Universität Bonn

Physikalisches Institut

A gateway to new physics: direct measurement of the top Yukawa coupling to the Higgs boson

Nello Bruscino

The top quark has the largest Yukawa coupling in the Standard Model, is the main contribution to the Higgs mass m_H corrections and defines the evolution of the Higgs effective potential with the energy, together with m_H. It can be directly measured through the associated production process $pp \to t\bar{t}H$. Two searches for this process are described in this thesis, using data collected with the ATLAS detector at the Large Hadron Collider.

The fully hadronic analysis is performed with data corresponding to an integrated luminosity of 20.3 fb^{-1} at a centre-of-mass energy of $\sqrt{s} = 8$ TeV and uses a boosted decision tree algorithm to discriminate between signal and background: the dominant multijet background is estimated using a data-driven method. An upper limit of 6.4 (5.4) times the Standard Model cross section is observed (expected) at 95% confidence level and a best-fit value of 1.6 ± 2.6 for the signal strength $\mu_{t\bar{t}H} = \sigma^{\text{obs}}_{t\bar{t}H}/\sigma^{\text{SM}}_{t\bar{t}H}$ is measured.

The multilepton analysis uses data collected at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 36.5 fb^{-1}. Events with exactly three leptons are selected and a boosted decision tree is also exploited. The major sources of background are estimated using a simultaneous fit technique, which determines their normalisations in three control regions. An upper limit of 2.3 (1.7) times the Standard Model cross section is observed (expected) at 95% confidence level and a best-fit value of $0.68^{+0.89}_{-0.68}$ for $\mu_{t\bar{t}H}$ is measured.

Physikalisches Institut der
Universität Bonn
Nussallee 12
D-53115 Bonn

BONN-IR-2017-04
Mai 2017
ISSN-0172-8741

A gateway to new physics: direct measurement of the top Yukawa coupling to the Higgs boson

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

von
Nello Bruscino
aus
San Giuseppe Vesuviano, Neapel (Italien)

Bonn, 23.03.2017

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn `http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

1. Gutachter: Dr. Markus Cristinziani
2. Gutachter: Prof. Dr. Norbert Wermes

Tag der Promotion: 19.05.2017
Erscheinungsjahr: 2017

Contents

CHAPTER 1

Introduction

The Standard Model (SM) of elementary particles elegantly describes the fundamental components of matter and characterises their dynamics and interactions at a high level of accuracy, with the exception of gravity. The theory predicts the existence of a scalar particle, the Higgs boson, to explain how the particles acquire mass. Many experiments tried to spot this elusive particle over the last 50 years, without success. Finally, in 2012 both, the ATLAS and CMS experiments at the LHC announced the discovery of a new particle with a mass around 125 GeV, compatible with the theoretical and experimental expectations for a SM Higgs boson. This was an important milestone towards the comprehension of nature and opened the way for many dedicated searches aiming for the characterisation of the novel particle (mass width, spin-parity, decay and production fractions, couplings with other SM particles). Any deviation from the SM predictions would reveal the presence of new physics processes - the evidence of a more comprehensive theory beyond the Standard Model.

To date, no significant deviations from the SM predictions are observed and no convincing signal in favour of the existence of such physics beyond the Standard Model is found. This begs for two questions:

> “Have we got, eventually, the ultimate theory of nature?”
> “If not, where should we search for new physics?”

The answer to the first question is easy and negative. Even though many measurements endorse the validity of the Standard Model, the theory does not address a number of important open questions, like the observation of neutrino flavour oscillations, the matter-antimatter asymmetry in the universe and the existence of dark matter. Moreover, the theory does not solve the so-called “hierarchy problem”, also known as naturalness, according to which the Higgs mass receives radiative corrections through boson and fermionic loops, which are 16 orders of magnitude larger than the Higgs mass at tree level (the largest contribution arising from the coupling to the top quark). All these problems seem to point to a more general theory, of which the Standard Model is merely a part.

The second question, on the other hand, does not have an obvious answer. What is clear from the theory point of view, is that some type of new physics should appear at the Planck scale, where gravity becomes important. However, a possible lower scale of new physics can be inferred by assuming that there is no new physics up to the Planck scale and see if the Standard Model runs into any contradiction. In this context, a critical observable is the top Yukawa coupling to the Higgs boson: it plays an important role in the one-loop corrections of the Higgs boson self-coupling constant; at large scales the self-coupling constant becomes negative, changing the vacuum structure of the Higgs effective potential, whose electroweak minimum can become unstable, and demanding for an extension of the model to solve the problem.

The measurement of the associated production of the Higgs boson with a top quark pair ($t\bar{t}H$), is, then, of particular importance, since it provides a direct access to the top Yukawa coupling. No evidence of such process has been observed yet at the LHC, due to its low production cross section (two orders of magnitude smaller than the inclusive Higgs cross section) and the complex final states. The work documented in this thesis concerns the search for the $t\bar{t}H$ production in two different final states: the fully hadronic and the multilepton final states.

The fully hadronic channel is the $t\bar{t}H$ channel with the largest branching fraction, taking advantage of the the top-quark decay mode into hadrons ($\approx 60\%$) and the Higgs boson into a pair of b quarks ($\approx 60\%$). The absence of triggering leptons, however, brings the multijet processes into being an overwhelming background, with a cross section six orders of magnitude higher than that of $t\bar{t}H$, even in the most sensitive regions. The multilepton channel, on the other hand, has a lower branching ratio, mostly targeting the $H \to WW^*$, ZZ^* and $\tau\tau$ decay modes, but a much higher signal purity. Albeit substantial dissimilarities in terms of selection and background composition, multilepton and fully hadronic searches follow a similar strategy: for both analyses, a multivariate classifier has been developed, in order to provide better distinction between the signal and background than single kinematic variables. Furthermore, various signal and control regions are exploited to reduce the impact of background uncertainties on the search sensitivity.

The work of this thesis has been performed within the ATLAS Collaboration, one of the largest scientific collaborative efforts of CERN. The content is organised in five Chapters; the specific contribution given by the author is reflected in the level of detail provided within each Chapter.

Chapter 2 gives an overview on the theoretical framework of the Standard model and describes the Higgs mechanism of symmetry breaking. An outline about the top-quark physics is also given, together with a description of the Higgs boson discovery and the main mechanisms of production and decay. Finally, the crucial role that the top Yukawa coupling plays to infer about a possible scale of new physics and the evolution of the effective Higgs potential is detailed.

Chapter 3 introduces the LHC accelerator and provides a description of the ATLAS experiment and the algorithms used to identify and reconstruct physics objects (electrons, muons, jets, ...), eventually employed for physics analyses.

In Chapter 4, an introduction to the analysis strategy of the two $t\bar{t}H$ searches is provided: the full analysis chain is outlined, and the motivations for employing a multivariate approach are given. The Chapter also provides a description of the frequentist formalism adopted for the profile likelihood approach and the tools to interpret the final results.

Finally, the two $t\bar{t}H$ searches, the main topic of this thesis, are described in detail in two dedicated Chapters, 5 and 6. Each Chapter provides a brief preface about the advantage and the shortcomings of the analysis and the description of the dataset and the MC simulations used. The definition of the physics objects, the event selection and the background composition are then discussed. The multivariate approach used to extract signal informations is eventually described and the final results are presented.

CHAPTER 2

The Standard Model Higgs boson

This Chapter gives an outline of the Standard Model of particle physics. It is a gauge theory describing the fundamental components of matter and their interactions, and represents our current understanding of the world at the level of elementary particles. The Standard Model was formulated in the 1960s and 70s [1–6], and since then it has been tested to an ever increasing level of precision. This Chapter also provides a description of the electroweak symmetry breaking mechanism and an overview of the top-quark physics. A description of the theoretical constraint to the Higgs boson mass and its discovery is given together with an outline of the main mechanisms of production and decay. Finally, the reasons of the crucial role played by the top Yukawa coupling to infer about a possible scale of new physics and its connection to the evolution of the effective Higgs potential are discussed.

2.1 Basic concepts

The Standard Model describes the fundamental building blocks of nature: matter is composed by three families of quarks, three of leptons and interact through mediators referred to as gauge bosons. The leptons and the quarks are fermions (spin-1/2 particles) obeying to Fermi-Dirac statistics; the force mediators are instead integer spin particles and thus obey to Bose-Einstein statistics. An antiparticle is associated to each elementary particle. Each family of leptons is composed by a charged lepton and its neutrino. The charged leptons differ in their masses, which are increasing in every generation with respect to the previous one. Table 2.1 shows the three lepton generations along with their mass and charge. The lightest charged leptons, the electron (e), is stable, while the second and third generation of charged leptons, the muon (μ) and the tau (τ), are unstable and decay to other particles.

There are six "flavours" of quarks with fractional electric charge forming three generations of increasing mass: "up" and "down", "charm" and "strange", "top" and "bottom", denoted by the first letter of their names. Quarks are the fundamental constituents of hadrons and come in three different colour states: "red" (R), "blue" (B), "green" (G); the colour for the strong interactions plays the same role of the charge for the electromagnetic interactions. Their mass values or mass limits are shown in Table 2.1.

The gauge bosons are the mediators of the fundamental interactions in nature (the electromagnetic, the weak, the strong and the gravitational). The electromagnetic force is carried by spin-1 photons (γ) and acts between electrically charged particles. The weak interaction is approximately 1000 times weaker than the electromagnetic force and is ruled by three gauge vector bosons, $W^{\pm}$ and Z, which are massive with spin 1; the weak force is responsible for phenomena like nuclear β-decays and absorption and emission of neutrinos. The strong interaction, responsible of holding together nuclei, is roughly 100 times stronger than the electromagnetic force. Its gauge bosons acting between quarks are eight

Generation	lepton/quark	charge $[Q/e]$	mass [GeV]
First	e	-1	0.511×10^{-3}
	ν_e	0	$< 0.225 \times 10^{-6}$ (95% C.L.)
Second	μ	-1	0.106
	ν_μ	0	$< 0.19 \times 10^{-3}$ (90% C.L.)
Third	τ	-1	1.777
	ν_τ	0	< 0.182 (95% C.L.)
First	u	$+\frac{2}{3}$	2.2×10^{-3}
	d	$-\frac{1}{3}$	4.7×10^{-3}
Second	c	$+\frac{2}{3}$	1.27
	s	$-\frac{1}{3}$	96×10^{-3}
Third	t	$+\frac{2}{3}$	173.2
	b	$-\frac{1}{3}$	4.18

Table 2.1: The six leptons and six quark flavours form three generations of leptons and quarks. The quoted masses are the averages or limits at a certain confidence level (C.L.) according to Reference [7].

massless, spin-1 particles called gluons (g). Finally, the gravitational interaction appears between all types of massive particles and is by far the weakest (about 10^{38} times weaker than the electromagnetic force); thus it has a negligible impact on microscopic particle interactions. The graviton (G) is postulated to be the gauge boson. All gauge bosons are presented in Table 2.2 along with their charge, mass and the respective interaction type.

Boson	charge $[Q/e]$	mass [GeV]	interaction
G	0	$< 6 \times 10^{-32}$	gravitational
γ	0	0	electromagnetic
$W^\pm$	± 1	80.4	weak
Z	0	91.2	weak
g	0	0	strong

Table 2.2: There are six bosons for the four fundamental forces. The quoted masses are the averages or limits set according to Reference [7].

An extensive effort has been spent in the last century in order to unify the four fundamental interactions by expressing them as different manifestations of a single field. This unification is partially achieved by the Standard Model. The Standard Model is a renormalisable quantum field theory which describes the electromagnetic, weak and strong interactions based on a combination of local gauge symmetry groups $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, where:

- $SU(3)_C$ refers to a colour local symmetry, with a corresponding gauge invariance and the associated 8 gauge bosons without mass (gluons), that hold quarks together through the strong force[1]; The letter C in $SU(3)_C$ stands for colour.

---

[1] The strong interactions of coloured quarks and gluons, developed by David Politzer, Frank Wilczek and David Gross, are described by the gauge field theory called Quantum Chromodynamics (QCD).

- $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ refers to the electroweak symmetry group, which unifies electromagnetic and weak interactions (denoted Electroweak theory[2]). In the $\mathrm{SU}(2)_L$ symmetry group the subscript L stand for "left" and indicates that the symmetry involves only left-handed fields. The $\mathrm{U}(1)_Y$ symmetry group, however, involves both states of chirality, left and right. The generators of $\mathrm{SU}(2)_L$ are given by the weak isospin (I), while the weak hypercharge (Y) is the generator of $\mathrm{U}(1)_Y$. The electric charge (generator of the $\mathrm{U}(1)_{\mathrm{em}}$) is connected to the weak isospin and the weak hypercharge by $Q = I_3 + \frac{Y}{2}$, where I_3 is the third component of the weak isospin.

The SM explains three of the four fundamental forces in a single theory.

The SM Lagrangian can be split in two parts: the QCD Lagrangian, which describes the strong interactions, and the electroweak Lagrangian, which describes electromagnetic and weak interactions:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} \,. \tag{2.1}$$

The observed masses of the EW mediators require mass terms in the Lagrangian which would violate the $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ symmetry. Therefore, it is necessary to introduce a mechanism of "spontaneous $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ symmetry breaking" to solve the inconsistency between the SM theory and the observations, and give mass to vector bosons. The Brout–Englert–Higgs mechanism [4–6] is the minimal mechanism of spontaneous symmetry breaking and predicts a scalar particle, the Higgs boson, whose mass is a free parameter of the theory; its field permeates the vacuum and interacts with particles; the mass acquired by a particle is proportional to the strength of its interaction with the Higgs field.

2.1.1 Quantum Field Theory

Quantum Field Theory (QFT) describes the behavior of particles and their interactions, extending quantum mechanics from single localised particles to fields. Unlike classic quantum mechanics where the state of a system is described by a wave function ψ, in quantum field theory, particles are defined as excitations of a local field $\phi(x)$. In classical mechanics, the Lagrangian density $\mathcal{L}$ describes the properties and the interactions of the field $\phi(x)$ as a function of $\phi(x)$ and its space-time derivatives

$$\mathcal{L}(x) = \mathcal{L}(\phi, \partial_\mu \phi) \,. \tag{2.2}$$

The evolution of a system occurs along a path for which the action (S) is stationary

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

which leads to the Euler-Lagrange equation that describes the motion of the field

$$\partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right] - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \,. \tag{2.4}$$

A gauge symmetry is any continuous transformation of the field that does not affect δS and consequently, does not change the equations of motion: such transformations are called symmetry groups of the system.

---

[2] The Electroweak theory (EW) is not a unification theory: it is not characterised by one gauge symmetry with one unified coupling constant. In fact, the $\mathrm{U}(1)_Y \otimes \mathrm{SU}(2)_L$ group is the product between two gauge groups, whose relation between coupling constants is not predicted by the theory. The theory was developed by Sheldon Lee Glashow, Steven Weinberg and Abdus Salam [1–3] in the 1960s and describes the dynamics of the W^+, W^-, Z and γ bosons. The Standard Model is not a unification theory either: it is the product between three gauge groups. Some theories, named Grand Unification Theory (GUT), seek to unify these three groups $G \supset \mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$.

Based on the Euler-Lagrange equation, the transformation defined as

$$\phi \rightarrow \phi + \varepsilon\, \Delta\phi\,, \tag{2.5}$$

where ε is an infinitesimal arbitrarily fixed constant, is a symmetry group of the system if the Lagrangian is invariant under the transformation up to a divergence

$$\mathcal{L} \rightarrow \mathcal{L} + \varepsilon\, \partial_\mu J^\mu\,. \tag{2.6}$$

According to Noether's Theorem, every symmetry produces a conservation law and, every conservation law represents a symmetry. Therefore, if the symmetry above holds, the current $j_\mu(x) = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\Delta\phi - J^\mu$ is conserved, meaning $\partial_\mu j^\mu = 0$.

2.1.2 Quantum Electrodynamics

Quantum Electrodynamics (QED) was the first relativistic quantum field theory to be developed. It is an Abelian gauge theory describing the dynamics and interactions of fermions and the electromagnetic field. The Lagrangian of the fermion field ψ needs to satisfy the "local gauge invariance" principle, i.e. has to be invariant under the local gauge transformation,

$$\psi \rightarrow U\psi = e^{\mathrm{i}\alpha(x)}\psi\,, \tag{2.7}$$

where $\alpha(x)$ is an arbitrary function, depending on the space-time coordinates. Such phase transformations belong to the unitary Abelian group U(1).

The Lagrangian density describing the field of a fermion with mass m and satisfying the local gauge invariance can be written as:

$$\mathcal{L} = [\mathrm{i}\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi] + e\bar{\psi}\gamma^\mu A_\mu\psi \tag{2.8}$$

where $\bar{\psi} = \psi^\dagger\psi^0$, and γ^μ are the 4×4 Dirac matrices, which satisfy the anticommutation relation $\{\gamma^\mu,\gamma^\nu\} = \gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}$, and $g^{\mu\nu}$ is the metric tensor; e corresponds to the elementary charge and A_μ is a new field, named "gauge field", transforming under the law $A_\mu \rightarrow A_\mu + \frac{1}{e}\partial_\mu\alpha$.

The covariant derivative D_μ is defined as $D_\mu = \partial_\mu - \mathrm{i}eA_\mu$ and transforms as $D_\mu\psi \rightarrow e^{\mathrm{i}\alpha(x)}D_\mu\psi$. This Lagrangian describes the interaction between electrons, ψ, and the electromagnetic field, A_μ, and includes also solutions for an antiparticle, the positron. It is important to remark that the introduced gauge field has to be massless in order to satisfy the gauge principle. Finally, the full Lagrangian density can be written as

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(\mathrm{i}\gamma^\mu D_\mu - m)\psi \tag{2.9}$$

where the electromagnetic field strength tensor $F^{\mu\nu}$ is defined as $F^{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

2.1.3 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is a non-Abelian theory with 8 generators, each of them corresponding to a mediator; it describes the behaviour of the quarks and the mediators of the strong interaction, the gluons. As in QED, the structure of QCD is derived from local gauge invariance by replacing the U(1) group used in QED with a SU(3) group of transformations on the quark colour fields. The free Lagrangian density is

$$\mathcal{L} = \sum_j \bar{q}_j(\mathrm{i}\gamma^\mu\partial_\mu - m_j)q_j \tag{2.10}$$

where $q_j = (q_r, q_b, q_g)_j^T$ and j runs over the six quark flavours. Hereafter, summation over flavours is implied.

The Lagrangian is then required to satisfy the SU(3) local gauge invariance,

$$q(x) \to Uq(x) = e^{-\mathrm{i}g\alpha_a(x)T_a}q(x) \tag{2.11}$$

where U is a 3×3 unitary matrix, g corresponds to the strong coupling constant and α_a to arbitrary parameters; $T_a = \frac{\lambda_a}{2}$ with $a = 1, ..., 8$, are the generators of the SU(3) group, where λ_a are the Gell-Mann matrices, a set of linearly independent traceless 3×3 matrices. The final QCD Lagrangian density becomes

$$\mathcal{L} = -\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a + \bar{q}(\mathrm{i}\gamma_\mu D^\mu - m)q\,, \tag{2.12}$$

where $G^a_{\mu\nu}$ is the gluon field tensor

$$G^{\mu\nu}_a = \partial_\mu G^\nu_a - \partial_\nu G^\mu_a - gf^{abc}G^\mu_b G^\nu_c\,, \tag{2.13}$$

and f^{abc} are the structure constants of the SU(3) group. The non-Abelian character of the theory, resulting in the last term of Equation 2.13, introduces additional terms in the Lagrangian corresponding to self-interaction between the gauge boson fields (*i.e.*, three and four gluon vertices). It should be noted that gluons carry the colour charge of the strong interaction as do quarks.

2.2 Electroweak theory

The weak interaction manifests itself in the emission or absorption of W and Z bosons and describes, for instance, the decay of muons, top quarks or neutrons. The Electroweak theory describes how electromagnetic processes and the weak charged current are invariant under the weak hyper-charge U(1) and the weak isospin SU(2) transformations. As QED and QCD, also this theory relies on the gauge invariance principle and combines the electromagnetic and weak interactions as two distinct manifestations of the same force.

The EW theory groups the fermion fields in left-handed and right-handed ones

$$\psi_L = \frac{1}{2}(1 - \gamma^5)\psi, \qquad \psi_R = \frac{1}{2}(1 + \gamma^5)\psi \tag{2.14}$$

where $\frac{1}{2}(1 \pm \gamma^5)$ are the chirality operators and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. In this description the left-handed fields ψ_L are doublets with isospin $I = \frac{1}{2}$ and the right-handed fields ψ_R are singlets with isospin $I = 0$:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \qquad \begin{pmatrix} c \\ s \end{pmatrix}_L \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \qquad \begin{pmatrix} t \\ b \end{pmatrix}_L \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \tag{2.15}$$

$$u_R,\ d_R,\ e_R, \qquad c_R,\ s_R,\ \mu_R, \qquad t_R,\ b_R,\ \tau_R. \tag{2.16}$$

The U(1) transformations simply changes left-handed doublets and right-handed singlets by a phase factor $e^{i\alpha^a(x)\frac{Y}{2}}$. Moreover, the left-handed doublets transform as

$$\psi_L \rightarrow e^{i\beta^a(x)\frac{\tau^a}{2}}\psi_L \tag{2.17}$$

under the SU(2)$_L$ symmetry, where $a = 1, 2, 3$, τ^a refers to the Pauli matrices and $\frac{\tau^a}{2}$ are the generators of the SU(2) group.

By applying the gauge invariance principle four gauge fields are introduced: three gauge fields (isotriplet), W^i_μ, associated to SU(2)$_L$, interact only with the left-handed components, whereas one gauge field (singlet), B_μ, associated to U(1)$_Y$, interact with states of either chirality. Thus, the Lagrangian term describing the interaction between the fermions and the four gauge fields is

$$\mathcal{L}_{\text{int}} = -\psi_L\gamma^\mu\left(g\frac{\tau_a}{2}W^a_\mu + g'\frac{Y}{2}B_\mu\right)\psi_L - \bar{\psi}_R\gamma^\mu\left(g'\frac{Y}{2}B_\mu\right)\psi_R, \tag{2.18}$$

where $W^a_{\mu\nu}$ and $B_{\mu\nu}$ are field tensors defined as

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\varepsilon^{ijk}W^j_\mu W^k_\nu \tag{2.19}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.20}$$

and ε^{ijk} are the structure constants of SU(2).
The coupling of the W^a_μ is visible only to the left-handed components.

It should be noticed that no quadratic terms of the gauge fields are present in the Lagrangian: the EW theory considers the gauge bosons as massless, despite measurements that confirm the existence of massive electroweak mediators, the Z and W bosons. As described in the next Section, the solution consists in introducing an additional field, the Higgs field, that gives mass to this bosons by breaking the electroweak symmetry.

Furthermore, in order to conserve the gauge invariance, terms related to the fermion masses are also absent in the Lagrangian density: they will be generated by introducing a gauge invariant interaction between the Higgs field and the fermions, i.e. the Yukawa interaction.

Finally, the complete electroweak Lagrangian will be

$$\mathcal{L}_{\text{EW}} = \mathcal{L}_{\text{f+g}} + \mathcal{L}_H + \mathcal{L}_{\text{Yukawa}} \tag{2.21}$$

where $\mathcal{L}_{\text{f+g}}$, $\mathcal{L}_H$, $\mathcal{L}_{\text{Yukawa}}$ are the fermion and gauge, Higgs and Yukawa terms, respectively.

2.3 The Higgs mechanism: symmetry breaking

Although the electroweak theory elegantly describes the weak and electromagnetic interactions by means of four massless boson mediators, it contradicts the experimental observations. The mediators of the weak interaction are in fact measured with masses of $m_{W^\pm} = 80.4$ GeV and $m_{Z^0} = 91.2$ GeV [7]. Adding a mass component in the Lagrangian density is forbidden, because it would lead to a violation of the gauge invariance, and the same applies for the fermions.

A proposed solution is the so-called “Higgs mechanism”, a theory that predicts the existence of an homonymous field, the Higgs field, which permeates the universe[3]. The interaction with this field gives

---

[3] According to the Higgs mechanism, the minimal Higgs field is a doublet in SU(2) space, has a non-zero U(1) hypercharge and is a SU(3) colour singlet.

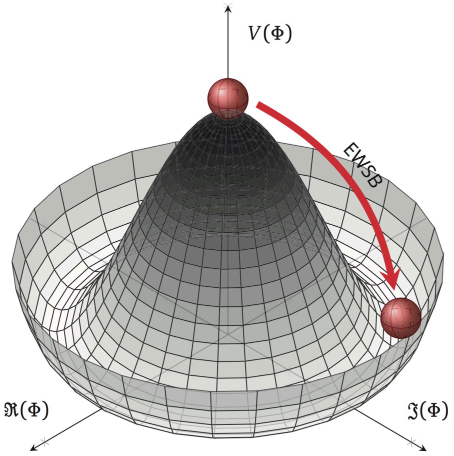


Figure 2.1: The shape of the Higgs potential $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$.

mass to the gauge bosons and fermions, by breaking the electroweak symmetry.

To achieve this, an additional SU(2) isospin doublet of complex scalar fields with $Y = 1$ is introduced,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{2.22}$$

and its corresponding term in the Lagrangian is

$$\mathcal{L}_H = (D_\mu\phi)^\dagger(D^\mu\phi) - V(\phi), \tag{2.23}$$

where $V(\phi)$ is the potential term

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 = \mu^2\phi^2 + \lambda\phi^4. \tag{2.24}$$

The parameter λ of the potential is assumed to be positive. To determine the ground state, ϕ_0, the potential is minimised. For $\mu^2 > 0$ the potential V assumes a unique minimum at $\phi_0 = 0$ and the ground state is symmetric under SU(2). For $\mu^2 < 0$, instead, the shape of the potential is modified, as shown in Figure 2.1, and V assumes a non-trivial minimum $\phi_0^2 = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}$. The vacuum expectation value[4] for ϕ^2 in the physical vacuum state is non-zero and corresponds to a circumference with radius $\phi_0^2 = \frac{v^2}{2}$ in the complex plane $\mathrm{Re}(\phi) - \mathrm{Im}(\phi)$. Among all possible ground states, a reference minimum for the local gauge transformation can be chosen, without any loss of generality

$$\phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.25}$$

The existence of several minima breaks the $\mathrm{SU(2)}_L$ symmetry (spontaneous symmetry breaking), since minima are not symmetric anymore and are altered by SU(2) transformations. By expanding around the reference minimum the scalar doublet ϕ can be rewritten as

$$\phi(x) = e^{\frac{i\xi_a(x)\tau^a}{2u}}\begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}. \tag{2.26}$$

---

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

where $H(x)$ is the physical scalar Higgs field and $\xi_a(x)$ ($a = 1, 2, 3$) are new real fields. By exploiting the invariance of the Lagrangian under SU(2) gauge transformations, the $\xi_a(x)$ can be removed ("gauged away") from the Lagrangian

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \tag{2.27}$$

Since $D^\mu = \partial^\mu + \mathrm{i}gW_a^\mu\frac{\tau^a}{2} + \mathrm{i}g'\frac{1}{2}B^\mu$, where g is the SU(2) and g' the U(1) coupling constants, the kinetic term of the Lagrangian $\mathcal{L}_H$ component is

$$(D_\mu\psi)^\dagger(D^\mu\psi) \rightarrow \frac{1}{2}\partial^\mu H\partial_\mu H + \frac{1}{8}g^2(v+H)^2|W_\mu^1 + W_\mu^2|^2 + \frac{1}{8}(v+H)^2|g'W_\mu^3 - gB_\mu|^2. \tag{2.28}$$

The charged physical fields $W^\pm$ are defined as $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp \mathrm{i}W_\mu^2)$. In order to be orthogonal to each other, the neutral physical fields (the photon and the Z boson fields) are defined as

$$Z_\mu = \frac{g'W_\mu^3 - gB_\mu}{\sqrt{g'^2 + g^2}} \text{ and} \tag{2.29}$$

$$A_\mu = \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g'^2 + g^2}}\,. \tag{2.30}$$

By introducing the "weak mixing angle" θ_w

$$\cos\theta_\mathrm{w} = \frac{g'}{\sqrt{g'^2 + g^2}},\ \sin\theta_\mathrm{w} = \frac{g}{\sqrt{g'^2 + g^2}} \tag{2.31}$$

the neutral fields can be rewritten as

$$Z_\mu = -B_\mu\sin\theta_\mathrm{w} + W_\mu^3\cos\theta_\mathrm{w} \text{ and} \tag{2.32}$$

$$A_\mu = B_\mu\cos\theta_\mathrm{w} + W_\mu^3\sin\theta_\mathrm{w}\,. \tag{2.33}$$

The masses of the gauge bosons are inferred from the quadratic terms (in the field) in Equation 2.28: $M_W = \frac{gv}{2}$ and $M_Z = \frac{\sqrt{g'^2+g^2}v}{2}$, while the photon remains massless. The masses of the weak bosons are linked to each other through the electroweak mixing angle:

$$\frac{M_W}{M_Z} = \cos\theta_\mathrm{w}\,. \tag{2.34}$$

2.3.1 Fermion mass through Yukawa coupling

As mentioned in Section 2.2, also the masses of fermions can be generated through the spontaneous breaking of the $\mathrm{SU(2)}_L \otimes \mathrm{U(1)}_Y$ gauge symmetry: this introduces a Yukawa term, which describes the interaction between the fermion and Higgs fields. For a single generation

$$\mathcal{L}_\text{Yukawa} = -Y_\ell\bar{L}_L\phi\ell_R - Y_d\bar{Q}_L\phi d_R - Y_u\bar{Q}_L\tilde{\phi}u_R + \text{h.c.} \tag{2.35}$$

where $L_L = (\nu_\ell, \ell)_L^T$ and $Q_L = (u, d)_L^T$ are the left-handed lepton and quark doublets, ℓ_R, u_R and d_R are right-handed singlets, $\tilde{\phi} = -\mathrm{i}\tau_2\phi^*$ corresponds to the charge conjugate of the Higgs doublet, ℓ is the charged lepton, and Y_ℓ, Y_d, Y_u are three matrices containing the Yukawa coupling constants between the fermions and the Higgs boson.

Replacing the value of the Higgs vacuum state from Equation 2.27 in the Yukawa terms involving the vacuum expectation value, $\mathcal{L}_{\text{Yukawa}}$ becomes:

$$\mathcal{L}_{\text{fermion mass}} = -(\bar{d}'_L M_d d'_R + \bar{u}'_L M_u u'_R) + \text{h.c.}\,, \tag{2.36}$$

where $M_{u,d} = (\frac{v}{\sqrt{2}})Y_{u,d}$, lepton terms are neglected for simplicity, u' and d' are quark weak eigenstates. By unitary transformations $U_{L,R}$ and $D_{L,R}$ the Yukawa coupling matrices can be diagonalised from the weak eigenstates u' and d' to the mass eigenstates u and d:

$$\begin{aligned} \begin{pmatrix} u' \\ c' \\ t' \end{pmatrix}_{L.R} &= U_{L,R} \begin{pmatrix} u \\ c \\ t \end{pmatrix} \\ \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{L.R} &= D_{L,R} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \end{aligned} \tag{2.37}$$

such that the quark quadratic terms become diagonal

$$\begin{aligned} U_R^{-1} M_u U_L &= \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \\ D_R^{-1} M_d D_L &= \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} \end{aligned}. \tag{2.38}$$

The mismatch between the weak eigenstates and the mass eigenstates leads to transitions between quark generations through flavour changing interactions. The flavour changing interactions are proportional to the Cabibbo-Kobayashi-Maskawa (CKM) matrix elements. The CKM matrix can be written as

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{2.39}$$

where diagonal elements predominantly lead to flavour changing interactions. The definition of the CKM matrix, up to a non-eliminable phase, leads to CP violation[5]. Analogously a mixing matrix can be introduced also for the neutrino sector.

The interaction with the Higgs field gives mass to the fermions,

$$m_f = \frac{v}{\sqrt{2}} y_f\,, \tag{2.40}$$

---

[5] The Charge-Parity (CP) symmetry is a combination of the charge conjugation symmetry and the parity symmetry. The CP symmetry states that the laws of physics should be the same if a particle or a system of particles are interchanged with the respective antiparticles (C symmetry), and when its spatial coordinates are inverted ("mirror" or P symmetry).

and the coupling constant y_f between the Higgs boson and the fermion results proportional to the mass of the fermion. For this reason fermion couplings are very different from each other (from $m_\nu < 1$ eV up to $m_t = 174$ GeV). From the free Lagrangian of Higgs boson,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - \lambda v^2 H^2 - \lambda v H^3 \frac{\lambda}{4} H^4 \,, \tag{2.41}$$

we obtain the self-interaction terms (cubic and quadratic) and the mass term $m_H^2 = 2\lambda v^2$.

The vacuum expectation value is also related to the Fermi constant G_F as $v = (\sqrt{2G_F})^{\frac{1}{2}} \approx 246$ GeV, whereas the constant λ is unknown. Thus the Higgs boson mass is a free parameter in the SM theory, but some constraints can be derived by theoretical and experimental considerations, as discussed in Section 2.5.1.

2.4 Top-quark physics

Since its discovery in 1995 by the CDF [8] and DØ [9] collaborations, the top quark has been subject of extensive studies. It belongs to the third generation of quarks, together with the bottom quark, and is the heaviest particle in the Standard Model with a mass of 173.34 ± 0.76 GeV [10]. Because of the large mass, the top quark has a Yukawa coupling y_t very close to one: such particular value suggests that the top quark might play a special role in more general theories proposed to solve some limitations of the Standard Model (see Sections 2.6 and 2.7).

Given the large value of its width (1.41 ± 0.17 GeV [7]), the top quark has a very short lifetime ($\sim 5.0 \times 10^{-25}$ s), a twentieth of the timescale for strong interactions. This means that the top quark decays before any hadronisation effect can take place: This unique feature allows to directly detect spin information, which is transferred to its decay products, undiluted by non-perturbative effects.

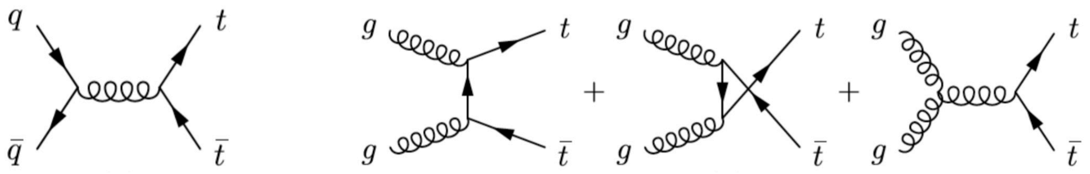
Precise measurements of top-quark production and decay can confirm QCD calculations at high scales as well as reveal indirect effects of new physics. Furthermore, $t\bar{t}$ pairs (plus additional jets) represent the main background in many searches for new physics and a good control on its modelling is decisive.

The dominant top-quark production mode at LHC is, indeed, as a pair of top and anti-top quarks via the strong interaction, followed by single top, which is suppressed at the LHC and stems from electroweak interaction with b quarks or antiquarks in the initial state. The $t\bar{t}$ production and decay modes will be described in the following Sections since they are a main ingredient of this work: both the signal and the major background processes are characterised by the associated production of a boson (Higgs, W or Z) with a $t\bar{t}$ pair.

2.4.1 Top-quark pair production

At hadron colliders $t\bar{t}$ pairs are produced by $q\bar{q}$ annihilation or gluon fusion, as shown in Figure 2.2. Dominant process at the Tevatron is the $q\bar{q}$ annihilation ($\approx 85\%$ of $t\bar{t}$ cross section), since collisions happens mainly between the valence quarks from the proton and the antiproton. Depending on the centre-of-mass energy of the LHC, 80-90% of the $t\bar{t}$ pairs are produced through gluon fusion, which dominates for both pp and $p\bar{p}$ colliders with increasing energies.

The theoretical $t\bar{t}$ production cross section as function of the centre-of-mass energies are shown in Figure 2.3 as well as measurements; the computation is made at next-to-next-to-leading order (NNLO) in α_s and with next-to-next-to-leading logarithm (NNLL) soft-gluon resummation [11–16]. For a mass of



(a) $q\bar{q}$ annihilation

(b) gluon-gluon fusion

Figure 2.2: Diagrams for the $t\bar{t}$ production at LO.

m_t = 173.2 GeV[42], the cross section is

$$\sigma_{t\bar{t}}(8\ \text{TeV}) = 247.7^{+13.1}_{-14.3}\ \text{pb}$$

$$\sigma_{t\bar{t}}(13\ \text{TeV}) = 816.0^{+39.5}_{-44.7}\ \text{pb}$$

where the uncertainties arise from variations of the renormalisation and factorisation scales and uncertainties in the parton distribution functions.

