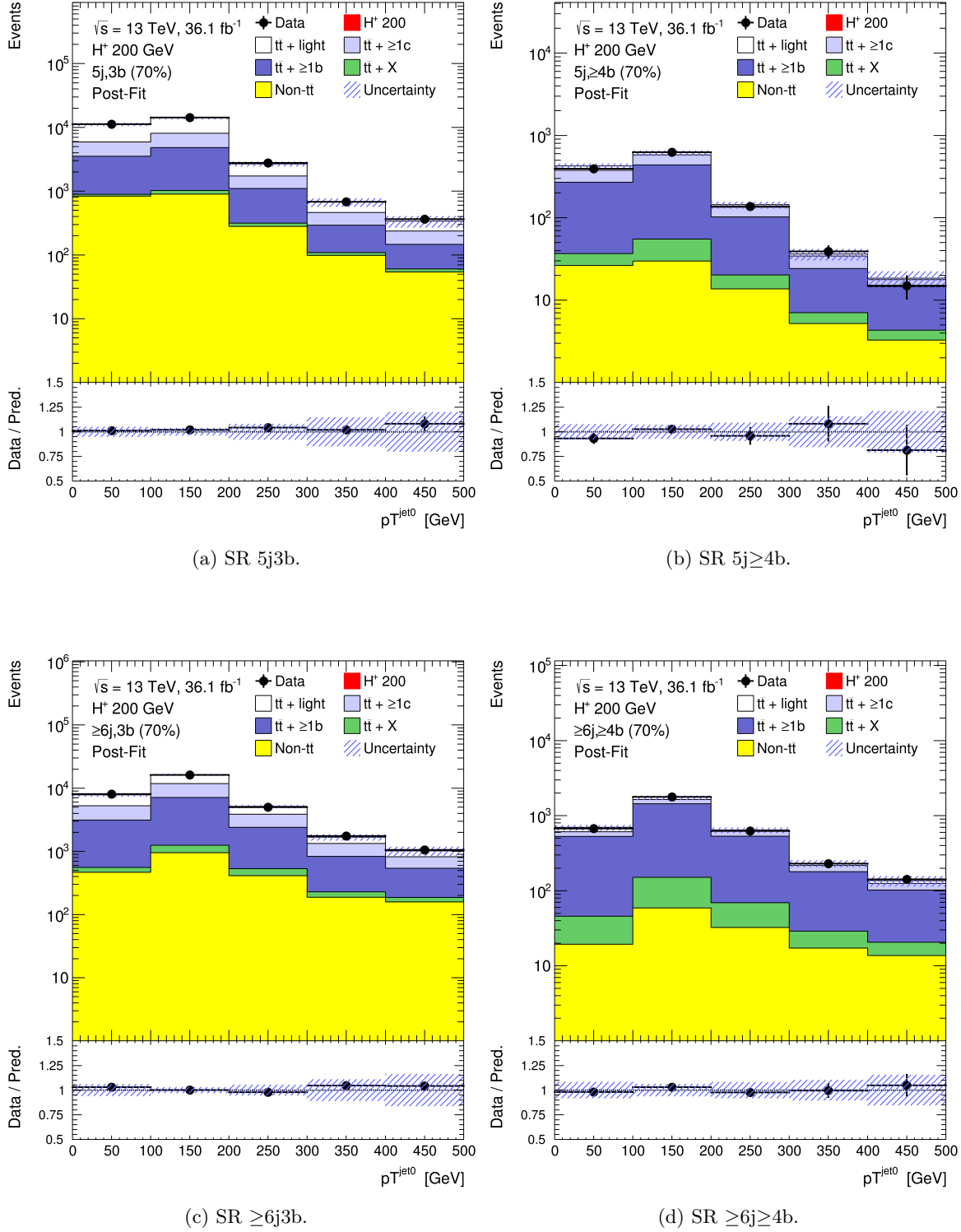


Figure C.1: Post-fit distribution of the leading-jet  $p_T$  in the signal regions, for the fit performed with the 200 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.

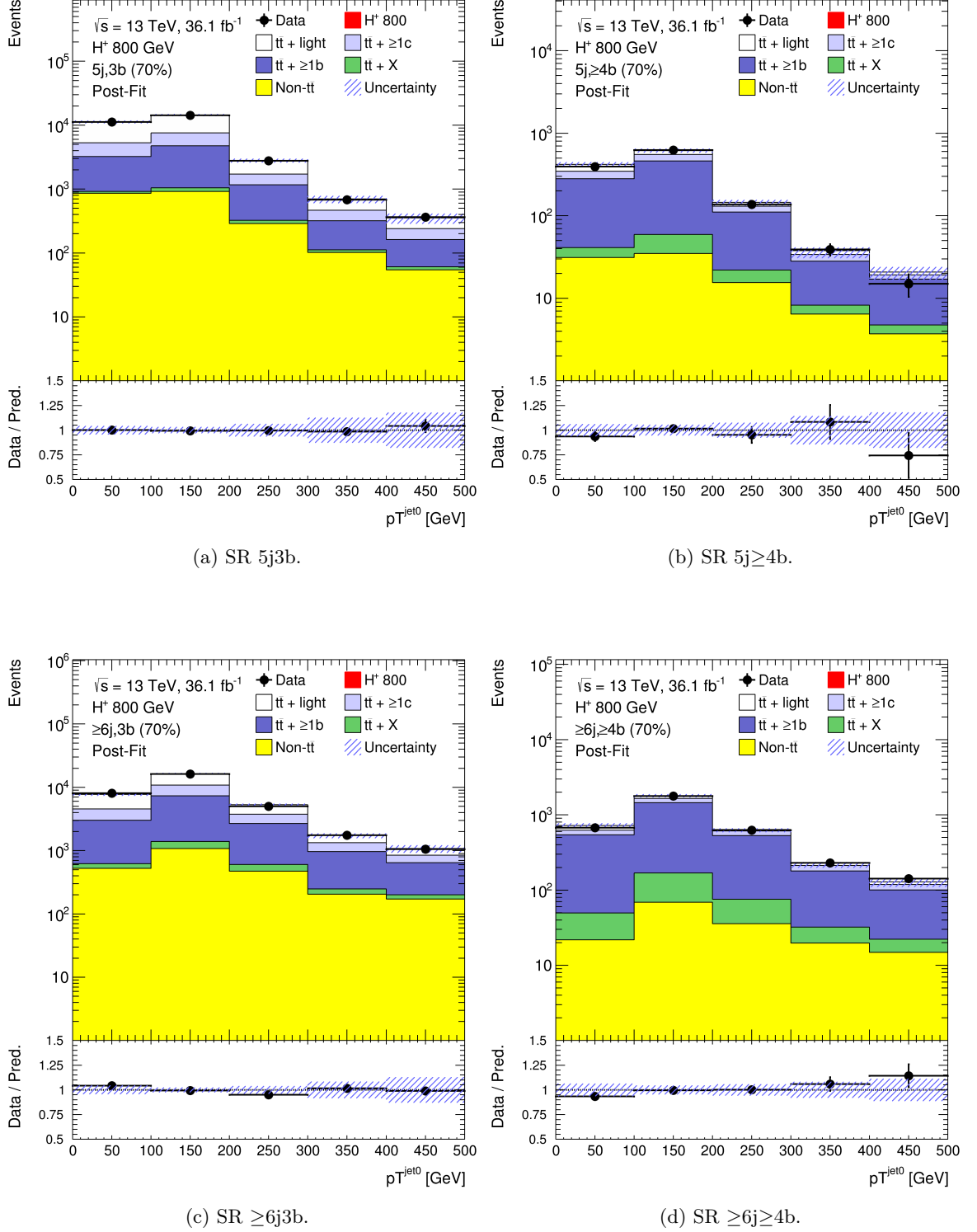


Figure C.2: Post-fit distribution of the leading-jet  $p_T$  in the signal regions, for the fit performed with the 800 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.

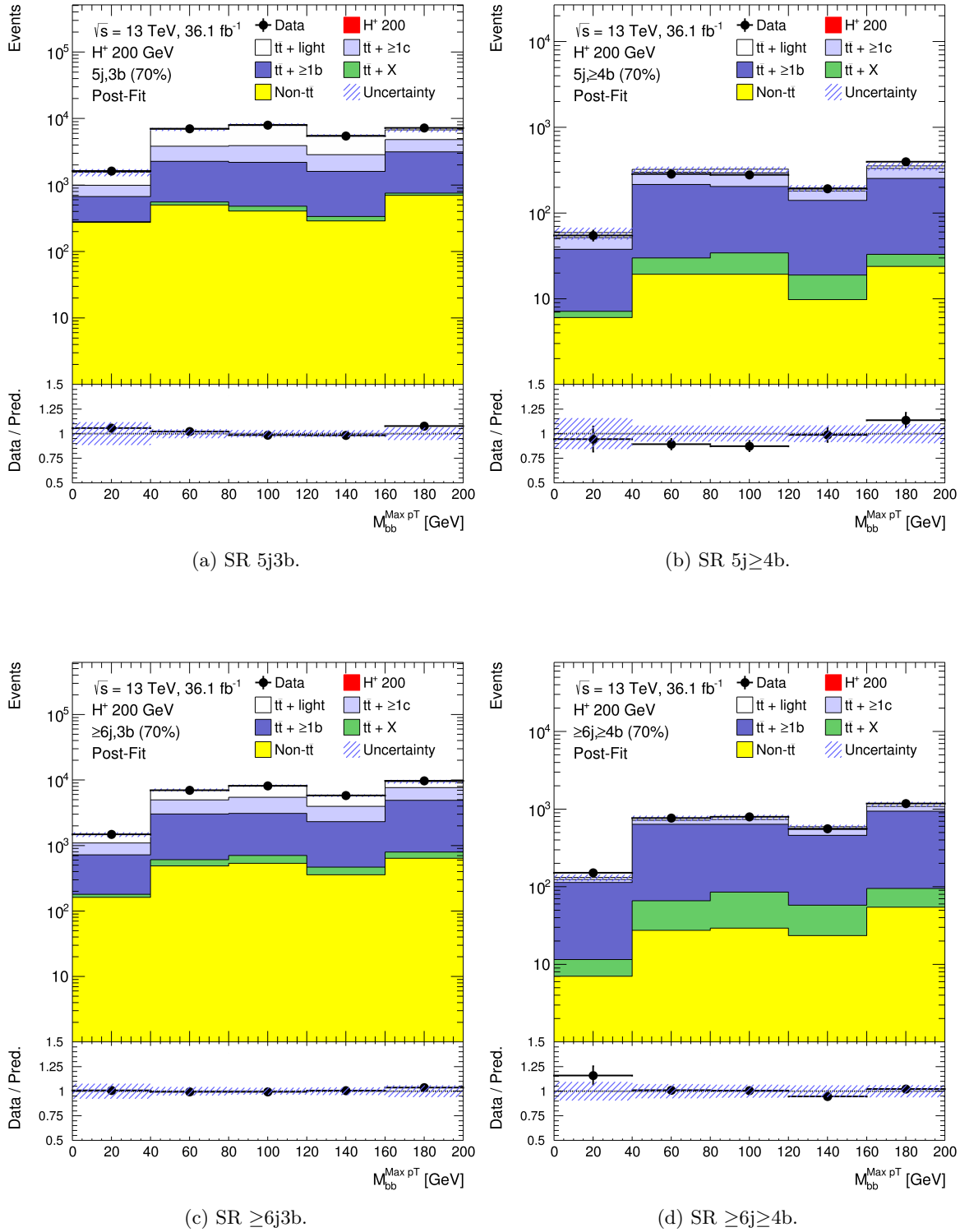


Figure C.3: Post-fit distribution of the  $M_{bb}^{\text{Max}(p_T)}$  variable in the signal regions, for the fit performed with the 200 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.