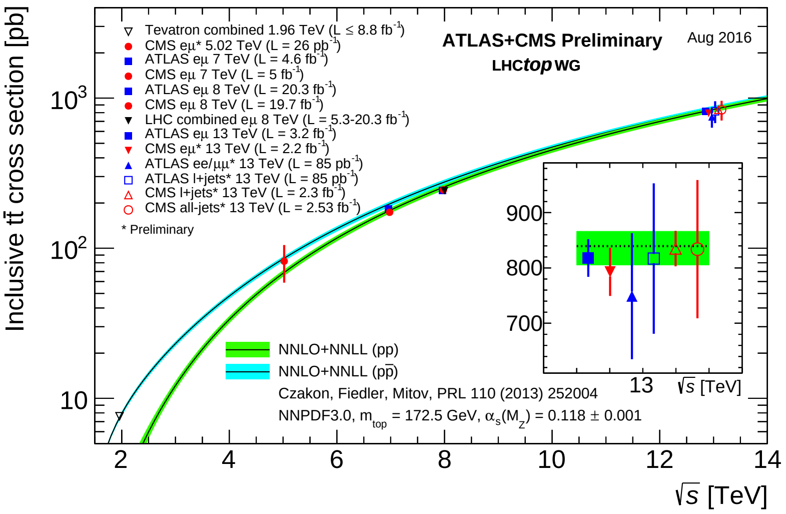


Figure 2.3: Measured and predicted $t\bar{t}$ production cross sections from Tevatron energies in $p\bar{p}$ collisions to LHC energies in pp collisions [17].

2.4.2 Top-quark pair decay

The top-quark decays almost exclusively into a b quark and an on-shell W boson: the bottom quark hadronises in most of the cases, while the W boson can decay either hadronically (68%) or leptonically ($W \to \ell\nu_\ell$, 32%). Thus, a $t\bar{t}$ final state is determined by the number and the flavour of the leptons produced by the two W bosons present in the event. Figure 2.4 displays the different $t\bar{t}$ signatures:

- fully hadronic, with the largest branching ratio (≈ 46%), both W bosons decay hadronically resulting in a final state with at least six jets;
- dilepton, with ≈ 6.4% branching ratio, both W bosons decay leptonically;
- lepton+jets or single lepton, when only one W boson decays into leptons, it corresponds to 35% of the total decays.

Events containing hadronically decaying τ's are considered separately given the particular experimental signature, while the leptonically decaying one are included in the corresponding single or dilepton channels.

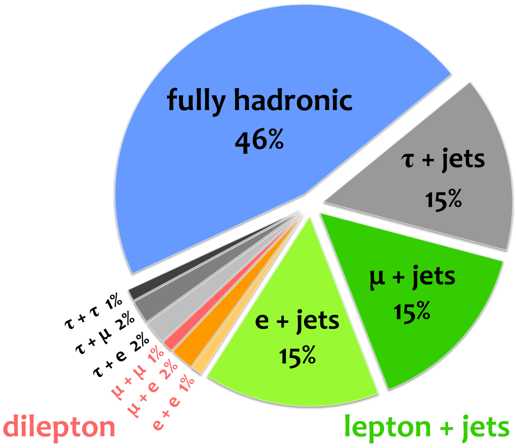


Figure 2.4: Pie chart representing the branching ratios (BR) of a top-antitop quark pair. The light blue represents the fully hadronic BR of 46% (56% when including hadronic decaying τ), the dileptonic BR (without τ lepton) is shown in 3 shades of red, with a total of 4% (6.4% when including leptonic decaying τ), and in 2 shades of green the lepton (e or μ) + jets BR of 30% (35% when including leptonic decaying τ).

2.5 The Higgs boson

Before the discovery of the Higgs boson in 2012 [18, 19], all the parameters of the SM including the coupling constants, gauge boson and fermion masses, and quark mixing angles, had been determined experimentally, except for the Higgs mass. Albeit considered a free parameter of the Standard Model, interesting theoretical constraints have been inferred from considerations on the energy range where the SM is valid, i.e. before the perturbation regime ceases and new phenomena arise: these constraints rely on perturbative unitarity [20–22], triviality [23], vacuum stability and fine-tuning. Additional experimental constraints have been also set by direct searches performed at LEP[6] [24] , Tevatron [25] and LHC, and by indirect searches like precision electroweak measurements [26, 27].

Finally, with the discovery of the Higgs boson an important step towards the comprehension of nature has been done, moving the focus on the characterisation of such particle and the measurement of its interaction with other SM particles.

2.5.1 Theoretical constraints on the Higgs mass

A short description of the theoretical considerations contributing to the mass constraint is provided in this Section.

Perturbative unitarity

Fermi's theory of weak interactions violates the unitarity condition when predicting cross sections at very high energies (comparable with the Fermi scale): for instance, if the cross section of the WW elastic scattering process in the SM is calculated only with the three upper diagrams shown in Figure 2.5 (leading order processes involving γ and Z exchange diagrams and W-W self-interaction diagrams), the

---

[6] The Large Electron–Positron Collider (LEP) was the largest electron-positron accelerator ever constructed, colliding electrons with positrons at energies of 209 GeV from 1989 until 2000.

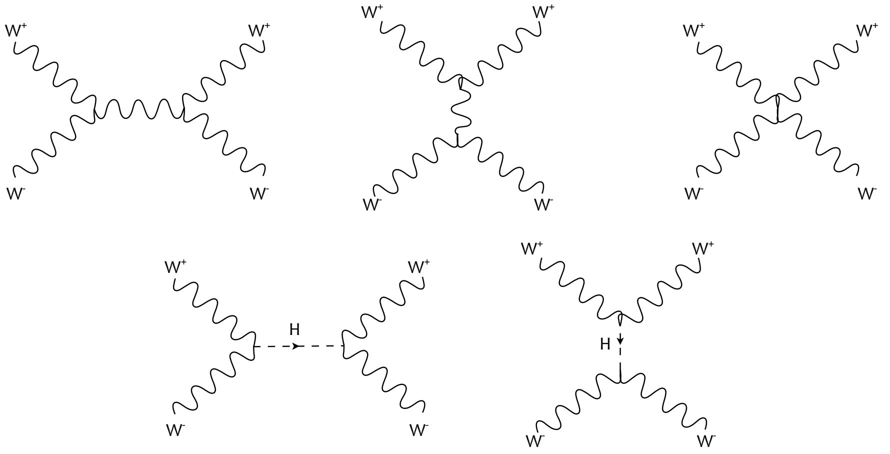


Figure 2.5: Feynman diagrams for WW elastic scattering: the first three diagrams correspond to the electroweak WW interactions, while the last two correspond to interactions mediated by the Higgs boson.

contribution of the longitudinal W and Z to the total amplitude increases with the energy, violating the unitarity at that point. The introduction of the contribution from the Higgs intermediate state in the s and t channels shown in the two bottom diagrams of Figure 2.5 restores the unitarity, leading at high energy scale to a constant amplitude proportional to the Higgs self-coupling constant, $\lambda = \frac{m_H^2}{2v^2}$. Anyhow, unitarity holds only if λ does not diverge, which means $\lambda \lesssim 1$ [20–22]:

$$m_H \lesssim 1.2\ \text{TeV}\,.$$

If the Higgs boson did not exist or its mass was larger, the unitarity has to be restored introducing new physics or using non-perturbation theories.

Triviality and vacuum stability

Accounting for the radiative corrections at one loop of the Higgs boson only to the Higgs quartic coupling λ, the coupling shows a dependence on the logarithm of the squared energy scale, μ^2.

$$\lambda(\mu^2) = \lambda(v^2)\left[1 - \frac{3}{4\pi^2}\lambda(v^2)\log\left(\frac{\mu^2}{v^2}\right)\right]^{-1} \tag{2.42}$$

For very small energies ($\mu^2 \ll v^2$) $\lambda(\mu^2)$ vanishes, leading to a trivial, non interacting theory. At very large energies, instead, the self-coupling can become infinite. An energy cut-off Λ_c can be defined as the Landau pole of $\lambda(\mu^2)$

$$\Lambda_c = \nu \exp\left(\frac{2\pi^2}{3\lambda\nu^2}\right) = \nu \exp\left(\frac{4\pi^2\nu^2}{3m_H^2}\right), \tag{2.43}$$

below which λ remains finite and the theory remains perturbative; the greater m_H is, the quicker a non-perturbative regime is established. Asking the perturbative regime to be valid up to an energy threshold can be translated into an upper limit, $m_{\text{max}}^{\text{Landau}}$, on the Higgs mass: $\Lambda_c \sim M_{\text{PL}} \simeq 1.2 \times 10^{19}$ GeV implies $m_{\text{max}}^{\text{Landau}} = \mathcal{O}(175\ \text{GeV})$, while $\Lambda_c \gtrsim 1$ TeV implies a weaker constraint $m_{\text{max}}^{\text{Landau}} = \mathcal{O}(800\ \text{GeV})$ [23].

On the other hand, adding terms from gauge bosons and fermions to the running of λ can be translated into a lower bound. Since the Higgs coupling is proportional to the mass of interacting particle, only loops involving massive vector bosons and the top quark are relevant in λ

$$\lambda(\mu^2) = \lambda(v^2) + \frac{1}{16\pi^2}\left\{ -12\frac{m_{\text{top}}^4}{v^4} + \frac{3}{16}\left[2g^4 + (g^2 + g'^2)^2\right]\right\} \ln\frac{\mu^2}{v^2}\,. \tag{2.44}$$

Driven by the top-quark loops, the Higgs quartic coupling λ tends to become negative, making the vacuum unstable and new minima of the potential appear, because of the renormalisation-group-improved potential $V(\phi) = \mu^2\phi^2 + \lambda(\mu = \phi)\phi^4$, where the self-coupling dependence on the energy scale ϕ is intentionally emphasised.

Two types of stability bounds are then possible, with $m_H = \sqrt{2\lambda}(\mu)v$: if $m_H > m_{\text{min}}^{\text{stability}}$, the electroweak vacuum is absolutely stable, whereas if $m_{\text{min}}^{\text{meta}} < m_H < m_{\text{min}}^{\text{stability}}$, then it is metastable with the life-time exceeding that of the universe. Figure 2.6 depicts possible behaviours for the whole Higgs boson mass range–Landau pole, stable, or unstable electroweak vacuum. Figure 2.7, instead, illustrates the SM effective potential V for the Higgs field for $m_H > m_{\text{min}}^{\text{stability}}$ (left and middle) and $m_H < m_{\text{min}}^{\text{stability}}$ (right).

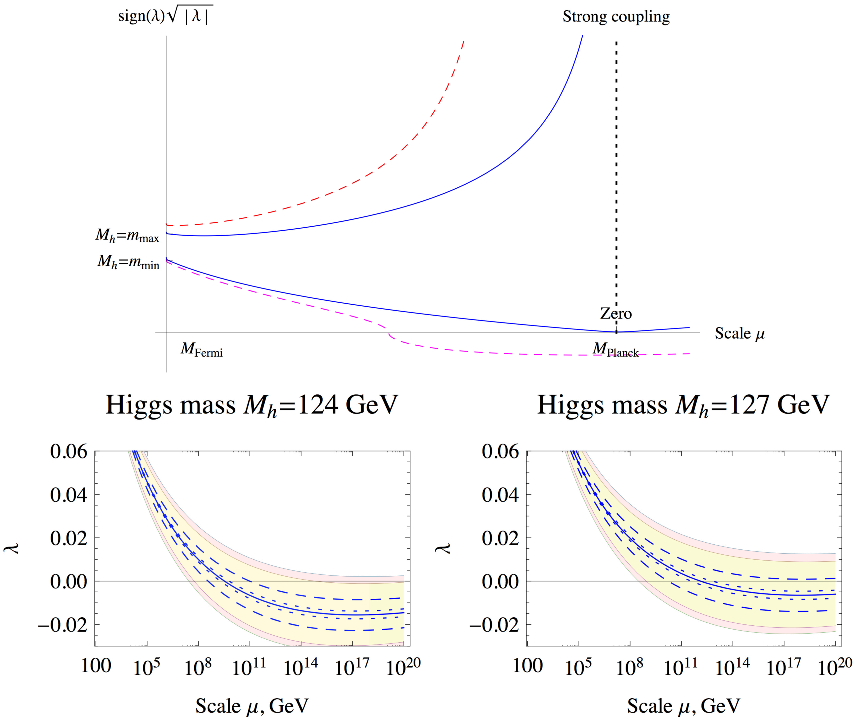


Figure 2.6: Upper plot: different patterns of the behaviour of the Higgs self-coupling with energy. For $m_H > m_{\text{max}}^{\text{Landau}}$ the Landau pole appears at energies below the Planck scale (triviality). If $m_H < m_{\text{min}}^{\text{stability}}$ the scalar constant becomes negative at energies below the Planck mass, and the electroweak vacuum becomes metastable [28]. Lower plots: detailed behaviour for low Higgs boson masses, with dashed (dotted) lines corresponding to the experimental uncertainty in the top-quark mass m_t (strong coupling constant α_s), and the shaded yellow (pink) regions correspond to the total experimental error and theoretical uncertainty, with the latter estimated as 1.2 GeV (2.5 GeV) [28].

Numerically, $m_{\text{min}}^{\text{meta}} \simeq 113$ GeV [29], while $m_{\text{min}}^{\text{stability}} \simeq 130$ GeV [30]. The existence of the Higgs boson with a mass smaller than $m_{\text{min}}^{\text{meta}}$ would provide an undisputed argument in favour of the existence of new physics between the Fermi and Planck scale. However, already since LEP we know that this is not the case.

Figure 2.8 shows the constraints due to the triviality (also called non-perturbativity) and vacuum

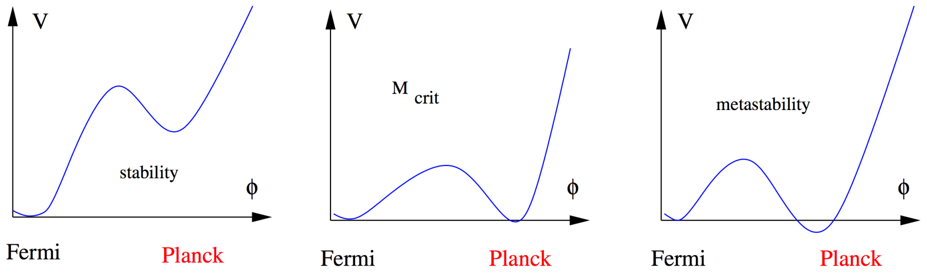


Figure 2.7: The form of the effective potential for the Higgs field ϕ which corresponds to the stable (left), critical (middle) and metastable (right) electroweak vacuum. The form of the effective potential is tightly related to the energy dependence of the Higgs self-coupling constant $\lambda(\mu)$: the potential is negative almost in the same domain where $\lambda(\phi) < 0$ [29].

stability in the phase diagram of the SM Higgs potential, taking into account the values for the Higgs mass measured by ATLAS and CMS [18, 19]. the remarkable coincidence for which the SM appears to live right at the border between the stability and instability regions is illustrated in the left plot; the right plot zooms into the relevant region and shows that there is significant preference for metastability of the SM potential. By taking into account all uncertainties, the stability region is disfavoured by present data at the 2σ level [30]. For $m_H < 126$ GeV, stability up to the Planck mass is excluded at a one sided 98% C.L..

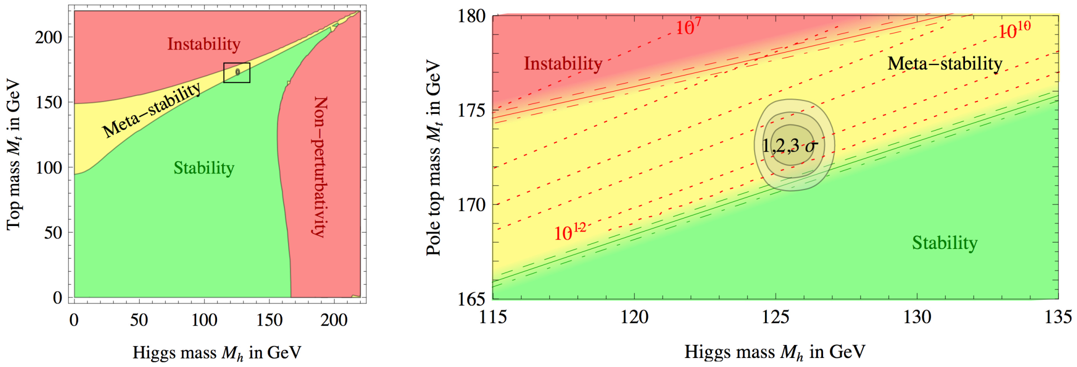


Figure 2.8: Regions of absolute stability, metastability and instability of the SM vacuum in the top-Higgs masses plane (left) and zoom in the region of the preferred experimental range of m_H and m_t (the gray areas denote the allowed region at 1, 2, and 3σ). The three boundary lines correspond to $\alpha_s(m_Z) = 0.1184 \pm 0.0007$, and the grading of the colours indicates the size of the theoretical uncertainty. The dotted contour-lines show the instability scale Λ in GeV [30].

2.5.2 Discovery

According to the Standard Model, the Higgs boson is a neutral particle with spin zero and its mass is a free parameter to be determined experimentally. A huge effort at the LHC, in both the ATLAS and CMS experiments, led to the discovery of a new particle compatible with the Higgs boson [18, 19] announced on 4 July 2012. The discovery was based on data recorded by ATLAS and CMS at a centre-of-mass energy of $\sqrt{s} = 7$ TeV in 2011 and $\sqrt{s} = 8$ TeV in 2012.

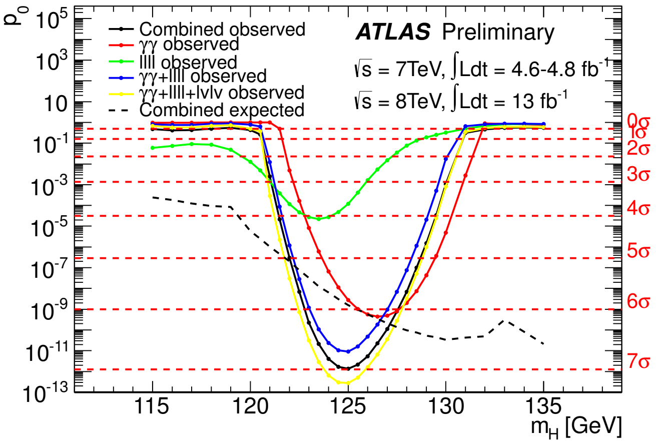


Figure 2.9: The local probability p-value for a background-only experiment to be more signal-like than the observation as a function of m_H for various combinations: $H \to \gamma\gamma$ (red line); $H \to ZZ^{(*)} \to 4\ell$ (green line); combination of $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to 4\ell$ (blue line); combination of $H \to \gamma\gamma$, $H \to ZZ^{(*)} \to 4\ell$ and $H \to WW^{(*)} \to \ell\nu\ell\nu$ (yellow line) and the combination of all channels, including $H \to b\bar{b}$ and $H \to \tau\tau$ (black line). The dashed black curve shows the median expected local p-value under the hypothesis of a Standard Model Higgs boson production signal at that mass for the combination of all channels. The horizontal dashed lines indicate the p-values corresponding to significances from 0 to 7σ [18].

The Higgs-like boson was observed in the decay channels $\gamma\gamma$, ZZ and WW with a mass of 125-126 GeV: in particular, using $\gamma\gamma$ and ZZ decay channels, its mass was measured by ATLAS and CMS as $m_H = 126.0 \pm 0.4(\text{stat.}) \pm 0.4(\text{syst.})$ GeV and $m_H = 125.3 \pm 0.4(\text{stat.}) \pm 0.5(\text{syst.})$ GeV, respectively. The discrepancy between the data and the background-only hypothesis, larger than 5σ, allowed to claim the discovery of a new boson. Figure 2.9 shows the p-value for various combinations.

The Higgs-like boson was found to be compatible with the SM 0^+ hypothesis when compared with other J^P hypotheses[7] [31]. The alternative hypotheses are excluded with the $1 - CL_s$ confidence levels shown in Table 2.3:

Table 2.3: List of spin-parity hypotheses together with their exclusion confidence level [31].

Exclusion in favour of 0^+	
Tested hypothesis	$1 - CL_s(\%)$
0^-	99.6
1^+	99.4
1^-	96.9
2^+_m	81.8
2^-	88.4

2.5.3 Production and decay modes

The most important production and decay mechanisms are described in this Section.

Production mechanisms

At the LHC, the SM Higgs boson can be generated through four main processes, illustrated in Figure 2.10: the gluon fusion (ggF) through a loop of heavy quarks is the dominant mechanism, followed by the vector boson fusion (VBF), where the Higgs boson is produced through the fusion of virtual bosons emitted

---

[7] J^P stands for the spin J and parity P hypothesis tested: the charge parity C is trivially positive, since the boson decays into the $\gamma\gamma$ final state.

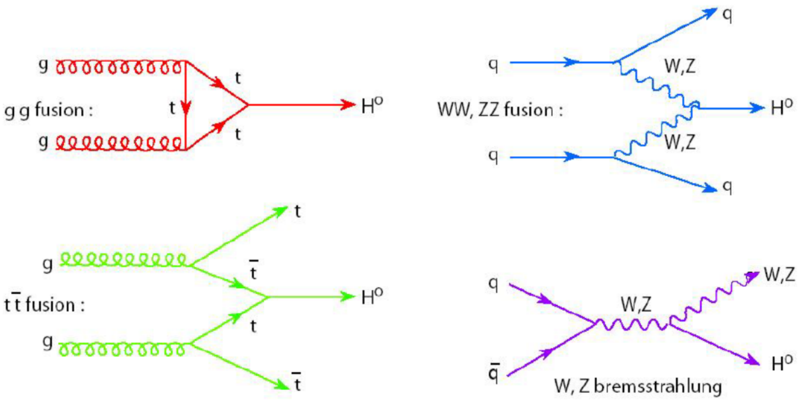
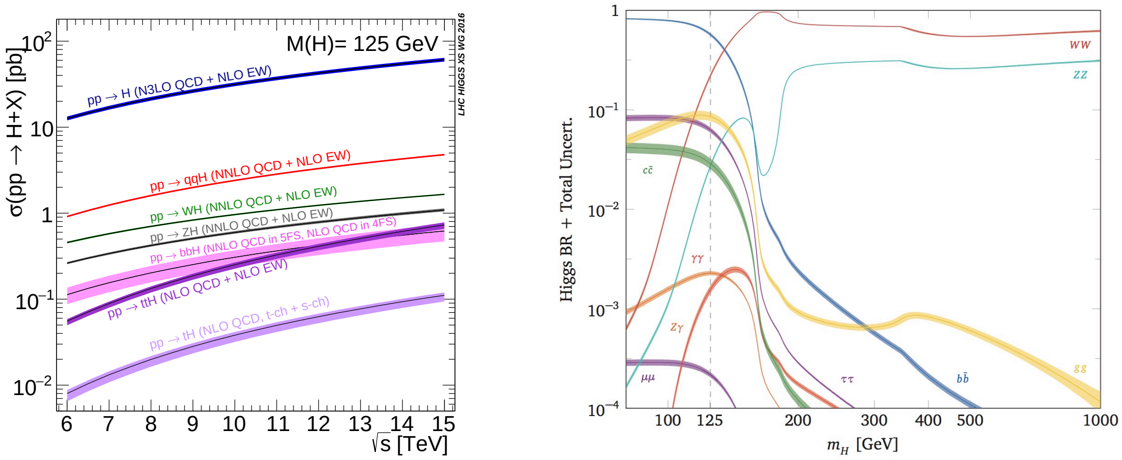


Figure 2.10: Leading order Feynman diagrams for the main production modes of the SM Higgs boson at LHC.

by incoming partons. The production in association with a vector boson gives a relevant contribution only in a low Higgs mass hypothesis, in contrast to the Tevatron's $p\bar{p}$ collider where the valence quarks and antiquarks enhance this production mode. The production in association with a top-quark pair is the rarest Higgs production mode, but its cross section increases by four times from 8 TeV to 14 TeV, more than the other processes. Figure 2.11(a) summarises the expected cross section as a function of the centre-of-mass energies [32, 33].



(a) Higgs production modes

(b) Higgs decay modes

Figure 2.11: (a) Standard Model Higgs boson production cross sections as a function of the centre-of-mass energies. The tH production cross section accounts for t-channel and s-channel only (no tWH production) [32, 33]. (b) The decay branching ratios of the Higgs boson, as a function of its mass [34].

Higgs boson decays

Once the Higgs mass is fixed, its decay and the branching ratios are uniquely determined (Figure 2.11(b)). The branching ratios are predicted at next-to-next-to leading order (NNLO) [35, 36] with QCD and electroweak corrections. The decay modes can be divided into two groups: decays to bosons and decays to fermions.

$H \to ZZ/WW$
Higgs decays into pairs of vector bosons; the $H \to ZZ^{(*)}/\gamma^{(*)} \to 4\ell$, where $\ell = \mu, e$, is also known as the "golden" channel due to the cleanness of the signature and the fully reconstructed Higgs mass. $H \to WW^{(*)}$ is the second dominant decay mode; its relevant final states are $\ell\nu\ell\nu$ and $\ell\nu q\bar{q}$. Because of the high missing energy (high p_{T} neutrinos), the only mass that can be reconstructed is the transverse mass of the system.

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

$H \to q\bar{q}$
At $m_H \leq 125$ GeV, the dominant mode for quark pair production is $H \to b\bar{b}$. Due to the large QCD multijet background, the pure gluon fusion decay mode is difficult to detect, but other associated production modes can be exploited. Another relevant fermion decay mode in the low mass region is the $qqH \to \tau^+\tau^-$: the τ's subsequently decay in the semileptonic mode, leptonic mode, or into pairs of hadrons.

2.6 Experimental success and shortcomings of the Standard Model

The Standard Model gives a convincing description of the fundamental structure of the observable matter. It requires 18 independent parameters:

- the masses of 6 quarks and 3 leptons,
- 3 gauge couplings (e, θ_{w} and α_s),
- 3 mixing angles and 1 CP violating complex phase of the unitary CKM matrix,
- the Higgs boson mass and the vacuum expectation value v.

Experiments during the last 30 years at lepton and hadron colliders have tested the theory in many ways. Several measured quantities and global fits confirm the accurate predictions of the SM over 12 orders of magnitude: a typical example is the higher order corrections of the electron gyromagnetic ratio [37]. The Higgs boson represents the last brick of the Standard Model.

2.6.1 Beyond Standard Model theories

Although many experimental measurements endorse the validity of the Standard Model, the theory does not address a number of open questions:

- the large set of parameters (18), that needs to be determined by measurement;
- the EW theory cannot be considered as unification theory, since the symmetry group is not one, but the product of two different groups with independent coupling constants, g and g', where $g'/g = \tan\theta_W$ is determined experimentally;
- the SM assumes neutrinos to be massless, while observations of flavour oscillations can only be explained by massive neutrinos, via a mixing of the EW eigenstates[8].

---

[8] The seesaw mechanism incorporates neutrino masses into the SM by introducing heavy Majorana neutrinos, whose masses are inversely coupled to the light SM neutrino masses, hence motivating their small values of $O(1\text{ eV})$. Despite extensive searches, the experimental proof is still pending.

- the so-called “hierarchy problem” (naturalness), according to which the Higgs mass receives radiative corrections through boson and fermionic loops, which are quadratically divergent with the cut-off (that represents the scale beyond which new physics needs to be considered). A typical scale is the Planck scale where the corrections are $\sim 10^{30} m_H^2$, many orders larger than the Higgs mass at tree level;
- the theory is not able to explain some symmetries like the proton stability and the conservation of the baryonic number (missing symmetry similar to $\mathrm{U}(1)_Q$ or $\mathrm{SU}(3)_C$);
- no motivation to the existence of exactly 3 generations of leptons and quarks (fermionic problem);
- the SM is not able to predict/include some results from astroparticle physics and cosmology, like the matter-antimatter asymmetry;
- ordinary matter (well described by the SM) constitutes only 4-5% of the universe energy density, according to cosmological observations. Observations of galaxy rotation profiles evidenced a large amount of undetected non-SM matter, very weakly interacting with SM matter, referred to as dark matter (24% of the universe). The remaining 71% is ascribed to a constant vacuum energy density, required to explain the accelerated expansion of the universe. The SM does not consider this energy densities at all.

These problems seem to suggest that the SM is merely part of a more general theory, like a Grand Unified Theory (GUT), which unifies the strong and EW sector at a high energy scale, or a Theory of Everything, including the General Relativity as well. The most popular of these theories is the Supersymmetry (SUSY) [38]. It relates fermions and bosons by introducing operators that turns fermionic states into bosonic states and vice versa. This extension predicts the existence of new supersymmetric partners of the SM particles (higgsinos, squarks, sleptons, goldstinos, neutralinos, charginos, and gluinos) and provides an explanation for the hierarchy problem and the dark matter.

We implicitly assume that there is only one ϕ doublet, but there is no reason except for simplicity. Two or more doublets could exist: this is translated into the existence of multiple neutral and charged Higgs bosons. A particularly important model, the Two Higgs Doublet Model (2HDM) [39], assumes the existence of two separate scalar complex doublets: $\phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix}$ and $\phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix}$ with vacuum expectation values $\langle \phi_1 \rangle = \begin{pmatrix} 0 \\ v_1 \end{pmatrix}$ and $\langle \phi_2 \rangle = \begin{pmatrix} 0 \\ v_2 \end{pmatrix}$.

With a Higgs boson mass calculated at a per-mill accuracy level [40], the Standard Model is fully determined: future global fits and precision measurements will verify the internal consistency of the model [41] or, in case of discrepancies, point to beyond SM theories.

Next Section generously draws from Reference [42], which explains why the top Yukawa coupling y_t can help to get an idea about a possible scale of new physics.

2.7 Instability of the universe connected to the top Yukawa coupling

The “message” from the LHC can be, so far, interpreted as follows: the SM is a self-consistent, weakly coupled effective field theory all the way up to the Planck scale (M_{PL}); no significant deviations from the SM predictions are observed and no convincing signal in favour of existence of new physics beyond the SM is found.

A possible scale of new physics can be defined by determining the energy where the SM becomes theoretically inconsistent or contradicts observations. Since the SM is a renormalisable theory, the problems can appear only because of the renormalisation evolution of coupling constants, *i.e.* when the model enters strong coupling at that scale (and the coupling constants become infinite), or additional

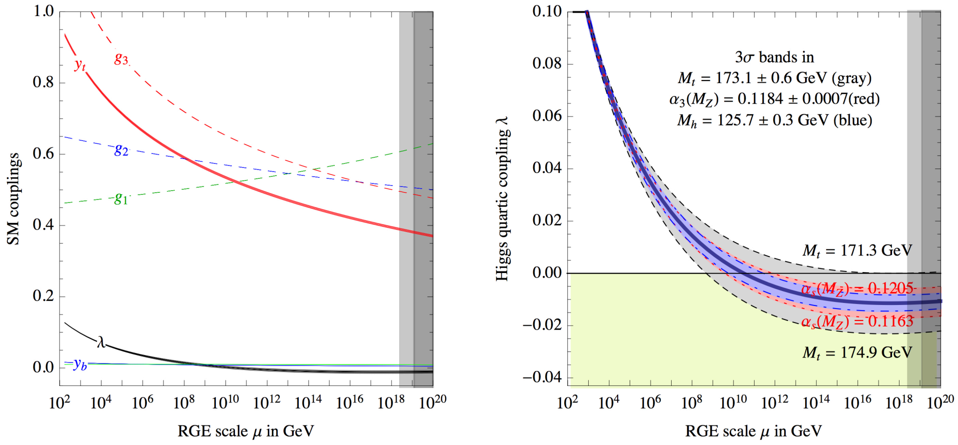


Figure 2.12: Left: SM renormalisation group evolution of the gauge couplings $g_1 = 5/3g'$, $g_2 = g$, $g_3 = \alpha_s$, of the top and bottom Yukawa couplings (y_t,y_b), and of the Higgs quartic coupling λ. All couplings are defined in the MS scheme. The thickness indicates the $\pm 1\sigma$ uncertainty. Right: renormalisation group evolution of λ varying m_t and α_s by $\pm 3\sigma$ [30].

minima of the effective potential develop changing the vacuum structure. The most dangerous constant[9] is the Higgs boson self-coupling constant with the renormalisation group evolution at one loop [42] (Figure 2.12)

$$16\pi^2 \frac{d\lambda}{d\ln\mu} = 24\lambda^2 + 12\lambda y_t^2 - 9\lambda(g^2 + \frac{1}{3}g'^2) - 6y_t^4 + \frac{9}{8}g^4 + \frac{3}{8}g'^4 + \frac{3}{4}g^2 g'^2 \,, \tag{2.45}$$

which depends on the interplay between the positive contributions of the bosons ($\propto g^4$) and the negative contribution from the top quark ($\propto y_t^4$).

Before the discovery of the Higgs the instability constraints were shown as a function of the Higgs mass m_H, with other parameters of the SM fixed by experiments, as also shown in Section 2.5.1. The mass of the Higgs boson is measured to be within the interval $m_{\text{min}}^{\text{meta}} < m_H < m_{\text{max}}^{\text{Landau}}$, allowed by theoretical constraints, meaning that our vacuum is metastable with the life-time exceeding that of the universe by many orders of magnitude [30], and that the SM without gravity is a weakly coupled theory even for energies exceeding the Planck scale, also by many orders of magnitude.

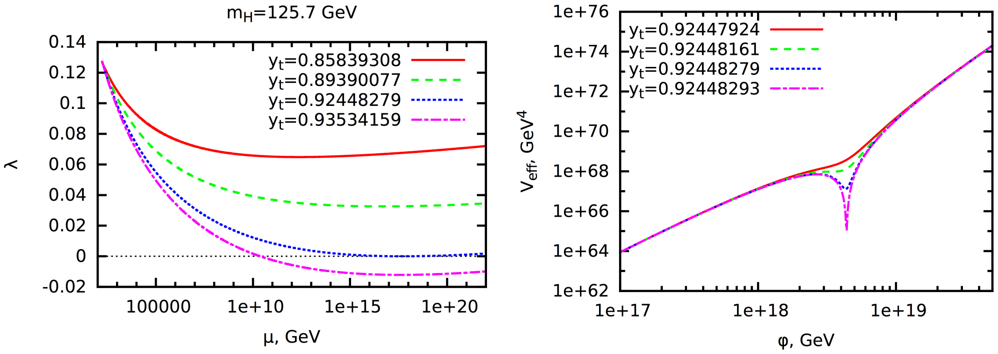
Thus, it seems that no hint about the scale of new physics can be inferred from these considerations. However, an alternative way to proceed is to assume that there is no new physics up to the Planck scale and see if the model runs into any contradiction.

One can start from the SM without gravity and consider effective potential for the Higgs field. Since it has the largest Yukawa coupling to the Higgs boson, the contribution of the top quark to the potential is very important (Equation 2.45): it comes with the minus sign and is responsible for appearance of the extra minimum of the effective potential at large values of the Higgs field. All parameters of the SM are fixed to their experimental values except the top Yukawa coupling , which is at present the most uncertain one for the problem under consideration.

The renormalisation-group evolution of the Higgs coupling λ as function of the energy scale μ for various top-quark Yukawa couplings is shown in Figure 2.13(a). Close to the “critical” value of the top

---

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



(a) (b)

Figure 2.13: (a) Renormalisation group running of the Higgs coupling constant λ for the Higgs mass $m_H = 125.7$ GeV and several values of the top quark Yukawa $y_t(\mu = 173.2 \text{ GeV})$; (b) a very small change of the top Yukawa coupling y_t converts the monotonic behaviour of the effective potential for the Higgs field to that with an extra minimum at large values of the Higgs field [30].

Yukawa coupling the effective potential behaves as illustrated in Figure 2.13(b); the critical value y_{crit} is defined as the value at which our electroweak vacuum is degenerate with a new one at a certain energy scale Λ. Then, three scenarios are possible:

- $y_t < y_{\text{crit}} - \epsilon$, the most cosmologically safe case, as our electroweak vacuum is unique;
- $y_t > y_{\text{crit}} - \epsilon$, a new minimum develops at large values of the Higgs field, however, in the interval $y_t \in [y_{\text{crit}} - 1.2 \times 10^{-6}, y_{\text{crit}}]$ the electroweak vacuum is the global minimum, so that the evolution of the universe should lead the system to our vacuum rather than the vacuum with a large Higgs field;
- $y_t > y_{\text{crit}}$, the new minimum is deeper than ours, meaning that our vacuum is not stable anymore. In this conditions the value of the potential barrier which separates our electroweak vacuum from that one at large values of the Higgs field plays a role[10]. If the lifetime of the fluctuation through the potential barrier exceeds the age of the universe, it is conceivable to think that the presence of another vacuum is not important; our vacuum is metastable and the happy ending is quite plausible;
- $y_t > y_{\text{crit}} + \eta$, the life-time of our vacuum is smaller than the age of the universe, which thus becomes unstable. A fluctuation of the Higgs field should have already driven the system to another vacuum.

The parameters ϵ and η strongly depend on the accuracy of loop corrections to $V(\phi)$. The most accurate results are $\epsilon = 1.2 \times 10^{-6}$ and $\eta = 0.04$, where $y_{\text{crit}} + 0.04$ corresponds roughly to the top-quark mass 178 GeV [42].

The stability bound y_{crit} is found to be

$$y_t^{\text{crit}} = 0.9244 + 0.0012 \cdot \frac{m_H/\text{ GeV} - 125.7}{0.4} + 0.0012 \cdot \frac{\alpha_s(m_Z)/\text{ GeV} - 0.1184}{0.0007}, \tag{2.46}$$

where the QCD coupling α_s is considered at the Z boson mass [42].

In practice, if the measurement of the top Yukawa coupling will give $y_t < y_{\text{crit}} + \eta$, the embedding of the SM with new physics in cosmology does not lead to any troubles and thus no information on the scale of new physics can be derived. If $y_t > y_{\text{crit}} + \eta$ the Higgs self-coupling λ becomes negative, indicating

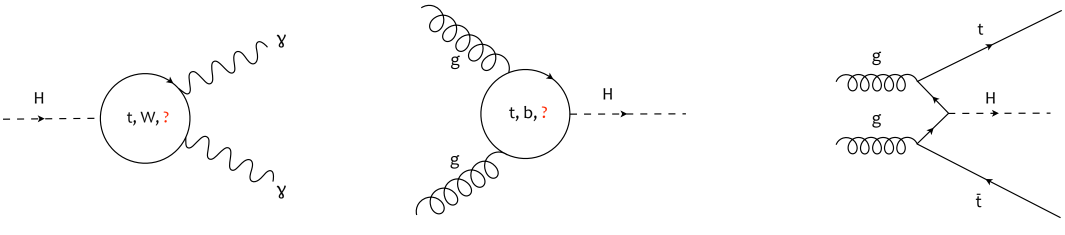
---

[10] The energy density corresponding to this extreme is gauge-invariant and does not depend on the renormalisation scheme.

an instability at the scale Λ: to make it positive for all energies, something new should intervene at the scale around or below Λ. Several papers argued that this is exactly what is going to happen [43, 44]: this statement is only true if the potential for the Higgs field is not modified by the gravitational effects or by the presence of some new physics at the inflationary energy scale (10^{16} GeV).

2.7.1 Direct measurement of the top Higgs Yukawa coupling

The Yukawa coupling y_t of the Higgs boson to the top quark is experimentally accessible by measuring the top-quark mass or the cross section of the gluon fusion (ggF) production process and the $H \to \gamma\gamma$ decay mode, where a sizeable contribution comes from a top-quark loop, shown in Figures 2.14(a) and 2.14(b). However, these cases require the assumption that no new physics contributes with additional induced loops in the measurement of y_t. Currently, the only process where y_t can be accessed directly is the production of a top-quark pair in association with a Higgs boson ($t\bar{t}H$), shown in Figure 2.14(c).



(a) (b) (c)

Figure 2.14: Example Feynman diagram for (a) the effective photon vertex, $\gamma\gamma H$, (b) the effective gluon fusion vertex, ggH, and (c) the $t\bar{t}H$ process.

The results of the searches for the Higgs boson are usually expressed in terms of the signal strength parameter μ, which is defined as the ratio of the observed to the expected number of signal events. The latter is calculated using the SM cross section.

The ATLAS and CMS collaborations have searched for the production of $t\bar{t}H$ in pp collisions at the LHC using data collected at a centre-of-mass energy of 7 and 8 TeV, with analyses sensitive to $H \to WW^*$, $\tau\tau$, $b\bar{b}$, and $\gamma\gamma$ decays [45–49]. The combined $t\bar{t}H$ signal strength measured by the CMS Collaboration, obtained by merging searches in several final states, is $\mu = 2.8 \pm 1.0$ [48]. The ATLAS Collaboration has searched for a $t\bar{t}H$ signal in events enriched in Higgs boson decays to two massive vector bosons or τ leptons in the multileptonic channels [45], finding $\mu = 2.1^{+1.4}_{-1.2}$, for $t\bar{t}H$ ($H \to b\bar{b}$) [46] in final states with at least one lepton obtaining $\mu = 1.5 \pm 1.1$, and for $t\bar{t}H$ ($H \to \gamma\gamma$) [47] measuring $\mu = 1.3^{+2.6}_{-1.7}$. The combination of ATLAS and CMS results yields a best fit of $\mu_{t\bar{t}H} = \sigma/\sigma_{\rm SM} = 2.3^{+0.7}_{-0.6}$, with the excess over the Standard Model expectation ($\mu_{t\bar{t}H} = 1$) driven primarily by the multilepton final states [50].

There are also measurements at 13 TeV, since Summer 2016. The combined $t\bar{t}H$ signal strength measured by CMS is $\mu = 2.0 \pm 0.8$ [51], while the combined value measured by ATLAS is $\mu = 1.7 \pm 0.8$ [52]. The ATLAS result in the multileptonic channels is $\mu = 2.5^{+1.3}_{-1.1}$ [53], $\mu = 2.1 \pm 1.1$ in final states with at least one lepton [54], and $\mu = -0.3^{+1.2}_{-1.0}$ for $t\bar{t}H$ ($H \to \gamma\gamma$) [55]. The excess is stil driven by the multilepton final states.

CHAPTER 3

The ATLAS experiment at the Large Hadron Collider

The European Organisation for Nuclear Research (CERN) is one of the largest and renowned centres for scientific research in the world. Its main activity is the study of the basic constituents of the universe, in order to understand the fundamental mechanisms that rule it. For this purpose, very complex and up-to-date scientific instruments are employed.

The CERN Laboratory is located in between the Franco-Swiss border, nearby Geneva. Founded in 1954 to create an European scientific centre of excellence, it was one of the first Europe's joint projects, with currently 22 member states. Particle accelerators and detectors are the instruments used at CERN: an accelerator produces collisions between particles (protons, ions) at very high energy, and detectors reconstruct and record the products of these collisions.

This Chapter introduces the Large Hadron Collider accelerator (LHC) and provides a short description of the ATLAS experiment: the detector components allow to trigger on and collect the data events needed for physics analyses. The software algorithms used for identifying and reconstructing physics objects (tracks, vertices, electrons, muons, jets) are also described. Finally, an overview on the simulation datasets (Monte Carlo simulation) used in ATLAS is provided.

More detailed descriptions of the structure and functions of each ATLAS subsystem can be found in References [56, 57] and [58].

3.1 The Large Hadron Collider

The Large Hadron Collider [59] at CERN is the highest energy collider ever built, dedicated to accelerating and colliding protons. It was designed to provide proton-proton (pp) collisions with a centre-of-mass energy of 14 TeV and an instantaneous luminosity of 10^{34} $\text{cm}^{-2}\text{s}^{-1}$, and lead ion collisions at a centre-of-mass energy of 2.76 TeV per nucleon and an instantaneous luminosity of 10^{27} $\text{cm}^{-2}\text{s}^{-1}$.

As illustrated in Figure 3.1, the rotating proton beams of the LHC collide in four interaction points (the four yellow dots), where four particle detectors have been build in order to analyse the products of the high-energy collisions. Data recorded by the ATLAS ("**A** **T**oroidal **LHC** **A**pparatu**S**") detector [58] are used in this thesis. The other three detectors are CMS (Compact Muon Solenoid), ALICE (A Large Ion Collider Experiment) and LHCb (b stands for beauty). In addition, there are smaller experiments, such as TOTEM and LHCf, installed at some distance from the interaction points to study the production of particles in the forward region (along the beam direction).

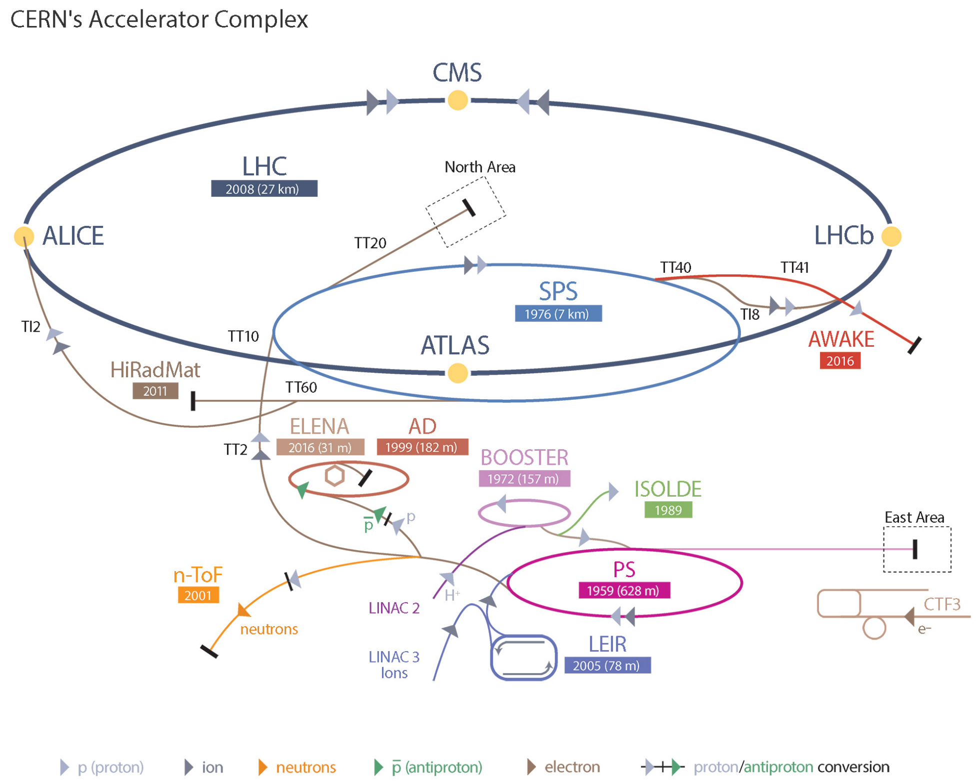


Figure 3.1: The layout of the LHC and the CERN accelerator complex acting as the injector chain for the LHC [59].

3.1.1 Machine design

Housed in the tunnel built between 1984 and 1989 for the LEP (**L**arge **E**lectron–**P**ositron Collider), the LHC is a 27 km long superconducting hadron collider. The tunnel is located between 45 m and 170 m below the ground surface, between the Jura mountains and the Geneva airport.

LHC magnets are made with niobium-titaniun (NbTi) cables and are cooled to less than 2 K with superfluid helium, in order to reach superconductivity (9.2 K) despite the high currents (11.850 A) and large magnetic fields (8.33 T). The large magnetic fields bend the 7 TeV proton beams around the LHC ring: in addition dipole, quadrupole and higher order magnets are used respectively to bend, correct and shrink the beam into the small area where collisions take place.

In order to have two counter circulating proton beams along the same circumference, oppositely oriented magnetic fields are needed. Due to the limited space, only a single cryogenic structure fits in the tunnel. The issue is solved employing a very complex twin-bore design, having both proton rings in the same cryostat. Figure 3.2 shows an example of the LHC twin-bore dipole magnets.

This advanced collider is designed to accelerate protons to an energy of 7 TeV, starting from an initial energy of 450 GeV. The already existing CERN accelerators system is used to accelerate protons up to 450 GeV, and inject them into the LHC ring later on.

Beams are injected into the LHC in a series of bunches of 1.15×10^{11} protons and every beam is designed to have 2808 circulating proton bunches. The bunches are arranged in “trains” of 72 bunches, with 25 ns spacing within the train, and 12 empty bunches between two trains.

Collisions between circulating beams occur at every “bunch crossing”, resulting in a peak collision

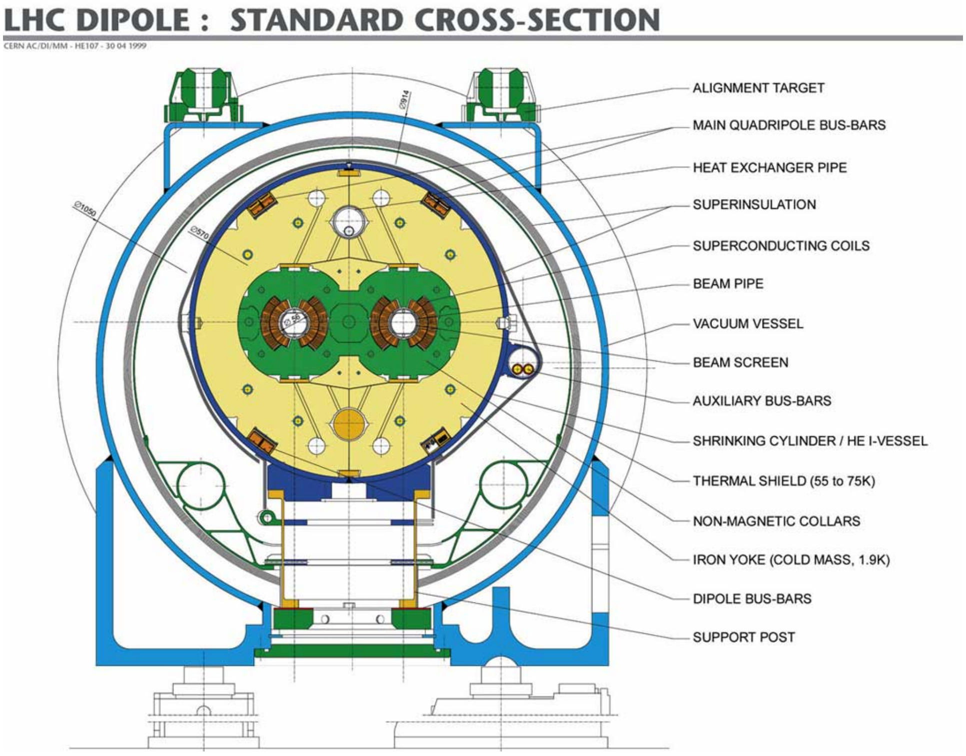


Figure 3.2: An example of an LHC dipole magnet with the twin bore design [60].

rate of 40 MHz. The beams are squeezed to a transverse size of $\sim$ 17 μm at the interaction point (IP) to maximise the pp collision rate. Near the IP, the two beams are kept together in a single beam pipe, for approximately 140 m in each direction. To avoid unwanted collisions in the pipe, the beams stay on parallel orbits. When the beams are ready for colliding at the interaction point, the separation is removed.

The rate at which collisions occur depends on the instantaneous luminosity $\mathcal{L}$ and the collision cross section σ, related by:

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma \,. \tag{3.1}$$

The rate at which a particular physics process occurs depends on the cross section for the process in question. Since many of the interesting physics processes at the LHC have small cross sections, it is important to maximise the luminosity as much as possible. The instantaneous luminosity is given by:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\mathrm{rev}} F \gamma_{\mathrm{r}}}{4\pi \varepsilon_n \beta^*} \,, \tag{3.2}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam,f_{rev} the revolution frequency; F is a geometric function to account for the crossing angle between the beams (since they are generally not collided head on); γ_{r} stands for the relativistic Lorentz factor $(1 - v^2/c^2)^{-1/2}$; ε_n is the beam emittance, a measure of the uniformity the momentum of particles in the beam, while β^* is a measure of how narrow the beam is at the interaction point.

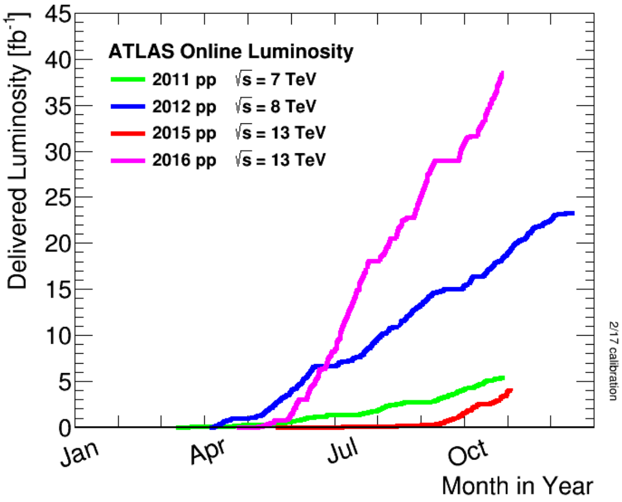


Figure 3.3: Cumulative luminosity versus day delivered to ATLAS during stable beams and for high energy *pp* collisions [61].

Parameter	Nominal	2011 Operation	2012 Operation	2015-2016 Operation
Proton Energy	7 TeV	3.5 TeV	4 TeV	6.5 TeV
N_b	1.15×10^{11}	1.5×10^{11}	1.6×10^{11}	1.18×10^{11}
n_b	2808	1380	1380	2736
Bunch spacing[ns]	25	50	50	25
β^*[m]	0.55	1.0	0.6	0.4
$\varepsilon_n[\mu m]$	3.75	1.9 – 2.3	1.7 – 3.0	2.6 – 3.5
Peak $\mathcal{L}$[cm^{-2}s^{-1}]	1.0×10^{34}	3.6×10^{33}	7.7×10^{33}	1.37×10^{34}

Table 3.1: LHC operational parameters. A comparison is made of the nominal design parameters, and those used in 2011-2012 operations and 2015-2016 operations [59].

3.1.2 The LHC operation in Run 1 and Run2

The LHC began operation in November 2009 with collisions at a centre-of-mass energy of 900 GeV, with the centre-of-mass energy rising to 2.36 TeV by the end of that year. In 2010 the centre-of-mass energy was successfully increased to 7 TeV. During the years 2010 and 2011 the LHC continued to run at $\sqrt{s}$ = 7 TeV, with the instantaneous luminosity steadily increasing. In 2010 and 2011 the LHC delivered 48.1 pb^{-1} and 5.43 pb^{-1} of integrated luminosity to ATLAS. In 2012 the centre-of-mass energy was increased to 8 TeV, and the instantaneous luminosity further increased, leading to a total integrated luminosity delivered to ATLAS in 2012 of 22.8 fb^{-1}.

The second phase of LHC, named Run 2, started in 2015, after a long shutdown, with collisions at a centre-of-mass energy of 13 TeV. The peak instantaneous luminosity achieved was 13.7×10^{33}cm^{-2}s^{-1}, at a bunch crossing of 25 ns. The total luminosity delivered so far is 43.1 fb^{-1}. Figure 3.3 shows the delivered luminosities as function of time for the 2011-2016 period. In Table 3.1 details of the LHC operational parameters, together with the nominal design values, are given.

3.2 The ATLAS detector

Built to study both *pp* and ion-ion interactions, ATLAS is one of two general purpose particle physics detectors at the LHC. The high centre-of-mass energy and the high luminosity of the LHC *pp* collisions

Detector Component	Design Resolution	η Coverage	
		Measurement	Level 1 Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	±2.5	None
EM Calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	±3.2	±2.5
Hadronic Calorimetry			
Barrel and End-Cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	±3.2	±3.2
Forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$	$3.1 < \lvert\eta\rvert < 4.9$
Muon Spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	±2.7	±2.4

Table 3.2: Performance goals of the ATLAS detector. Units of p_T and E are GeV.

allow for the study of physics at the TeVscale. The detector has been designed to allow several types of research:

- the search of the Higgs boson and the measurement of its properties;
- Supersymmetry searches;
- high precision tests of QCD, flavour physics and electroweak interactions;
- measurements of the properties of the top quark;
- searches for new vector bosons and searches for extra-dimensions (the so-called exotic searches).

The high luminosity allows for the study of rare processes, but involves also a complicated scenario. At the design luminosity, 10^9 inelastic collisions occur per second, resulting in multiple scattering. The mean number of interactions per crossing was observed to be ~ 20 during the 2012 data taking period. The detector has been designed to cope with these high “pileup” conditions, as well as to be capable of operating in the high radiation environment arising from the high luminosity. The detector must be able to distinguish processes of interest from the background: many of the physics processes of interest often occur at very small rates with respect to extremely high QCD background rates.

In order to deal with these challenges, ATLAS was designed to have:

- full azimuthal coverage, for missing transverse energy measurement, and large acceptance in pseudo-rapidity;
- high granularity, to cope with high particle fluxes and overlapping events;
- precision tracking, to provide high charged particle momentum resolution and reconstruction efficiency, and to allow observation of secondary vertices to identify b-hadrons and τ-leptons;
- precise electromagnetic calorimetry, for electron and photon identification;
- full-coverage hadronic calorimetry, for accurate jet and missing transverse energy measurements;
- high muon identification efficiency, momentum resolution and charge determination over a wide range of momentum;
- efficient triggering on low transverse-momentum objects.

In Table 3.2 the main performance goals are given.

Figure 3.4 shows a scheme of the ATLAS detector: it consists of an inner tracking detector (ID), which is surrounded by electromagnetic calorimeters, hadronic calorimeters and a muon spectrometer. The inner detector is immersed in a solenoidal field of 2 T to allow for momentum measurement. The muon spectrometer is also immersed in a magnetic field, provided by an air-core toroid system which generates strong bending power over a large volume with a minimum of material, allowing to minimise multiple-scattering effects. A three-level trigger system is used to select events to read out and is described in Section 3.3. The various sub-systems are described in detail in the following Sections.

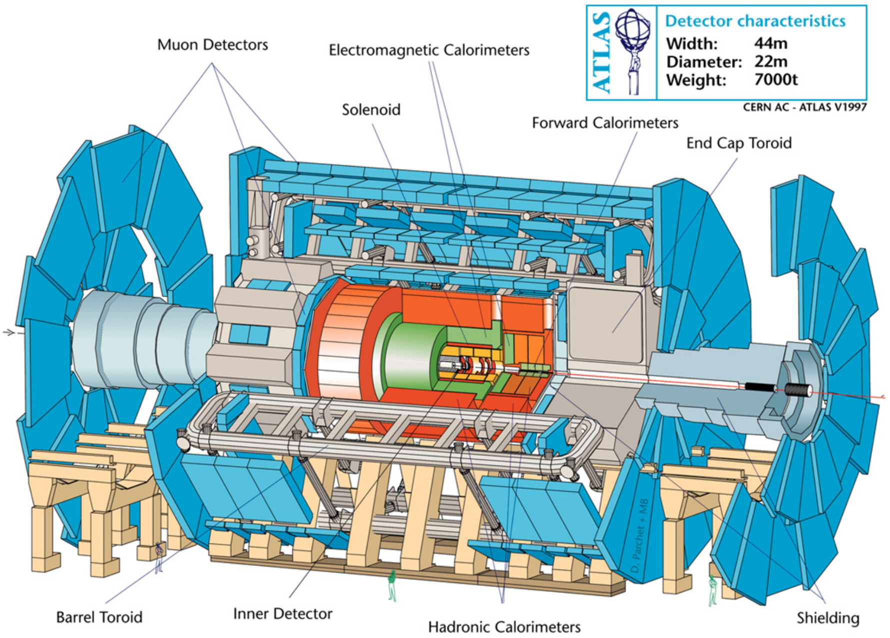


Figure 3.4: Cut-away view of the ATLAS detector. The various detector sub-systems are labelled [62].

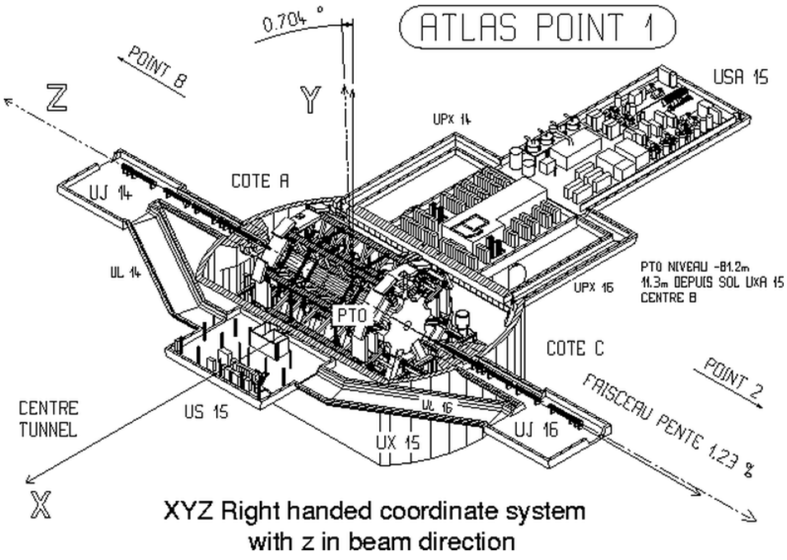


Figure 3.5: The coordinate system in the ATLAS detector. The general tilt of the LEP/LHC tunnel causes the y-axis to be slightly different from vertical [58].

3.2.1 Coordinate system

oThe ATLAS coordinate system is a right-handed Cartesian system with the origin located at the nominal interaction point, as illustrated in Figure 3.5. The z-axis lies along the beam line, while the x-y plane is transverse to the beam line, with positive x pointing into the centre of the LHC ring and positive y pointing upward.

A cylindrical coordinate system is employed when referring to the coordinates of a physics object in the detector. In this system θ denotes the polar angle, r and ϕ denote the radius and the azimuthal angle in the x-y plane. The two angles, θ and ϕ, are measured respectively from the positive z-axis and from the positive x-axis. The θ angle is usually converted in the pseudo-rapidity,

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

which approaches the rapidity

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

in the limit where $E \gg m$. The pseudo-rapidity is 0 in the transverse plane and infinity along the z axis, with $\eta = 1$ at 45° from the axis. The rapidity and the pseudo-rapidity are natural variables for describing angles in a system where the initial z-momentum is unknown, since the difference in rapidity between two particles is invariant under boosts along the z axis. A commonly used quantity is the angular distance between objects in the ϕ/η plane, defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The energy and momentum of outgoing particles, E and p, are often projected onto the transverse plane: the transverse momentum conservation can be required, since the initial component is known to be zero, whereas the initial component along the z axis is not known. The transverse momentum is then defined as $p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}$, and the transverse energy as $E_\mathrm{T} = E\sin\theta$.

In a solenoidal magnetic field charged particles follow a helical trajectory, which can be computed at each point in space using five parameters: r, z, ϕ, θ, q/p. The variables r, z, ϕ and θ are the cylindrical coordinates already described, while the variables q and p are the charge of the track and its momentum, respectively, and q/p represent the bending of the track. Other parameterisations focus on the track parameters closest to the interaction point, like d_0, defined as the transverse impact parameter representing the transverse distance to the beam axis at the closest approach point, and z_0, the longitudinal impact parameter at the closest approach point.

3.2.2 Magnetic field

Four superconducting magnets are used to provide the magnetic field for bending charged tracks[1]. The magnetic field in the inner detector is supplied by a solenoid producing a 2 T field in the z-direction. The muon spectrometer magnetic field, instead, is provided by three air-core toroid magnets producing fields between 0.5 T and 4 T in the ϕ direction. As a result, the z and ϕ fields bend tracks in the ϕ direction in the inner detector, and in the η direction in the muon spectrometer.

Both magnets are made of aluminium (Al) stabilised niobium-titanium (NbTi) cables cooled to 4.5 K. To reduce the material thickness and the resulting energy loss of tracks, the solenoid has a thickness of ~ 0.66 radiation lengths (10 cm) only, and it is situated in the same cryostat as the electromagnetic calorimeter. For the same reasons, the toroid structure is inspired by an air-core design: muons can traverse the magnetic field without crossing any of the superconducting coils. The size of the toroid provides a large bending volume for muons and results in a better lever arm for the muon spectrometer tracking.

The layout of the magnet system is shown in Figure 3.6: The green cylinder in the centre is the solenoid magnet, while the blue and read ovals are the three air-core toroids, for the barrel and the end-cap regions.

---

[1] The momentum component of a charged track that is perpendicular to a uniform magnetic field can be estimated by measuring its bending radius R in a magnetic field B: $p[\mathrm{GeV}] = 0.3B[\mathrm{T}]R[\mathrm{m}]$.

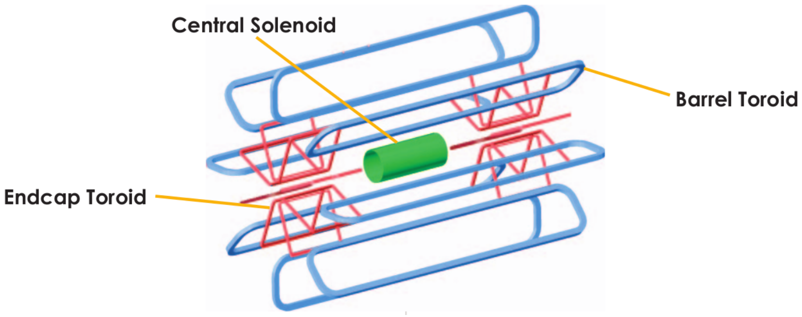
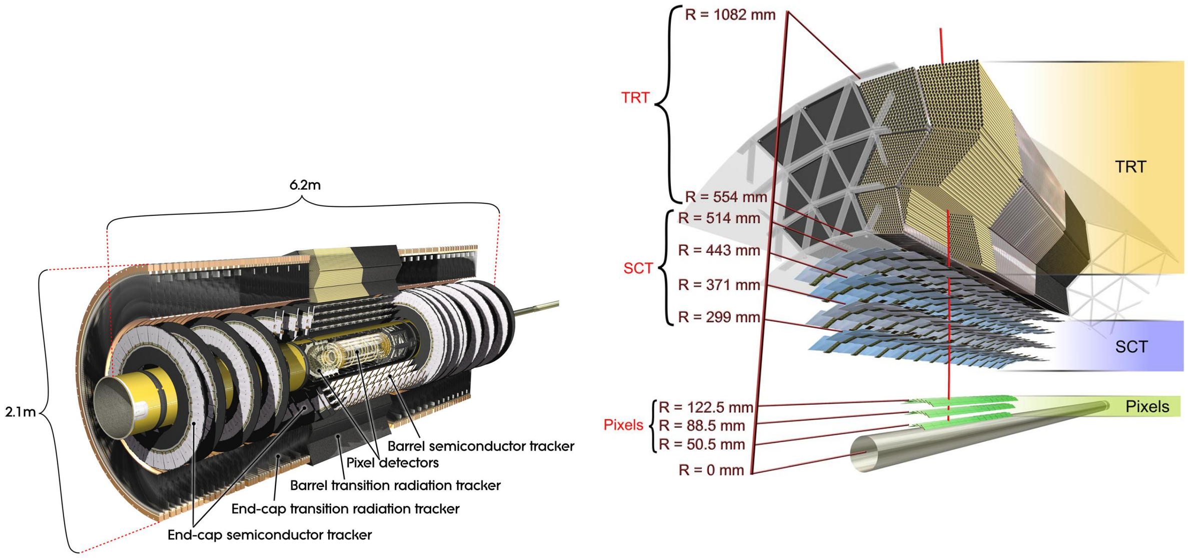


Figure 3.6: Layout of the ATLAS magnet system.



(a) (b)

Figure 3.7: (a) Layout of the ATLAS Inner detector of Run 1; the IBL is not shown since this additional pixel layer was introduced in Run 2. (b) A 3D zoomed view of the ATLAS Inner detector, consisting of three subdetectors: the Pixel Detector, the SemiConductor Tracker and the Transition Radiation Tracker.

3.2.3 Inner detector

The inner detector reconstructs the paths of charged particles, as tracks, while they traverse a 2 T solenoidal magnetic field. Tracks are reconstructed through high-resolution position measurements known as hits, and multiple tracks are combined then to reconstruct vertices. With a radius of 1.1 m and a length of 7 m, the ID has a cylindrical shape around the interaction point. Three technologies are exploited to measure track hits: an innermost pixel detector composed of silicon pixels, an intermediate silicon strip detector (SCT) and an outermost transition radiation tracker built from small drift tubes (TRT). Figure 3.7(a) shows a three-dimensional illustration of the ID layout, while Figure 3.7(b) includes a more detailed layout.

The inner detector provides accurate and efficient tracking for charged particles with $p_{\mathrm{T}} > 0.5$ GeV within $|\eta| < 2.5$, with a transverse momentum resolution of 1% (5%) for track p_{T} of 0.5 GeV (100 GeV), as specified in Table 3.2. Besides single-track measurements, multiple tracks can be combined to reconstruct primary vertices from pp collisions and secondary vertices from the decays of long-lived

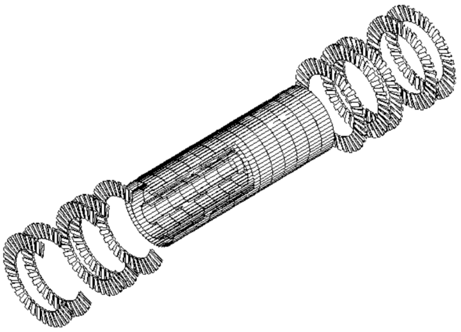


Figure 3.8: Layout of the ATLAS Pixel detector.

particles.

The ID results in a combination of different technologies, briefly summarised in the following Sections.

Silicon pixel tracker and the Insertable B-Layer (IBL)

The pixel detector, as shown in Figure 3.8, is the closest component to the beam. Formed of layers of silicon pixels, it is designed to have a very high granularity for resolving primary and secondary interaction vertices. It is composed by three cylindrical layers in the barrel region, closed by an endcap consisting of three disks at each end. The B-layer, the closest layer to the beam pipe, positioned at a radius of 50.5 mm, plays an important role in detecting secondary vertices for the identification of b-jets.

Because of the high radiation dose that it will receive at this position, it was expected to be replaced after five years of operation. The time needed to replace it would have been more than one year because of the long cooling down time of the activated material inside the detector. It was instead decided to install a fourth pixel layer inside the existing detector, the Insertable B-Layer (IBL) at a radius of 33 mm from the beam axis, in 2014 [63, 64]. This new B-layer fits into the pixel detector thanks to a smaller-radius beampipe. The new pixel layer provides an additional space point very close to the interaction point, which keeps the performance of the tracking while the older B-layer continues to degrade.

Particles with $|\eta| < 2.5$ will traverse four layers of the detector; in most case producing four space-points. The pixel detector allows for a resolution of $\sigma_\phi = 10\,\mu\text{m}$ in the bending direction (ϕ), and $\sigma_{z,R} = 115\,\mu\text{m}$ in the z (barrel) or R (end-cap) direction.

Semiconductor Tracker (SCT)

The SCT is a silicon strip detector, composed of four barrel layers and two end-caps consisting of nine disks each. The barrel layers consist of 2112 separate modules; each endcap consists of 988 modules, disposed in such a way that a particle will pass through four layers of the detector.

Each SCT module is made of two layers of single sided p-in-n silicon chips. Charged particles passing through the depletion region of the module junction produce electron hole pairs, which are swept apart by the bias voltage. The electrons are collected on the top of the chip, producing a signal which can be read out.

The spatial resolution of the detector is $\sigma_\phi = 17\,\mu\text{m}$ in the bending direction (ϕ), and $\sigma_{z,R} = 580\,\mu\text{m}$ in the z (barrel) or R (end-cap) direction.

Transition Radiation Tracker (TRT)

The Transition Radiation Tracker is a straw drift tube tracker, with additional particle identification capabilities from transition radiation. It is composed of modules formed from bundles of 4 mm diameter straws, filled with a gas mixture consisting of 70% Xe, 27% CO_2 and 3% O_2. The charge is collected through a tungsten wire that runs down the centre of the tube. In the barrel the straws are parallel to the beam axis and are electrically divided into two halves at $|\eta| = 0$ and read out at either end. This subdivision leads to an inefficiency along a length of 2 cm at the centre of the TRT. In the endcaps the straws are radially disposed.