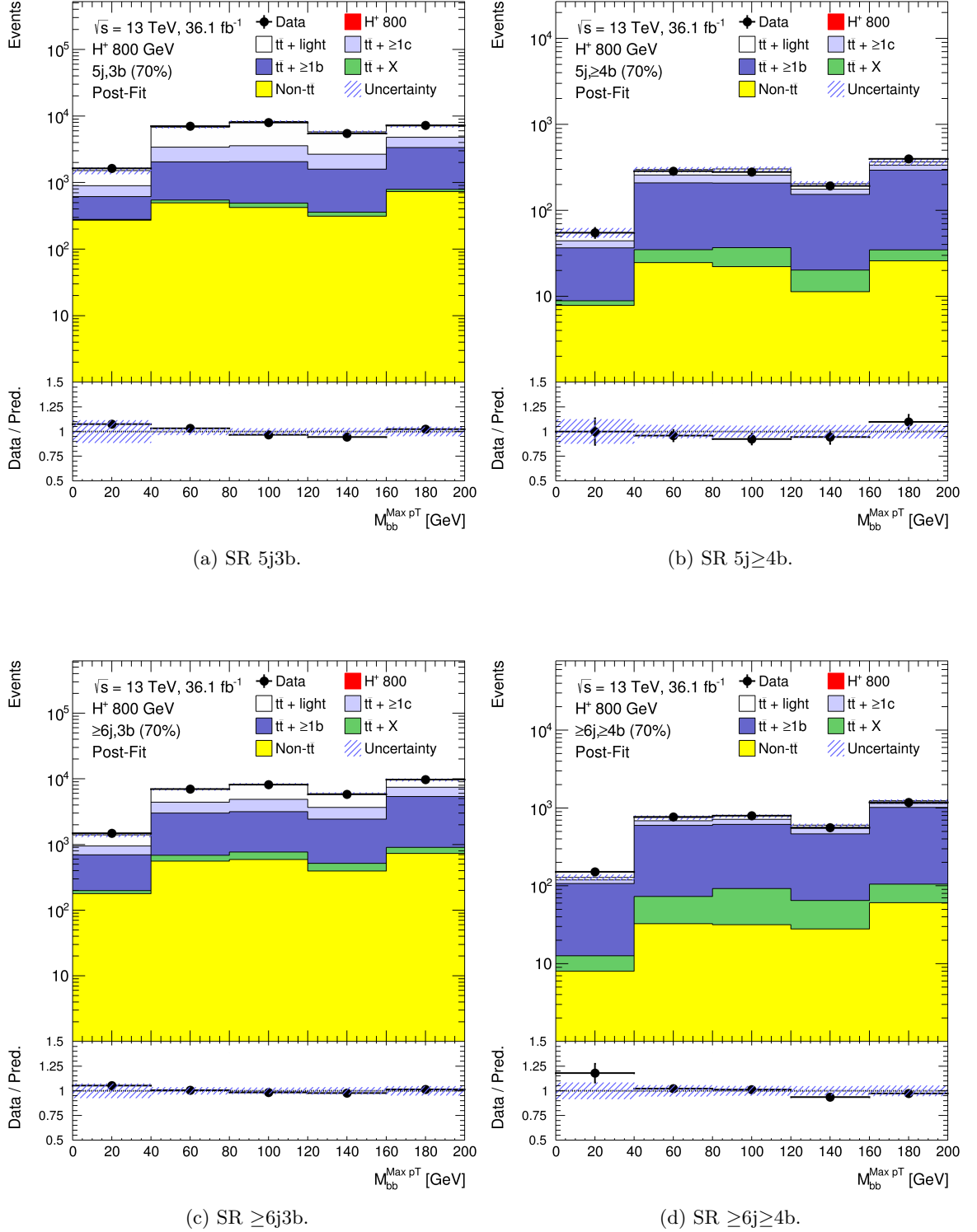


Figure C.4: Post-fit distribution of the  $M_{bb}^{\text{Max}(p_T)}$  variable in the signal regions, for the fit performed with the 800 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.

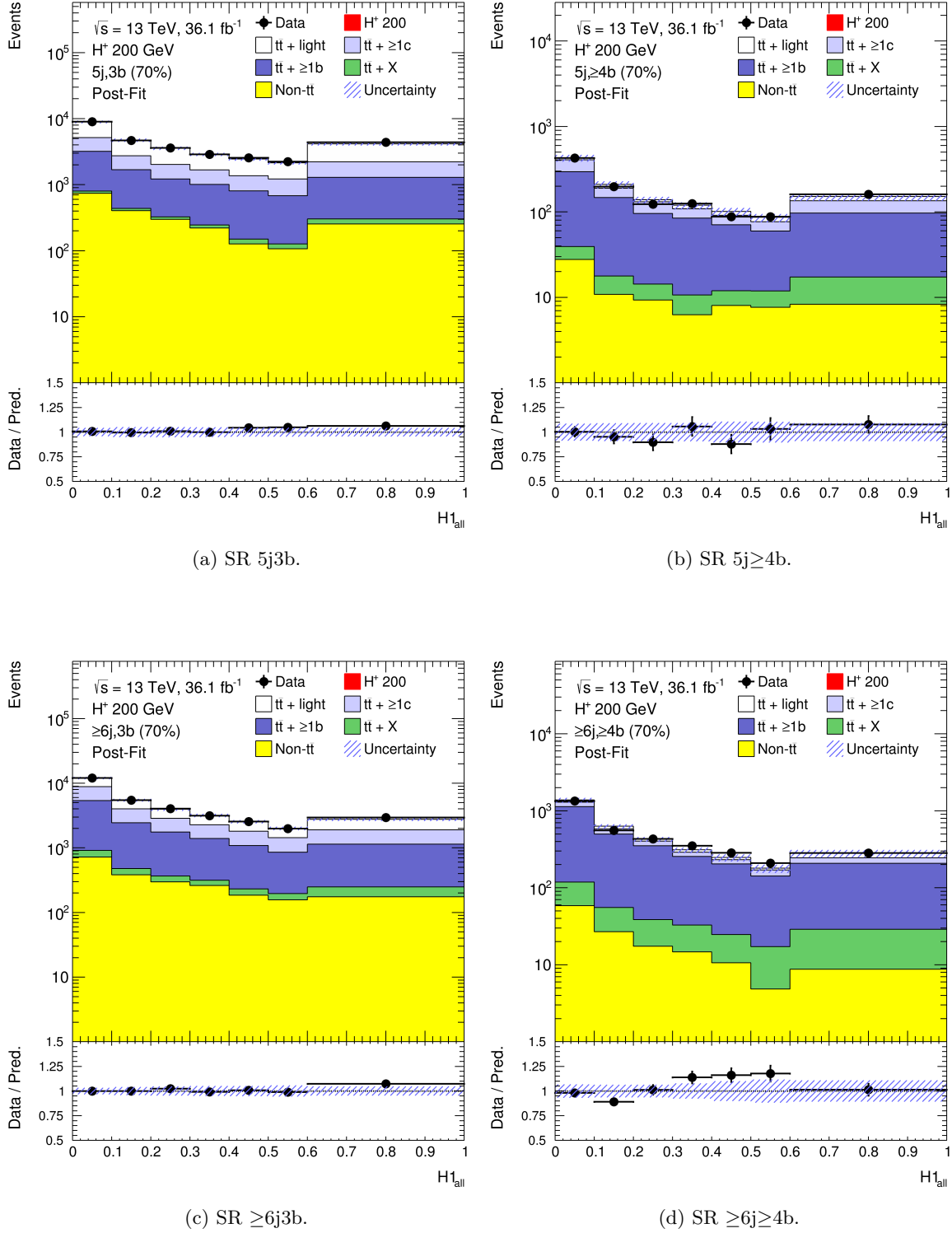


Figure C.5: Post-fit distribution of the  $H_1^{\text{all}}$  variable in the signal regions, for the fit performed with the 200 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.

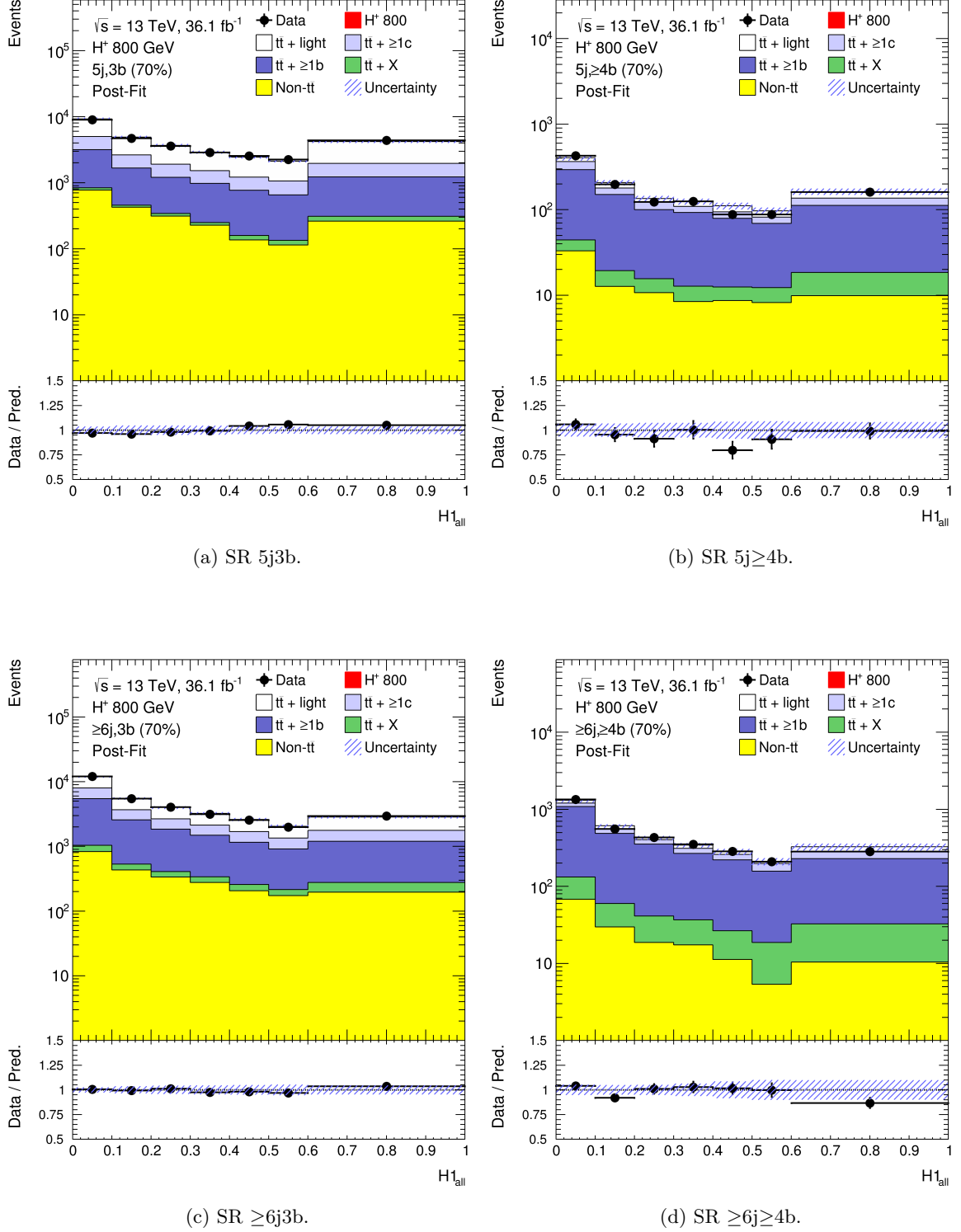


Figure C.6: Post-fit distribution of the  $H_1^{\text{all}}$  variable in the signal regions, for the fit performed with the 800 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.