Charged particles with $p_T > 0.5$ GeV and $|\eta| < 2.0$ will traverse at least 36 straws, except in the barrel to endcap transition region ($0.8 < |\eta| < 1.0$) where only 22 straws will be traversed. In the bending direction (ϕ) the spatial resolution is $\sigma_\phi = 130\,\mu$m. Despite the low resolution compared to the silicon trackers, and the lack of a measurement in the z direction, the TRT contributes significantly to the pattern recognition and momentum resolution thanks to the large number of measurements and longer measured track length.

3.2.4 Calorimetry

The ATLAS calorimeter systems sit outside the inner detector and its magnetic field. The purpose of the calorimeter is to measure the energy and position of particles. A particle entering the calorimeter produces a “shower” of secondary particles; the energy of this shower is then measured. ATLAS uses sampling calorimeters, where different materials, sandwiched together in layers, are used to initiate the shower development (absorption) and to measure the energy of its constituents. This allows for a more compact design and, hence, better shower containment. The position measurement is obtained by segmenting the calorimeter in the z and ϕ directions. Different absorbers are required depending on whether the particle interacts via the electromagnetic or the strong force; consequently, the showers that develop feature different properties.

ATLAS calorimeters are divided into two distinct subsystems, the electromagnetic calorimeter and the hadronic calorimeter. An electromagnetic shower consists of electrons, positrons and photons, and is normally fully contained in the calorimeter; therefore, it can be fully detected. Hadronic showers involve many more particle types, including neutrons, muons and neutrinos (which escape detection), and tend to be longer and wider, often spilling out of the calorimeter: thus, the energy of the shower is not fully detected and a calibration of the energy response is needed. It is important for the calorimeter to provide good containment of electromagnetic and hadronic showers, not only for the purposes of energy measurement, but also to allow a good missing transverse energy measurement, and to prevent punch-through into the muon system.

A cutaway view showing the location of the various calorimeter elements is shown in Figure 3.9. The calorimeters cover the range $|\eta| < 4.9$. Over the η range of the inner-detector, the electromagnetic calorimeter gives fine granularity to allow precise measurement of electrons and photons. The hadronic calorimeter is more coarsely segmented, but is sufficient to meet the requirements of the jet and missing transverse energy measurement.

Electromagnetic Calorimeters

The electromagnetic (EM) calorimeter (also referred to as the LAr) uses liquid argon as the active detector material, and lead as an absorber. Charged particles in the shower ionise the liquid argon, where the electrons drift to copper electrodes in the presence of an electric field.

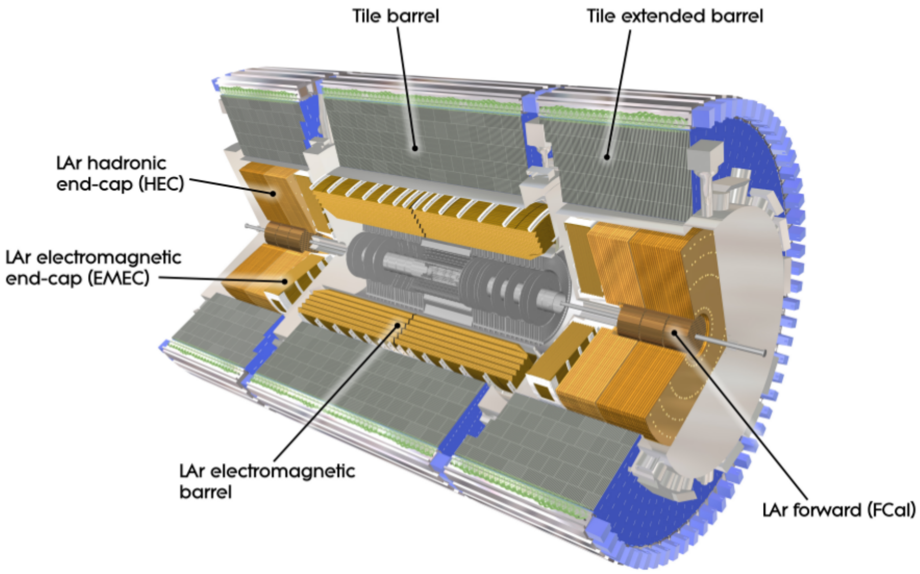


Figure 3.9: Cut-away view of the ATLAS calorimeter system [58].

The LAr consists of two half barrels, extending to $|\eta| < 1.475$ (with a 4 mm gap at $z = 0$), and two coaxial wheels on each side (named the EMEC), the first covering $1.375 < |\eta| < 2.5$ and the second covering $2.5 < |\eta| < 3.2$. Additional material needed to instrument and cool the detector creates a "crack" region at $1.375 < |\eta| < 1.52$, where the energy resolution is significantly degraded.

The barrel calorimeter has an accordion structure in order to avoid azimuthal cracks and to provide full ϕ symmetry, as shown in Figure 3.10. The accordion structure is made of the lead absorber, with the liquid argon filling the 2.1 mm gaps between the absorbers.

Hadronic Calorimeter

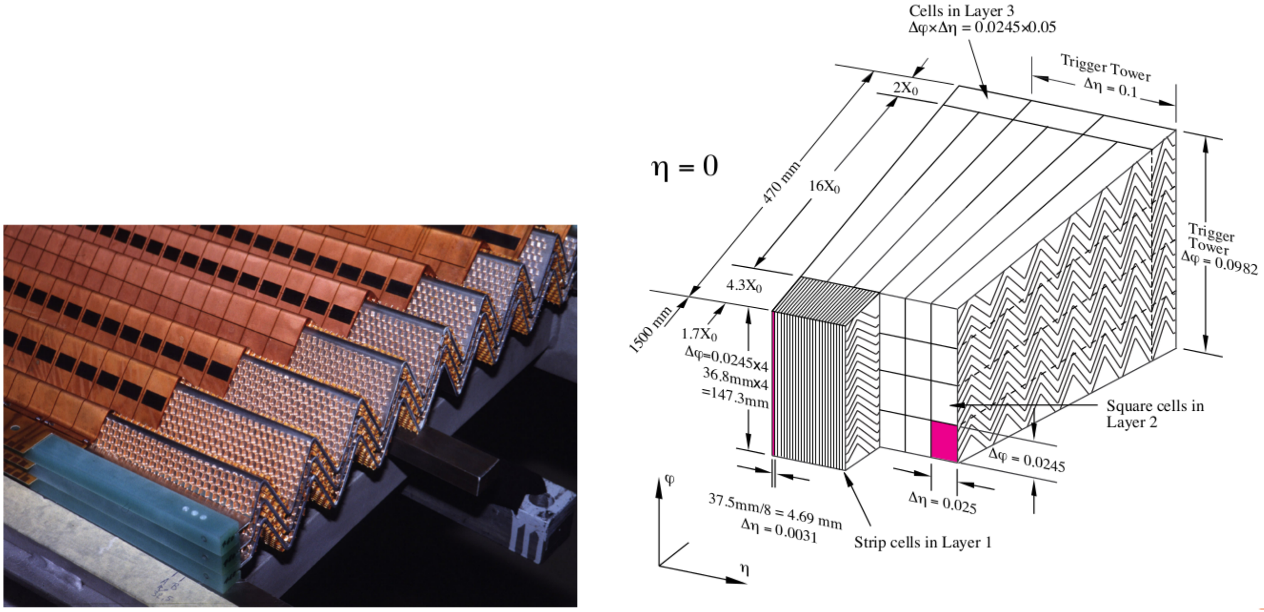
The hadronic calorimeter consists of a plastic scintillator tile calorimeter (referred to as the tile calorimeter) covering $|\eta| < 1.7$ and a liquid argon endcap calorimeter, referred to as the HEC, covering $1.5 < |\eta| < 3.2$ and illustrated Figure 3.9.

The tile calorimeter consists of a barrel covering $|\eta| < 0.8$ and two extended barrels covering $0.8 < |\eta| < 1.7$, and is located immediately behind the EM calorimeter. The active material consists of 3 mm thick layers of the plastic scintillator placed perpendicular to the beam direction, sandwiched between steel absorbers. The scintillators are connected at each end to readout photomultiplier tubes by wavelength-shifting fibres. The fibres are grouped together to form readout cells, giving projective towers in η.

The HEC consists of two wheels per endcap located directly behind the EMEC and sharing the same cryostat. Each wheel has two layers of cells. The HEC covers $1.5 < |\eta| < 3.2$ and so overlaps with the tile calorimeter on one side and the FCAL on the other, thus avoiding cracks in the transition regions.

Forward Calorimeter

The forward calorimeter (FCAL) covers $3.1 < |\eta| < 4.9$. To reduce the neutron flux, the FCAL begins 1.2 m away from the EM calorimeter front face. Due to the high particle fluxes and energies in the forward region, the calorimeter must contain relatively long showers in the small volume allowed by design constraints, and thus must be very dense. The FCAL is divided into three compartments. The first



(a) (b)

Figure 3.10: A photo (a) and a diagram (b) of the ATLAS liquid argon calorimeter, showing the accordion structure and the different granularity in the different layers [58].

is designed for electromagnetic measurements, and uses copper as a passive material with liquid argon as active material. The other two compartments are designed for hadronic measurements, and use tungsten as a passive material, chosen for its high density to provide containment and minimise the lateral spread of hadronic showers.

3.2.5 Muon spectrometer

The ATLAS muon spectrometer (MS) [65] adopts detector technologies for accurate momentum and direction measurements of muons with momenta from ~ 6 GeV up to a few TeV. In the same energy range, the spectrometer provides excellent triggering performance for muons. It fully covers the calorimeter system and occupies a large part of the ATLAS cavern. The bending power of the toroid magnets, located in the barrel and end-cap sides, allow to bend the muon trajectories over a large distance. In order to measure the curvature of the tracks, a very good hit resolution is needed. Figure 3.11 illustrates the complete layout of the toroid magnets and the muon spectrometer. As shown in Table 3.2, its design benchmark is a 10% momentum resolution for 1 TeV muon tracks.

The four subsystems composing the muon spectrometer rely on four detector technologies. Two of them, the Resistive Plate Chambers (RPC) in the barrel region and the Thin Gap Chambers (TGC) in the end-cap region, provide trigger signals. Those are fast tracking detectors with time resolution smaller than the bunch spacing, 25 ns. The Monitored Drift Tubes (MDT) provide high precision measurements in the bending direction over most of the detector acceptance; however, in the forward region, where the particle flux is too high for the MDT chambers, Cathode Strip Chambers (CSC) are employed.

In the Figure 3.12(a) the layout of the MS in the barrel region can be seen in the transverse plane: three concentric layers are used, each one with sixteen sectors of MDT and RPC chambers covering the ϕ region, based on the eight-fold symmetry toroid. Figure 3.12(b) shows, in the longitudinal plane, three

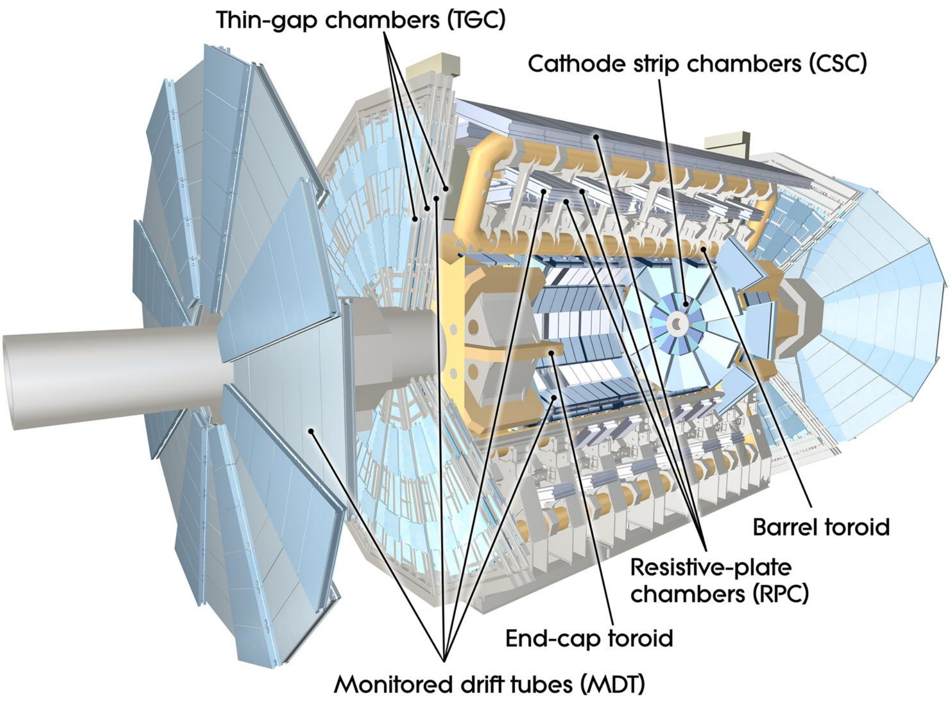


Figure 3.11: A schematic of the ATLAS muon spectrometer [58].

layers of MDT and TGC chambers employed in the end-cap; the position of the CSC chambers in the forward region is also shown. The first two layers of MDT and RPC are immersed in the barrel toroid, whereas in the end-cap region the toroid is located between the first two MDT layers, before the three TGC layers. An additional TGC layer (not labelled in the Figure) lies to the right of the innermost MDT layer, in the end-cap, providing a ϕ-coordinate measurement used for tracking. More details can be found in Reference [65].

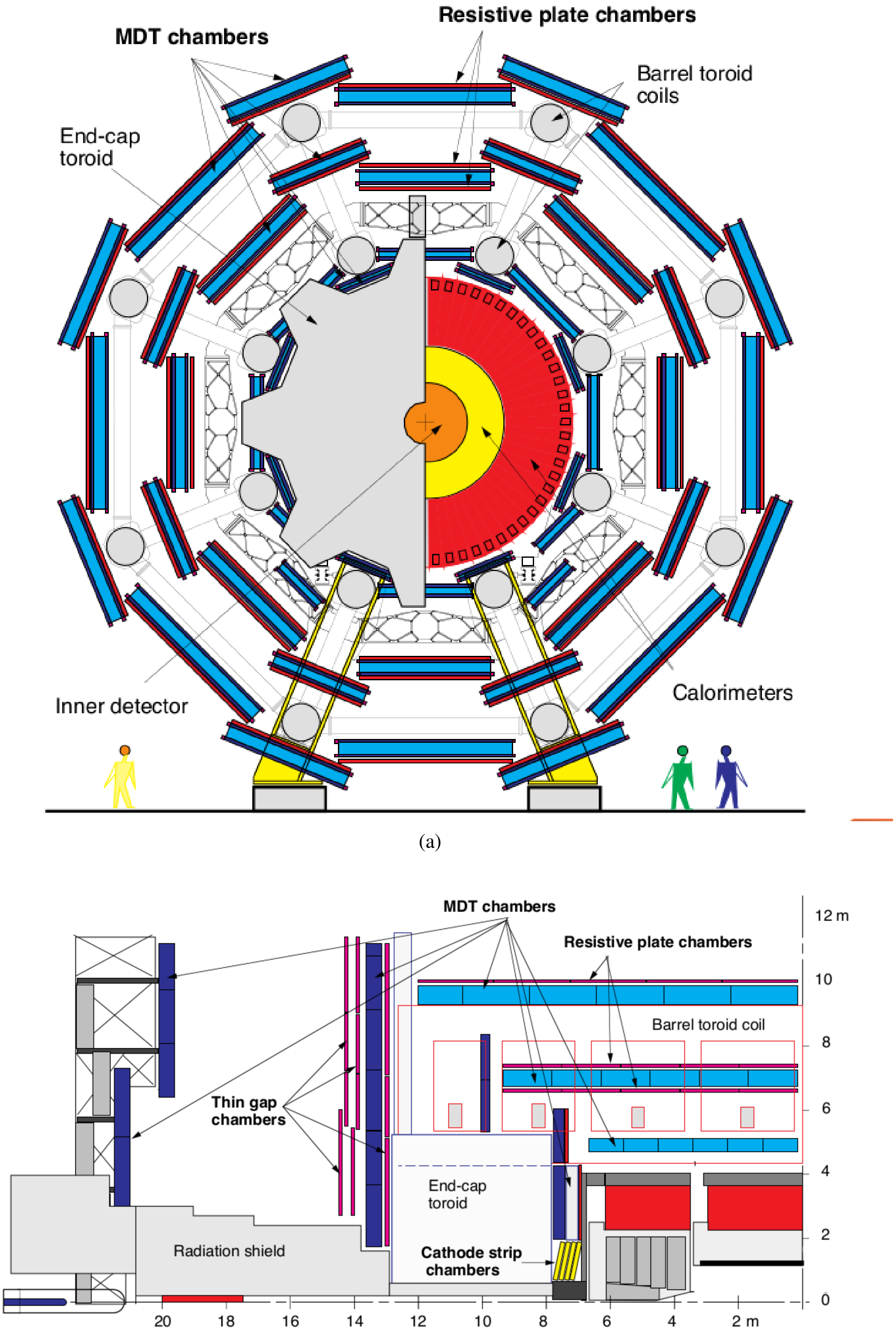
3.3 Trigger and data acquisition

The ATLAS trigger and data acquisition (TDAQ) system [56] identifies and records interesting events. As shown in Figure 3.13, it analyses events at three successive levels of increasing complexity:

- the first level (L1), implemented using custom-made electronics;
- the second level (L2), implemented using computers and networking equipment;
- the third level (Event Filter or EF), which performs more sophisticated calculations and forms, together with L2, the High Level Trigger (HLT).

The L1 trigger looks for signatures from large p_{T} electrons, photons, muons, jets and τ leptons decaying into isolated hadrons. It is also designed to select events with high $E_{\mathrm{T}}^{\mathrm{miss}}$ and high total transverse energy ($\sum E_{\mathrm{T}}$). The L1 has to take decisions in less than 2.5 µs to cope with the high rate of LHC collisions (40 MHz): as a result, it accesses informations with reduced-granularity from the calorimeters and the muon detector. The informations coming from the ID tracking system are not used at this stage, since the time needed to reconstruct tracks and vertices would exceed the latency of the L1 system.

The L1 system reduces the input rate to ~ 100 kHz, which is then handed to the L2 system, which performs decisions within 40 ms and passes the informations to the EF at a 3.5 kHz rate. The EF analyses



(a)

(b)

Figure 3.12: Layout of the muon spectrometer (a) in the x-y plane and (b) in the y-z plane [65]: The lower illustration shows a quarter of the y-z plane only, with the interaction point located in the lower right corner.

the L2-triggered events and provides an output rate of ∼ 400 Hz. The HLT system is able to analyse multiple events in parallel, achieving high input and output rates. The L2 decision uses only information located in the Region-of-Interest (or RoI, the specific region where the L1 trigger originated); this results in a much faster response compared to the EF one. The EF, instead, performs a "full-angle" analysis of the event and produces a response. Both systems have access to the high-granularity and full-precision informations from the calorimeters, muon spectrometer and inner detector, and are software-based.

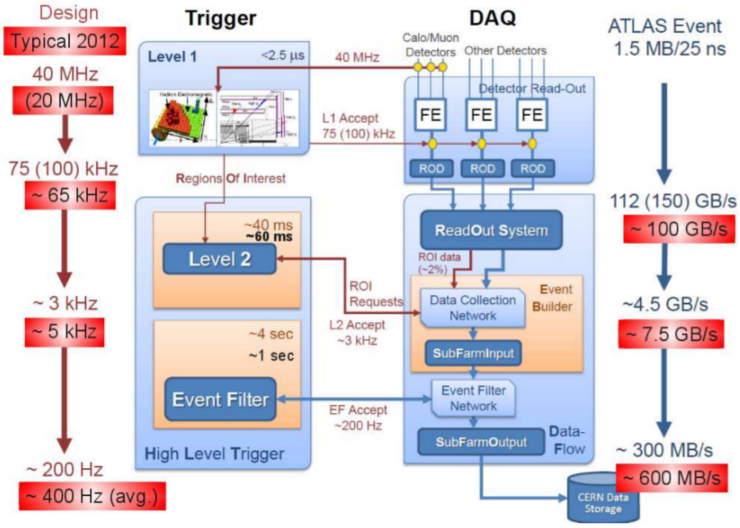


Figure 3.13: An overview of the ATLAS trigger and DAQ system. The design and 2012 typical trigger rates at each level are shown on the left, and the design and 2012 typical output bandwidths are shown on the right [58].

The Data Acquisition (DAQ) system monitors the recording of the data on the storage disks. If the L1 system triggers an event, the DAQ moves the event data from the detector electronics to detector-specific Read-Out Drivers (ROD): the informations are encoded in a common format. If the L2 trigger and EF trigger pass, the event data are merged together and recorded to disk, respectively.

For Run 2, the trigger system have been upgraded during the long shutdown of LHC. The upgrade includes changes to the L1Calo trigger, the introduction of new L1 topological trigger modules, improvements in the L1Muon system and the merging of the two-level HLT system into a single event-filter farm. This allows to cope with the increased trigger rates while maintaining or even improving efficiencies to select relevant physics processes. More information is provided in Reference [66].

3.4 Physics object reconstruction

For further physics interpretation, tracks and interaction vertices are reconstructed in the inner detector and the muon spectrometer, while clusters of energy deposits are identified in the calorimeter systems. These informations are then combined to reconstruct particles like electrons, muons, photons, jets and tau leptons, and to measure properties of the event such as the missing transverse mementum.

3.4.1 Tracking

When traversing the inner detector, particles describe an approximately helical path, due to the influence of the homogeneous magnetic field, and leave hits by interacting with the various detector components that they traverse, as described in Section 3.2.3. The particle tracks are then reconstructed from these hits in order to identify and measure particles, in a procedure known as tracking. Given the high collision energies of the LHC and, consequently, the high levels of pileup, there will typically be hundreds of hits in the inner detector per bunch crossing. Thus, the tracking algorithm has to correctly associate hits with tracks and reconstruct the track parameters, taking into account ionisation energy losses (especially for electrons), radiation energy losses from bremsstrahlung and possible multiple scatterings. A detailed description of the ATLAS tracking is given in Reference [67].

The main tracking algorithm is called "inside-out" tracking and begins in the inner detector layers. The algorithm forms space-points from the measurements in the silicon detectors: a space-point corresponds to a hit in the pixel detector, while the SCT space-points correspond to hits in both sides of the module. Then, by combining space-points in the three pixel detector layers and the first layer of the SCT, track seeds are constructed and are used to build physics objects through the rest of the detector elements. A Kalman Filter [68] is used to follow the trajectory and add hits to the track.

As next step, ambiguity needs to be resolved, since several track candidates could share the same hits; the track is refitted with a more precise χ^2 fit [69] and a score is assigned to each track, based on the fit quality $\chi^2/\mathrm{N}_{\mathrm{dof}}$, the number of hits of the track and the presence of overlapping hits or "holes" (missing hits). Then ambiguities are solved by choosing the track with the largest score, while, tracks with a score below a certain threshold are dropped.

The tracks are then extended into the TRT and, by using the full information of the three detectors, they are refitted once again. If the extended track is better than the silicon only track, according to the fit quality, the extended one is kept.

Since tracks from photon conversions or decays of long lived particles do not produce hits in the inner layers of the detector, the inside-out tracking procedure fails to find them. Therefore, a complementary tracking procedure, the "outside-in" tracking, solves the problems by starting from the TRT and working inwards, following an analogous procedure to the "inside-out" tracking.

3.4.2 Vertex finding

The reconstruction of the interaction vertices is important for the knowledge of the particle origin from the primary interaction vertex, and for the determination of the longitudinal and transverse impact parameters, used to distinguish leptons from photon conversions and from secondary decays in jets. The vertex-finding process is run after the reconstruction of inner detector tracks. The algorithm associates tracks with vertices and fit them to obtain the best vertex positions.

The standard ATLAS approach to vertex-finding is the so-called "finding-through-fitting" [70, 71]. Tracks are preselected according to their consistency with the interaction region, and used to create a single seed vertex. An iterative fit between the vertex and the tracks is carried out, and tracks are assigned a weight depending on their consistency with the vertex: the process stops when the fit converges. The excluded tracks (called "outliers") are used to build a second vertex seed. A fit is performed using the two vertices, and again outlier tracks are used to fit a new vertex. The procedure stops when none of the remaining outliers fits with any vertex give a χ^2 probability of more than 1%.

3.4.3 Electron reconstruction and identification

Electrons are reconstructed by using tracking and calorimeter information. Tracks are reconstructed according to the description of Section 3.4.1, while a cluster based algorithm uses energy deposits in the electromagnetic calorimeter to form energy clusters. The algorithm divides the η-ϕ space into a grid of $N_\phi \times N_\eta$ elements ($\Delta\phi \times \Delta\eta = 0.025 \times 0.025$). Then, energies of all cells are summed across the longitudinal layers, and a fixed-size window (nominally $N_\phi \times N_\eta = 5 \times 5$) is used to define a "pre-cluster". The final EM clusters are constructed through the seed found in the pre-cluster. Finally, the cells within a given η-ϕ range of the seed are included into the final cluster.

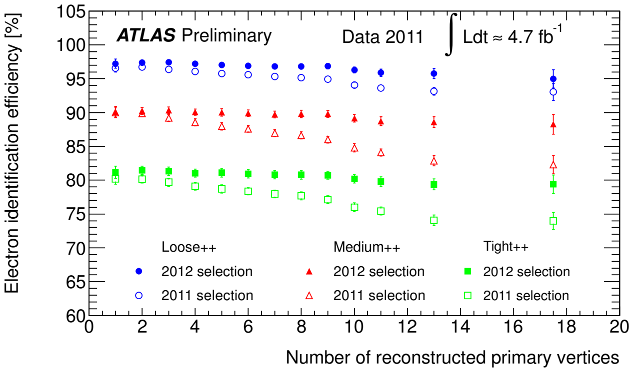


Figure 3.14: Electron identification efficiency in 2011 (open markers) and 2012 (solid markers) as a function of the number of reconstructed primary vertices in the event. The blue circles show the efficiency for the Loose++ selection, the red triangles for the Medium++ and the green squares for the Tight++ [73].

A track-matched cluster is identified as an electron. The cluster energy is also corrected by taking into account the scintillator losses in the crack regions and the leakage outside the EM calorimeter, while the tracks are refitted by considering bremsstrahlung.

The reconstructed electron candidates contain a high contamination from electrons from photon conversions, non-isolated electrons from decays in jets and jets faking electrons. To identify prompt electrons, a cut-based or a multivariate identification can be performed. Discriminating variables are the shape of the electromagnetic shower, the quality and length of the inner detector track and the track-calorimeter matching. Several reference sets of cuts are usually provided, like three cut-based in Run 1 (Loose, Medium and Tight) and three likelihood-based (LH) in Run 2 (LooseLH, MediumLH ad TightLH). The likelihood-based sets exploit a likelihood discriminant which combines shower shape and track information. Those working points are designed to give progressively greater background rejection, at the expense of some efficiency. The expected jet rejection of the three Run 1 points are 500, 5000 and 50000, respectively [72].

Electron identification efficiency

Figure 3.14 shows the electron identification efficiency, as a function of the number of reconstructed primary vertices, in 2011 and 2012 data. The 2011 identification requirements did not allow for constant efficiency performance: the efficiency reduces by more than 5% in events with 18 reconstructed primary vertices with respect to events with a single primary vertex. In Figure 3.15 the Loose++ identification efficiency is shown as a function of E_{T}, using the 2011 requirements. There are differences between data and MC efficiencies, at the level of a few percent, which can be attributed mainly to mismodelling of the shower-shape variables in the Monte Carlo. Scale-factors, parameterised as a function of η and E_{T}, are therefore applied to the simulation in order to correct the reconstruction efficiency.

3.4.4 Muon reconstruction and identification

The muon reconstruction is based on the combination of accurate measurements in the muon spectrometer and the inner detector [75–77]. Four categories are defined, according to the available information in the detector subsystems:

- combined muons, combination of an MS track with an ID track. They have an acceptance limited by the ID at $|\eta| < 2.5$;
- segment-tagged muons, combination of an ID track with an MS track segment. MS track segments are segments that did not form full MS tracks and are reconstructed in a single MS station. The track parameters to describe the reconstructed muon are taken from the ID track only;

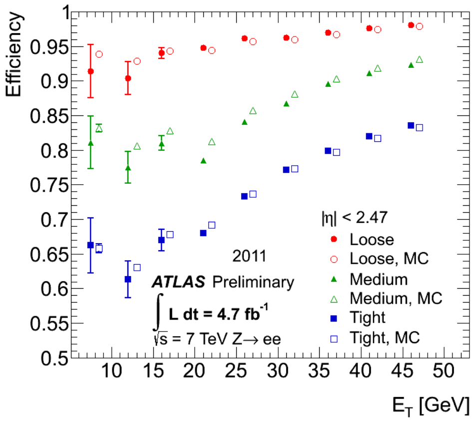


Figure 3.15: Efficiency of the Loose++ identification requirements as a function of the cluster transverse energy. The solid points indicate data based measurements whilst the open points indicate predictions from Monte Carlo. The different markers indicate the method used to measure the efficiency [74].

- stand-alone muons, whom track reconstruction is based only on MS measurements. Thus, they exist over the full acceptance of the MS, $|\eta| < 2.7$.
- calorimeter-tagged muons, in which ID tracks are matched to calorimeter deposits consistent with a minimum ionising muon. The MS information is not used in this case.

The muon identification is performed by applying quality requirements that suppress fake muons, while selecting prompt muons with high efficiency and/or guaranteeing a robust momentum measurement. Some of the variables used in muon identification are the relative difference between the p_{T} measurements in the ID and MS and the number of hits in the ID and MS. Four muon identification selections (Medium, Loose, Tight, and High-p_{T}) are provided [77].

Muon identification efficiency

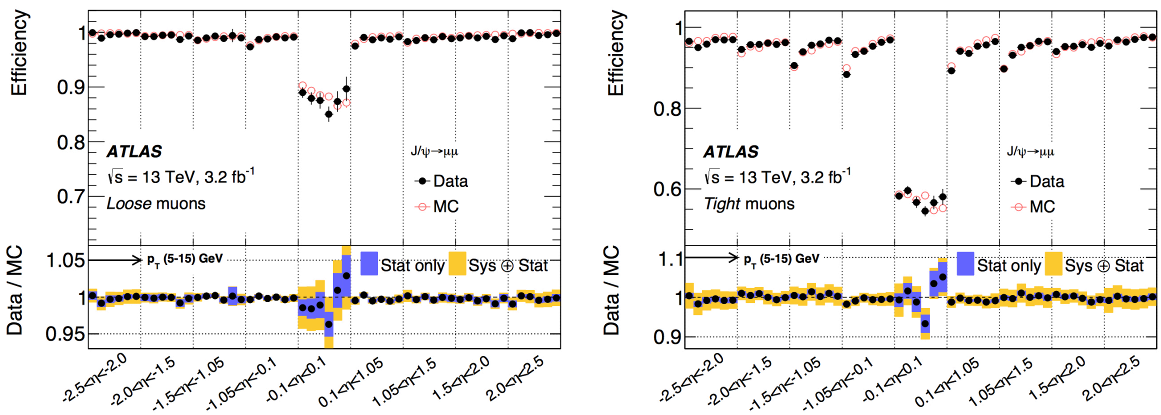
Figures 3.16 show the observed reconstruction efficiency for Loose and Tight muons reconstructed in simulation and in the 2015 data: a tag-and-probe technique on J/ψ boson decays is used to determine the efficiencies [77]. The efficiency drops significantly in the low coverage region, $|\eta| < 0.1$. A good data/MC agreement is observed, thus the MC scale factors, applied in order to reproduce the efficiency observed in data, are close to unity. The charge misidentification rate for muons is negligible.

3.4.5 Jet reconstruction and identification

Quarks and gluons produced in particle interactions hadronise and produce a collimated spray of particles known as a jet, by the mechanism of gluon radiation and splitting. The objective of jet reconstruction is to combine those particles in order to obtain a physics object describing the characteristics of the initial parton. For the jet reconstruction, an algorithm of association of multiple energy deposits in the calorimeters is necessary: the algorithm clusters them into a single jet (clustering) and combines their four-momenta.

The default jet clustering algorithm in ATLAS is called anti-k_t [78].It combines objects according to the distance parameters $d_{i,j} = \min(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}) \cdot \frac{\Delta R}{R}$ and $d_{i,\mathrm{beam}} = p_{\mathrm{T},i}^{-2}$, where $p_{\mathrm{T},i}$ is the transverse momentum of object i and ΔR is the distance between objects i, j, as defined above. The parameter R controls the size of the jet.

The four-momentum of the jet is simply the sum the four-momenta of the constituent objects. This method conserves energy and momentum, allowing a meaningful definition for the jet mass.



(a) (b)

Figure 3.16: Muon reconstruction efficiency in different η regions measured in $J/\psi \to \mu\mu$ events for (a) Loose and (b) Tight muon selections. Within each η region, the efficiency is measured in six p_T bins (5–6, 6–7, 7–8, 8–10, 10–12, and 12–15 GeV). The error bars on the efficiencies indicate the statistical uncertainty. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties [77].

The hadronic calorimeters in ATLAS are "non-compensating", which means that they do not compensate for the different energy response to electromagnetically interacting particles and hadronically interacting ones. Moreover, they do not account for energy lost from the hadronic shower, due to leakage of the shower out of the calorimeter or the production of secondary particles. Therefore, a calibration of the hadronic calorimeter response is absolutely necessary for a proper use of the jet informations. For instance, the topological clusters are corrected to the EM scale: their energy is tuned such that the response to purely electromagnetic showers is correct. The Jet Energy Scale (JES) correction is, instead, applied to jets which are constructed from the EM scale clusters, in order to cure the non-compensating nature of the calorimeters, the leakage outside of the calorimeters, the effects of dead material and energy loss (due to particles deflected out of the jet by the magnetic fields).

b-tagging

The identification of jets coming from b-quarks fragmentation, the so-called b-tagging, is crucial for analyses looking for one or more b-quarks in the final state. Above 10 GeV, long lived b-hadrons ($\tau \sim 1.6$ ps), produced in the hadronisation of b-quarks, decay sufficiently faraway from the production vertex ($d = 1.8$ mm at $p_T \approx 20$ GeV): a displaced secondary vertex can be resolved in the detector and used to infer the b-jet-ness of the jet.

If a secondary vertex is identified within a jet, its distance to the primary vertex and the the mass of all particles associated to the vertex are the quantities used for the identification. b-jets can also be identified exploiting the high multiplicity of charged tracks and the high p_T of decay particles from B hadrons. Without the need to correctly reconstruct the secondary vertex, the impact parameter (IP) of the tracks belonging to the jet is another useful quantity. It represents the transverse distance of a track to the primary vertex. The primary vertex position is essential in b-tagging to measure the impact parameter.

Currently, the algorithm of b-tagging in ATLAS is based on a neural network, called MV1 (Multivariate Tagger) for Run 1 [79] and MV2 for Run 2 [80], whose input informations are the weights of several algorithms for b-jet identification. The most important are:

b-jet efficiency [%]	b-jet efficiency [%]	b-jet efficiency [%]	b-jet efficiency [%]
60	21	93	1900
70	8.1	26	440
77	4.5	10	140
85	2.6	3.8	28

Table 3.3: Operating points for the MV2c20 b-tagging algorithm, including benchmark numbers for the efficiency and rejections rates. The statistical uncertainties on the rejection and efficiency estimates are negligible and thus are not shown [81].

- IP3D, based on the distribution of impact parameter significance, projected on the transverse and longitudinal planes;
- SVx, based on secondary vertex reconstruction;
- JetFitterCOMBNN, fit along the flight direction of b-hadrons, combining this information with informations from the previous tagger by a neural network.

If the MV weight exceeds a certain threshold, the jet is tagged as b-jet. Several working points are provided for different efficiencies in identifying real b-quark jets and rejecting c- and light quark jets. Table 3.3 shows the four working point of the “MV2c20” tagger: the MV2c20 algorithm is defined as the output of a boosted decision tree, with the training performed assigning b-jets from $t\bar{t}$ as signal, and a mixture of 80% light-flavour jets and 20% c-jets as background.

3.4.6 Missing transverse momentum

The missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) corresponds to the momentum which was not reconstructed in the transverse plane by detector elements. Assuming that any imbalance observed in the transverse plane is caused by unobserved objects, it is a very useful quantity to infer about exotic particles or neutrinos. However, $E_{\mathrm{T}}^{\mathrm{miss}}$ can also be affected by mismeasured objects or “gaps” in the detector. A good knowledge of all objects considered for the vectorial sum is required for its reconstruction.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ algorithm [82] in ATLAS makes use of jets and electrons to take advantage of their precise calibration:

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\mathrm{electrons}} \vec{E}_{\mathrm{T}}^{e} - \sum_{\mathrm{muons}} \vec{p}_{\mathrm{T}}^{\vec{\mu}} - \sum_{\mathrm{jets}} \vec{p}_{\mathrm{T}}^{\mathrm{jet}} - \sum_{\mathrm{clusters}} \vec{E}_{\mathrm{T}}^{\mathrm{cluster}} . \tag{3.3}$$

3.5 Simulation of physics processes

To obtain informations from the data recorded by ATLAS, the precise comparison with the theoretical predictions is needed: it quantifies the agreement between data and the Standard Model or possible new physics models. Monte Carlo (MC) event generators are employed to simulate events for signal and background processes. A MC simulation program generates events based on theoretical probability distributions of interaction between particle and decay, predicting the statistical behaviour of the events collected by the detector. MC generators are also used to simulate the interaction of particles with the detector materials and to estimate detector acceptance, efficiency and resolution.

Figure 3.17 shows a schematic view of the different steps involved in the simulation for hadron colliders: the parton[2] distribution, the hard process, the parton shower and the hadronisation. An important aspect

---

[2] The parton model was proposed by Richard Feynman in 1969 as a way to analyse high-energy hadron collisions. Any hadron (for example, a proton) can be considered a composition of a number of point-like constituents, termed “partons”. Later, with

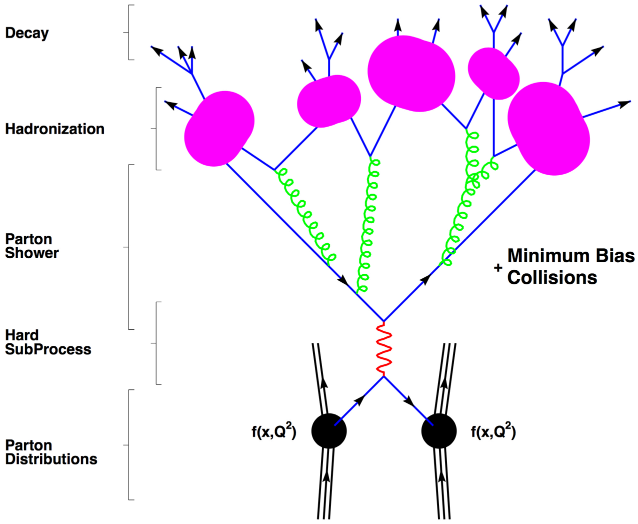


Figure 3.17: The basic structure of a generated event, including showering and hadronisation, is shown schematically [83].

in the simulation of pp collisions is, indeed, the possibility to factorise the different energy scales in the collision process. This allows to compute the hard interaction to a fixed order in perturbation theory and describe the subsequent softer scales with phenomenological models.

As a first step, a proton-proton collision is simulated by using parton distribution functions (PDF): partons are generated with a momentum fraction x at an energy scale Q, according to the PDF. Different types of PDF are used for the MC datasets employed in this thesis; they will be specified in each analysis Chapter.

Then, these two partons collide and undergo an interaction with a large momentum transfer. The interaction is computed at fixed order in perturbation theory, using the so-called "parton-level generators" to simulate specific final states. The parton-level generators can be separated in two main groups: the process-specific and the arbitrary-process generators. Process-specific generators are optimised to simulate a limited list of processes, while the latter can generate any tree-level SM process, but they spend a larger computation time, because of the lack of optimisation. Thus, a balance between the level of complexity and accuracy is necessary to well simulate a process. In this thesis, arbitrary-process generators are mostly used: MadGraph [84], MadGraph5_aMC@NLO [85], Powheg-box [86–88] and Sherpa [89].

Since the partons produced by the collisions are charged particles, they will emit gluons, which will split into quark-antiquark pairs or radiate further gluons, leading to a particle cascade (parton showering). The parton showering continues until the partons reach the hadronisation scale energy at approximately 1 GeV. Here, the hadronisation phase starts and partons combine into colourless hadrons. Several phenomenological models are exploited to describe the hadronisation and the following decay of hadrons into the final state particles: non-physical parameters are usually calibrated using experimental data; each set of these calibration parameters is named "tune". In the end, the final states particles interact with the detector and are reconstructed with the same reconstruction and identification algorithms used for data.

---

the experimental observation in favor of the quark model and the confirmation of asymptotic freedom in QCD, partons were matched to quarks and gluons. The parton model remains a justifiable approximation at high energies.

CHAPTER 4

Analysis strategy and statistical framework

This Chapter provides an introduction to the analysis strategy of the two $t\bar{t}H$ searches presented in this dissertation. Section 4.1 motivates the employment of a multivariate approach and the profile likelihood method, and it also outlines the analysis chain. The descriptions of the Boosted Decision Tree classifier, its optimisation and validation are given in Section 4.2. Section 4.3 introduces the frequentist formalism for the template fit procedure, the profile likelihood approach and its advantage to reduce the impact of systematic uncertainties on the final result. The statistical procedure to determine the signal significance and the upper limit on the $t\bar{t}H$ production cross section is also discussed.

4.1 Analysis strategy: motivation and description

A multivariate approach can be particularly useful to distinguish potential signal events from the background, when single variables do not exhibit a clear separation power. It maximises the amount of information that can be extracted from several input variables by combining them into one output discriminant, resulting in a non-trivial selection in the variables' phase space. Common examples of multivariate classifiers are likelihood estimators, Neural Networks (NN) or Boosted Decision Trees (BDT).

MVA techniques are particularly decisive in $t\bar{t}H$ searches for different reasons. In the fully hadronic final state (i.e., without leptons), described in Chapter 5, large backgrounds dominate even the most sensitive regions, making the measurement of the $t\bar{t}H$ cross section challenging. The best signal discrimination is therefore essential for this analysis. In the multilepton $t\bar{t}H$ final state, as described in Chapter 6, standard observables (angular distances, global event variables, ...) show good separation: a non-basic variable like MVA discriminants provides even better distinction between the signal and background. For both analyses, two types of BDT classifiers (Adaptive BDT and Gradient BDT) are considered and are described in the next Section; the choice of such classifiers resulted from optimisation studies. Both analyses employ the Toolkit for Multivariate Data Analysis (TMVA) [90], a standalone package that provides a ROOT-integrated [91] machine learning environment for the evaluation of sophisticated multivariate classification techniques.

Both analyses are based on the profile likelihood fit approach, directly incorporating in the fit model the effect of systematic uncertainties, such as the imperfect detector calibration. The approach exploits various regions with different signal fractions for a more effective statistical combination. Including subsidiary regions in the model allows to improve the knowledge on background normalisation and to constrain the systematic uncertainties.

This fitting procedure is adopted in many physics analyses: once the most suitable discriminator (the most performing BDT, for instance) is determined, the template distributions for signal and backgrounds are compared to the data distribution in order to extract the signal normalisation. Section 4.3 describes the general idea behind the profile likelihood fit as well as its formalism and implementation.

In order to test the effectiveness of the BDT classifier and the stability of the combined fit approach, the analysis strategy holds on four sequential benchmarks:

- *determination of the best discriminator*, through an optimisation procedure according to the analysis' needs. The optimisation determines the minimal set of variables to construct the best classifier. The deterioration of the BDT performance (e.g., due to "overtraining" effects) requires careful studies;
- *validation of variables modelling*, needed to check the presence of mismodelling issues that would disprove the analysis and could bring an undesirable BDT mismodelling.
- *"blind" fit*, performed by excluding the regions/bins where larger signal-to-background ratio more is expected. This intermediate step is necessary to exercise the fitting procedure without generating potential bias due to signal excess in data and to validate the modelling of the background. After ensuring the robustness of the analysis strategy, the "unblind" fit can be performed;
- *unblind fit*, performed with all events collected, which provides the final result.

4.2 Boosted Decision Trees: a multivariate discriminator

A *decision tree* is a multivariate classifier with a binary tree structure [90]. It takes many decisions (nodes) based on one variable at a time until a particular condition is fulfilled (*e.g.* the maximal allowed depth of the tree or the minimum percentage of training events required in a node is reached). In this way the phase space is divided into several regions (leaves), as schematically illustrated in Figure 4.1: this phase of learning of the classifier is often referred to as "training" of the decision tree. Each region is classified as signal-like or background-like according to the type of the majority of training events in such final leaf.

Figure 4.2 illustrates three classifiers of increasing performances in a 2-dimensional phase space, *i.e.* with 2 variables. The red and blue points stand for signal and background events, the black lines correspond to the classification, respectively. Although the rightmost classifier appears to be the most effective, it is also the most subject to misclassification caused by statistical fluctuations in the signal and background datasets (the so-called "overtraining").

Decisions during the training are taken according to various separation criteria [90]: these are symmetric with respect to the event classes, since a cut that selects mostly background is as valuable as one that selects signal. All criteria have a maximum separation index where the samples are fully mixed, *i.e.*, at purity[1] $p = 0.5$, and fall off to zero when the sample consists of one event class only. The standard criteria are the default Gini Index, defined as $p \cdot (1 - p)$, and the statistical significance, defined by $S / \sqrt{S + B}$.

At each node of the decision tree the procedure of training always cuts on a single variable in order to optimise the increase in the separation index between the parent node and the sum of the indices of the two daughter nodes, weighted by their relative fraction of events. The cut values are optimised by scanning over the variable range with a granularity option (*nCuts*). The decision tree training stops when either the maximal depth is reached or the number of events in one leaf node goes below a certain threshold.

In principle, decisions splitting events could continue until each leaf node contains only one event. Such decision tree could be thought as the perfect classifier, but it would be strongly overtrained. To

---

[1] The purity of a decision is defined as the fraction of events belonging to the same event class, after the decision.

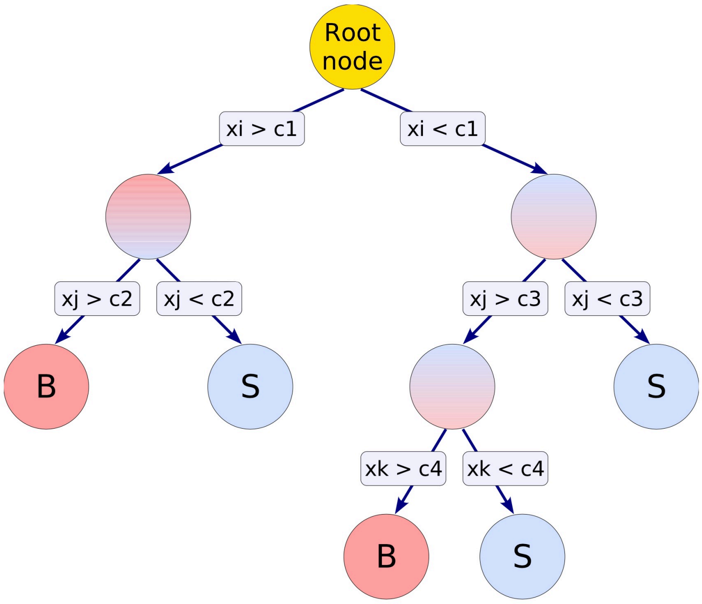


Figure 4.1: Schematic view of a decision tree. Starting from the root node, a sequence of binary decisions using the discriminating variables x_i is applied to the data. Each decision uses the variable that at this node gives the best separation between signal and background. The same variable may thus be used at several nodes, while others might not be used at all. Depending on the majority of events that end up in the leaves at the bottom of the tree, leaves are labelled “S” for signal and “B” for background. Illustration from Reference [90].

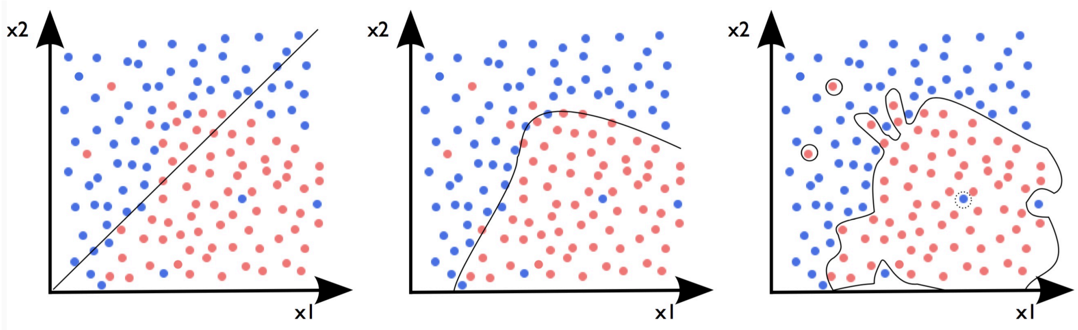


Figure 4.2: The red (blue) points stand for signal (background) events in a 2-dimensional phase space spanned by two variables x_1 and x_2. The black line denotes the classification of the phase space into two regions for (a) a linear classifier, (b) a more complex classifier and (c) a classifier suffering from overtraining. Illustration from Reference [92].

avoid overtraining long decision trees must be “pruned”[2] after training or the tree depth limited during training. The extension of these concepts from one trained tree to many trees (a forest) is known as boosting: the boosting algorithm is a procedure that combines many “weak” classifiers[3] to achieve a final powerful classifier, the *Boosted Decision Tree* (BDT), which is given by an average of the individual

---

[2] The tree pruning corresponds is the removal of insignificant nodes in the decision tree, that does not deteriorate the separation power of the classifier.

[3] “Weak” classifiers (or weak learners) are classifiers performing only slightly better than a random classifier. These are classifiers that have some clue on how to predict the right labels, but are not as much as strong classifiers have like, e.g., Naive Bayes, Neural Networks or SVM (Support Vector Machine).

decision trees [93, 94]. Compared to a single tree, boosting stabilises the response of the decision trees and considerably enhances the performance. Events misclassified in the previous training receive higher weights (higher importance) in the next one, so that the new decision tree might differ from the previous one. Several boosting algorithms are available in TMVA, including:

- AdaBoost, the Adaptive Boost algorithm; it trains an army of "decision stumps" each focusing on one part of the characteristics of the data. The "decision stump", one of the simplest weak classifiers, is a one-level decision tree, which selects a threshold for one feature and splits the data on that threshold. AdaBoost focuses on events misclassified in the previous tree and assigns a higher weight to them. For each stump tree the misclassification rate ε is determined, from which a boosting weight α is derived according to $\alpha = \frac{1-\varepsilon}{\varepsilon}$ and applied to the samples used for the next tree iteration. The final boosted classification is

  $$y_{\text{boost}}(\mathbf{x}) = \frac{1}{N_{\text{trees}}} \sum_{i}^{N_{\text{trees}}} \ln(\alpha_i) \cdot h_i(\mathbf{x}),$$

  with $h_i(\mathbf{x}) = \pm 1$ as individual classifiers for signal and background respectively, with $\mathbf{x}$ being the tuple of input variables, and N_{trees} being the number of trees in the collection. Small (large) values for $y_{\text{boost}}(\mathbf{x})$ indicate a background-like (signal-like) event.
- Gradient boost uses the model response $F(\mathbf{x})$ (a weighted sum of weak classifiers like $y_{\text{boost}}(\mathbf{x})$). The boosting procedure employs a loss-function $L(F, y) = \ln(1 + e^{-2F(\mathbf{x})y})$ to adjust the parameters of the weak classifiers, in order to minimise the deviation between the model response $F(\mathbf{x})$ and the true value y obtained from the training sample is minimised. The Gradient boost algorithm slightly differs from the one used in the AdapBoost method and has the advantage of being more robust with respect to fluctuations. The final classification corresponds to the minimal model response $F(\mathbf{x})$.
- Bagging denotes a re-sampling technique, where a classifier is repeatedly trained using re-sampled training events, such that the combined classifier represents an average of the individual classifiers. A priori, bagging does not aim at enhancing a weak classifier in the way adaptive or gradient boosting does, and is thus not a boosting algorithm in a strict sense.
- Randomised trees, where each tree is grown in such a way that at each split only a random subset of all variables is considered. Moreover, each tree in the forest is grown using only a (re-sampled) subset of the original training events.

4.2.1 Variables entering the BDT

Variables used to train a classifier like a BDT can usually be grouped in four main classes, reflecting their connotation:

- variables about individual physics objects, like the four-momentum components of a jet, the number of tracks associated to a b-tagged jet, the isolation of a lepton;
- global event variables, such as S_{T} (the modulus of the vector sum of the physics objects p_{T}) and $H_{\text{T}5}$ (the scalar sum of the jet p_{T} starting from the fifth jet in p_{T} order);
- event-shape variables, such as linear combinations of the eigenvalues of the momentum tensor (centrality, sphericity and aplanarity); the distributions are very discriminative between top-like signal and other backgrounds [95], and can be very effective because they are less sensitive to the loss of jets by reconstruction inefficiencies;
- pairs of objects, like $(E_{\text{T}1} + E_{\text{T}2})/\sum E_{\text{T}}^{\text{jets}}$ (the sum of the transverse energies of the two leading jets divided by the sum of the transverse energies of all jets), $m_{j\ell}^{\text{max}}$ (the largest invariant mass of all

jet-lepton combinations), and the minimum ΔR between jets.

Driven by the presence of many decay products in the $t\bar{t}H$ final state, a Pseudo Matrix-Element ($\mathcal{P.M.E.}$) discriminator has been developed: it is a simplified matrix-element variable, which considers some of the variables described above and their correlations. The discriminator exploits all possible combinations of physics objects to reconstruct, even partially, the signal and background final states. It relates to the probability of an event to be a signal candidate compared to the probability of being a background candidate, and is described in the next Section.

4.2.2 A simplified matrix-element: the Pseudo Matrix-Element approach

The matrix-element method [96] makes the most complete use of the kinematic information in an event, by linking directly theoretical calculations and observed quantities. It defines the probability $P^i(\mathbf{x}|\alpha)$ of the event to be consistent with the physics process i described by a set of parameters α, given an observation defined by the four-momentum vectors of all reconstructed objects, $\mathbf{x}$, in the final state. The probability $P^i(\mathbf{x}|\alpha)$ is defined as:

$$P^i(\mathbf{x}|\alpha) = \frac{(2\pi)^4}{\sigma_i^{\mathrm{exp}}(\alpha)} \int \mathrm{d}p_{\mathrm{A}} \mathrm{d}p_{\mathrm{B}}\, \mathbf{f}(p_{\mathrm{A}})\, \mathbf{f}(p_{\mathrm{B}})\, \frac{\left|\mathcal{M}_i(\mathbf{y}|\alpha)\right|^2}{\mathcal{F}}\, W(\mathbf{y}|\mathbf{x})\;\, \mathrm{d}\Phi_N(\mathbf{y}) \tag{4.1}$$

where the integral stands for a numerical integration over the entire phase space of the initial and final state particles. In this Equation, $\mathbf{x}$ and $\mathbf{y}$ represent the four-momentum vectors of all final state particles at reconstruction and parton level, respectively; the flux factor $\mathcal{F}$ and the Lorentz-invariant phase space element $\mathrm{d}\Phi_N$ describe the kinematics of the process; the transition matrix-element $\mathcal{M}_i$ is defined by the Feynman diagrams of the hard process; the transfer functions $W(\mathbf{y}|\mathbf{x})$ map the detector quantities $\mathbf{x}$ to the parton level quantities $\mathbf{y}$. Finally, the cross section σ_i^{exp} normalises P^i to unity taking acceptance and efficiency into account.

Due to the complexity and high dimensionality of the integration, simplifications and approximations are needed to obtain results with reasonable computing time. Furthermore, detector resolution effects can considerably spoil the final matrix-element accuracy, while acceptance effects can lead to partially reconstructed final states, that cannot be handled by a pure matrix method.

A simplified version of the matrix method, the Pseudo Matrix-Element ($\mathcal{P.M.E.}$) method, can deal with partial event reconstruction, because it aims for the presence of partially reconstructed resonances and combines reconstructed observables in a single discriminator.

The $\mathcal{P.M.E.}$ approach is adopted in both $t\bar{t}H$ analyses described in this work: inspired by the same idea, the two discriminators (referred as $\mathcal{D}$) are separately implemented in order to exploit the features of the two different final states. The multilepton discriminator can be considered as a more elaborated and effective version of the fully hadronic one. A detailed description of the two discriminators is provided in Sections 5.6.2 and 6.6.2. Ultimately, $\mathcal{D}$ is used as an input variable of the BDT along with other kinematic variables and shows the largest discrimination power in both analyses.

4.2.3 BDT optimisation

Several figures of merit and approaches can be used to the performances of the final BDT classifier. The “ranking score” can be used, for instance, to quantify the importance of a variable in the BDT and remove it without significantly affecting the BDT performance. The preliminary ranking score derived by the TMVA toolkit is independent of the algorithm and relies on the inherent discriminating power of a given variable. Later this is superseded by algorithm-specific rankings. The BDT-specific ranking

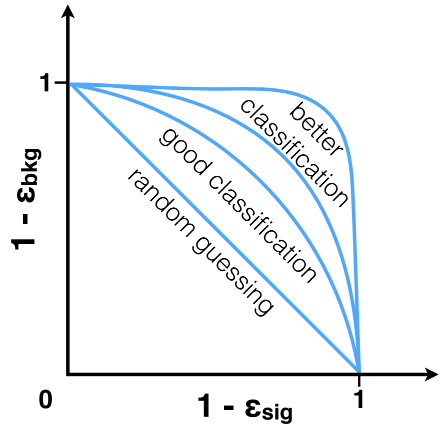


Figure 4.3: The sketch of 3 classifiers in the ROC plane. Performing classifiers increase the integral below the ROC curve.

score is obtained by identifying first all the nodes in which the variable is used to split a decision tree and summing up their purity. This measure of the variable importance can be used for a single decision tree, as well as for a forest.

Besides the individual separation power, the correlations between the input variables can also be used to keep or reject variables. Strongly correlated input variables do not increase the discriminating power of the BDT significantly compared to the case when only one of those is kept. The removal of variables can be driven by other considerations: possible mismodelling of single variables or mismodelled correlations in simulated event that could bias the BDT response. Moreover, variables exhibiting large generator dependence can be excluded as well.

There is no unique way to choose the best set of discriminating variables to enter the BDT classifier. Several optimisation strategies can be exploited: for instance, variables can be selected and ranked according to their algorithm-specific ranking value, adding the highest ranked variable until adding more variables does not significantly improve the separation between signal and background. The final number of variables used usually results from a compromise between the performance of the BDT and the practical aspect of the validation of all used variables.

The Receiver Operating Characteristic (ROC) curve expresses graphically the performance of a classifier in terms of the signal purity $(1 - \varepsilon_{\mathrm{bkg}})$ as a function of the signal efficiency $\varepsilon_{\mathrm{sig}}$, as illustrated in Figure 4.3: the better the classifier, the higher its ROC integral. The ROC curve, or its integral, represents the most common way to quantify performance and improvements of a classifier, but other figures can be exploited, according to the analysis' needs, as done for the fully hadronic $t\bar{t}H$ search in Section 5.6.3.

After the optimisation step, it is very important to check whether the training procedure suffers of possible overtraining. The overtraining leads to a seeming increase in the performance over the objectively achievable one, if measured on the training sample, and to an effective performance decrease when measured with an independent test sample. This is due to the fact that features exploited in the training sample are not helpful, instead, to separate signal and background in general. A convenient way to detect overtraining and to measure its impact is therefore to split the signal and background samples in two (training and testing samples). The training sample is used for the BDT training. Finally, the performance results between training and testing samples (using ROC integrals, for instance) are compared.

Besides the ROC curve, overtraining effects can also be spotted by employing the so-called Kolmogorov-

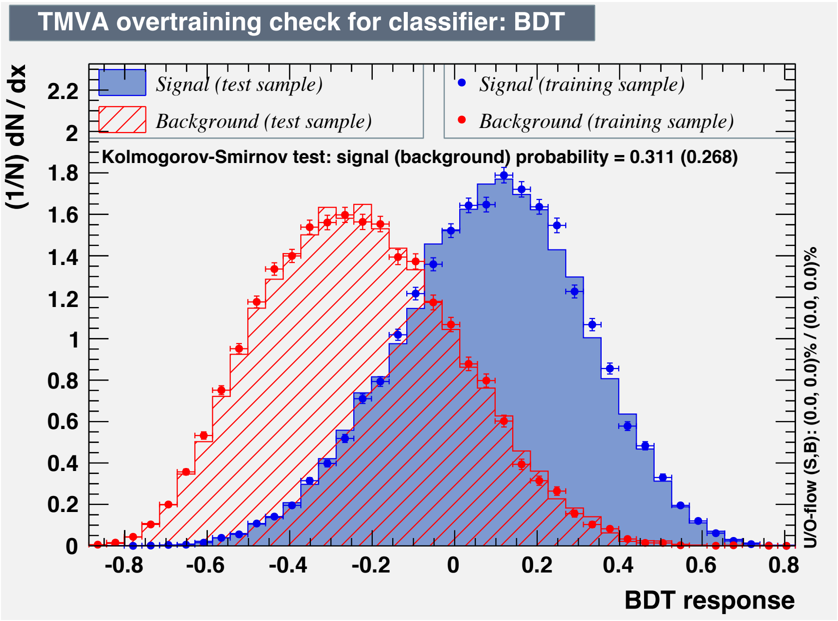


Figure 4.4: Illustration of the response of the boosted decision trees. The blue histogram indicates the BDT output for signal used in training the trees, the blue points indicate the BDT output distribution from the signal test subsample. In an analogous way, the red histogram (dots) indicates the BDT output distribution for background used in training (testing). The K-S test results are also provided: no overtraining is observed for such BDT classifier.

Smirnov (K-S) test [97] on the BDT classifier. The Kolmogorov-Smirnov test measures the maximum distance between the cumulative distribution of two samples (train and test) to assess whether their distributions differ. The null distribution of the K-S statistic is calculated under the null hypothesis that the two samples are drawn from the same distribution. The probability that the maximum K-S distance λ_α is larger than the observed may be calculated using K-S statistics for each parameter α that has the meaning of statistical significance level. When $\lambda < \lambda_\alpha$ then the two distributions are equivalent with $1 - \alpha$ significance level: a typical significance value is 95%, which means $1 - \alpha = 0.05$. The TMVA toolkit reports the K-S test converted in significance levels $(1 - \alpha)$ for overtraining checks; values larger than 0.05 imply good agreement between the train and test samples, whereas smaller values would entail some overtraining issues.

An example of BDT overtraining test is illustrated in Figure 4.4 for one of the $t\bar{t}H$ fully hadronic signal regions: the BDT training and test responses are respectively shown in blue (red) histogram and points for the signal (background) process. The K-S test results exceed the 0.05 threshold, confirming the absence of overtraining effects.

4.2.4 Validation of the variables entering the BDT

Once the best set of variables is determined, a careful validation of each variable is required, especially for the BDT discriminator in the most signal-sensitive regions. It is a necessary step to demonstrate that the MC simulation provides a good description of all the input variables: a mismodelled variable can impact the BDT modelling and will produce unreliable fit results. Namely, the validation helps to prove the stability and the robustness of the approach adopted.

Expected distributions are compared to data in all the regions: to quantify the level of agreement a K-S test can be performed. At the first stages of the analysis, the validation is performed in a “blind mode”: an “anti-BDT” cut is applied to exclude bins where the expected S/B ratio exceeds a certain percentage (for instance, 5% of the bin yield), in order to remove signal contamination. The aim of blinding is to not cause human bias in the analysis due to premature observations made in the most sensitive regions. After the so-called “unblinding” step those events are visible and to be validated as well.

The validation is performed before the fit procedure (“pre-fit”) and after (“post-fit”). The fit, performed on the final BDT discriminants, improves the knowledge on some background normalisations and is also expected to increase the level of agreement for each variable.

4.3 Statistical method: the profile likelihood for hypothesis testing

The frequentist approach using the profile likelihood as statistical test is widely used in particle physics to establish discovery or exclusion of an hypothesis [98]. It tests the compatibility of the observed data with a hypothesis H or provides frequentist confidence intervals for a certain measurement.

To summarise the outcome of a search, the compatibility of the observed data with a given hypothesis H is quantified by computing the p-value [98]: it corresponds to the probability, under assumption of H, to find data of equal or greater incompatibility with the predictions of H. The p-value is based on a test statistic q and is expressed as:

$$p_H = \int_{q_{\text{obs}}}^{\infty} f(q|H)\, dq\,, \tag{4.2}$$

where q_{obs} is the observed value of the statistic q in the data and $f(q|H)$ denotes the probability density function (p.d.f.) of q under the hypothesis H. For the moment the details of how the test statistic q is defined are left open.

The smaller the p-value, the less the data is compatible with the hypothesis. The conventional 95% confidence level (C.L.) of exclusion is defined as $1 - \alpha = 95\%$ if the p-value satisfies $p_H < \alpha$, where α equals to 0.05.

Generally, to prove the validity of one hypothesis against another, the two hypotheses are defined as:

- H_0, the null hypothesis, describing the established physics processes and usually referred to as background-only hypothesis (b);
- H_1, the alternative hypothesis, also describing the additional physics model to probe, and referred to as signal-plus-background ($s + b$) hypothesis.

If $f(q|s + b)$ is the p.d.f. of q under the assumption of signal-plus-background hypothesis, then its p-value (also known as CL_{s+b}) can be expressed as (Figure 4.5)

$$p_{s+b} = P(q \geq q_{\text{obs}}|s + b) = \int_{q_{\text{obs}}}^{+\infty} f(q|s + b)\, dq\,, \tag{4.3}$$

and the p-values of the background-only hypothesis (named also CL_b) as

$$p_b = P(q \leq q_{\text{obs}}|b) = \int_{-\infty}^{q_{\text{obs}}} f(q|b)\, dq\,. \tag{4.4}$$

In particle physics, besides the p-value, also its conversion in terms of significance is used. The equivalent significance Z is defined such that a Gaussian-distributed variable found Z standard deviations

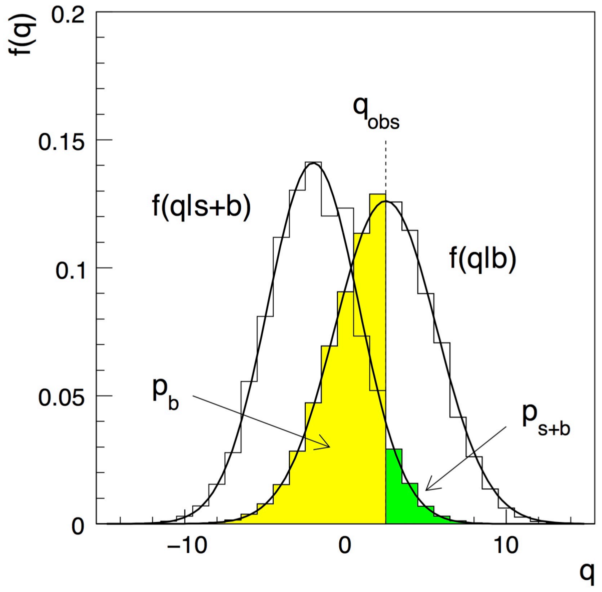


Figure 4.5: Distributions of the test variable q under the hypotheses $s + b$ and b, with the respective p-value [98].

above its mean has an upper-tail probability equal to p[4]. That is,

$$Z = \Phi^{-1}(1 - p) \tag{4.5}$$

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

The particle physics community regards the rejection of the background-only hypothesis with a significance of at least $Z = 5$ as an appropriate level to constitute a discovery, corresponding to $p = 2.87 \times 10^{-7}$. For purposes of excluding a signal hypothesis, a threshold p-value of 0.05 (*i.e.*, 95% C.L.) is often used, which corresponds to $Z = 1.64$.

The sensitivity of an experiment is quantified by the expected significance in the assumption of the different alternative hypotheses. For example, the sensitivity to discover a given signal process $s + b$ is characterised by the expectation value, under the assumption of $s + b$, of the value of Z obtained from a test of b.

The following Sections closely follow the discussion in Reference [98], in order to provide a general overview of the frequentist formalism adopted and the expertise to interpret fit results.

4.3.1 Formalism of the profile likelihood ratio

While $f(x|H)$ refers to the probability density for the observable x for a single event, the probability density for a dataset $\boldsymbol{x}$ with many events is described by the probability $L(\boldsymbol{x}|H)$.

The likelihood function is derived by reinterpreting the probability density function $L(\boldsymbol{x}|H)$ of a set of data $\boldsymbol{x}$ given a hypothesis H as the probability function of a certain hypothesis H given the data. The likelihood function should not be interpreted as a probability density for H.: In particular, it does not have the property that it normalises to unity, unlike the probability density function.

In particular, the probability density function denotes the value of L as a function of $\boldsymbol{x}$ given a fixed hypothesis H (*i.e.*, its parameters); likelihood function denotes the value of L as a function of H given a

---

[4] This relation can also be defined by using a two-sided fluctuation of a Gaussian variable, with a 5σ significance corresponding to $p = 5.7 \times 10^{-7}$. We take the one-sided definition above as this gives $Z = 0$ for $p = 0.5$.

fixed value of $\boldsymbol{x}$. The full structure of $L(\boldsymbol{x}|H)$, with both assumption $\boldsymbol{x}$ and hypothesis H fixed, is referred to as model.

In the $t\bar{t}H$ searches the signal-plus-background hypothesis corresponds to the physics predicted by the SM, whilst the background-only hypothesis excludes the Higgs sector. A parameter μ (called parameter of interest, POI) can be introduced to embody both hypotheses in the likelihood, $L(\boldsymbol{x}|\,\mu)$, and reflects the strength of the signal process,

$$\mu = \frac{\sigma_{t\bar{t}H}}{\sigma^{\mathrm{SM}}_{t\bar{t}H}} \tag{4.6}$$

where $\mu = 0$ corresponds to the background-only hypothesis and $\mu = 1$ to the expected signal-plus-background hypothesis.

Since the measurement can be affected by systematic and statistical uncertainties, their effect can be described by introducing nuisance parameters (NP), $\boldsymbol{\theta} = (\theta_1, \theta_2, ...)$, in the likelihood function $L(\boldsymbol{x}|\,\mu, \boldsymbol{\theta})$.

To test a hypothesised value of μ the profile likelihood ratio $\lambda(\mu)$ is defined as

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}\,, \tag{4.7}$$

whose maximum $\mu^{\max}$ is an estimator of the signal strength. In the numerator $\hat{\hat{\boldsymbol{\theta}}}$ denotes the value of $\boldsymbol{\theta}$ that maximises L for the specified μ, called conditional maximum-likelihood (ML) estimator of $\boldsymbol{\theta}$, and therefore it is a function of μ. The denominator, instead, is the maximised (unconditional) likelihood function, *i.e.*, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the values maximasing

A complete description of the construction of the fit model is provided in Section 4.4.

4.3.2 Test statistic t_μ and p-value

The profile likelihood ratio $\lambda(\mu)$ varies between 0 and 1, with $\lambda(\mu)$ near 1 implying good agreement between the data and the hypothesised value of μ. The compatibility of the observed data with a given hypothesis is usually quantified using a more convenient test statistic

$$t_\mu = -2\ln\lambda(\mu). \tag{4.8}$$

Higher values of t_μ thus correspond to increasing incompatibility between the data and μ.

Maximising $\lambda(\mu)$ or, analogously, minimising the statistic t_μ define the value of the signal strength μ more compatible with data. The error on μ is assessed by scanning the values of t_μ as a function of μ: the values where t_μ decreases by 0.5 with respect to the best estimate define the 68% C.L. error band.

A hypothesised value of μ can be tested by using the statistic t_μ directly as measure of discrepancy between the data and the hypothesis. As described in Section 4.3, the p-values quantifies the level of disagreement:

$$p_\mu = \int_{t_{\mu,\,\mathrm{obs}}}^{\infty} f(t_\mu|\mu)\, dt_\mu\,, \tag{4.9}$$

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

The compatibility of a hypothesis μ when data originate from a different model $\mu' \neq \mu$ is fundamental to determine the sensitivity of an experiment. As an illustration, distributions $f(t_\mu|\mu)$ and $f(t_\mu|\mu')$ are shown in Figure 4.6: $f(t_\mu|\mu')$ is shifted to higher values of t_μ, showing, as expected, less compatibility with the tested hypothesis μ. The p-value corresponding to the median t_μ assuming the alternative value

μ' (med$[t_\mu|\mu']$) determines the sensitivity of the experiment. As the p-value is a monotonic function of t_μ, it is equal to the median p-value assuming μ'.