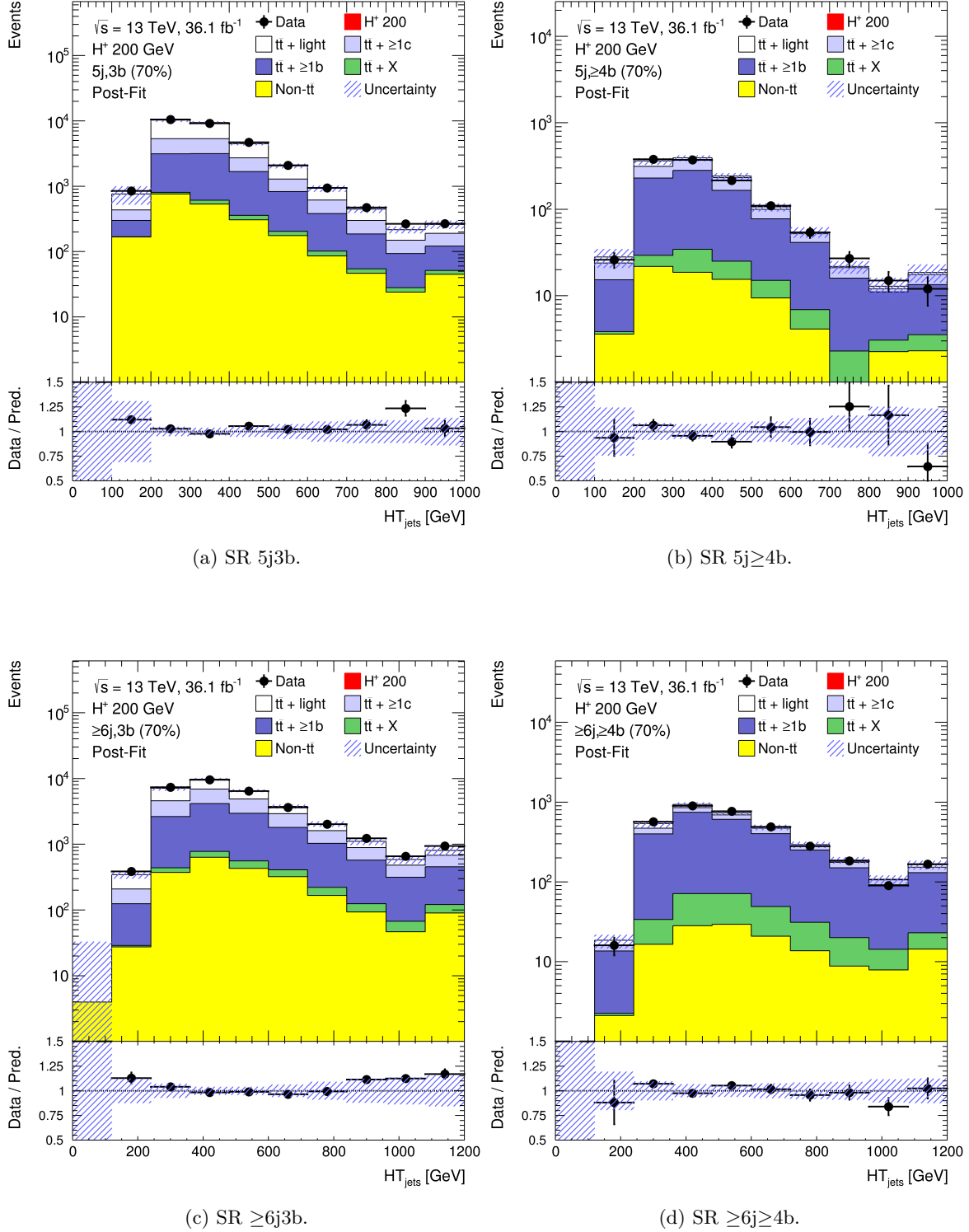


Figure C.7: Post-fit distribution of the  $H_T^{\text{jets}}$  variable in the signal regions, for the fit performed with the 200 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.



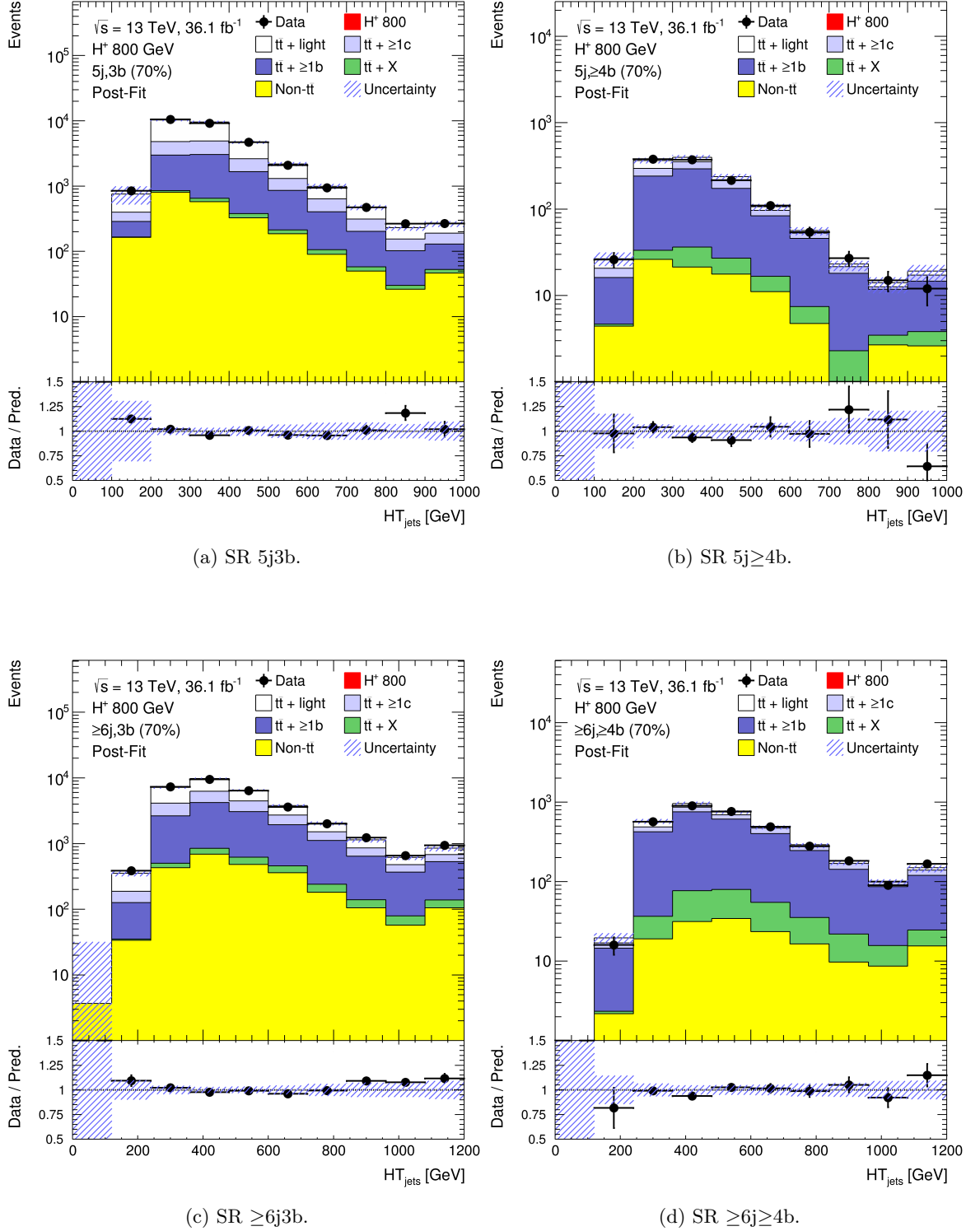


Figure C.8: Post-fit distribution of the  $H_T^{\text{jets}}$  variable in the signal regions, for the fit performed with the 800 GeV signal hypothesis. The lower panel of each plot shows the ratio between the observed and predicted yields. The error bands include all systematic and statistical uncertainties. The signal strength is fit to negative values therefore the signal is not visible.

	5j2b	$\geq 6j2b$	5j3b	5j $\geq 4b$	$\geq 6j3b$	$\geq 6j\geq 4b$
$t\bar{t}$ + light	$233\,000 \pm 14\,600$	$140\,000 \pm 13\,600$	$12\,500 \pm 1440$	$98.0 \pm 40.3$	$8880 \pm 1440$	$251 \pm 156$
$t\bar{t} + \geq 1c$	$47\,100 \pm 14\,800$	$51\,400 \pm 13\,600$	$6520 \pm 1980$	$290 \pm 99.2$	$9070 \pm 2420$	$423 \pm 219$
$t\bar{t} + \geq 1b$	$14\,800 \pm 2930$	$16\,900 \pm 3570$	$7500 \pm 1390$	$727 \pm 137$	$11\,300 \pm 2000$	$2480 \pm 357$
$H^+$	$-289 \pm 911$	$-210 \pm 670$	$-134 \pm 430$	$-14.9 \pm 49.8$	$-142 \pm 458$	$-35.3 \pm 116$
$t\bar{t}W$	$350 \pm 48.8$	$543 \pm 79.3$	$29.4 \pm 4.24$	$0.684 \pm 0.230$	$70.6 \pm 11.7$	$4.71 \pm 1.16$
$t\bar{t}Z$	$395 \pm 47.9$	$721 \pm 92.6$	$78.4 \pm 10.6$	$12.8 \pm 2.29$	$183 \pm 24.4$	$50.5 \pm 7.11$
Single-top $Wt$	$9000 \pm 2530$	$5330 \pm 1880$	$694 \pm 219$	$25.0 \pm 13.8$	$629 \pm 271$	$56.8 \pm 33.9$
Single-top t-channel	$3470 \pm 494$	$1480 \pm 317$	$217 \pm 44.1$	$14.3 \pm 5.95$	$164 \pm 51.3$	$20.8 \pm 8.68$
Other-top	$320 \pm 34.2$	$182 \pm 23.8$	$27.9 \pm 2.70$	$3.43 \pm 0.614$	$46.1 \pm 11.2$	$14.2 \pm 4.99$
Diboson	$392 \pm 198$	$331 \pm 170$	$25.8 \pm 13.6$	$2.31 \pm 2.26$	$36.1 \pm 18.6$	$3.86 \pm 2.27$
$W$ + jets	$7520 \pm 3340$	$4710 \pm 2060$	$444 \pm 196$	$15.2 \pm 8.97$	$429 \pm 191$	$27.5 \pm 12.9$
$Z$ + jets	$1970 \pm 635$	$1180 \pm 395$	$103 \pm 35.2$	$4.66 \pm 1.75$	$121 \pm 41.6$	$9.77 \pm 3.71$
$t\bar{t}H$	$255 \pm 24.9$	$523 \pm 53.9$	$131 \pm 13.8$	$31.5 \pm 4.28$	$317 \pm 33.7$	$119 \pm 16.6$
$tH$	$19.6 \pm 2.43$	$26.7 \pm 3.62$	$10.6 \pm 1.35$	$2.23 \pm 0.324$	$15.7 \pm 2.11$	$5.09 \pm 0.770$
Fakes el	$9540 \pm 5110$	$8980 \pm 4960$	$415 \pm 482$	$11.3 \pm 8.12$	$480 \pm 365$	$3.54 \pm 4.54$
Fakes mu	$2970 \pm 1460$	$1680 \pm 828$	$212 \pm 114$	$0 \pm 0$	$260 \pm 138$	$0 \pm 0$
Total	$331\,000 \pm 7100$	$233\,000 \pm 5810$	$28\,800 \pm 923$	$1220 \pm 75.4$	$31\,900 \pm 843$	$3430 \pm 177$
Data	334813	234053	29322	1210	32151	3459

Table C.1: Post-fit yields for the fit performed with the 200 GeV signal hypothesis. The errors include both statistical and systematic uncertainties. The "other-top" category includes contributions from  $tWZ$ , 4-top,  $t\bar{t}WW$  and the single-top  $s$ -channel.