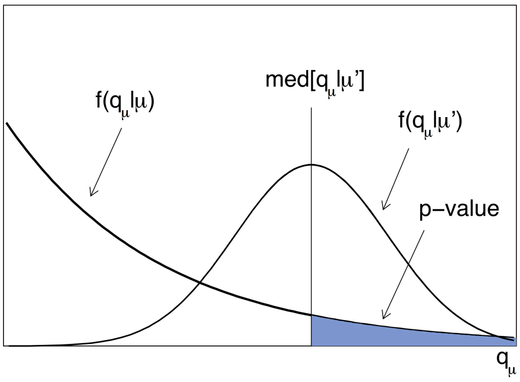


Figure 4.6: Illustration of the p-value corresponding to the median of t_μ (called q_μ in the figure) assuming a strength parameter μ' [98].

Since the presence of a new signal will, usually, increase the mean event rate beyond the background expectation, an additional constraint to the signal strength can be set. To take this into account, an alternative test statistic $\tilde{t}_\mu$ can be defined: if data results in $\hat{\mu} < 0$, the best level of agreement between the data and any physical value of μ occurs for $\mu = 0$. Therefore the profile likelihood ratio $\lambda(\mu)$ can be redefined as

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \text{for } \hat{\mu} < 0\,, \\ \frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \text{for } \hat{\mu} \geq 0\,, \end{cases} \tag{4.10}$$

where $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ and $\hat{\hat{\boldsymbol{\theta}}}(0)$ are the conditional ML estimators of $\boldsymbol{\theta}$ given a strength of 0 and μ, respectively, and

$$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu)\,. \tag{4.11}$$

Again, the level of agreement can be quantified with the p-value defined in Equation 4.9.

4.3.3 Test statistic q_0 for a discovery

The special case of the statistic $\tilde{t}_\mu$ when $\mu = 0$, is used to reject the $\mu = 0$ hypothesis and leads to the discovery of a positive signal. The test statistic for discovery is defined as

$$q_0 = \tilde{t}_0 = \begin{cases} -2\ln\lambda(0) & \text{for } \hat{\mu} < 0\,, \\ 0 & \text{for } \hat{\mu} \geq 0\,, \end{cases} \tag{4.12}$$

where $\lambda(0)$ is the profile likelihood ratio for $\mu = 0$ as defined in Equation 4.7.

The statistic q_0 could have been defined as special case of t_μ, when $\mu = 0$. In that case, anyhow, the $\mu = 0$ hypothesis would have been rejected for either an upward or downward fluctuation of the data. This is appropriate only if the presence of a new phenomenon can either increase or decrease the final yield, as in experiments looking for neutrino oscillations, where the signal hypothesis predicts a greater or lower event rate than the no-oscillation hypothesis.

The p.d.f. $f(q_0|1)$ of the q_0 statistic is used to compute the median expected significance of the experiment, as described in a more general case in the previous Section. It quantifies how the analysis is sensitive to the signal-plus-background hypothesis ($\mu = 1$) assuming the background-only model ($\mu = 0$).

4.3.4 Test statistic q_μ for limit setting

To compute an upper limit on the strength parameter μ, we consider the test statistics q_μ, defined as

$$q_\mu = \begin{cases} -2\ln\lambda(\mu) & \text{for } \hat{\mu} \leq \mu\,, \\ 0 & \text{for } \hat{\mu} > \mu\,, \end{cases} \tag{4.13}$$

where $\lambda(\mu)$ is the profile likelihood ratio as defined in Equation 4.7. Setting $q_\mu = 0$ when $\hat{\mu} > \mu$ is motivated by the idea that values of the signal strength μ smaller than the one preferred by data, $\hat{\mu}$ cannot be excluded. Thus this region is not considered in the test.

An additional remark regards q_0, which has a different definition and does not correspond to a special case of q_μ when $\mu = 0$: q_0 is zero if the data fluctuate downward ($\hat{\mu} < 0$), while q_μ is zero if the data fluctuate upward ($\hat{\mu} > \mu$).

As for the signal significance, in order to compute the expected and observed upper limit the probability distribution of the test statistic $f(q_\mu|\mu')$ needs to be evaluated. For example, the 95% C.L. expected upper limit on the hypothesis $\mu = 0$, using a data model $\mu' = 1$, correspond to the value μ'' such that the p-value of $f(q_0|1)$ using the median of $f(q_{\mu''}|1)$ is equal to 5%. The observed upper limit, instead, corresponds to the value μ'' such that the p-value of $f(q_0|1)$ using $q_{\mu'',\mathrm{obs}}$ is equal to 5%.
The "signal-injected" upper limit, instead, is obtained by computing the p-value of the test statistic $f(q_1|1)$, instead of $f(q_0|1)$ (*i.e.*, by testing the signal-plus-background hypothesis).

When a positive signal strength is expected, $\mu \geq 0$, the statistic $\tilde{q}_\mu$ can be used instead of q_μ:

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \text{for } \hat{\mu} \leq \mu\,, \\ 0 & \text{for } \hat{\mu} > \mu\,, \end{cases} = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \text{for } \hat{\mu} < 0\,, \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & \text{for } 0 \leq \hat{\mu} \leq \mu\,, \\ 0 & \text{for } \hat{\mu} > \mu\,, \end{cases} \tag{4.14}$$

In some numerical examples $\tilde{q}_\mu \sim q_\mu$, but the use of q_μ leads to important simplifications.

4.3.5 Approximation of t_μ: the Asimov dataset for expected results

In order to compute the p-value related to a certain hypothesis of μ, the full distribution $f(t_\mu|\mu')$ of the test statistic t_μ needs to be determined. This can be done by constructing several pseudo-experiments of the hypothesis μ' and calculating the statistic t_μ. Such MC methods are computationally expensive: for instance, having a detailed description of the tails of a distribution up to a $p_0 \sim 10^{-7}$ necessitates of roughly 10^8 simulated pseudo-experiments.

An approximation of the $f(t_\mu|\mu')$ distribution can be obtained in the limit of large statistics (called "asymptotic limit") by using Wald's approximation [99]. Wald's approximation of the test statistic t_μ, in the case of a single parameter of interest, is

$$t_\mu \simeq -2\ln\lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N})\,. \tag{4.15}$$

where $\hat{\mu}$ follows a Gaussian distribution with a mean μ' and standard deviation σ, and N refers to the data sample size. In the asymptotic limit the standard deviation σ can be estimated from the so-called "Asimov dataset" [98], described below.

Neglecting the term $\mathcal{O}(1/\sqrt{N})$ in Equation 4.15, the statistic t_μ follows a non-central χ^2 distribution for one degree of freedom,

$$f(t_\mu; \Lambda) = \frac{1}{2\sqrt{t_\mu}} \frac{1}{\sqrt{2\pi}} [e^{-\frac{1}{2}(\sqrt{t_\mu} + \sqrt{\Lambda})^2} + e^{-\frac{1}{2}(\sqrt{t_\mu} - \sqrt{\Lambda})^2}] , \tag{4.16}$$

where the non-centrality parameter Λ is:

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2} . \tag{4.17}$$

For the special case $\mu' = \mu$, $\Lambda = 0$ and t_μ becomes a χ^2 distribution for one degree of freedom, a result shown by Wilks [100]. Figure 4.7 illustrates an example of the level of agreement between the distribution of q_0 predicted by the asymptotic approximation (black curve), which is independent of the predicted background, and the pseudo-experiments method (histograms): sufficiently good agreement is obtained for large background yields ($b \geq 10$). The asymptotic approximation is used for the two searches described in this thesis work in order to compute the relevant p-values.

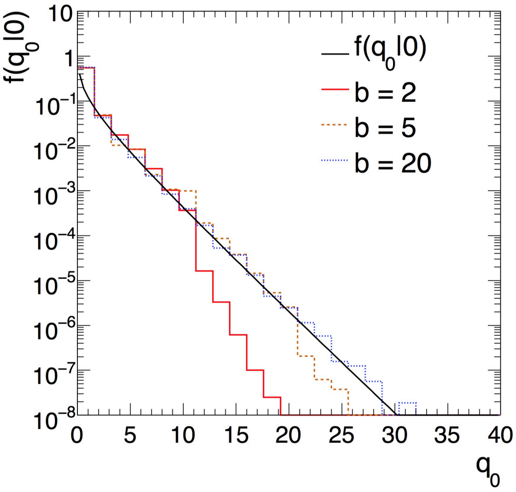


Figure 4.7: The p.d.f. $f(q_0|0)$ for a counting experiment. The solid curve shows $f(q_0|0)$ in the asymptotic limit and the histograms are from pseudo-experiments using different values of the background b [98].

The Asimov dataset is an artificial dataset which replaces the ensemble testing performed with MC pseudo-experiments with a single “representative” dataset. This dataset is defined in such a way to return exactly the true value for each estimated parameter ($\hat{\mu} = \mu'$, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$), and determines the asymptotic parametrisation of $f(t_\mu; \Lambda)$. This dataset is also used to determine the error bands on the median expected limit of μ.

The Asimov dataset allows to study the constraints on the nuisance parameters that could be obtained with the expected data distributions and statistical uncertainties. Any difference in constraints of a given nuisance parameter between the result of the Asimov dataset fit and the data helps to diagnose possible over-constraints from data in the fit.

4.4 Likelihood construction and details about the fit model

Let’s suppose that for each event in the data an observable x is measured and used to construct a histogram $\boldsymbol{n} = (n_1, ..., n_N)$ with entries. The expectation value of n_i can be written as

$$E[n_i] = \mu s_i + b_i \tag{4.18}$$

where s_i and b_i correspond to the number of expected signal and background events, respectively, in the i^{th} bin. The predictions s_i and b_i are affected by systematic and statistical uncertainties, as described in Section 4.3.1. The effect of these uncertainties is incorporated in the likelihood function through the nuisance parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_P)$, whose variations change the signal and background predictions $s_i(\boldsymbol{\theta})$ and $b_i(\boldsymbol{\theta})$. The nuisance parameters values are not known a priori but they can be fitted to the data: the additional flexibility introduced to parametrise systematic effects increases the capability of the model to describe the data, but results in a loss in sensitivity.

In the fit model a nuisance parameter can vary within an allowed range, determined by auxiliary studies, and described by a p.d.f. $\rho(\theta|\tilde{\theta})$, where $\tilde{\theta}$ represents the auxiliary measurement[5]. The $\rho(\theta|\tilde{\theta})$ function is usually called penalty term or prior function on θ and is reinterpreted, according to the Bayes' theorem, as the posterior p.d.f. arising from the measurement $\tilde{\theta}$:

$$\rho(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta) \tag{4.19}$$

where $\pi_\theta(\theta)$ function is the "hyperprior" for those measurements (*i.e.*, the initial degree of belief in θ) and is often chosen to be uniformly distributed. This assumption implies that if $p(\tilde{\theta}|\theta)$ is a Poissonian, $\rho(\theta|\tilde{\theta})$ is a gamma distribution, and if $p(\tilde{\theta}|\theta)$ is Gaussian, $\rho(\theta|\tilde{\theta})$ is normal (*i.e.*, Gaussian) or log-normal.

In addition, subsidiary measurements can be made to constrain the nuisance parameters and can be included in the fit model: for instance the distribution of a kinematic variable in a signal depleted region, mainly influenced by one source of background, can be exploited. It provides a set of measurements $\boldsymbol{m} = (m_1, ..., m_M)$ for the number of entries in each of the M bins. The expectation value of m_i can be written as

$$E[m_i] = u_i(\boldsymbol{\theta}) \tag{4.20}$$

where the u_i are calculable quantities depending on the parameters $\boldsymbol{\theta}$, as the number of predicted events in the background dominated region. This kind of measurement provides additional information on the background normalisation parameter and also possibly on the signal and background shape parameters (discussed in the next Section).

According to the model, data follow a Poisson distribution around its expected number of events. Therefore the full likelihood L for the model to describe the observed data is the product of Poisson probability terms over all bins N and M of each distribution used in the fit, and the nuisance term:

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k} \prod_{l=1}^{P} \rho(\theta_l)\,. \tag{4.21}$$

where $\rho(\theta_l)$ represents the functional form of the priors for the l^{th} nuisance parameter. The Equation 4.21 assumes all systematic uncertainties are uncorrelated so that their p.d.f.'s can factorise in the profile likelihood definition.

4.4.1 Treatment of systematic uncertainties

The functional forms of the prior $\rho(\theta|\tilde{\theta})$ used in the fit model are typically one of the following:

- Gaussian function, which is the common assumption for most systematic uncertainties and is adopted for systematic uncertainties that change the shape of the histograms in the fit model

$$\rho(\theta) = (2\pi\sigma)^{-1/2} e^{-(\theta-\tilde{\theta})^2/2\sigma^2}\,. \tag{4.22}$$

---

[5] An analogous approach is chosen for theoretical uncertainties although they are not associated with an auxiliary measurement.

The Gaussian form allows for positive and negative values of θ and disfavours large variation from the independently measured value $\tilde{\theta}$.

- log-normal function, which is used for normalisation systematics, since the parameter θ can assume only positive values
$$\rho(\theta) = (2\pi)^{-1/2}(\ln(\sigma)\,\theta)^{-1} e^{-\ln^2(\theta/\tilde{\theta})/2\ln^2(\sigma)} . \tag{4.23}$$
For small values of σ the log-normal function approximates a Gaussian distribution.
- Gamma function, which is employed to describe the statistical uncertainties due to the finite number of simulated events
$$\rho(\theta) = \frac{1}{\alpha}\frac{(n/\alpha)^k}{k!} e^{-n/\alpha} \tag{4.24}$$
where $n = \alpha \cdot N$ is related to the rate α and the number of simulated events N.

The prior distribution is defined by the auxiliary measurement of the NP and its uncertainty: in order to easily compare the fitted θ with the initial value $\tilde{\theta}$, the NP can be redefined as centred at zero with a width of one:
$$\theta' = \frac{\theta - \tilde{\theta}}{\sigma} . \tag{4.25}$$
The presence of nuisance parameters, in general, broadens the profile likelihood as function of μ and reflects a loss of information about μ due to the systematic uncertainties.

On the other hand, the profile likelihood approach can increase the knowledge of some systematic uncertainties by fitting the data: a better data-MC agreement can be achieved through the shift (pull) of a given systematic uncertainty, as well as a reduction of the original uncertainty. Moreover, nuisance parameters can evidence correlations among themselves during the likelihood maximisation, inducing further reduction of the effect of the total systematic uncertainties.

The profile likelihood model adopts a continuous parameterisation of the effect of the systematic uncertainties on the used templates $\boldsymbol{n}(\boldsymbol{\theta})$ and $\boldsymbol{m}(\boldsymbol{\theta})$. The nominal templates (with all systematic uncertainties turned off, $\theta' = 0$) and those with a $\pm 1\sigma$ variation of the uncertainty are given in forms of histograms; an interpolation procedure determines the continuous template distribution. Different approaches are adopted for the interpolation of the uncertainty acting on the normalisation and the shape of the template, even though regulated by the same nuisance parameter [101].

The interpolation of the shape distributions into a continuous function of θ is performed through the vertical morphing technique [102]: the contents of each histogram bin is adjusted as a linear or quadratic function of θ, as illustrated in Figure 4.8, and is treated independently.

For NP's that affect only the normalisation of the templates, the interpolation approach adopts a 6^{th} order polynomial function in the range $[-1\sigma, +1\sigma]$, whilst an exponential parametrisation is used outside: in this way a smooth parametrisation is achieved, without discontinuities and the exponential term prevents the template normalisation from becoming negative when the NP is extremely pulled. When only one variation is available (e.g., the jet energy resolution or the parton shower related uncertainties), the effect of the variation is symmetrised.

The fit model incorporates also the statistical effects of the templates, due to the finite amount of simulated events. A single NP is defined for each bin of the histogram allowing for the bin to vary within its statistical uncertainties, according to a simplified version of the approach proposed by Barlow and Beeston [101, 103]. If the statistical uncertainty per bin is small, the prior of this NP approximates a Gaussian distribution with mean of 1 and a σ equal to the relative statistical uncertainty of the content of all the templates in that bin.

It needs to be remarked that the binning of the templates used in the fit model is a subtle parameter of

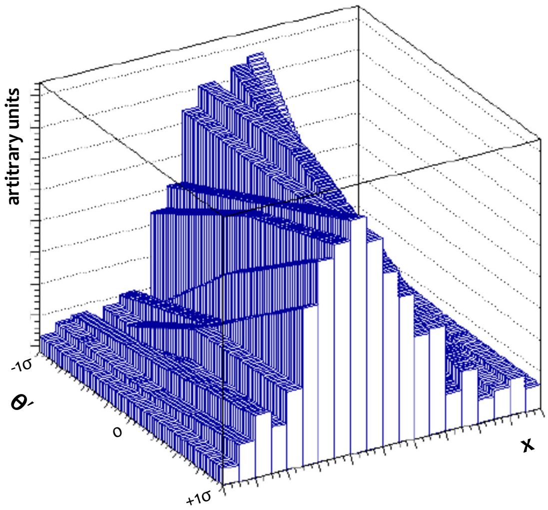


Figure 4.8: Sketch of an interpolation in vertical template morphing which transforms the discrete $\pm 1\sigma$ shape distributions for a given systematic uncertainty into a continuous function of θ.

the fitting procedure: its choice stems from a compromise between the best separation for the classifier shape and the available amount of events in data and in the Monte Carlo samples. An extremely narrow width of the bins, for instance, could cause a lack of background events in the signal-enriched side of the classifier distribution, inducing bias in the signal extraction. Conversely, a large width definition could reduce the discriminating power of the classifier by mitigating the shape differences between signal and background.

Moreover, the available number of events in Monte Carlo is crucial for a reliable estimate of the systematic effects on the shape of the discriminant. Expected performances are strongly affected by the binning variation that could lead to drastic distortion of some important shape systematics. In general a careful check of all the systematic variations per sample per analysis region is performed: an optimal binning is then chosen using the Asimov dataset to take into account those effects.

CHAPTER 5

Search for the Higgs boson decaying into $b\bar{b}$ in association with hadronically decaying top quarks

This Chapter describes the search for the $t\bar{t}H$ production mode in a fully hadronic final state. It is worth mentioning that this is the first fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) search ever performed and data collected by ATLAS during 2012 at a centre-of-mass energy of 8 TeV have been employed. The analysis has been published in Reference [104].

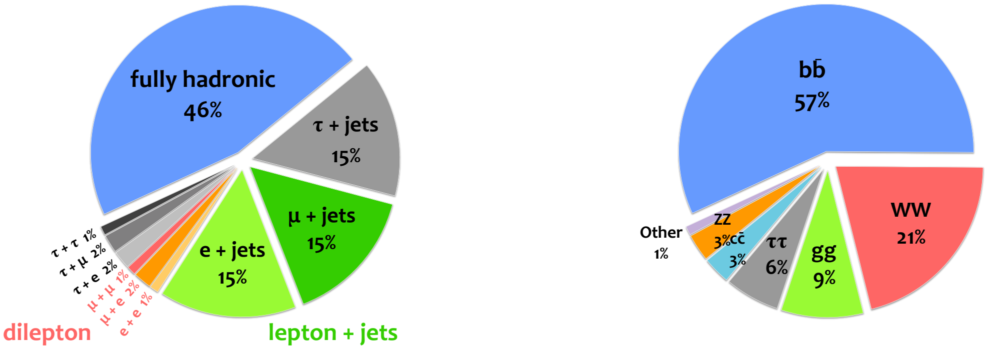
The general aspects of the analysis are introduced in Section 5.1, by discussing the main advantages and shortcomings of a fully hadronic final state. The Monte Carlo samples and the state-of-the-art cross sections used to perform the analysis are then summarised in Section 5.2. Since the signal signature at Born level is characterised by eight quarks, four of which are b quarks, the analysis requires at least 5 jets at preselection level and no leptons, as described in Section 5.3, together with the definition of the physics objects. The background is dominated by the non-resonant production of multijet events, which is known to be difficult to model by Monte Carlo simulation: a data-driven method is therefore adopted by extrapolating its contribution from a control region with the same number of jets, but a lower multiplicity of jets identified as containing b-hadrons (b-tagged jets) than the analysis region. The parameters used for the extrapolation are measured from an extraction region and checked using Monte Carlo simulation. The description and validation of such data-driven methods and the estimation of the subdominant background processes are documented in Section 5.4.

Preselected events are subsequently categorised according to the multiplicity of jets and b-tagged jets, in order to maximise the signal sensitivity, as discussed in Section 5.5. A BDT classifier, based on event shape and kinematic variables, is trained, optimised and used to discriminate the signal from the background, as discussed in Section 5.6. The systematic uncertainties related to detector effects, the signal modelling and the data-driven method used to evaluate the non-resonant multijet production are presented in Section 5.7. Finally the measurement of the top Yukawa coupling is performed through a fit to the BDT-based discriminant distribution in the different analysis regions, as described in Section 5.8.

5.1 The fully hadronic channel

The fully hadronic $t\bar{t}H$ analysis presents several advantages, but also shortcomings: a fully hadronic final state allows to increase the signal acceptance because of the large decay branching fractions of the top quark into hadrons ($\approx 60\%$) and the Higgs boson into a pair of b quarks ($\approx 60\%$). Pie charts in Figure 5.1

summarise the $t\bar{t}$ and the Higgs boson branching fractions. Moreover the absence of missing energy due to leptonic decays in the final state allows for a complete reconstruction of all decay products of the $t\bar{t}H$ system, not taking into account detector and reconstruction inefficiencies. The absence of triggering leptons is instead a major disadvantage: the multijet processes becomes overwhelming, with a cross section six orders of magnitude higher than the $t\bar{t}H$, even in the most sensitive regions.



(a) Top quark pair branching fractions

(b) Higgs decay fractions at $m_H = 125$ GeV

Figure 5.1: (a) Pie chart representing the branching ratios (BR) of a top-antitop quark pair. The light blue represents the fully hadronic BR of 46% (56% when including hadronic decaying τ), the dileptonic BR (without τ lepton) is shown in 3 shades of red, with a total of 4% (6.4% when including leptonic decaying τ), and in 2 shades of green the lepton (e or μ) + jets BR of 30% (36% when including leptonic decaying τ). (b) Pie chart representing the branching ratios of the Higgs boson, assuming a mass of 125 GeV.

5.2 Dataset and Monte Carlo event generation

5.2.1 Dataset

The analysis uses a set of data events collected with the ATLAS detector from pp collisions at the LHC during 2012 at a centre-of-mass energy of 8 TeV. The recorded data correspond to an integrated luminosity of 21.3 fb^{-1}, with a total uncertainty of 2.8%[1] [105]. The time evolution of the total integrated luminosity collected by ATLAS in 2011 and 2012 is shown in Figure 5.2.

The 2012 data are separated in eleven periods taken between April 4th and December 16th in order to organise and separate different running conditions, such as trigger configurations or beam settings. They are further filtered through one of the so-called Good Run Lists (GRL) [106], which collect only events belonging to sets of lumiblocks[2] , where all the subdetectors are fully functional. Only 6% of the events do not satisfy this requirement, as shown in blue in Figure 5.2. The final 2012 data correspond to a total integrated luminosity of 20.3 fb^{-1}. Finally, all events in data are required to have triggered a combination of single jet triggers: triggers applied in this analysis are discussed in Section 5.3.1.

---

[1] The luminosity uncertainty for the 8 TeV data has been further reduced to less than 2%. However, the analysis was already published at that time and results have not been updated to the new uncertainty.

[2] A luminosity block (lumiblock) is the unit of time in the ATLAS data-taking defined as the maximal period where all the data-taking configurations are constant. In general the duration of a luminosity block is of the order of 1 minute.

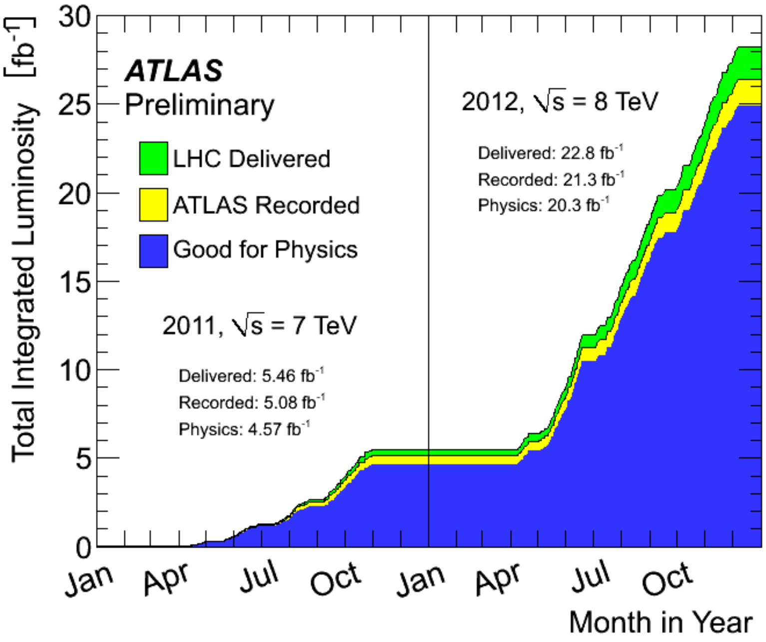


Figure 5.2: Integrated luminosity versus time delivered to (green), recorded by (yellow) ATLAS, and certified to be good quality data (blue) during stable beams and for pp collisions at 7 and 8 TeV centre of mass energy in 2011 and 2012 [61].

5.2.2 Signal and background modelling

The $t\bar{t}H$ signal process is modelled using matrix-element calculations obtained from the HELAC-Oneloop package [107] with next-to-leading order (NLO) accuracy in α_{s}. POWHEG-BOX [86–88] serves as an interface to the MC programs used to simulate the parton shower and hadronisation. The samples created using this approach are referred to as POWHEL samples [108]. They include all SM Higgs boson and top-quark decays and use the CT10 [109] parton distribution function (PDF) sets with the factorisation (μ_{F}) and renormalisation (μ_{R}) scales set to $\mu_{\mathrm{F}} = \mu_{\mathrm{R}} = m_t + m_H/2$. The POWHEL $t\bar{t}H$ samples use PYTHIA 8.1 [110] to simulate the parton shower with the AU2 underlying-event set of generator parameters (tune) [111], while HERWIG [112] is used to estimate systematic uncertainties due to the fragmentation modelling. For these $t\bar{t}H$ samples the cross-section normalisations and the Higgs boson decay branching fractions are taken from the NLO QCD and from the NLO QCD + EW theoretical calculations [113], respectively, and are reported in Table 5.1.

The dominant background to the fully hadronic $t\bar{t}H$ signal is the non-resonant multijet production, followed by $t\bar{t}$+jets production. Small background contributions come from the production of a single top quark and from the associated production of a vector boson and a $t\bar{t}$ pair, $t\bar{t}V$ ($V = W, Z/\gamma^*$). The multijet background is determined from data using a dedicated method described in Section 5.4.The other background contributions are estimated using MC simulations. Table 5.2 shows event generators and configurations used for simulating the signal and background processes.

Multijet events, simulated with PYTHIA 8.1 using the NNPDF2.3 LO [114] PDFs, are used for jet trigger studies and for the validation of the data-driven estimation of the background. Samples of single top quark events produced in the s- and Wt-channels are normalised to the approximate next-to-next-to leading order (NNLO) theoretical cross sections [115, 116] using the MSTW2008 NNLO PDF set [117, 118]. Overlaps between the $t\bar{t}$ and Wt final states are removed [119] by the diagram removal (DR) method [120]. The samples of $t\bar{t}V$ ($V = W, Z/\gamma^*$) events are normalised to NLO cross-sections [121, 122] and have QCD scale uncertainties of 15–20% and PDF+α_s uncertainties is ≈ 10%. Event samples for single top quark plus Higgs boson production, $tHqb$ and tWH, are also considered and the cross

Table 5.1: Production cross sections for signal $t\bar{t}H$, for m_H = 125 GeV, and various simulated background processes. The quoted uncertainties arise from variations of the renormalisation and factorisation scales and uncertainties in the parton distribution functions. Table from Reference [123].

Process	σ [pb]	State-of-the-art precision
$t\bar{t}H$	$0.129^{+0.012}_{-0.016}$	NLO
$t\bar{t}$+jets	253^{+13}_{-15}	NNLO+NNLL
single top Wt-channel	22.4 ± 1.5	aNNLO
single top t-channel	$87.7^{+3.4}_{-1.9}$	aNNLO
single top s-channel	5.61 ± 0.22	aNNLO
$t\bar{t}W$	0.232 ± 0.070	NLO
$t\bar{t}Z$	0.205 ± 0.061	NLO
$tHqb$	$0.0172^{+0.0012}_{-0.0011}$	NLO
tWH	$0.0047^{+0.0010}_{-0.0009}$	NLO

sections are computed using the MadGraph5_aMC@NLO generator [85] at NLO in QCD. These two processes together are referred to as tH. A summary of the cross-section values and their uncertainties for the signal as well as for the simulated background processes is given in Table 5.1.

5.2.3 Modelling of the $t\bar{t}$+jets process

The main $t\bar{t}$+jets sample is generated using the Powheg NLO generator with the CT10 PDF set. It is interfaced to Pythia 6.425 [124] with the Perugia2011C underlying-event tune; this combination of generator and showering programs is hereafter referred to as Powheg+Pythia. The sample is normalised to the theoretical calculation performed at NNLO in QCD as calculated with top++2.0 and includes resummation of next-to-next-to leading logarithmic (NNLL) soft-gluon terms [11–16]. It has uncertainties of 2.5% from parton distribution function uncertainties and $^{+2.5\%}_{-3.4\%}$ from QCD renormalisation and factorisation scale choices, as shown in Table 5.1.

Unlike NLO generators with higher accuracy (*e.g.*, Sherpa), Powheg is expected to describe jet multiplicity properly only for $t\bar{t}$ accompanied by up to two jets. Nevertheless, this generator, when interfaced with Pythia, turned out to provide a good description of the jet multiplicity in data, up to much higher jet multiplicities [125]. Also the heavy flavour fraction in $t\bar{t}$ production is well modelled, despite the fact that heavy flavour component originates only from the parton shower [46].

An alternative $t\bar{t}$+jets sample is also generated using fully matched NLO predictions with massive b-quarks [126] within the Sherpa framework with OpenLoops [89, 127], henceforth referred to as Sherpa. The Sherpa NLO sample is generated following the four-flavour scheme[3] using the Sherpa 2.0 pre-release and the CT10 PDF set. The renormalisation scale is set to $\mu_{\mathrm{R}} = \prod_{i=t,\bar{t},b,\bar{b}} E_{\mathrm{T},i}^{1/4}$, where $E_{\mathrm{T},i}$ is the transverse energy of parton i, and the factorisation and resummation scales are both set to $(E_{\mathrm{T},t} + E_{\mathrm{T},\bar{t}})/2$.

The prediction from Sherpa is expected to model the $t\bar{t} + b\bar{b}$ contribution better than Powheg+Pythia, since the latter MC produces $t\bar{t} + b\bar{b}$ exclusively via the parton shower. The $t\bar{t}$+jets events from Powheg +Pythia are categorised into three non-overlapping samples, $t\bar{t} + b\bar{b}$, $t\bar{t} + c\bar{c}$, and $t\bar{t}$ +light-jets, hereafter

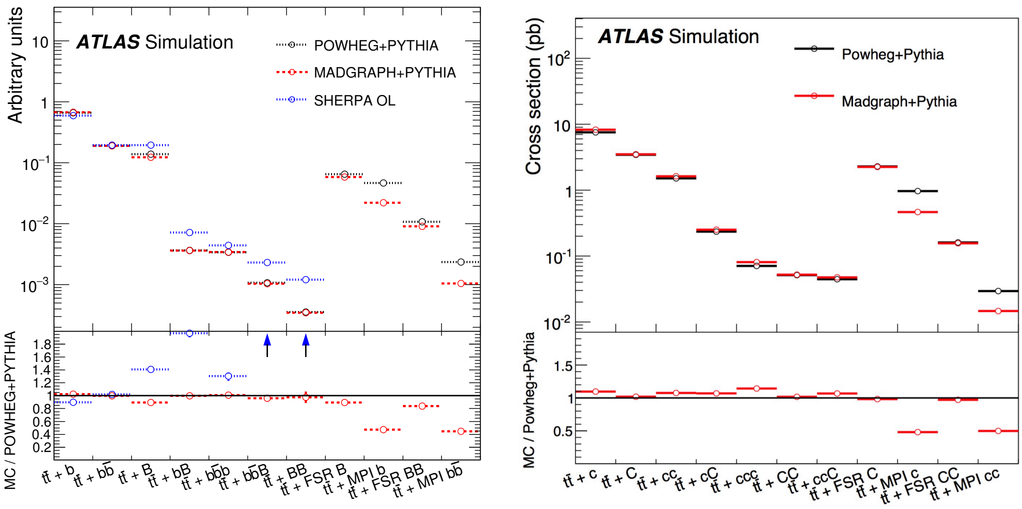
---

[3] The "massive" or four-flavour scheme [128] is the most straightforward way, from the conceptual point of view, to perform calculations of high-energy processes involving the production of bottom quarks. Since bottom quarks are significantly heavier than the proton and therefore can be only created in pairs (or singly in association with a top quark), they are considered to not contribute to the proton wave function. In practice, the scheme amounts to employ an effective theory with n_l light quarks, where the heavy quarks are decoupled and do not enter in the computation of the running coupling constant and in the evolution of the PDFs.

called $t\bar{t}$ + light, using a labelling procedure based on an algorithm that matches hadrons to particle jets and described in Section 5.2.3. Then, $t\bar{t} + b\bar{b}$ events from Powheg+Pythia are reweighted to reproduce the Sherpa NLO $t\bar{t} + b\bar{b}$ prediction. The reweighting is done at generator level using a finer categorisation to distinguish events where one particle jet is matched to two b-hadrons, or where only one b-hadron is matched. Two additional and independent reweighting procedures are also applied to improve the agreement of different variables: the first is based on the top-quark and the $t\bar{t}$ system p_{T} while the second on the ΔR and p_{T} of the dijet system not originating from the top-quark decay, when this is defined [46].

Unlike $t\bar{t} + b\bar{b}$, no fully matched NLO predictions exist for $t\bar{t} + c\bar{c}$ and $t\bar{t}$ + light events. A dedicated reweighting is therefore applied to the top-quark p_{T} spectra as well as to the p_{T} spectra of the $t\bar{t}$ system of $t\bar{t}$ + light and $t\bar{t} + c\bar{c}$ events in Powheg+Pythia, based on the ratio of data to simulation of the measured differential cross sections at $\sqrt{s} = 7$ TeV [129].

To gain more confidence in the description of the jet multiplicity and heavy flavour content of $t\bar{t}$+jets production, additional MadGraph [84] samples have been produced. They are used to further validate the modelling in Powheg generator by comparing to the LO prediction in MadGraph. Reasonable agreement between both generators is found [46] comparing several kinematic variables and can be seen in Figure 5.3. Figure 5.3(a), in particular, shows the relative contributions of different categories of $t\bar{t} + b\bar{b}$ in the different samples used: it demonstrates that Powheg+Pythia is able to reproduce reasonably well the $t\bar{t}$ +HF (heavy flavour jets, namely $b\bar{b}$ and $c\bar{c}$) content of the MadGraph $t\bar{t}$+jets sample, which includes a LO $t\bar{t}+b\bar{b}$ matrix-element calculation, as well as the NLO Sherpa prediction. All the systematic uncertainties associated to the $t\bar{t}$+jets modelling are discussed in Section 5.7.2.



(a) (b)

Figure 5.3: Relative contributions of different categories of (a) $t\bar{t} + b\bar{b}$ and (b) $t\bar{t} + c\bar{c}$ events in Powheg+Pythia, MadGraph +Pythia and Sherpa samples. The labels "tt+MPI" and "tt+FSR" refer to events where heavy flavour is produced via multiparton interaction (soft interactions involving spectator partons from the colliding protons) or final state radiation (a gluon radiated from the top decay products), respectively. These contributions are not included in the Sherpa OL (OpenLoops) calculation. An arrow indicates that the point in the ratio is off-scale. Uncertainties are from the limited MC sample sizes [46].