	5j2b	$\geq 6j2b$	5j3b	5j $\geq 4b$	$\geq 6j3b$	$\geq 6j\geq 4b$
$t\bar{t}$ + light	$246\,000 \pm 15\,400$	$155\,000 \pm 14\,100$	$14\,200 \pm 1660$	$155 \pm 59.5$	$10\,900 \pm 1750$	$298 \pm 146$
$t\bar{t} + \geq 1c$	$38\,100 \pm 13\,100$	$35\,500 \pm 13\,800$	$5640 \pm 1740$	$191 \pm 88.8$	$6640 \pm 2410$	$410 \pm 174$
$t\bar{t} + \geq 1b$	$14\,800 \pm 2870$	$18\,500 \pm 3450$	$7180 \pm 1220$	$763 \pm 103$	$11\,600 \pm 1790$	$2460 \pm 293$
$H^+$	$-54.6 \pm 86.3$	$-115 \pm 168$	$-33.1 \pm 53.3$	$-4.68 \pm 7.72$	$-91.3 \pm 138$	$-30.6 \pm 47.2$
$t\bar{t}W$	$352 \pm 49.5$	$586 \pm 85.6$	$31.0 \pm 4.66$	$0.696 \pm 0.242$	$79.3 \pm 13.6$	$5.59 \pm 1.47$
$t\bar{t}Z$	$390 \pm 47.3$	$777 \pm 99.4$	$77.9 \pm 10.7$	$12.5 \pm 2.29$	$201 \pm 27.2$	$54.2 \pm 8.53$
Single-top $Wt$	$9450 \pm 2530$	$6270 \pm 2200$	$740 \pm 220$	$26.3 \pm 14.1$	$749 \pm 307$	$58.8 \pm 30.8$
Single-top $t$ -channel	$3840 \pm 548$	$1760 \pm 359$	$252 \pm 53.2$	$18.0 \pm 6.34$	$220 \pm 71.4$	$26.2 \pm 8.24$
Other-top	$351 \pm 37.1$	$209 \pm 26.0$	$30.1 \pm 3.31$	$3.56 \pm 0.641$	$50.9 \pm 12.4$	$15.4 \pm 5.51$
Diboson	$431 \pm 217$	$384 \pm 195$	$25.1 \pm 13.1$	$2.97 \pm 2.12$	$40.2 \pm 20.5$	$3.38 \pm 1.92$
$W$ + jets	$9330 \pm 4030$	$5660 \pm 2460$	$477 \pm 210$	$23.1 \pm 11.9$	$532 \pm 238$	$36.9 \pm 17.9$
$Z$ + jets	$2390 \pm 768$	$1420 \pm 466$	$129 \pm 43.9$	$4.08 \pm 1.58$	$151 \pm 51.3$	$11.1 \pm 4.07$
$t\bar{t}H$	$244 \pm 23.4$	$563 \pm 56.0$	$127 \pm 14.1$	$30.2 \pm 4.62$	$337 \pm 36.9$	$127 \pm 20.2$
$tH$	$20.2 \pm 2.47$	$28.5 \pm 3.74$	$10.7 \pm 1.44$	$2.13 \pm 0.363$	$16.7 \pm 2.30$	$5.55 \pm 0.937$
Fakes el	$7220 \pm 4190$	$6800 \pm 4120$	$384 \pm 387$	$11.7 \pm 8.06$	$444 \pm 355$	$3.66 \pm 4.22$
Fakes mu	$2790 \pm 1360$	$1570 \pm 770$	$202 \pm 110$	$0 \pm 0$	$247 \pm 131$	$0 \pm 0$
Total	$336\,000 \pm 6630$	$234\,000 \pm 5090$	$29\,400 \pm 768$	$1240 \pm 57.2$	$32\,200 \pm 682$	$3480 \pm 133$
Data	334813	234053	29322	1210	32151	3459

Table C.2: Post-fit yields for the fit performed with the 800 GeV signal hypothesis. The errors include both statistical and systematic uncertainties. The "other-top" category includes contributions from  $tWZ$ , 4-top,  $t\bar{t}WW$  and the single-top  $s$ -channel.

Figure C.9 and Figure C.10 show the pulls and constraints of some of the most important systematic uncertainties for fits performed with the 200 and 800 GeV signal hypothesis. The black marks correspond to the fits performed on data, as described in Chapter 6. The red marks correspond to fits performed on "blinded" data, i.e. removing the bins of the fitted distributions for which the ratio between the expected signal and the SM background is larger than 10%. Furthermore, the fit is performed with a pure background model, i.e. fixing the signal strength to 0. The two versions show similar pulls and post-fit uncertainties, pointing to a solid MC modelling. Some differences are visible for few  $t\bar{t} + \text{jets}$  nuisance parameters, but not larger than 1 standard deviation.

The blue marks are the result of a fit performed on Asimov pseudo-data (see Section 2.1 for details regarding Asimov datasets). The corresponding uncertainties can be compared to the black error bars in order to validate the post-fit constraints, which appear very similar in the two cases. This confirms that the distributions of the nuisance parameters are well-behaved and that there are no artificial constraints due to statistical fluctuations.

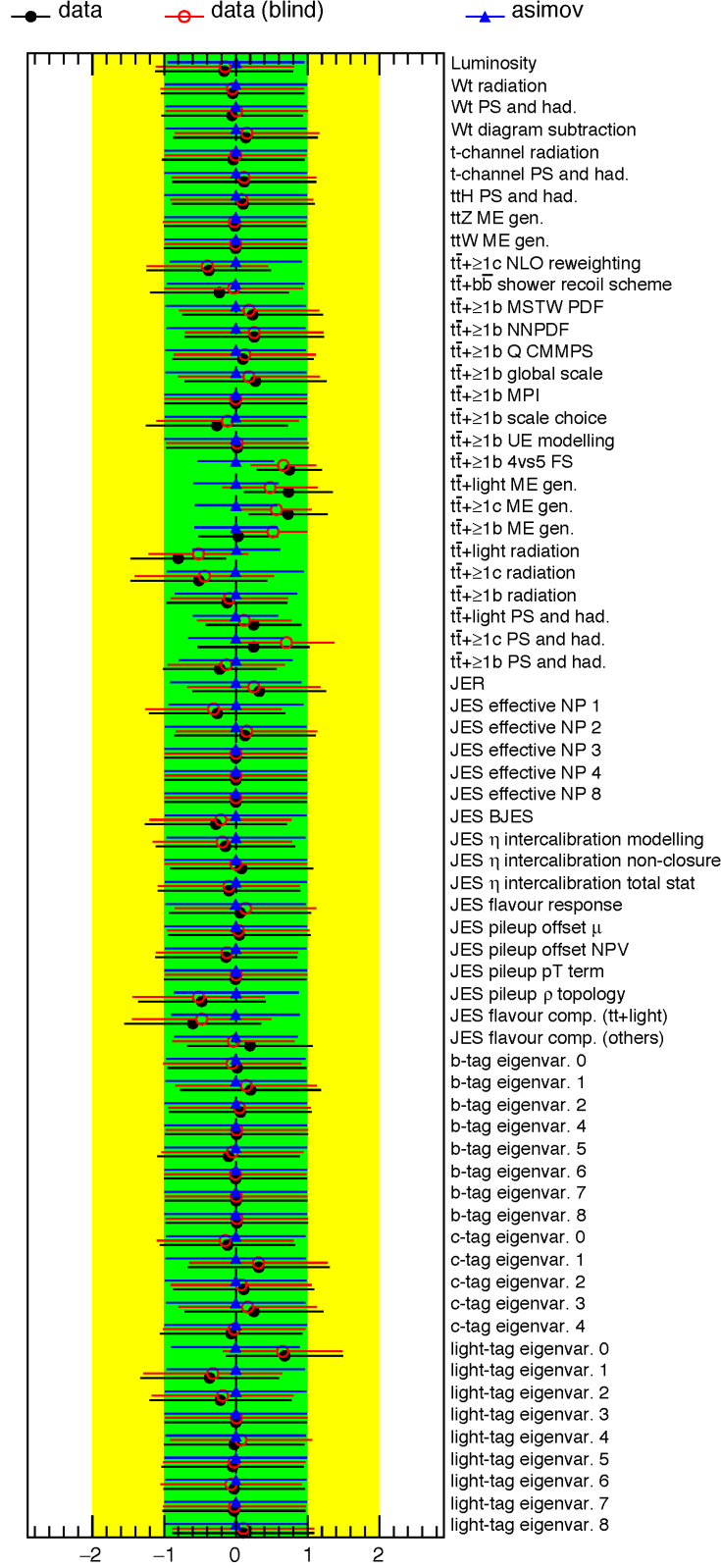


Figure C.9: Pulls and constraints for the fit performed with the 200 GeV signal hypothesis. The black marks correspond to the fit performed on data, as described in Chapter 6. The blue marks correspond to an Asimov fit. The red marks correspond to the fit performed on data while blinding the most signal-like bins ( $S/B > 0.1$ ) and using a background-only model ( $\mu$  fixed to 0).

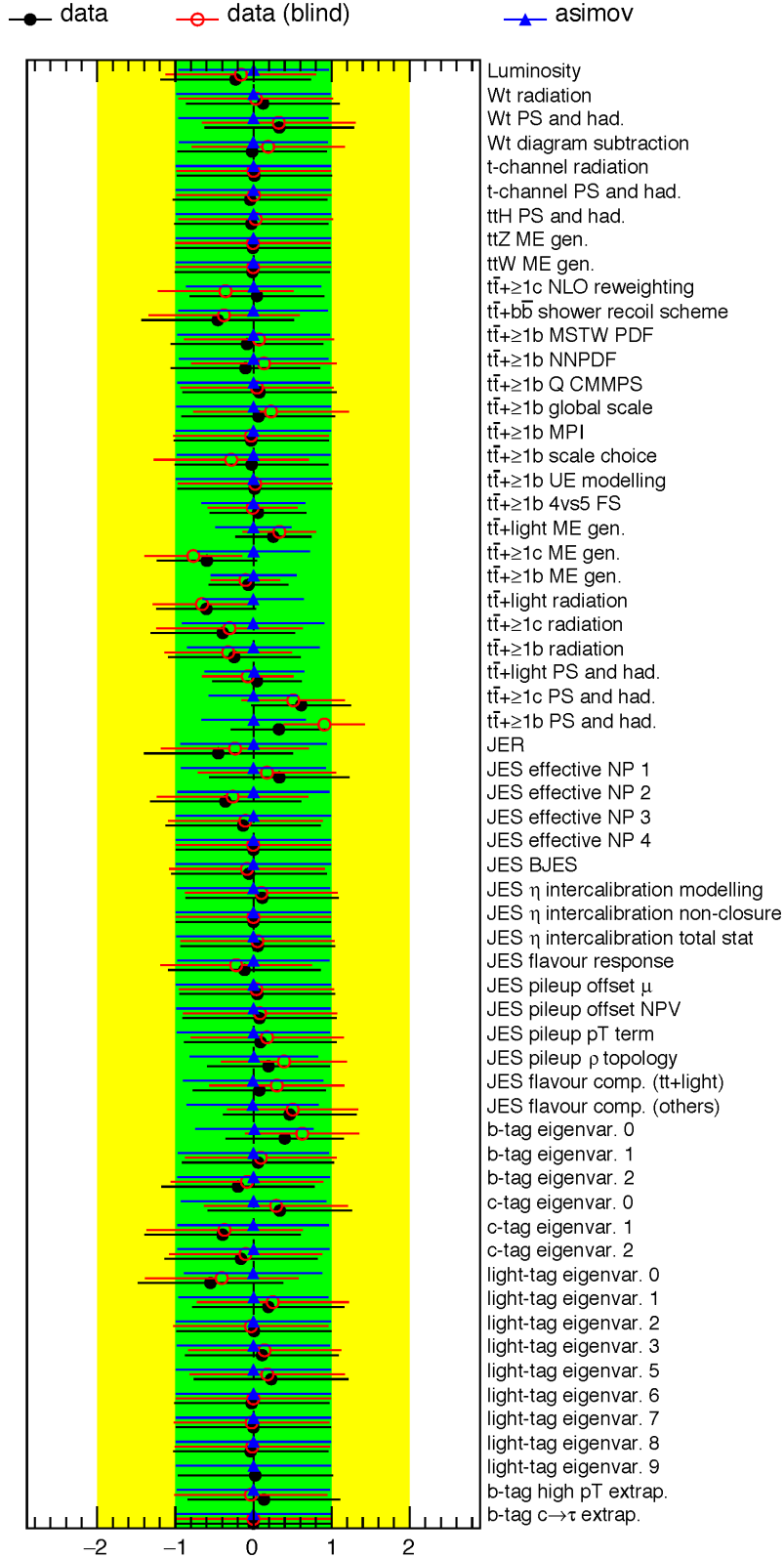
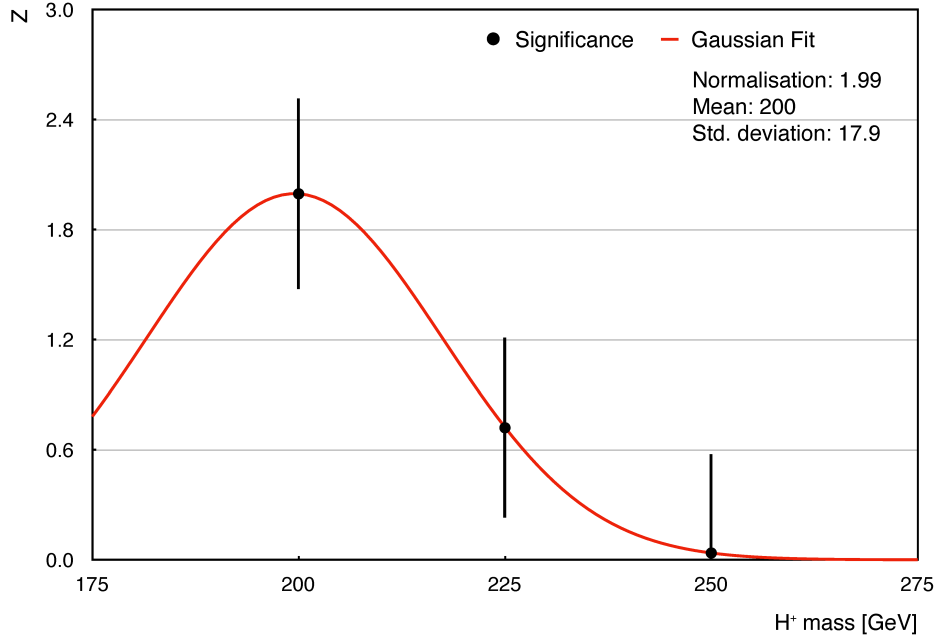


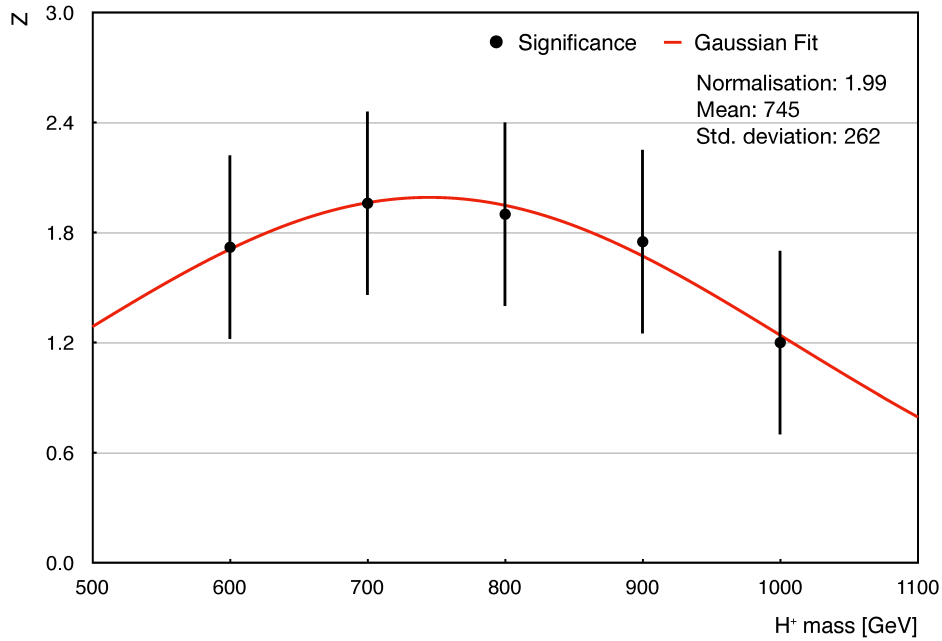
Figure C.10: Pulls and constraints for the fit performed with the 800 GeV signal hypothesis. The black marks correspond to the fit performed on data, as described in Chapter 6. The blue marks correspond to an Asimov fit. The red marks correspond to the fit performed on data while blinding the most signal-like bins ( $S/B > 0.1$ ) and using a background-only model ( $\mu$  fixed to 0).

---

Figure C.11 shows the results of the injection tests performed with the 200 and 800 GeV signal hypothesis. In order to test the stability of the fits between simulated mass points, a signal normalised to the expected upper limit cross section is injected in an Asimov dataset which is then fit with the corresponding model and with models associated to the neighbouring mass points. The fitted signal strength is expected to coincide with the injected signal strength when the model and the Asimov dataset are created with the same signal. When the model makes use of a different signal hypothesis, the fitted signal strength can be different. When the 200 GeV signal is injected in the Asimov dataset, the 225 GeV model is still able to fit a non-zero signal strength. Similarly, for an injected signal of 800 GeV, both the 700 GeV and the 900 GeV models are partially able to detect the signal. This justifies the linear interpolation used for the limit setting in Chapter 6. A Gaussian distribution (plotted as a red solid line) is fit on the observed significance  $Z$ . The standard deviation of the distribution is a proxy for the experimental resolution of the analysis.



(a) 200 GeV.



(b) 800 GeV.

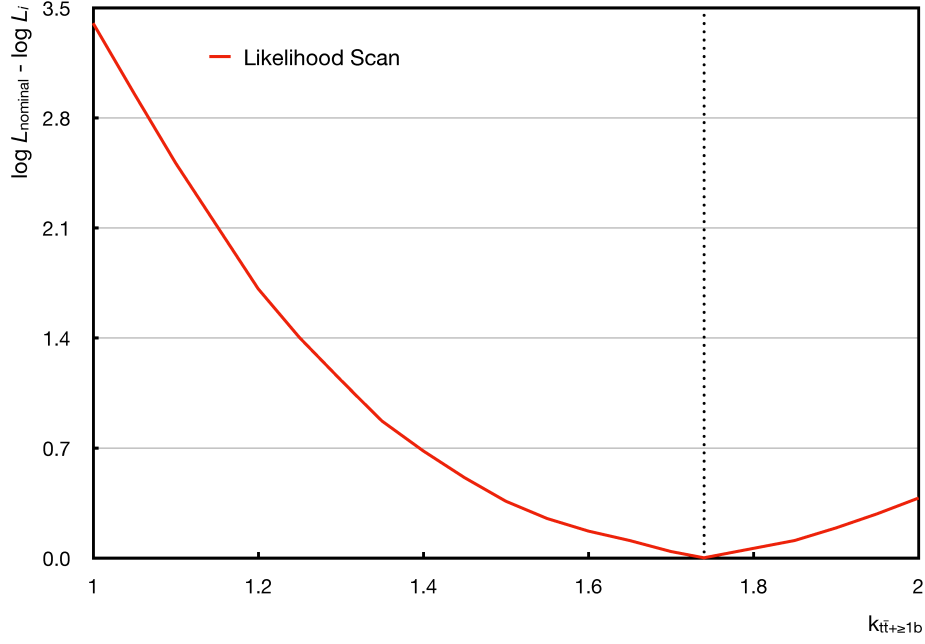
Figure C.11: Injection tests for (a) the 200 and (b) the 800 GeV (b) signal hypotheses.  $Z$  is the significance, defined as the ratio between the signal strength and its error. A Gaussian distribution fit to the observed significance is shown as a red solid line. The normalisation, mean and standard deviation of the Gaussian are reported in the legend. The standard deviation is a proxy for the experimental resolution of the analysis.



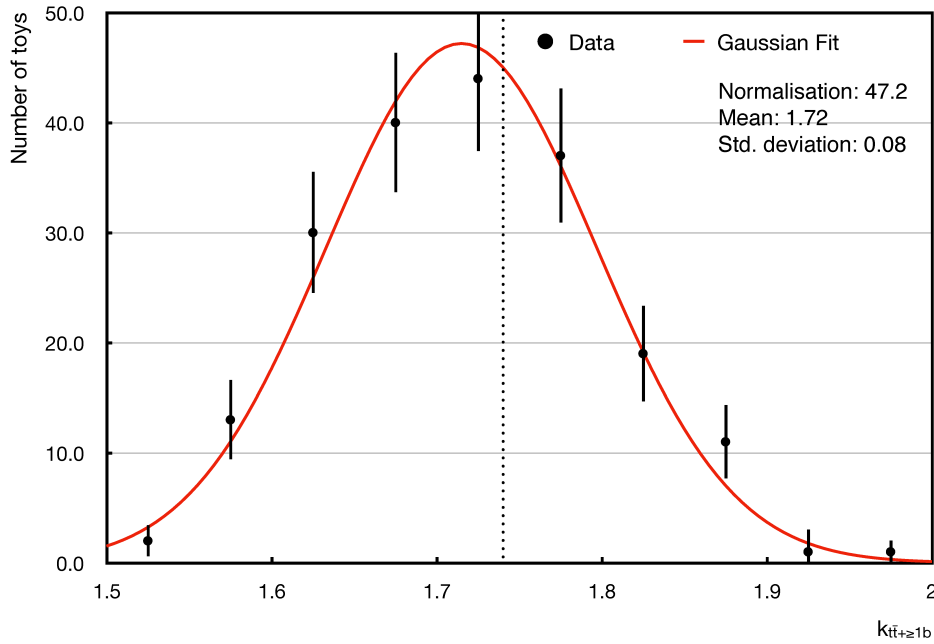
---

Dedicated tests were conducted on the fit performed with the 300 GeV mass hypothesis in order to investigate the large pull observed for the  $t\bar{t} + \geq 1b$  normalisation. Figure C.12 (a) shows the result of a likelihood scan, obtained by fixing  $k_{t\bar{t} + \geq 1b}$  to values between 1.0 and 2.0, and repeating the fit. The test allows to check whether the result of the fit corresponds to a real minimum of the log-likelihood or multiple local minima are available. A single minimum is visible at the nominal value of 1.74.

A bootstrap test was also performed in order to test the dependency of the fit on statistical fluctuations in the distribution of  $t\bar{t} + \geq 1b$  events. Such test consists of generating 200 statistically independent models ("toys") by reweighting each  $t\bar{t} + \geq 1b$  event with a weight extracted from a Poisson distribution with mean 1. The full fit procedure is then repeated for each of the models, obtaining the distribution of Figure C.12 (b) for the  $k_{t\bar{t} + \geq 1b}$  parameter. The vertical line corresponds to the minimum extracted by the nominal fit. The value is close to the peak of the distribution, as expected. Larger distances from the peak would have indicated that the nominal fit was the result of an "unlucky" configuration and that the results were heavily susceptible to fluctuations in the shape of the  $t\bar{t} + \geq 1b$  distribution. The standard deviation of the distribution provide an estimate of the statistical uncertainty of  $k_{t\bar{t} + \geq 1b}$  due to the finite size of the sample.



(a) Likelihood scan.



(b) Bootstrap test.

Figure C.12: (a) Likelihood scan and (b) bootstrap test, for the fit performed with the 300 GeV mass hypothesis. The vertical line corresponds to the nominal value of  $k_{t\bar{t}+\geq 1b}$ . The Likelihood scan was performed by fixing  $k_{t\bar{t}+\geq 1b}$  to values between 1.0 and 2.0, and repeating the fit. The result of the bootstrap test corresponds to the distribution of the  $k_{t\bar{t}+\geq 1b}$  parameter obtained by fitting 200 independent toys.