Table 5.2: Configurations used for event generation of signal and background processes. If only one parton distribution function (PDF) is shown, the same one is used for both the matrix-element (ME) and parton shower generators; if two are shown, the first is used for the matrix-element calculation and the second for the parton shower. "V" refers to production of an electroweak boson (W or Z/γ^*). "Tune" refers to the underlying-event tune of the parton shower generator. "Pythia 6" refers to version 6.425; "Pythia 8" refers to version 8.1; "Herwig" refers to version 6; "Herwig++" refers to version 2.7; "Sherpa" refers to version 2.0; "MadGraph" refers to version 5; "MadGraph5_aMC@NLO" refers to version 2.2.1; "Jimmy" refers to version 4.31. Samples using Pythia 6 and Pythia 8 have heavy flavour hadron decays modelled by EvtGen 1.2.0 [136]. All samples include leading-logarithm photon emission modelled by PHOTOS [131], and Tauola 1.20 [132] to simulate τ decays, with the exception of the Sherpa samples. The QCD multijet background is estimated through a data-driven technique and validated by comparison with the Pythia 8 MC sample. Table from Reference [104].

Process	ME Generator	Parton Shower	PDF	Tune
$t\bar{t}H$	PowHel [108]	Pythia 8 [110]	CT10 [109]/ CTEQ6L1 [137, 138]	AU2 [111]
$tHqb$	MadGraph [84]	Pythia 8	CTEQ6L1	AU2
tWH	MadGraph5_aMC@NLO [85]	Herwig++ [139]	CT10/CTEQ6L1	Jimmy [130]
$t\bar{t}V$	MadGraph	Pythia 6[124]	CTEQ6L1	AUET2B
$t\bar{t}$+jets	Powheg -BOX [140]	Pythia 6	CT10/CTEQ6L1	Perugia2011C [141]
single top Wt-channel	Powheg -BOX [120, 142]	Pythia 6	CT10/CTEQ6L1	Perugia2011C
single top t-channel	AcerMC [143]	Pythia 6	CTEQ6L1/CTEQ6L1	Perugia2011C
single top s-channel	Powheg -BOX	Pythia 6	CT10/CTEQ6L1	Perugia2011C
QCD multijet	Pythia 8	Pythia 8	NNPDF2.3 LO [114]	AU2

Labelling of $t\bar{t}$+jets events

The main purpose of dividing the $t\bar{t}$+jets samples into categories is to evaluate for each subsample a systematic uncertainty during the fit and the proper reweighting to match the state-of-the-art configurations. The labelling into three non-overlapping categories ($t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$, and $t\bar{t}$ + light) is based on a hadron-matching algorithm to particle jets. The particle jets are jets reconstructed from stable truth particles in simulation. All particle jets with $p_{\mathrm{T}} > 15$ GeV and $|\eta| < 2.5$ which are matched using a $\Delta R < 0.4$ to a b-hadron with $p_{\mathrm{T}} > 5$ GeV, not originating from a top quark decay, are labelled as b. If the event has one or two matches, the event is labelled as $t\bar{t}+b$ or $t\bar{t}+b\bar{b}$. If the event has one particle jet matched to two b-hadrons, the event is given the label $t\bar{t}+B$, representing unresolved gluon splitting to $b\bar{b}$. The same procedure is repeated for c jets, if no b-hadron not-from-top match is found. All other events are labelled as $t\bar{t}$ + light. For the fit procedure, the subcategories of events with at least one b-hadron not-from-top match are labelled $t\bar{t}+b\bar{b}$ and events that have no b-hadron match but at least one c-hadron match are labelled $t\bar{t}+c\bar{c}$.

5.2.4 Common processing of MC samples

All samples using Herwig++ are also interfaced to Jimmy 4.31 [130] to simulate the underlying event. With the exception of Sherpa, all MC samples use Photos 2.15 [131] to simulate photon radiation and Tauola 1.20 [132] to simulate τ decays. The samples are then processed through a simulation [133] of the detector geometry and response using Geant 4 [134]. The single top sample produced in the t-channel is simulated with a parametrised calorimeter response [135].

All simulated events are processed through the same reconstruction software as the data. Simulated events are corrected so that the lepton and jet identification efficiencies, energy scales and energy resolutions match those in data, as described in Chapter 3.

5.3 Object and event preselection

The all-hadronic $t\bar{t}H$ final state is composed of jets originating from (u, d, s)-quarks or gluons (light jets) and jets from c- or b-quarks (heavy-flavour jets). Hence, the search is based on data collected using a multijet trigger, which requires at least five jets passing the Event Filter stage. Events are discarded if any jet with $p_{\mathrm{T}} > 20$ GeV is identified as:

- out-of-time activity from previous pp collisions;
- out-of-time activity from residual collisions in the beam pipe, but far from the main interaction point;
- calorimeter electronics noise.

Those quality criteria are designed for an efficiency for good jets of more than 99.8% for $p_{\mathrm{T}}^{\mathrm{jet}} > 20$ GeV [144].

At least one reconstructed primary vertex is required, with at least five associated tracks with $p_{\mathrm{T}} \geq 400$ MeV, and a position consistent with the luminous region of the beams in the transverse plane. If more than one vertex is found, the primary vertex is taken to be the one which has the largest sum of the squared transverse momenta of its associated tracks.

The general description of the procedure to identify single physics objects is given in Chapter 3; requirements resulted from an optimisation or from specific needs of the fully hadronic analysis are summarised and motivated in the following.

Jets are reconstructed with the anti-k_t algorithm, with a radius parameter $R = 0.4$ in the (η, ϕ) plane [78, 145, 146]. They are built from calibrated topological clusters of energy deposits in the calorimeters [58]. After energy calibration based on in-situ measurements [147], jets are required to have transverse momentum $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. To avoid selecting jets from additional pp interactions within the same bunch crossing (known as "pileup"), a loose selection is applied to the jet vertex fraction (JVF), defined as the ratio of the scalar sum of the p_{T} of tracks matched to the jet and originating from the primary vertex to that of all tracks matched to the jet. This criterion, JVF ≥ 0.5, is only applied to jets with $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$.

Electron candidates are reconstructed from energy deposits (clusters) in the electromagnetic calorimeter that are matched to a reconstructed track in the inner detector. They are required to have $|\eta_{\mathrm{cluster}}| < 2.47$, to be out of the crack region and to be well isolated. Since the single-lepton channel suffers significantly by backgrounds with non-prompt leptons (due to jet misidentification or leptons from semileptonic b/c-hadron decays), an η-dependent isolation cut is made, based on the sum of transverse energies of cells around the direction of each candidate, in a cone of size $\Delta R = 0.2$ (referred to as $E_{\mathrm{T}}^{\mathrm{cone20}}$). A further isolation cut is made on the scalar sum of the track p_{T} around the electron in a cone of size $\Delta R = 0.3$ (referred to as $p_{\mathrm{T}}^{\mathrm{cone30}}$).

Muon candidates are reconstructed from track segments in the muon spectrometer, and matched with tracks found in the inner detector. They are required to satisfy $|\eta| < 2.5$ and a p_{T}-dependent track-based isolation requirement that has good performance under conditions with a high number of jets from pileup, or in boosted configurations where the muon is close to a jet: the track p_{T} scalar sum in a cone of variable size $\Delta R < 10\ \mathrm{GeV}/p_{\mathrm{T}}^{\mu}$ around the muon must be less than 5% of the muon p_{T}.

The longitudinal impact parameter of the lepton track with respect to the primary vertex, z_0, is required to be less than 2 mm. A p_{T} cut at 25 GeV is applied on the lepton, driven by the threshold of the lowest unprescaled single lepton trigger available in 2012.

Ultimately all physics objects surviving the previous requirements are required to pass an overlap removal procedure (OLR), which prevents double-counting of objects (e.g., calorimetric deposits from electrons and muons misreconstructed as jets), and assures a resolved regime, where all objects are well separated in the ϕ–η space. During jet reconstruction the jet finding algorithm uses as input all the energy deposits in the calorimeter, thus isolated electrons can also be reconstructed as jets. The overlap removal procedure requires first the removal of any jet within a cone of size $\Delta R < 0.2$ around a reconstructed electron and the electron removal then within a $\Delta R = 0.4$ of a remaining jet. This requirement eliminates any perturbation due to close-by jets activity for the calculation of the electron isolation and shower shape variables. Muons are required to be separated by $\Delta R > 0.4$ from any selected jet, they are discarded otherwise. Finally, if a muon candidate shares the same track with a selected electron in the Inner Detector, the full event is dropped.

To ensure the orthogonality with the $t\bar{t}H$ leptonic analyses well identified isolated muons or electrons with $p_{\mathrm{T}} > 25$ GeV are discarded in order to avoid overlap with other $t\bar{t}H$ analyses. The five leading jets in p_{T} are required to have $p_{\mathrm{T}} > 55$ GeV with $|\eta| < 2.5$ and all other jets are required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$. Events are required to have at least six jets, of which at least two must be b-tagged.

Jets are b-tagged by means of the MV1 algorithm [79]. The working point used for this search corresponds to a 60% efficiency to tag a b-quark jet, a light-jet rejection factor of approximately 700 and a charm-jet rejection factor of 8. The tagging efficiencies obtained in simulation are adjusted to match the results of the calibrations performed in data [148].

5.3.1 The multijet trigger

After the reconstruction of a jet at the trigger level [149] ("online jet") a set of corrections are applied offline, so that the "offline jets" have a more precise calibration and better resolution. The difference in efficiency when objects are reconstructed online and offline has to be taken into account, especially in regions close to the trigger activation threshold. Thus the multijet trigger efficiency with respect to the offline selection needs to be derived for data and Monte Carlo and a Scaling Factor (SF) applied to correct the discrepancies of simulation compared with data.

Usually trigger studies need to identify an event variable and parametrise the trigger efficiencies with respect to this variable. While this is trivial for single object trigger, where the efficiencies are highly correlated to the energy (or p_{T}) of the most energetic offline object in the event, for multi-object trigger this is usually done by parametrising with respect to the lowest N^{th} leading objects.

In this analysis a very accurate procedure to assess trigger efficiency between data and MC is needed [123]. Instead of using the usual approach and parametrise with respect to an event-based quantity, it is assumed that the trigger behaviour of any event can be inferred by the properties of each offline reconstructed object, which individually contribute to the global efficiency. By taking into consideration the properties of all relevant reconstructed objects in the event, this approach tries to derive a scale factor that can be applied universally.

The idea is to derive a parametrisation of the multijet trigger efficiency, as convolution of single jet efficiencies, $\varepsilon_{\mathrm{trig}}(p_{\mathrm{T}}, \eta)$, validate the the method on different MC samples and eventually derive a scale factor, $\mathrm{SF}_{\mathrm{trig}}(p_{\mathrm{T}}, \eta) = \frac{\varepsilon^{\mathrm{data}}_{\mathrm{trig}}(p_{\mathrm{T}},\eta)}{\varepsilon^{\mathrm{MC}}_{\mathrm{trig}}(p_{\mathrm{T}},\eta)}$. Such single jets are associated with a complete jet trigger chain, i.e., a complete sequence of jets reconstructed at Level-1, Level-2 and EF satisfying the requirements described in Section 5.3. More details about the calculations of the trigger scale factors and related sudies to validate the method are documented in Reference [123].

5.4 Background estimation: the Tag Rate Function (TRF) method

Requiring a large number of b-tagged jets can cause a severe reduction of the available amount of Monte Carlo events and produce large statistical fluctuations in the resulting distributions for samples when a large number of b-tags is required. For instance, with a 70% b-tagging working point, the efficiency to b-tag at least three jets is $(70\%)^3 = 34\%$ per event. This reduction of simulated events can affect the sensitivity of the analysis, as the corresponding statistical fluctuations on the background templates could lead to unreliable estimates of the impact of systematic uncertainties. Furthermore, exclusion limits may be biased, depending on how the MC distributions fluctuate with respect to the data in the signal region.

In order to overcome this problem, a Tag Rate Function ($\mathrm{TRF}_{\mathrm{MC}}$) method is employed for all MC samples used in the analysis, with the aim of using all events in the pre-b-tagged sample to predict the normalisation and shape after b-tagging. Rather than "directly" tagging the jets, individually using the b-tagging weight, the normalisation and the shape of these distributions are predicted by calculating the efficiency that a jet with a given p_{T}, η and flavour [4] will be b-tagged [150]. The flavour-dependent parameterisation of the efficiency is crucial for this approach, since it allows to replace the exact b-tagging by the jet weight with the corresponding efficiency.

The method is validated by verifying that the predicted normalisations and shapes reproduce the expected ones, obtained for a given working point of the b-tagging algorithm.

For the multijet background, instead, a data-driven TRF method has been developed for this specific analysis. At the time the analysis strategy was developed, the available amount of statistics for simulated multijet events was too low and closure tests using the $\mathrm{TRF}_{\mathrm{MC}}$ were not conclusive. Therefore, a data-driven technique, the Tag Rate Function for multijet events ($\mathrm{TRF}_{\mathrm{MJ}}$) method, was used to estimate the multijet background, which is the dominant contribution in all the analysis regions.

Since the analysis regions consist of events with at least three b-tagged jets, the $\mathrm{TRF}_{\mathrm{MJ}}$ method needs to reasonably predict properties of these multijet events. Data events with exactly two b-tagged jets (fully dominated by multijet processes and outside of signal regions) can be employed to predict the multijet properties for the ≥ 3 b-tagged jets regions.

Similarly to the $\mathrm{TRF}_{\mathrm{MC}}$ method, the data-driven version uses τ, the probability (efficiency) of b-tagging a third jet in a sample of events with at least two b-tagged jets, to extrapolate the multijet background from the regions with lower b-tag multiplicity to the search regions with higher b-tag multiplicity, but otherwise identical event selection. Since the starting sample is data and the probability is also extracted from data, the method is fully data-driven.

The principle of this methods, and the various studies performed to validate them for other sub-leading backgrounds ($t\bar{t}$+jets), the $t\bar{t}H$ signal and the dominant multijet background are discussed in Reference [123].

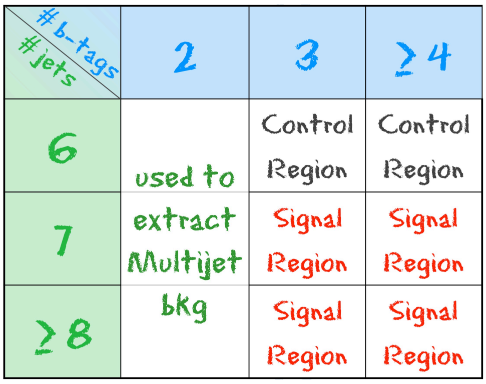
5.5 Event classification: definition of signal and control regions

To enhance the overall sensitivity, six analysis regions are considered for the fit, according to their jet and b-tag multiplicities: two control regions $(6j, 3b)$, $(6j, \geq 4b)$ and four signal regions $(7j, 3b)$, $(7j, \geq 4b)$, $(\geq 8j, 3b)$ and $(\geq 8j, \geq 4b)$. The two latter are expected to be the most sensitive regions and to contribute more than 50% of the total significance (see Table 5.3).

The scheme of how the analysis is organised and illustrated in Figure 5.4. Each row shows a specific jet multiplicity (6, 7, ≥ 8) and the columns show the b-tagged jet multiplicity (2, 3, ≥ 4). In the high b-tag multiplicity regions, namely where exactly 3 and ≥ 4 jets are b-tagged, a Boosted Decision Tree (BDT) is employed to separate signal and background, as described in the next Section.

---

[4] The flavour of a jet in MC events is defined by looking at partons with $p_{\mathrm{T}} > 5$ GeV within a $\Delta R < 0.3$ cone around the jet direction. If a b-quark is found, the jet is labelled as having been originated from a b-quark, otherwise c-quarks are considered for labelling; if no c-quarks are found either, a jet is labelled as a "light" jet.



#b-tags / #jets	2	3	≥4
6	used to extract Multijet bkg	Control Region	Control Region
7		Signal Region	Signal Region
≥8		Signal Region	Signal Region

Figure 5.4: Overview of the fully hadronic $t\bar{t}H$ analysis regions. The multijet background is defined in the exactly 2 b-tagged region and then extrapolated to higher b-tagged jet multiplicity regions by means of the $\mathrm{TRF}_{\mathrm{MJ}}$.

Table 5.3: Event yields from simulated backgrounds and the signal as well as data in each of the analysis regions prior to the fit (pre-fit). The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties in the yields for all samples but the multijet background. The multijet normalisation and its systematic uncertainty are determined by the fit, thus only its statistical uncertainty is quoted here. Since the numbers are rounded, the sum of all contributions may not equal the total value. The signal-to-background ratio, S/B, and the significance, $S/\sqrt{B}$, are also given. The tH background is not shown as it amounts to fewer than 1.5 events in each region.

	6j, 3b	6j, ≥4b	7j, 3b	7j, ≥4b	≥8j, 3b	≥8j, ≥4b
Multijet	16380 ± 130	1112 ± 33	12530 ± 110	1123 ± 34	10670 ± 100	1324 ± 36
$t\bar{t}$ + light	1530 ± 390	48 ± 18	1370 ± 430	45 ± 18	1200 ± 520	40 ± 23
$t\bar{t} + c\bar{c}$	280 ± 180	17 ± 12	390 ± 240	21 ± 15	560 ± 350	48 ± 33
$t\bar{t} + b\bar{b}$	330 ± 180	44 ± 26	490 ± 270	87 ± 51	760 ± 450	190 ± 110
$t\bar{t}V$	14.2 ± 6.3	1.8 ± 1.5	22.0 ± 9.0	3.5 ± 2.3	40 ± 15	8.0 ± 4.2
single top	168 ± 63	6.0 ± 3.7	139 ± 55	8.3 ± 4.6	110 ± 49	10.6 ± 5.9
Total background	18700 ± 480	1229 ± 48	14940 ± 580	1288 ± 66	13330 ± 780	1620 ± 130
$t\bar{t}H$ (m_H = 125 GeV)	14.3 ± 4.6	3.3 ± 2.1	23.7 ± 6.4	7.2 ± 3.3	48 ± 11	16.8 ± 6.1
Data events	18508	1545	14741	1402	13131	1587
S/B	< 0.001	0.003	0.002	0.006	0.004	0.010
$S/\sqrt{B}$	0.10	0.095	0.194	0.20	0.415	0.417

The event yields in the six analysis regions prior to the fit ("pre-fit") are summarised in Table 5.3. Figure 5.5 shows the $S/\sqrt{B}$ ratio, where S and B denote the expected signal (assuming SM cross sections and branching ratios, and $m_H = 125$ GeV) and background, respectively, in each region: the signal enriched regions are highlighted in red. The relative proportions of the background yields obtained in each topology are shown in Figure 5.6. Following the same convention, the ratio of $t\bar{t}H$ events labelled by their Higgs decay is presented in Figure 5.7.

5.6 Analysis method

The employment of an MVA technique is crucial for this analysis, given the very poor S/B ratio. However, the large number of reconstructed jets, the absence of missing particles in the final state (apart from acceptance inefficiencies) and the complex topology of the fully hadronic $t\bar{t}H$ state allow for the definition of a multitude of variables that can be exploited by an MVA classifier.

For this analysis an Adaptive BDT classifier is considered, as implemented in the TMVA Toolkit. To obtain maximal sensitivity, a dedicated BDT is defined and optimised in each of the six analysis regions

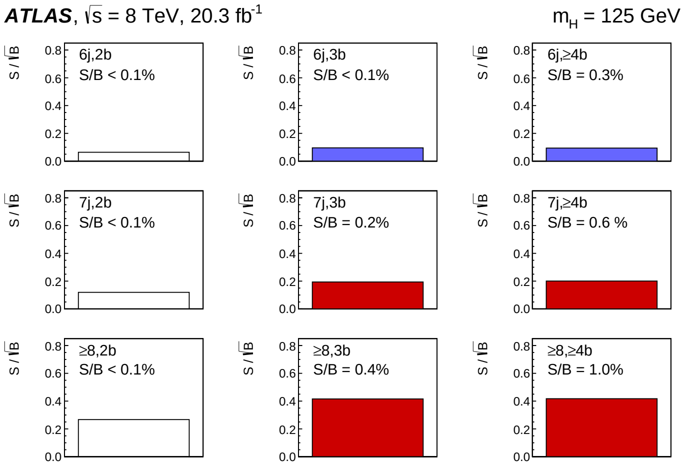


Figure 5.5: $S/\sqrt{B}$ ratio for each of the topologies under consideration after preselection (assuming SM cross sections and branching ratios at $\sqrt{s}$ = 8 TeV, and m_H = 125 GeV). Each row shows the plots for a specific jet multiplicity (6, 7, ≥ 8), and the columns show the b-tagged jet multiplicity (2, 3, ≥ 4). Signal regions are shown in red [123].

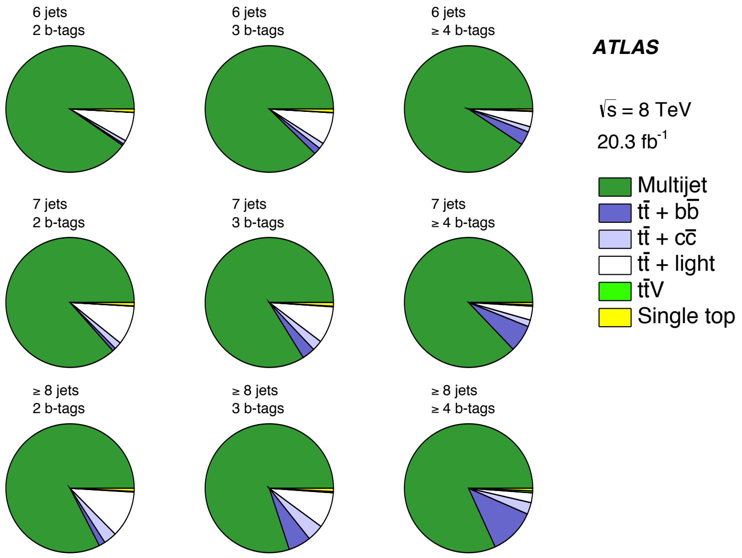


Figure 5.6: Pie-chart of the predicted signal and background events simulated by MC after preselection [123].

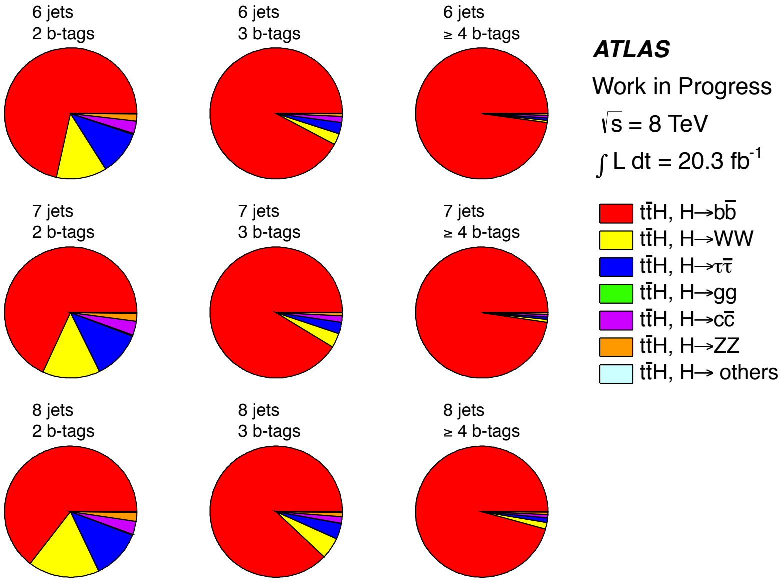


Figure 5.7: Pie-chart of the predicted $t\bar{t}H$ events simulated by MC after preselection, classified by their Higgs decay [123].

in terms of a figure of merit, that will be defined in the next Section, to discriminate the $t\bar{t}H$ signal from the total background.

Using the BDT output distributions, a hypothesis test is performed, based on the profile likelihood ratio, as described in Chapter 4. Details like the choice and validation of the input variables and several studies aiming to increase the sensitivity of the analysis are detailed in the following Sections.

5.6.1 Variables entering the BDT

A large number of variables is considered for optimizing the BDT; the p_T of the softest jet in the event is the only individual kinematic variable that enters the BDT directly. Concerning global event variables, the S_T (the modulus of the vector sum of all jet p_T) and $H_{\mathrm{T}5}$ (the scalar sum of the jet p_T starting from the fifth jet in p_T order) are considered. The used event-shape variables are the centrality, sphericity and aplanarity, whereas $(E_{\mathrm{T}\,1} + E_{\mathrm{T}\,2})/\sum E_\mathrm{T}^{\mathrm{jets}}$ (the sum of the transverse energies of the two leading jets divided by the sum of the transverse energies of all jets), m_{jj}^{min} (the smallest invariant mass of all dijet combinations), and the minimum ΔR between jets are used as object pairs.

In principle the full kinematics of the events can be reconstructed in the $(8j,4b)$ region only; however, some partial reconstruction can be attempted in the other regions as well.

The variables $\Delta R(b,b)^{p_\mathrm{T}^{\mathrm{max}}}$ (the ΔR between the two b-tagged jets with highest vector sum p_T) and $m_{bb}^{\Delta R(b,b)^{\mathrm{min}}}$ (the invariant mass of the two b-tagged jets with the smallest ΔR) are quite efficient in detecting the 2 b-tagged jets originated from the Higgs decay and exhibit a peak at the Higgs mass for the signal.

Analogously $m_{2\,\mathrm{jets}}$ (the mass of the dijet pair, which, when combined with any b-tagged jet, maximises the magnitude of the vector sum of the p_T of the three-jet system) and $m_{2\,\mathrm{b\text{-}jets}}$ (the invariant mass of the two b-tagged jets, which are selected by requiring that the invariant mass of all the remaining jets is maximal) try to track the hadronic top system and the Higgs decay products, respectively.

Two variables are also calculated as the invariant mass of three jets: $m_{\mathrm{top},1}$ is computed from the three jets whose invariant mass is nearest to the top-quark mass, taking into account the jet energy resolutions; the $m_{\text{top},2}$ calculation uses the same algorithm but excludes the jets that enter $m_{\text{top},1}$.

Finally, a Pseudo Matrix-Element discriminator $\mathcal{D}$ is employed: it relies on the assumption that in a fully hadronic final state at least one top quark, a hadronic W from top decay and a Higgs boson decaying into two b quarks can be recognised. It is described in the next Section.

5.6.2 A simplified matrix-element: the Pseudo Matrix-Element approach

The Pseudo Matrix-Element method can deal with partial event reconstruction: it exploits the presence of reconstructed resonances (masses) and combines them in a single discriminator.

Firstly we concentrate on the mass resonances that are present in the signal events. For a given resonance X, decaying to two jets, the signal probability $P^{\text{sig}}(m_{jj})$ is built within a mass window $w_X = 30$ GeV around the given particle mass:

$$P^{\text{sig}}(m_{jj}) = \begin{cases} s \cdot G(m_{jj}|m_X, \sigma_X), & \text{for } |m_{jj} - m_X| \leq w_X, \\ 1 - s, & \text{for } |m_{jj} - m_X| > w_X \end{cases} \tag{5.1}$$

where s is the signal efficiency to find a jet pair with an invariant mass within 30 GeV of m_X and the signal mass distribution is simply modelled with a Gaussian $G(m_{jj}|m_X, \sigma_X)$. The efficiency s is calculated from the signal Monte Carlo simulation and $P^{\text{sig}}(m_{jj})$ is normalised to unity.

For the top-quark resonance the three-particle mass, m_{jjb}, is used instead. The width of the Gaussian is set to $\sigma_X = 18$ GeV for all resonances[5]; this value corresponds to the expected experimental width of a Higgs boson with no combinatorial background.

Finally the expression for the complete signal probability P^{sig} is:

$$\begin{aligned} P^{\text{sig}}(m_{jj}, m_{jjb}, m_{bb}) = P^{\text{sig}}_W(m_{jj}|m_W, \sigma_X) \cdot P^{\text{sig}}_{\text{top}}(p_{\text{T},jjb}, m_{jjb}|m_{\text{top}}, \sigma_X) \cdot \\ \cdot P^{\text{sig}}_H(p_{\text{T},bb}, m_{bb}|m_H, \sigma_X). \end{aligned} \tag{5.2}$$

where the three terms refer to the W, top, and Higgs, resonances respectively, and the jets are used exclusively. For the top-quark and Higgs boson resonances the masses, m_{jjb} and m_{bb}, as well as the p_{T}, defined as the magnitude of the vector sum of the p_{T} of the jets, used to reconstruct the top quark, $p_{\text{T},jjb}$, or to reconstruct the Higgs boson, $p_{\text{T},bb}$, are used. In this case P^{sig} is defined as the product of the mass and p_{T} distributions and no correlation effects are taken into accout. More generally, the probability P^{sig} can be parametrised as function of any kinematic distribution (p_{T}, E, H_{T}, etc.) that helps in discerning signal from background events:

Since the multijet background dominates all the analysis regions, for the background hypothesis we neglect residual contributions and concentrate on the multijet description. Multijet backgrounds do not contain any mass resonance, leading to a background probability $P^{\text{bkg}}(m_{jj})$:

$$P^{\text{bkg}}(m_{jj}) = \begin{cases} b \cdot U(m_X - w_X, m_X + w_X), & \text{for } |m_{jj} - m_X| \leq w_X, \\ 1 - b, & \text{for } |m_{jj} - m_X| > w_X. \end{cases} \tag{5.3}$$

where b is the background efficiency to find a jet pair with an invariant mass within 30 GeV of m_X and $U(m_X - w_X, m_X + w_X)$ is a uniform distribution between $m_X - w_X$ and $m_X + w_X$, reflecting the absence of

---

[5] The Gaussian width of 18 GeV around the mass resonance has been fixed after several studies on the reconstructed invariant mass distributions for the three resonances in different analysis regions.

any mass resonance. Analogously:

$$P^{\text{bkg}}(m_{jj}, m_{jjb}, m_{bb}) = P_W^{\text{bkg}}(m_{jj}|m_W) \cdot P_{\text{top}}^{\text{bkg}}(p_{\text{T},jjb}, m_{jjb}|m_{\text{top}}) \cdot \\ \cdot P_H^{\text{bkg}}(p_{\text{T},bb}, m_{bb}|m_H). \tag{5.4}$$

A discriminator $\mathcal{D}$ is then calculated as the log-ratio $\ln(P^{\text{sig}}/P^{\text{bkg}})$ for all possible jet combinations: according to the Neyman–Pearson lemma [151] this ratio is the most powerful discriminant between signal and background processes. Finally the jet combination yielding maximum $\mathcal{D}$ of the event is chosen.

The variable is used as input variable to the BDT and results to be very discriminant in the most sensitive analysis regions. It should be noticed that, in this definition, signal-like and background-like events get positive and negative values of $\mathcal{D}$, respectively.

5.6.3 BDT optimisation

In the training of the Boosted Decision Tree the sum of all backgrounds is considered as total background: more precisely the sum of the data-driven multijet prediction, obtained through the application of the TRF_{MJ},and the background MC samples, with TRF_{MC} applied to increase the amount of events used to estimate the b-tagged contribution, are considered. The signal is purely determined from Monte Carlo samples, to which the TRF_{MC} is applied, as in the background samples.

Both the signal and the background samples are split according to the jet multiplicity and then each of them is randomly split in two samples of equal size, for training and testing purposes, in order to check possible effects of overtraining.

The variables entering the BDT are selected and ranked according to their separation power with an iterative procedure, which stops when adding more variables does not significantly improve the separation between signal and background ($< 1\%$ improvement). The choice of the Adaptive BDT comes after dedicated studies using a subset of 25 most discriminant variables, according to their algorithm-unspecific ranking score; the same applies for the optimisation of the decision tree options (minimum node size, maximal depth, granularity of cuts, randomisation) and the boosting (number of trees, β parameter of adaptive boosting).

Once the optimal setting is chosen, the BDT is first trained with a set of 90 input variables in order to identify the least useful ones: variables with the worst ranking score or strongly correlated with others are removed, keeping the more discriminant amongst them. The number of candidate variables gets reduced from 90 to 35.

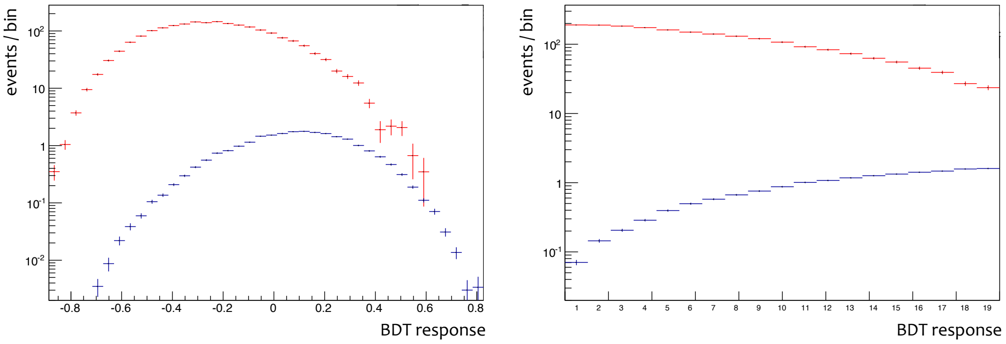
After the skimming, an iterative procedure starts in order to select the most performing set of variables for the BDT. Variables are added one by one to the BDT and are retained if they ranked highest by the TMVA tool, i.e. the best improvement in term of the figure of merit F is produced. The figure of merit F is calculated using the ratio of the signal and background histograms of the BDT response, at the expected integrated luminosity (20.3 fb^{-1}):

$$F = \sqrt{\sum_{i=1}^{n} \frac{S_i^2}{B_i}},$$

where S_i and B_i are, respectively, the number of signal events and the number of background events in the i^{th} bin.
Since the figure of merit F is crucial for the choice of the variables and is calculated $\sim 3 \cdot 10^3$ times, a proper BDT binning needs to be defined and determined in an automatic way, as illustrated in Figure 5.8.

The BDT response binning has been built to ensure:

- at least 10% of signal-plus-background events per bin;
- no empty bin for the background BDT distribution.



(a)

(b)

Figure 5.8: Response curve of the boosted decision tree to $t\bar{t}H$ signal (blue) and all background (red). The right hand plot shows the same, with bins of variable width, which are more suitable to calculate the figure of merit.

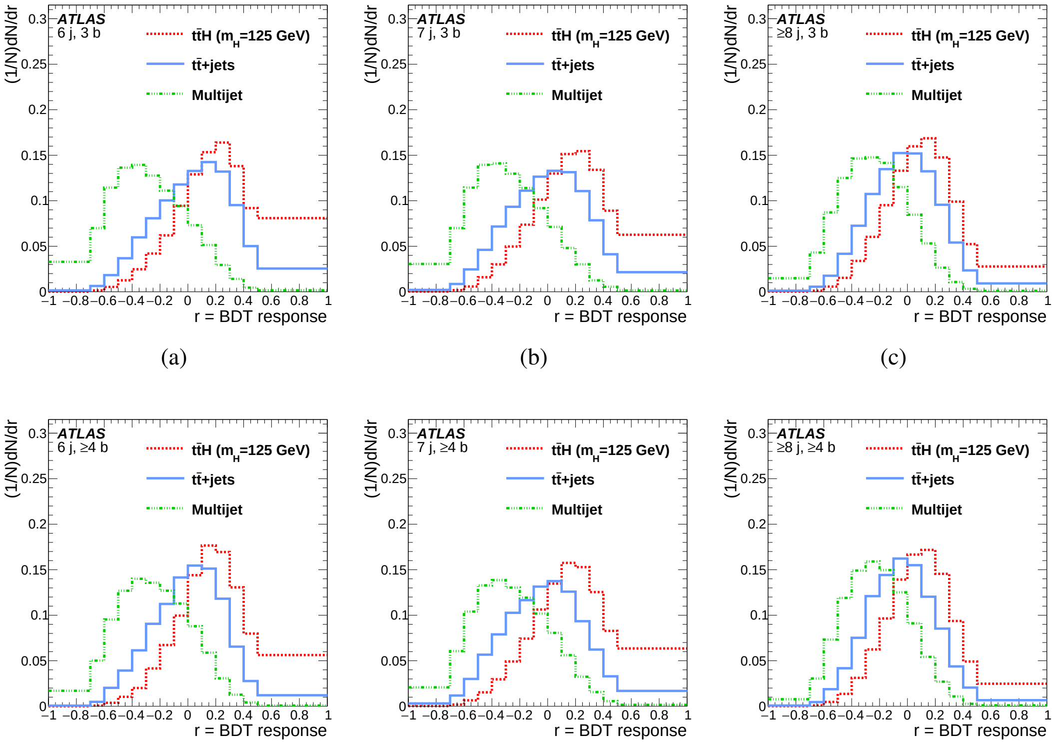
Discriminant variables are iteratively added to the BDT, until the figure of merit F value reaches a plateau with less than 1% variation: in most regions this occurs after adding the 11^{th} variable. At the end of the optimisation, 22 different variables enter the BDT of one or more analysis regions.