## Appendix D

### Additional studies on the BDT input variables

This chapter provides a comparison between different sets of input variables used for the training of the classification BDT. Version v2 of the classification BDT is used as reference for the following studies. Such version was trained with 17 input variables, including both the likelihood discriminant and two variables related to the reconstruction BDT (see Section 7.2.2). This version is compared with other two: v3, which adds ten variables associated to the reconstruction of jets with larger radius, and v4, which includes another set of ten kinematic variables on top of the previous ten.

The usage of jets with a larger radius is motivated by the fact that, at high masses, the decay products of the charged Higgs boson are expected to be very boosted, i.e. to have a large momentum. This is due to the large difference between the mass of the  $H^+$  and the masses of the top and bottom quarks, which is converted in kinetic energy. Consequently, the particles produced by the decay of the two quarks can be quite collimated and jets with large radius have a good chance of including the full final state of each of the decays. The reconstruction of the large-radius jets is performed with the anti- $k_t$  algorithm (see Section 4.5.1), with a radius set to  $r = 1.0$ . However, the algorithm is not applied to the topo-clusters of the calorimeters, but on the small-radius jets, which are used for further re-clustering [160, 161]. The benefit of performing the re-clustering on the small-radius jets is that their calibration is inherited by the re-clustered jets and no additional systematic uncertainties are needed. In order to minimise contaminations from pile-up, the re-clustered jets are required to include at least two small-radius jets and to have a mass larger than 50 GeV. Furthermore, small-radius jets are subtracted from each re-clustered jet if their  $p_T$  is lower than 5% of its total transverse momentum. Lastly, all large-radius jets are required to have an  $|\eta| \leq 2$ . Large-radius jets with a mass larger than 100 GeV and a transverse momentum larger than 300 GeV are defined "top-tagged".

The variables used for the training of version v3 include:

- the number of re-clustered jets in the event;
- the number of top-tagged re-clustered jets in the event;
- the invariant mass and  $p_T$  of the leading re-clustered jet;

- the invariant mass and  $p_T$  of the leading top-tagged re-clustered jet;
- the number of small-radius jets contained in the leading re-clustered jet;
- the number of  $b$ -tagged small-radius jets contained in the leading re-clustered jet;
- the invariant mass of the system made of the leading re-clustered jet and the leading  $b$ -jet (not included in the re-clustered jet);
- the invariant mass of the system made of the leading re-clustered jet and the leading light-jet (not included in the re-clustered jet);

The input variables used for the training of version v4 include the set above and the following:

- $\Delta R_{bb}^{\text{Max}(p_T)}$ : the angular distance between the two  $b$ -jets with the maximum  $p_T$ ;
- $M_{jj}^{\text{Min}(\Delta R)}$ : the invariant mass of the closest jets;
- $p_T^{\text{jet}_3}$ : the transverse momentum of the third jet (ordered from the highest to the lowest  $p_T$ );
- $H_4^{\text{all}}$ : the fourth Fox-Wolfram moment calculated using all leptons and jets in the event;
- $M_{bj}^{\text{Max}(p_T)}$ : the invariant mass of the system made of the jet and  $b$ -jet with the maximum  $p_T$ ;
- $M_{bj}^{\text{Min}(\Delta R)}$ : the invariant mass of the system made of the closest jet and  $b$ -jet;
- $H_T^{\text{all}}$ : the sum of the transverse momentum of all jets, leptons and neutrinos in the event;
- $M_{jj}^{\text{Max}(p_T)}$ : the invariant mass of the system made of the jets with the maximum  $p_T$ ;
- $\Delta R_{lj}^{\text{Min}(\Delta R)}$ : the angular distance between the lepton and its closest jet;
- $\Delta \eta_{jj}^{\text{Max}(\Delta \eta)}$ : the maximum difference in pseudorapidity between jets;

The three versions were trained (and tested) inclusively in the  $\geq 5j \geq 3b$  region, using 300 trees, with a maximum depth of 5. The performances of the three classification BDTs are compared in Figure D.1 (a) and Figure D.1 (b). The addition of the re-clustered jet variables has almost no impact on the separation power of the BDT, while some change is observed for version v4. The relative improvement is however smaller than 5% in both cases. The small change in performance is most probably due to the limited number of training events. This is particularly true for the variables related to the re-clustered jets. In fact, given the requirements on the mass and  $p_T$  of the large-radius jets, the percentage of background events including one or more re-clustered jet is very low ( $\sim 8\%$ ). Furthermore, the reconstruction BDT is already capable of performing a "re-clustering" and it has the advantage of being able to exploit the full kinematics of the small-radius jets, instead of their angular separation only.

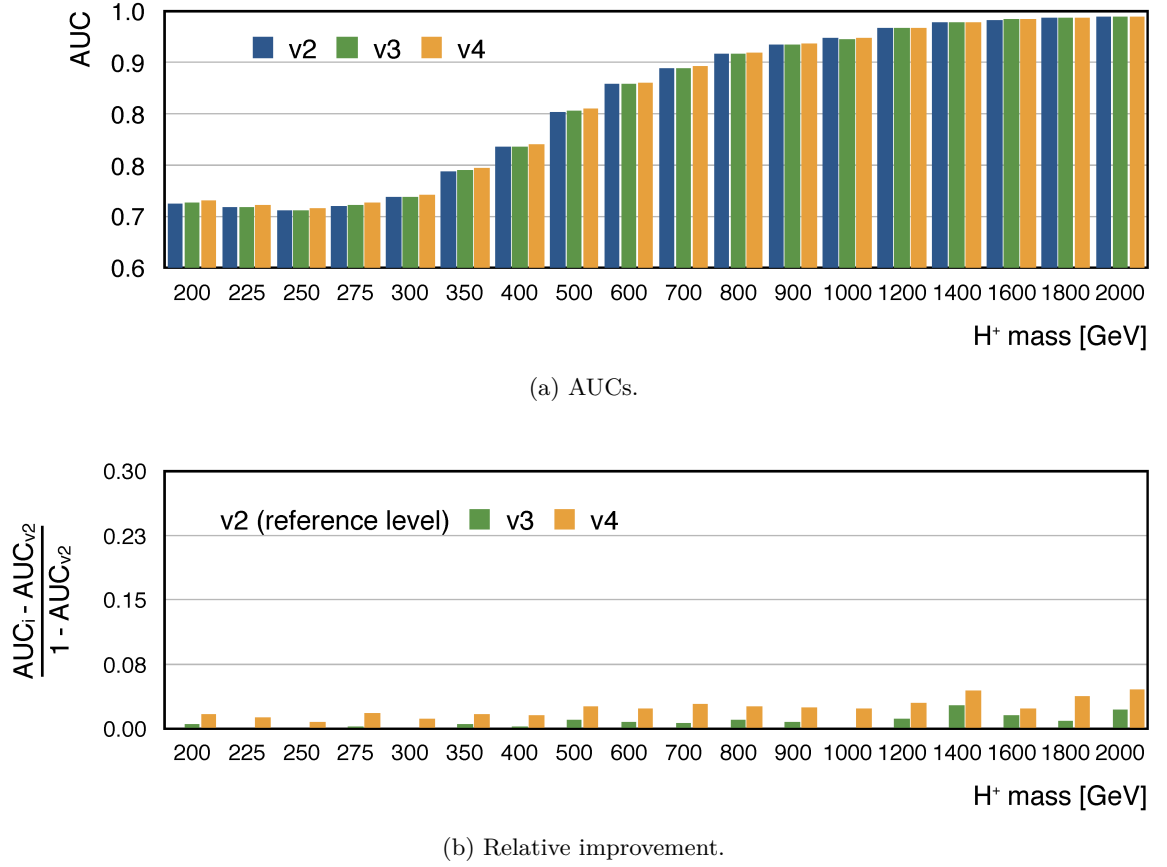


Figure D.1: Performance of the classification BDTs. Sub-figure (a) presents the AUC values obtained with versions v2, v3 and v4 of the BDT, while sub-figure (b) shows the relative improvement provided by version v3 and v4 with respect to v2. The AUCs are computed comparing the BDT distribution of the signal with the one produced by the full set of SM backgrounds, in the  $\geq 5j \geq 3b$  region.



## Appendix E

### Systematic uncertainties and adversarial networks

As mentioned in Section 6.4, the sensitivity of the analysis is limited by the large impact of the systematic uncertainties. The largest effects are produced by the Monte Carlo modelling systematic uncertainties, in particular by the ones related to the  $t\bar{t}$  + jets background, which contribute with both shape and normalisation. Reducing the difference in normalisation between the distribution of the nominal sample and the one of the systematic variations is only possible with a better understanding of the model. On the other hand, there are techniques that allow to control the shape effects.

The variables selected to perform the fit (or to train the classification BDT) are usually the ones with the largest signal-to-background separation. The separation power is however estimated with the nominal samples and does not take into account the shape of the systematic uncertainties. In principle, variables with small or no dependence on the nuisance parameters could outperform variables with a large separation power. To understand how big the impact of the shape of the systematic uncertainties is, one can take the fit performed with a given signal hypothesis and compare the signal strength obtained with the full set of systematic uncertainties (considering both shape and normalisation), with the result obtained dropping all shape contributions. At 400 GeV, a shape-plus-normalisation (Asimov) fit performed with the default classification BDT provides a signal strength  $\mu = 0 \pm 0.54$ . When all shape effects are removed, the signal strength becomes  $\mu = 0 \pm 0.37$ .

Removing all shape contributions is unrealistic, however creating a discriminant which combines the separation power of multiple variables, while staying invariant with respect to (some) nuisance parameters, could bring some real improvement to the sensitivity of the analysis. Recent studies have shown that adversarial neural networks (ANNs) are suitable for this purpose [162]. The following is a summary of the preliminary results obtained with ANNs in the  $H^+$  analysis.

A nice introduction to neural networks can be found in [163, 164]; this section only provides a quick overview of the fundamental concepts. Neural networks are computational models partly inspired by the architecture of human brains. They are based on a collection of nodes, or "neurons", which are organised in layers. Figure E.1 provides a schematic representation of a simple neural network composed of three layers. The first layer is known as the "input" layer, while the last is the "output" layer. All the layers in between are "hidden" layers. All neurons

receive a certain number of inputs  $x_1 \dots x_n$  and produce an output  $y = f(w_1x_1 + \dots + w_nx_n + b)$ . The inputs of the first layer are the training variables, while neurons of the hidden and output layers operate on the output of previous nodes. The function  $f$  is called "activation" function and it depends on the trainable parameters of the model, the weights  $w_i$  and the bias  $b$ .

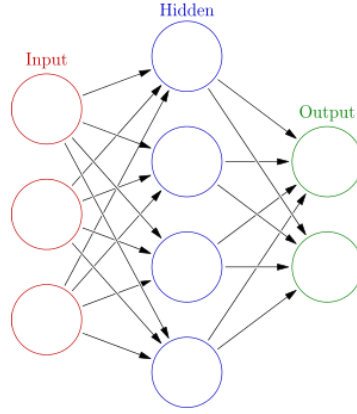


Figure E.1: Schematic representation of a neural network [165].

When used for classification problems, neural networks are trained to associate events to specific classes (signal or background, for example). The training proceeds as follows:

- for each event, the training variables are passed to the first layer of the network. Each node processes its inputs and passes the result to the following layer, until the information arrives to the output nodes.
- the predicted outputs are compared to the expected values and all weights and biases are corrected in order to produce a result that is closer to the expected. The update of a parameter  $\theta$  (weight or bias) proceeds accordingly to a gradient descent method:

$$\theta^{n+1} = \theta^n - \alpha \frac{\partial L(\mathbf{x}_i, \theta^n)}{\partial \theta}, \quad (\text{E.1})$$

where  $\theta^n$  is the array of parameters at the iteration  $n$ ,  $L$  is the loss function and  $\alpha$  is the learning rate.

- The loss function is used to evaluate the difference between  $\mathbf{y}_i$  and  $\hat{\mathbf{y}}_i$ , the expected and computed outputs, for a given input  $\mathbf{x}_i$ . An example of a very common loss function is the mean squared error:

$$L = \frac{1}{2N} \sum_i^N (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2. \quad (\text{E.2})$$

The derivative of the loss function  $\frac{\partial L(\mathbf{x}_i, \theta^t)}{\partial \theta}$  is computed first on the weights of the last node, where the expected output  $\mathbf{y}_i$  is known, and then back-propagated to the previous



nodes.

- The learning rate  $\alpha$  corresponds to the size of the step used by the gradient descent method to move towards the minimum of the loss function.

While the architecture shown in Figure E.1 is very simple, neural networks used in real applications can be much more complicated and include a much larger number of layers. This is often the case for adversarial neural networks. A schematic representation of the ANN used for the following studies is shown in Figure E.2. A classifier  $f$ , trained on the input variables  $X$ , produces an output  $f(X, \theta_f)$  that depends on the trainable parameters  $\theta_f$ . The corresponding loss function is  $L_f(\theta_f)$ . The output of the classifier corresponds to the probability for each event of being assigned to the signal class. A second neural network is then trained on the output of the first, to recognise if the set of input variables  $X$  was produced by a nominal sample  $\gamma_1$  or by a systematic variation  $\gamma_2 \dots \gamma_n$ . The parameters of the adversarial network  $\theta_r$  are tuned by means of the loss function  $L_r(\theta_r, \theta_f)$ . The goal of the training is to minimise the loss of the classifier  $L_f(\theta_f)$ , while maximising the loss of the adversarial term  $L_r(\theta_r, \theta_f)$ . This would produce a discriminant that is invariant with respect to the nuisance parameters used in the training, while preserving a high separation power between signal and background. In practical terms, this is achieved by defining a new loss function  $L_t = L_f - \lambda L_r$  that is minimised during the training process.  $\lambda$  is a parameter that adjusts the trade-off between the classifier separation-power and the (in)dependence with respect to the systematic uncertainties [162].

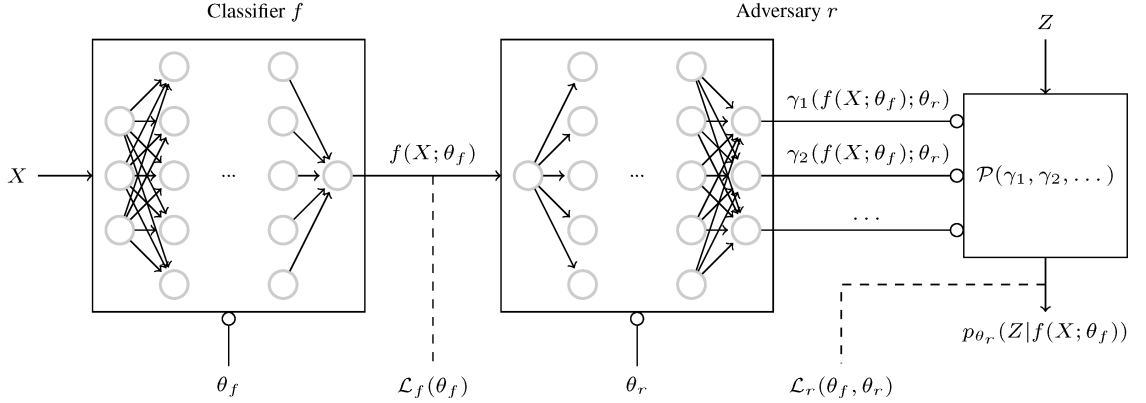


Figure E.2: Schematic representation of the adversarial neural network used for the training [162].

The method was tested assuming a signal mass of 400 GeV. The adversarial neural network was trained to distinguish between the signal and the inclusive  $t\bar{t}$  + jets background. The training variables are the same used for the BDT training in Section 6.2.1, with the addition of the number of jets and the number of  $b$ -jets. The architecture used for the neural network of the classifier  $f$  consists of 2 dense hidden layers of  $10 \cdot N_v$  and  $5 \cdot N_v$  nodes, where  $N_v$

is the number of input variables. During the training, 25% of the nodes were randomly deactivated at each iteration to reduce overtraining effects. A "sigmoid" activation function is used for the output layer, while a "relu" activation function is used for each of the hidden layers. The adversarial term consists of a single hidden layer of  $10 \cdot N_v$  nodes and a "softmax" activated output layer with 3 nodes. A stochastic gradient descent optimiser is used for both the classifier and adversarial terms, with a learning rate set to 0.01. Technical details regarding the activation functions and the library used for the training (Keras) are available in Ref. [166].

Two systematic samples are used together with the nominal  $t\bar{t} + \text{jets}$  sample during the training: the SHERPA sample, associated to the ME generator uncertainty, and the POWHEG + Herwig 7 sample, which is used for the PS and hadronisation uncertainty. The following results compare the performances obtained with four adversarial neural networks trained with  $\lambda = 0, 10, 20$  and 50.  $\lambda = 0$  means that the adversarial term is removed.

The AUCs obtained with the four configurations, using exclusively the nominal samples, are:

- AUC ( $\lambda = 0$ ): 0.733;
- AUC ( $\lambda = 10$ ): 0.733;
- AUC ( $\lambda = 20$ ): 0.730;
- AUC ( $\lambda = 50$ ): 0.714.

The NN without the adversarial term has the best performance, as expected. It is however more interesting to understand what happens to the systematic uncertainties: Figures E.3 to E.6 show the distribution of the systematic uncertainties related to the PS and ME generators obtained with the different discriminants. The black distribution corresponds to the nominal POWHEG + PYTHIA 8 sample, while the red distribution corresponds either to the SHERPA sample or to the POWHEG + HERWIG sample. The blue distribution is obtained by symmetrisation. Both sides of the uncertainties are smoothed. The difference in shape between the nominal and systematic distributions is evaluated by mean of a  $\chi^2$  test<sup>1</sup>. The plots report the corresponding  $p$ -value. The relative shape variation with respect to the nominal sample is consistently reduced for large values of  $\lambda$ . The version with  $\lambda = 50$  seems to produce the optimal shape for both systematic uncertainties. As expected, the method does not affect the normalisation of the systematic variations, which is the same in all cases.

It is also possible to estimate the effect of the different trainings on the signal strength sensitivity. A very simple model was used for this test: the fit was performed in the  $\geq 5j \geq 3b$  region, using an inclusive  $t\bar{t} + \text{jets}$  category and applying only the two systematic uncertainties used in the training. The resulting signal strengths are:

- $\mu$  ( $\lambda = 0.0$ ):  $0 \pm 0.39$ ;

---

<sup>1</sup>All distributions are scaled to the same normalisation before performing the  $\chi^2$  test in order to account for shape effects only.

- 
- $\mu (\lambda = 10.0)$ :  $0 \pm 0.37$ ;
  - $\mu (\lambda = 20.0)$ :  $0 \pm 0.37$ ;
  - $\mu (\lambda = 50.0)$ :  $0 \pm 0.35$ .

The NN with the largest  $\lambda$  is the one with the lowest error. The results are confirming what was observed in [162]: the sensitivity of the analysis can be improved by creating a discriminant which is independent from the nuisance parameters. However, some remarks need to be made:

- the full model described in Section 6 includes six regions and three separate  $t\bar{t} + \text{jets}$  categories, while only one region and one inclusive category was used for the test above. A much more refined training would be required to suppress the shape of the systematic uncertainties in the different regions. Furthermore, neural networks require a large number of events to produce robust models and some of the signal regions, as  $5j \geq 4b$ , only include few thousands events. Training the NNs separately per signal region is therefore not trivial.
- the gain obtained with ANNs is tiny and could be smoothed away when including more systematic uncertainties in the fit. In principle, it is possible to add more NPs to the training but the larger the number, the more complicated it is for the adversarial term to distinguish between them.

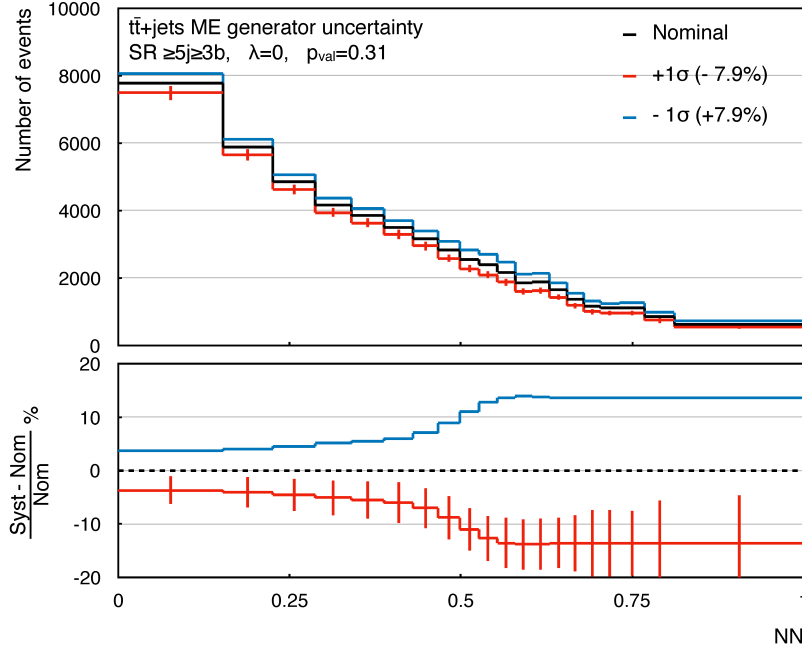
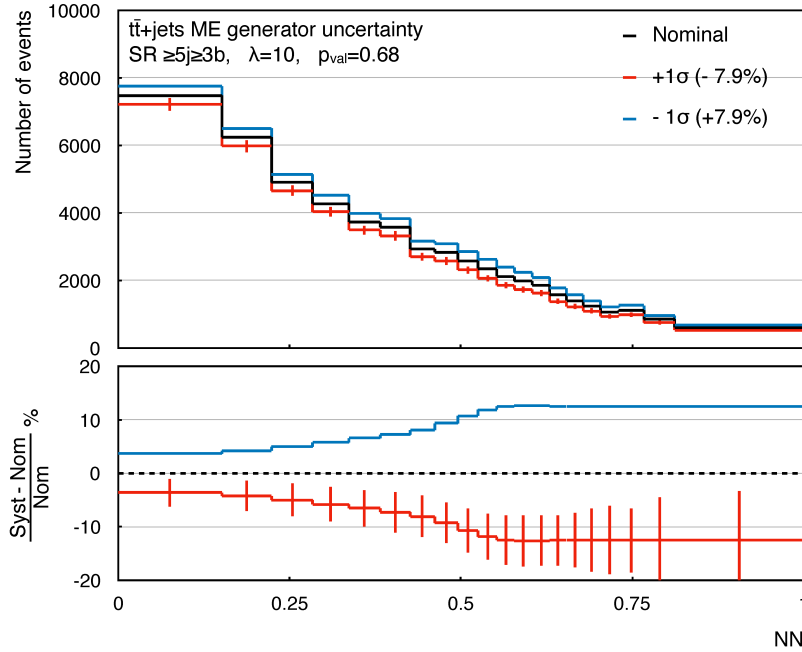
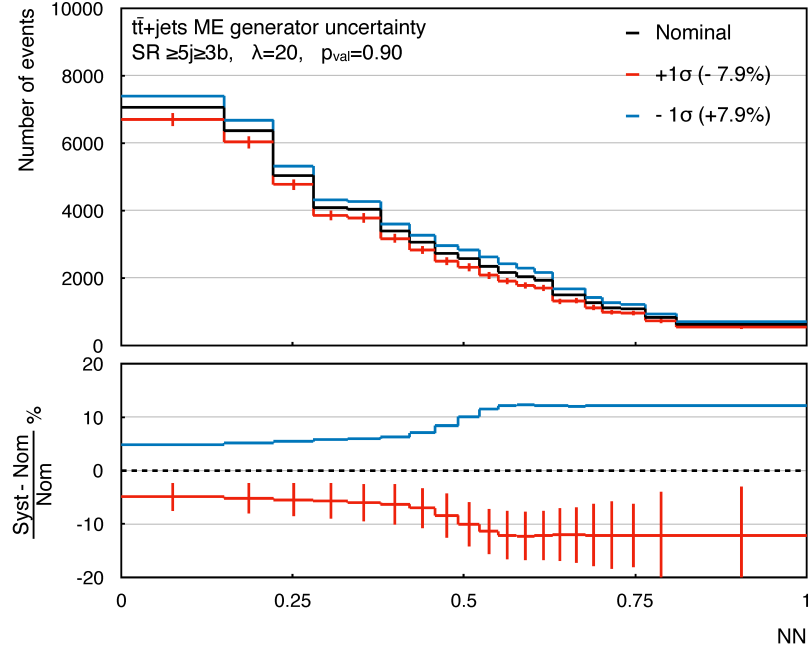
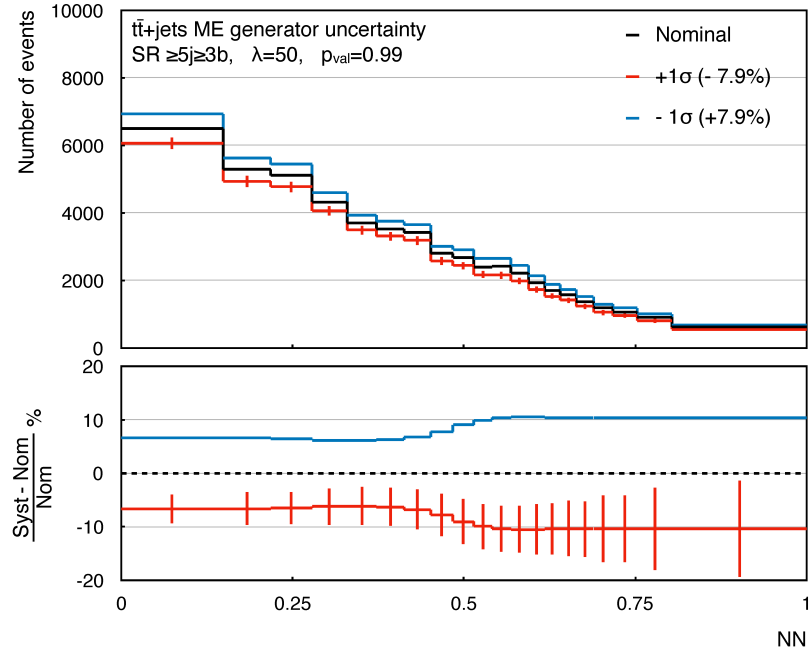

 (a) ME generator uncertainty,  $\lambda = 0$ .

 (b) ME generator uncertainty,  $\lambda = 10$ .

Figure E.3: Distribution of the NN discriminants for the ME generator uncertainty of the  $t\bar{t} + \text{jets}$  sample. Sub-figure (a) shows the NN trained without the adversarial term, while sub-figure (b) correspond to the NN trained with  $\lambda = 10$ . The black line corresponds to the nominal sample, while the red and blue lines correspond to the two sides of the systematic variation. The  $p$ -value shown in the legend is an estimate of the shape compatibility between distributions. Given the symmetry between the two sides of the systematic uncertainty, the statistical error is only reported for the  $+1\sigma$  variation.



(a) ME generator uncertainty,  $\lambda = 20$ .



(b) ME generator uncertainty,  $\lambda = 50$ .

Figure E.4: Distribution of the NN discriminants for the ME generator uncertainty of the  $t\bar{t}$  + jets sample. Sub-figure (a) shows the NN trained with  $\lambda = 20$ , while sub-figure (b) correspond to the NN trained with  $\lambda = 50$ . The black line corresponds to the nominal sample, while the red and blue lines correspond to the two sides of the systematic variation. The  $p$ -value shown in the legend is an estimate of the shape compatibility between distributions. Given the symmetry between the two sides of the systematic uncertainty, the statistical error is only reported for the  $+1\sigma$  variation.

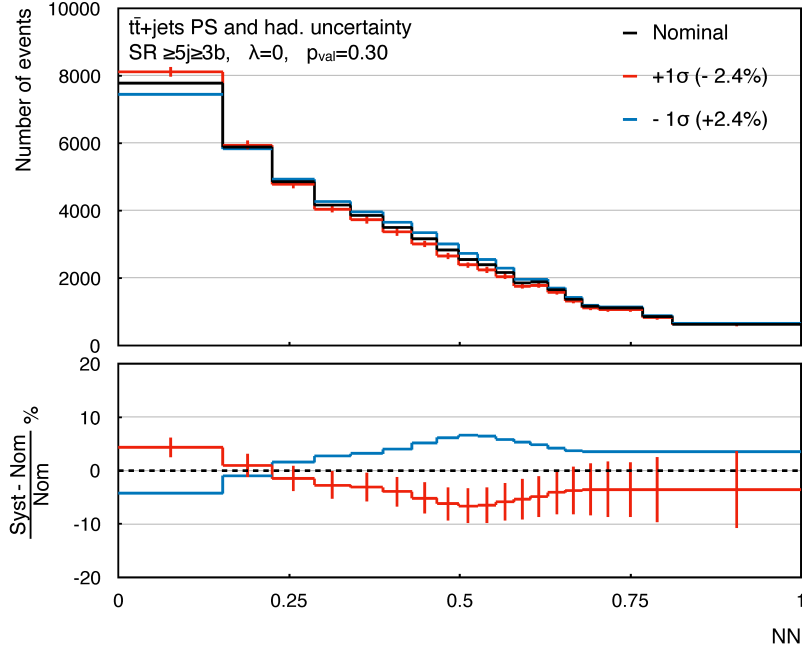
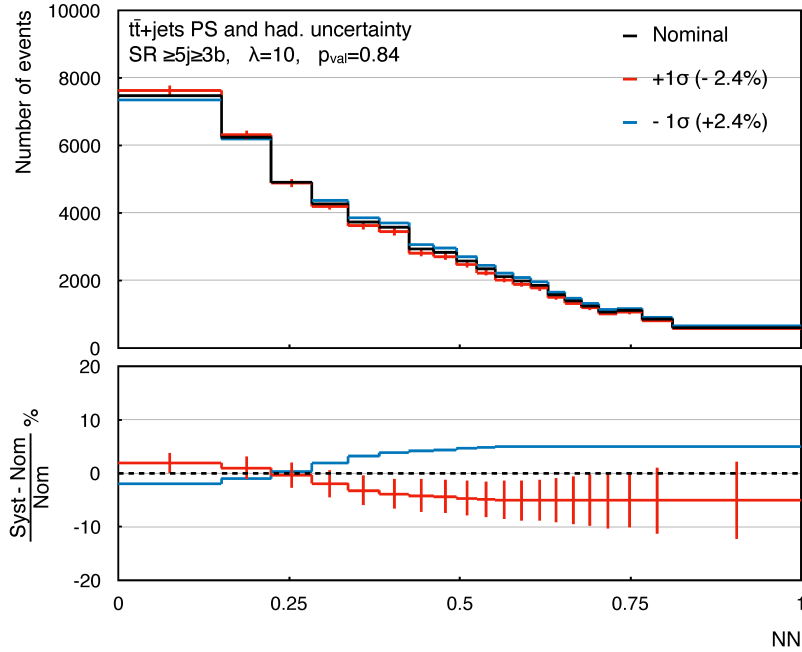
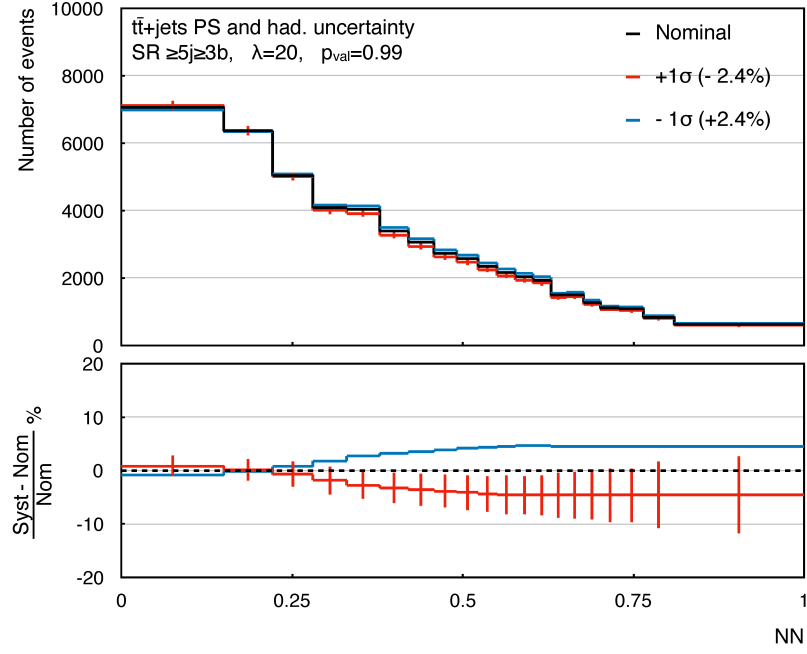
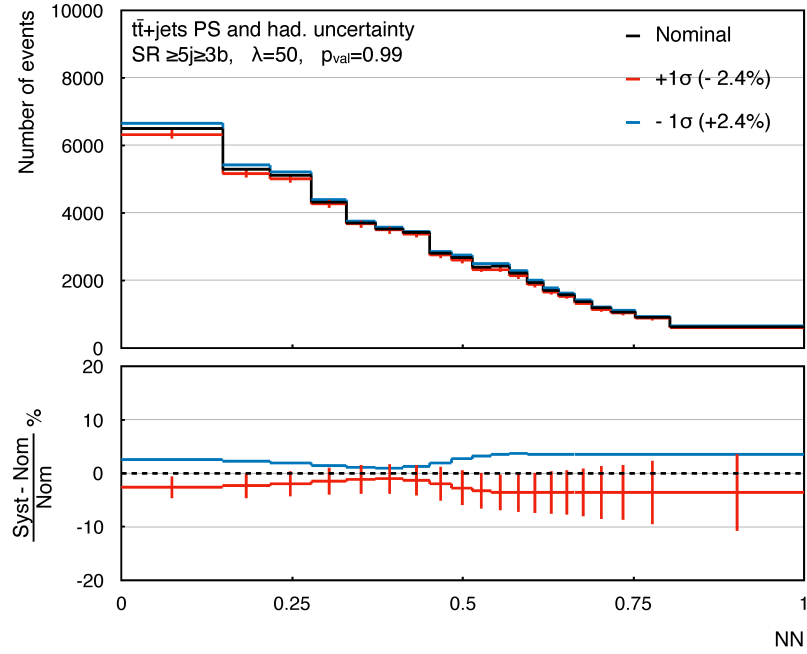

 (a) PS and had. uncertainty,  $\lambda = 0$ .

 (b) PS and had. uncertainty,  $\lambda = 10$ .

Figure E.5: Distribution of the NN discriminants for the PS and hadronisation uncertainty of the  $t\bar{t}$  + jets sample. Sub-figure (a) shows the NN trained without the adversarial term, while sub-figure (b) correspond to the NN trained with  $\lambda = 10$ . The black line corresponds to the nominal sample, while the red and blue lines correspond to the two sides of the systematic variation. The  $p$ -value shown in the legend is an estimate of the shape compatibility between distributions. Given the symmetry between the two sides of the systematic uncertainty, the statistical error is only reported for the  $+1\sigma$  variation.



(a) PS and had. uncertainty,  $\lambda = 20$ .



(b) PS and had. uncertainty,  $\lambda = 50$ .

Figure E.6: Distribution of the NN discriminants for the PS and hadronisation uncertainty of the  $t\bar{t}$ +jets sample. Sub-figure (a) shows the NN trained with  $\lambda = 20$ , while sub-figure (b) correspond to the NN trained with  $\lambda = 50$ . The black line corresponds to the nominal sample, while the red and blue lines correspond to the two sides of the systematic variation. The  $p$ -value shown in the legend is an estimate of the shape compatibility between distributions. Given the symmetry between the two sides of the systematic uncertainty, the statistical error is only reported for the  $+1\sigma$  variation.