From a practical point of view a further skimming is preferable, since the validation of those variables in all analysis regions would require very careful studies. In case of a single occurrence of a variable, this is simply replaced with another, which has a similar physical content, but features a higher occurrence. Mass-related variables are replaced by mass-related variables, distances by distances, etc. The dependence of the BDT performance on the chosen variables is also taken into account. This way a shorter list of best variables is obtained, without a degradation of the performance.

The set of variables entering the BDT and their definitions are listed in Table 5.4: the ranking of each variable in terms of separation power for each analysis region is also shown.

Table 5.4: List of variables used in the BDT in the six analysis regions. The numbers indicate the ranking of the corresponding variables, ordered by decreasing discriminating power. Variables not used in the BDT of a specific region are marked by a dash.

Variable	Definition	BDT rank					
		6j, 3b	6j, ≥ 4b	7j, 3b	7j, ≥ 4b	≥ 8j, 3b	≥ 8j, ≥ 4b
$\text{Centrality}_{\text{Mass}}$	Scalar sum of the jet p_{T} divided by the invariant mass of the jets	1	1	1	1	9	6
Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with all jets	–	11	–	–	6	–
S_{T}	The modulus of the vector sum of jet p_{T}	2	2	2	4	2	2
$H_{\text{T}\,5}$	Scalar sum of jet p_{T} starting from the fifth jet	8	–	–	7	–	–
m_{jj}^{min}	Smallest invariant mass of any combination of two jets	9	–	6	10	11	12
ΔR^{min}	Minimum ΔR between two jets	6	5	9	–	8	4
$p_{\text{T}}^{\text{softest jet}}$	p_{T} of the softest jet	–	6	10	–	–	10
$\Delta R(b,b)^{p_{\text{T}}^{\text{max}}}$	ΔR between two b-tagged jets with the largest vector sum p_{T}	11	–	7	5	5	3
$m_{bb}^{\Delta R(b,b)^{\text{min}}}$	Invariant mass of the combination of two b-tagged jets with the smallest ΔR	3	3	8	9	3	9
$\frac{E_{\text{T}\,1}+E_{\text{T}\,2}}{\sum E_{\text{T}}^{\text{jets}}}$	Sum of the E_{T} of the two jets with leading E_{T} divided by the sum of the E_{T} of all jets	5	8	4	2	7	5
$m_{2\,\text{jets}}$	The mass of the dijet pair, which, when combined with any b-tagged jet, maximises the magnitude of the vector sum of the p_{T} of the three-jet system	10	–	–	8	–	–
$m_{2\,\text{b–jets}}$	The invariant mass of the two b-tagged jets which are selected by requiring that the invariant mass of all the remaining jets is maximal	12	7	–	6	–	8
$m_{\text{top},1}$	Mass of the reconstructed top quark	13	10	–	–	4	11
$m_{\text{top},2}$	Mass of the reconstructed top quark calculated from the jets not entering $m_{\text{top},1}$	7	9	5	–	10	7
$\mathcal{D}$	The Pseudo Matrix-Element ($\mathcal{P.M.E.}$) discriminator, defined as the logarithm of the ratio of event probabilities under the signal and background hypotheses	4	4	3	3	1	1



(a) (b) (c)

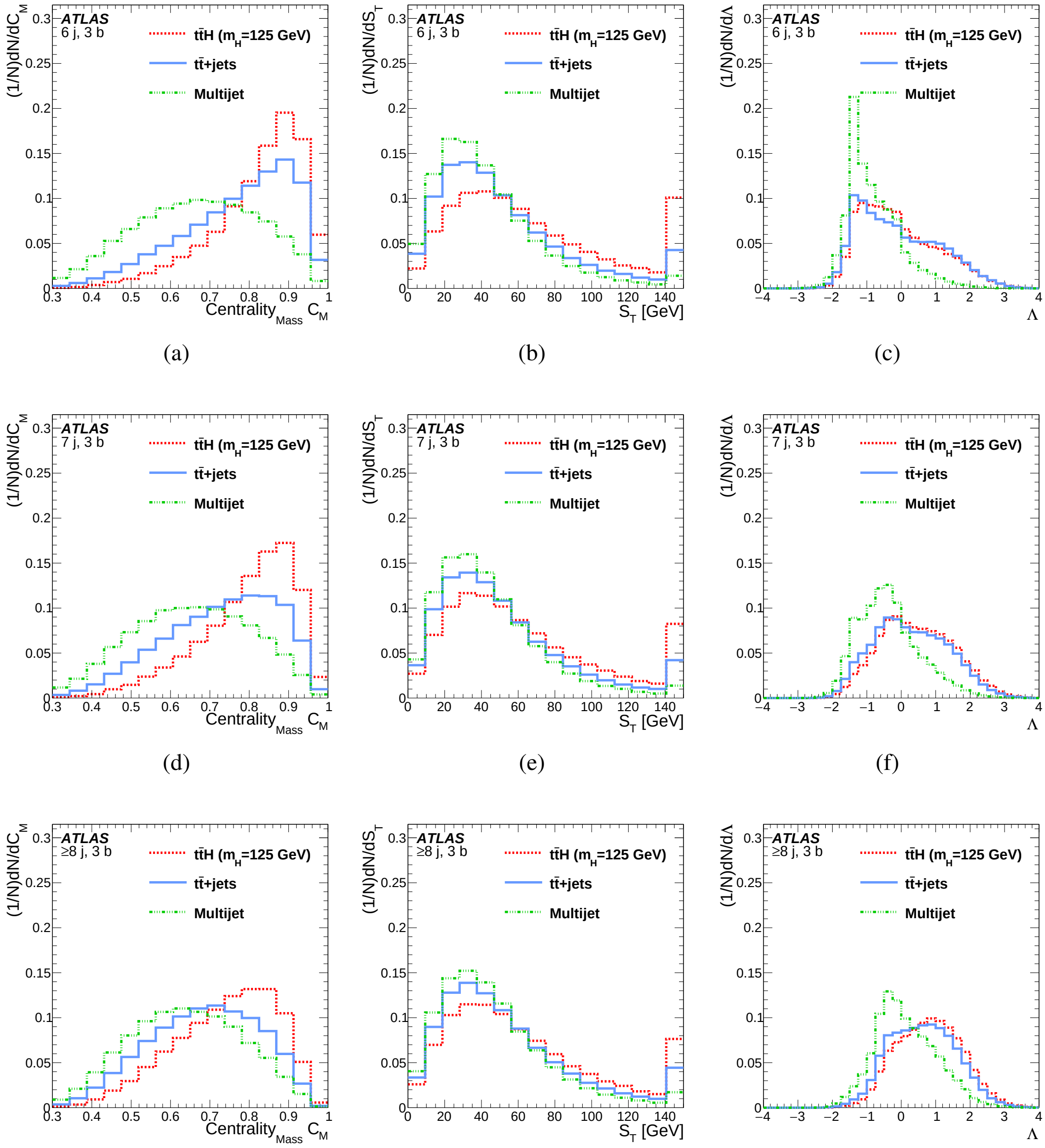
(d) (e) (f)

Figure 5.9: Response of the BDT algorithm for simulated signal (dashed red), $t\bar{t}$+jets background (solid blue) and multijet background (dotted green) events in the (top) regions with 3 b-tags ((a) 6, (b) 7 and (c) $\geq$ 8 jets) and in the (bottom) regions with $\geq$ 4 b-tags ((d) 6, (e) 7 and (f) $\geq$ 8 jets).

The distributions of the BDT outputs for simulated signal and background events are shown in Figure 5.9 for each analysis region. The Figure shows a better separation between signal and background for low jet multiplicities than for high jet multiplicities. This is explained by the number of possible jet permutations. The number of jet permutations increases giving the background more configurations to mimic the signal, reducing the discriminating power of the discriminator $\mathcal{D}$ and, consequently, the BDT.

The set of most discriminating variables used by the BDT in each region is shown in Figures 5.10 and 5.11. A K-S test is performed to check whether the training procedure suffers of overtraining: the separation plots for BDT training and test responses are shown in Figure 5.12 for the $\geq$ 8 jets regions and in Appendix A. No overtraining has been observed for signal or background hypotheses in any of the regions. Additional plots and the correlation matrices between variables for signal and background are shown in Appendix A.

Figure 5.13 shows the ROC curves of different BDTs in the six analysis regions: the green (black) curve refers to the BDT performance before (after) adding the discriminator $\mathcal{D}$ and the optimisation of the BDT, while the red curve corresponds to an intermediate step, when the discriminator was not introduced yet. Large improvements in the separation power are attained with the discriminator $\mathcal{D}$ and the iterative optimisation, especially in the most sensitive channels (with a multiplicity of $\geq$ 8 jets).



(a) (b) (c)

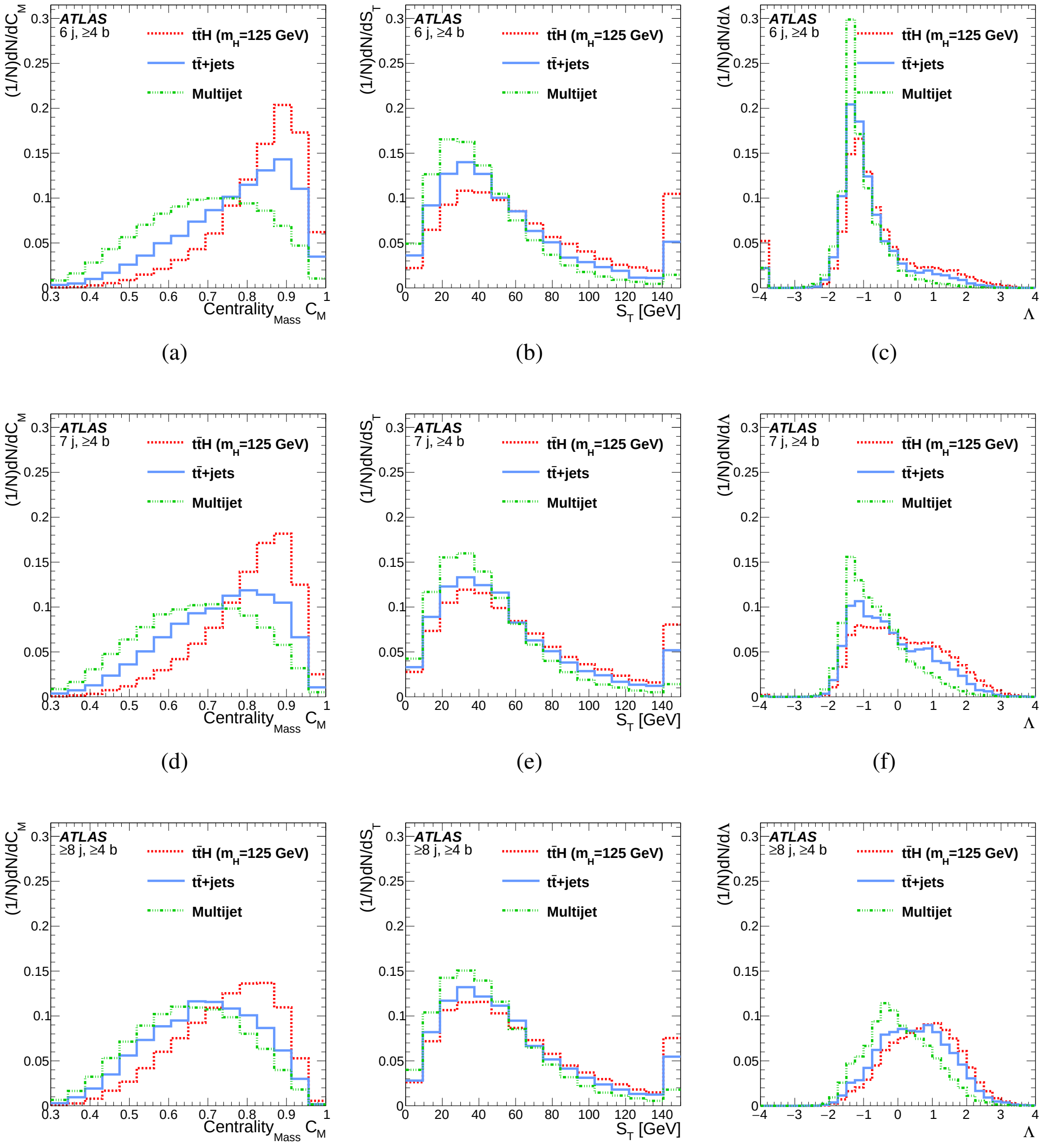
(d) (e) (f)

(g) (h) (i)

Figure 5.10: Distribution of the most discriminant variables entering the BDT in the 3 b-tag regions (($6j$,$3b$) in (a)-(c), ($7j$,$3b$) in (d)-(f) to ($\geq 8j$,$3b$) in (g)-(i)). The plots illustrate the separation power: the red histogram indicates the signal, the blue one the $t\bar{t}$+jets background, and the green one the multijet background. The definitions of the variables are listed in Table 5.4.

5.6.4 Validation of the variables entering the BDT

After the BDT optimisation, a careful validation of the modelling of all employed variables is performed. The modelling is validated by comparing the expected distributions with data in the signal regions and control regions. The large amount of events in the control regions like ($6j$,$3b$) and ($6j$, $\geq 4b$) can be used to check the modelling of the multijet background, which represents 95% of the expected events.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 5.11: Distribution of the most discriminant variables entering the BDT in the ≥ 4 b-tag regions (($6j,\geq 4b$) in (a)-(c), ($7j,\geq 4b$) in (d)-(f) to ($\geq 8j,\geq 4b$) in (g)-(i)). The plots illustrate the separation power: the red histogram indicates the signal, the blue one the $t\bar{t}$+jets background, and the green one the multijet background. The definitions of the variables are listed in Table 5.4.

The remaining four regions have larger contamination of $t\bar{t}$+jets followed by $t\bar{t}V$, besides the dominating multijet contribution. As a result, information on the modelling of multijet background events can be easily extracted, but not so for the $t\bar{t}$+jets background. More details about the fit and how control regions are exploited to infer the multijet background are provided in Section 5.8.

During the blind phase of the analysis, the modelling of the input variables has been tested in the four signal regions after applying an “anti-BDT” cut, which excludes bins where the expected S/B ratio

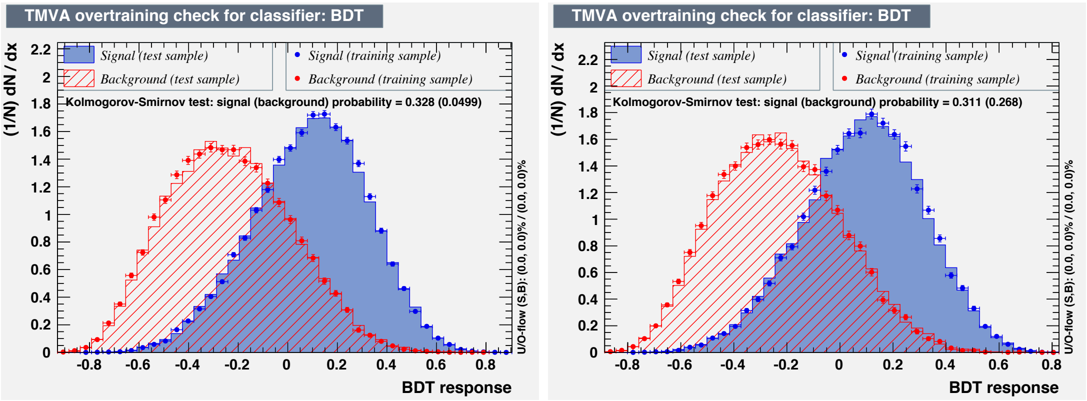


Figure 5.12: The response of the boosted decision trees in the 8-jet inclusive regions: (left) exactly 3 b-tags, (right) ≥ 4 b-tags. The blue histogram indicates the BDT output for signal used in the training of the trees; the blue points indicate the BDT output distribution from the signal test subsample. In an analogous way, the red histogram (dots) indicates the BDT output distribution for background used in training (testing).

Analysis region	anti-BDT cut
6 jets and 3 b-tags	not blind
6 jets and ≥ 4 b-tags	not blind
7 jets and 3 b-tags	0.5
7 jets and ≥ 4 b-tags	0.2
≥ 8 jets and 3 b-tags	0.3
≥ 8 jets and ≥ 4 b-tags	0.0

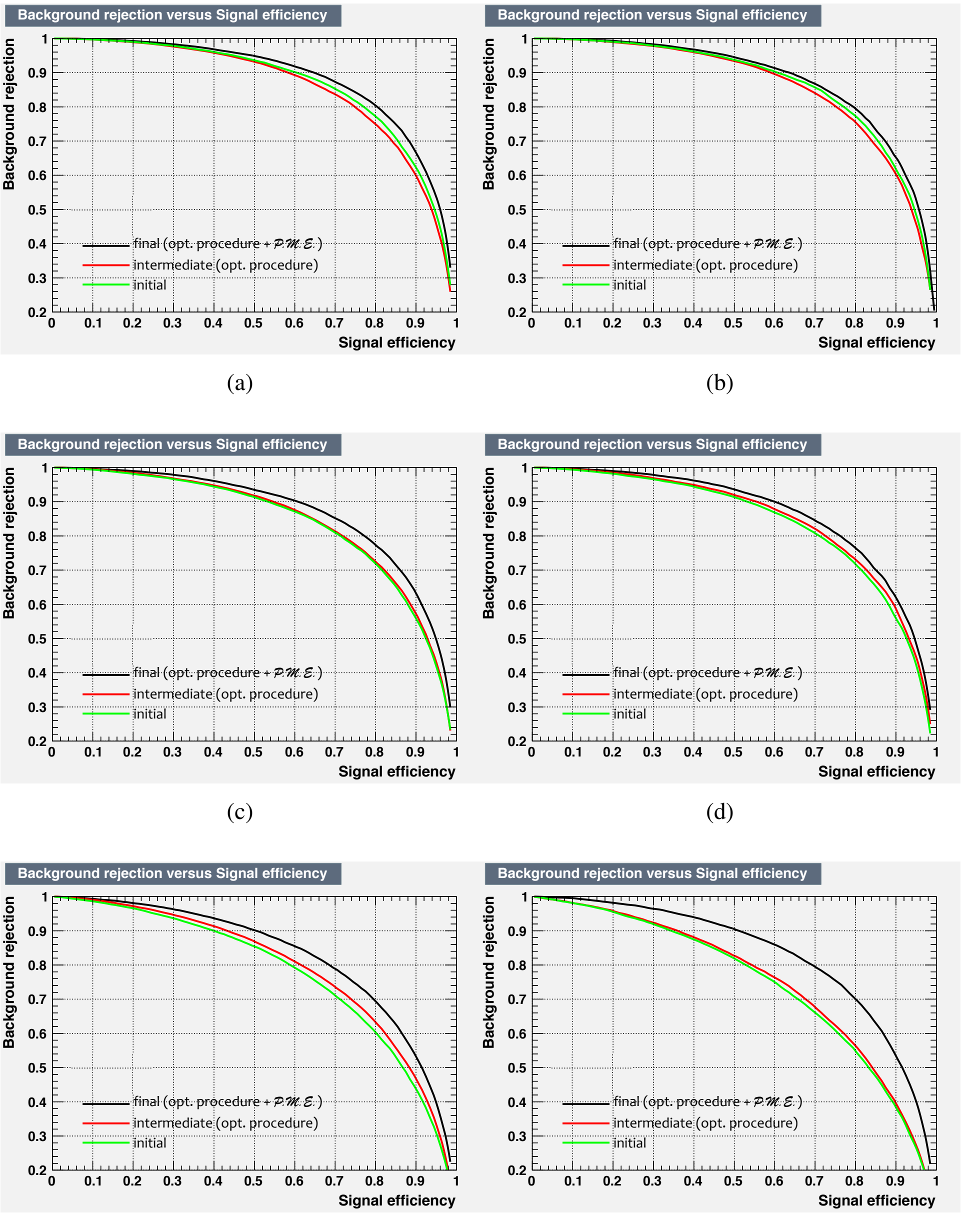
Table 5.5: Value of the BDT discriminant, for each region, below which less than 2% of the signal events are expected. This value is used for blinding.

exceeds 2%, such to remove any signal contamination. The results are shown in Table 5.5: after the cut, the expected signal contamination in each region reduces to less than 15%, while retaining 70% to 100% of the background events.

Figure 5.14 shows the data/MC comparison in the ($\geq 8j$, $\geq 4b$) region after unblinding, for six of the most sensitive variables among all regions. Additional comparisons for the remaining regions are shown in Appendix A. The uncertainties, that will be discussed in Section 5.7, are considered in the comparisons. Additional validation plots for the remaining BDT variables can be found in Appendix B. No significant shape disagreement is observed: a Kolmogorov-Smirnov test is performed to establish the level of agreement after normalizing the distributions for the sum of the background to the data. All the observed normalisation differences are within the total uncertainty.

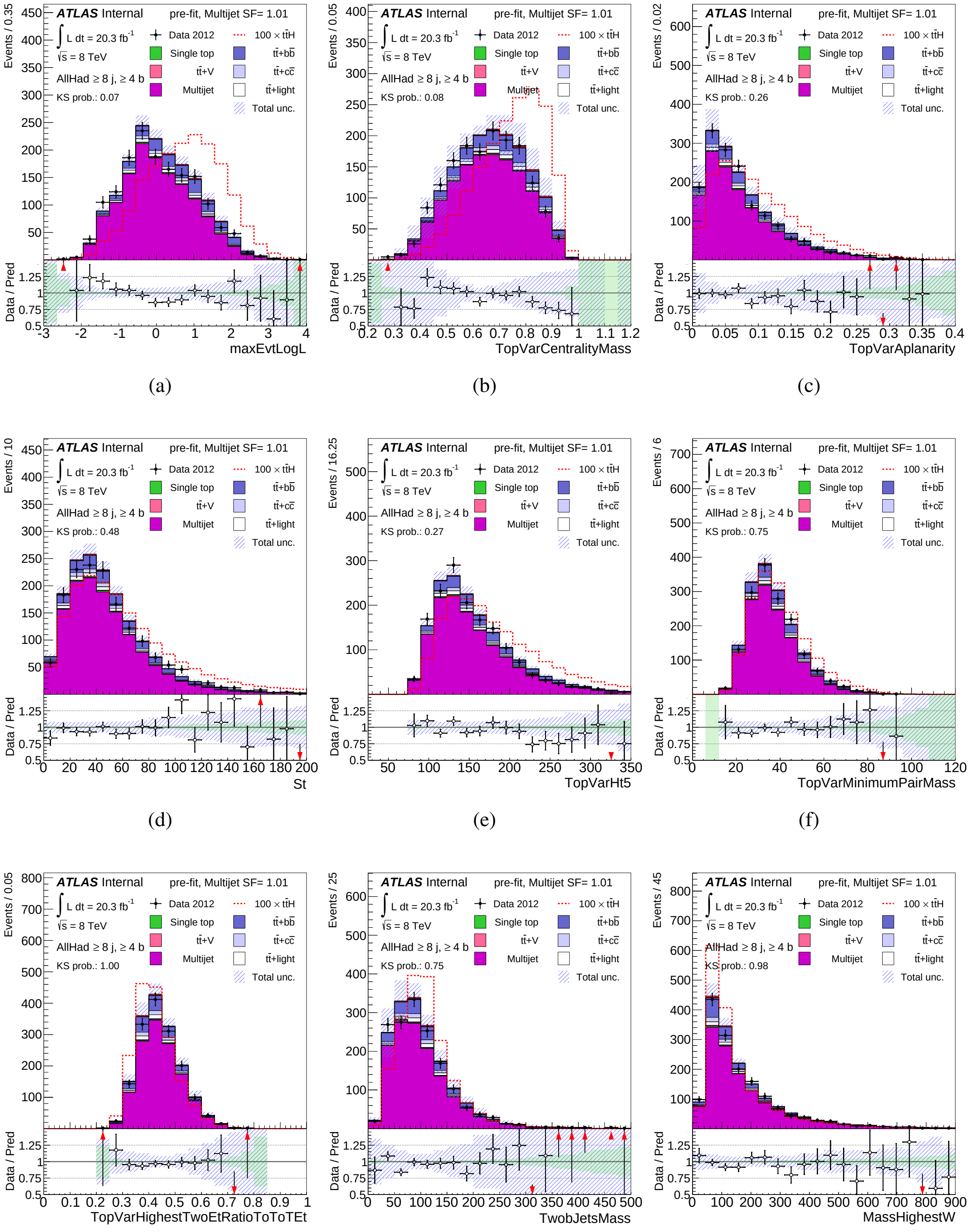
All the validation plots are shown before the fit procedure (“pre-fit”), which will be described in Section 5.8. The fit will further improve the knowledge of the background normalisation and reduce the systematic uncertainties. After the fit (“post-fit”) is performed on the final discriminants, the level of agreement for each variable is also verified and results adequate. Post-fit distributions can also be found in Appendix B.

Finally the distributions of the BDT classifiers in the ($6j$, $3b$), ($6j$, $\geq 4b$), ($7j$, $3b$), ($7j$, $\geq 4b$), ($\geq 8j$, $3b$) and ($\geq 8j$, $\geq 4b$) regions are validated. This is illustrated in Figures 5.20 and 5.21 in the (6–$8j$, $3b$) and (6–$8j$, $\geq 4b$) regions, respectively. In conclusion, a good level of agreement is found for the BDT distributions in all the analysis regions, prior to the fitting procedure and after.



(a) (b) (c) (d) (e) (f)

Figure 5.13: BDT performance improvement in the six analysis regions, in the ROC plane: ($6j$,$3b$) and ($6j$,$\geq 4b$) on top, ($7j$,$3b$) and ($7j$,$\geq 4b$) in the middle, and ($\geq 8j$,$3b$) and ($\geq 8j$,$\geq 4b$) on bottom. The ROC plane shows the signal purity ($1-\varepsilon_{\mathrm{bkg}}$) and the signal efficiency $\varepsilon_{\mathrm{sig}}$ of a classifier. Performing classifiers increase the integral below the ROC curve.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

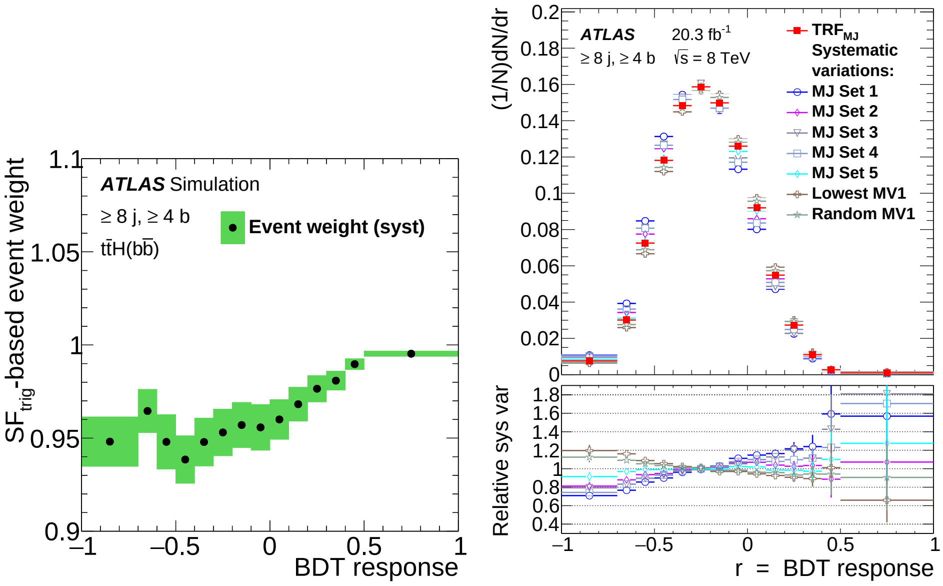
Figure 5.14: Data/MC comparison for a set of the most discriminating variables used in the BDT. These variables are highly ranked in at least one of the BDTs. Distributions are shown in the ($\geq 8j$, $\geq 4b$) analysis region before the fitting procedure ("pre-fit"). The uncertainty band contains the statistical and systematic contribution [123].

5.7 Systematic uncertainties

The sources of systematic uncertainty considered in this analysis can be grouped into six categories, as summarised in Table 5.6. Each systematic uncertainty is represented by an independent parameter, referred to as a nuisance parameter.

The uncertainties in the integrated luminosity, reconstruction of the physics objects, and the signal and background MC models follow the recommended treatment described in the next Sections. The uncertainties related to the jet trigger, as well as those related to the data-driven method to estimate the multijet background are discussed below. In total, 99 fit parameters are considered. The determination and treatment of the systematic uncertainties are detailed in this Section. Their impact on the fitted signal strength is summarised in Table 5.9 in Section 5.8.2.

The discussion on the systematic uncertainties closely follows the discussion in the ATLAS publication [104].



(a)

(b)

Figure 5.15: (a) Per event trigger scale factor $\mathrm{SF}_{\mathrm{trig}}$ (black dots) versus the BDT output of $t\bar{t}H$ events, shown with its corresponding systematic uncertainty (green band) for the (≥8j,≥4b) region. (b) Comparison of the BDT output of the multijet background predicted with different sets of $\mathrm{TRF}_{\mathrm{MJ}}$. The nominal $\mathrm{TRF}_{\mathrm{MJ}}$ is represented by the red points. The bottom panel shows the ratios of the alternative $\mathrm{TRF}_{\mathrm{MJ}}$ predictions to the nominal set [104].

5.7.1 Detector-related systematics

The systematic uncertainty in the luminosity determination for the data sample is 2.8%. It is derived following the same methodology as that detailed in Reference [105]. The evaluation of the luminosity scale is performed using several luminometers, and comparisons between these luminosity detectors are made to assess the accuracy, consistency and long-term stability of the results.

The trigger uncertainty is determined from the difference between $\varepsilon_{\mathrm{trig}}$, estimated using $t\bar{t}H$ and dijet MC events. Each jet in the event is weighted according to $\mathrm{SF}_{\mathrm{trig}}(p_{\mathrm{T}}, \eta)$, the uncertainty of which is

Table 5.6: Sources of systematic uncertainty considered in the analysis grouped in six categories. "N" denotes uncertainties affecting only the normalisation for the relevant processes and channels, whereas "S" denotes uncertainties which are considered to affect only the shape of normalised distributions. "SN" denotes uncertainties affecting both, shape and normalisation. Some sources of systematic uncertainty are split into several components. The number of components is also reported.

Systematic uncertainty source	Type	Number of components
Luminosity	N	1
Trigger	SN	1
Physics Objects		
Jet energy scale	SN	21
Jet vertex fraction	SN	1
Jet energy resolution	SN	1
b-tagging efficiency	SN	7
c-tagging efficiency	SN	4
Light-jet tagging efficiency	SN	12
Background MC Model		
$t\bar{t}$ cross section	N	1
$t\bar{t}$ modelling: p_{T} reweighting	SN	9
$t\bar{t}$ modelling: parton shower	SN	3
$t\bar{t}$ +heavy-flavour: normalisation	N	2
$t\bar{t}+c\bar{c}$: heavy-flavour reweighting	SN	2
$t\bar{t}+c\bar{c}$: generator	SN	4
$t\bar{t}+b\bar{b}$: NLO Shape	SN	8
$t\bar{t}V$ cross section	N	1
$t\bar{t}V$ modelling	SN	1
single top cross section	N	1
Data driven background		
Multijet normalisation	N	6
Multijet TRF$_{\mathrm{MJ}}$ parametrisation	S	6
Multijet H_{T} correction	S	1
Multijet S_{T} correction	S	1
Signal Model		
$t\bar{t}H$ scale	SN	2
$t\bar{t}H$ generator	SN	1
$t\bar{t}H$ hadronisation	SN	1
$t\bar{t}H$ parton shower	SN	1

propagated to the shape and normalisation of the BDT output distribution, as shown in Figure 5.15(a).

Since the analysis does only veto the presence of a lepton and does not make use of the missing transverse momentum, no uncertainties related to leptons or $E_{\mathrm{T}}^{\mathrm{miss}}$ are considered; the uncertainties in physics objects are related to the reconstruction and b-tagging of jets, only. The jet energy resolution (JER) and the jet energy scale (JES) uncertainties are derived combining the information from test-beam data and simulation [147]. The JES uncertainties are split into 21 uncorrelated components. The largest of these uncertainties is due to the jet-flavour composition. In general, the effect of jet energy scale uncertainties increases with the number of selected jets; several shape differences are possible, since the various components can selectively affect specific kinematic regions (low or high p_{T} jets, more forward regions). The JVF uncertainty, instead, is derived from $Z(\rightarrow \ell^{+}\ell^{-})+$ 1-jet events in data and simulation by varying the nominal cut value.

The uncertainty related to the b-tagging is modelled with six independent parameters, while four parameters model the c-tagging uncertainty [79]. These are eigenvalues obtained by diagonalising the matrix which parametrises the tagging efficiency as a function of p_{T}, taking into account bin-to-bin correlations. Twelve parameters, which depend on p_{T} and η, are used to parametrise the light-jet-tagging systematic uncertainties [152]. The per-jet b-tagging uncertainties are 3%–5%, about 10% for c-tagging and 20% for light-jet tagging. An additional uncertainty is assigned to the b-tagging efficiency for jets with $p_{\mathrm{T}} > 300$ GeV, because of the lack of events for an accurate calibration from data.

5.7.2 $t\bar{t}$-related systematics

A combined uncertainty of 6% is assigned to the $t\bar{t}$+jets production cross section, including modelling components due to the value of α_{s}, the PDF used, the process energy scale, and the top quark mass [109, 117, 118, 153, 154]. Other systematic uncertainties related to $t\bar{t}$+jets production are due to the modelling of parton showers and hadronisation.

As already described in Section 5.2.3, the $t\bar{t}$+jets background is split into three categories and treated separately: this allows to better model the different heavy-flavour components b/c in the additional jets, not originating from the top-quark decay products and to vary their percentage separately in the fit, using dedicated systematics.

The systematic uncertainties arising from the reweighting procedure to improve $t\bar{t}$ background description in the simulation (Section 5.2.3) have been extensively studied and adopted in this analysis. Nine largest uncertainties are associated to the top quark and $t\bar{t}$ p_{T} reweighting and represent approximately 95% of the total experimental uncertainty of the $t\bar{t}$ differential-cross-section measurement at 7 TeV [129]. Each component is applied to the inclusive $t\bar{t}$+jets process; two additional uncertainties on the p_{T} modelling are considered as well for $t\bar{t}$ +HF ($t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$), since no specific measurement existed for the top kinematics in $t\bar{t}$ +HF before the publication of this analysis.

The largest uncertainties in the $t\bar{t}$ background description arise from radiation modelling, the choice of generator to simulate $t\bar{t}$ production, the JES, JER, and flavour modelling. These systematic uncertainties are considered for the $t\bar{t}$ + light and $t\bar{t}+c\bar{c}$ components.

Four additional systematic uncertainties in the $t\bar{t}+c\bar{c}$ estimate are derived from the simultaneous variation of the factorisation and renormalisation scales in MadGraph+Pythia. For the $t\bar{t}+b\bar{b}$ background, three scale uncertainties are evaluated by varying the renormalisation and resummation scales. The shower recoil model uncertainty and two uncertainties due to the PDF choice in the Sherpa+OpenLoops NLO calculation are also taken into account. Additional uncertainties are included to account for multiparton interaction and final state radiation $t\bar{t}+b\bar{b}$ production, not included in the Sherpa+OpenLoops calculation.

An uncertainty of 50% is assigned to the $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ components of the $t\bar{t}$+jets cross section, treated as uncorrelated, and is derived by comparing Powheg+Pythia with a NLO result based on Sherpa +OpenLoops. The uncertainty in the $t\bar{t}+b\bar{b}$ contribution represents the dominant systematic effect in this analysis.

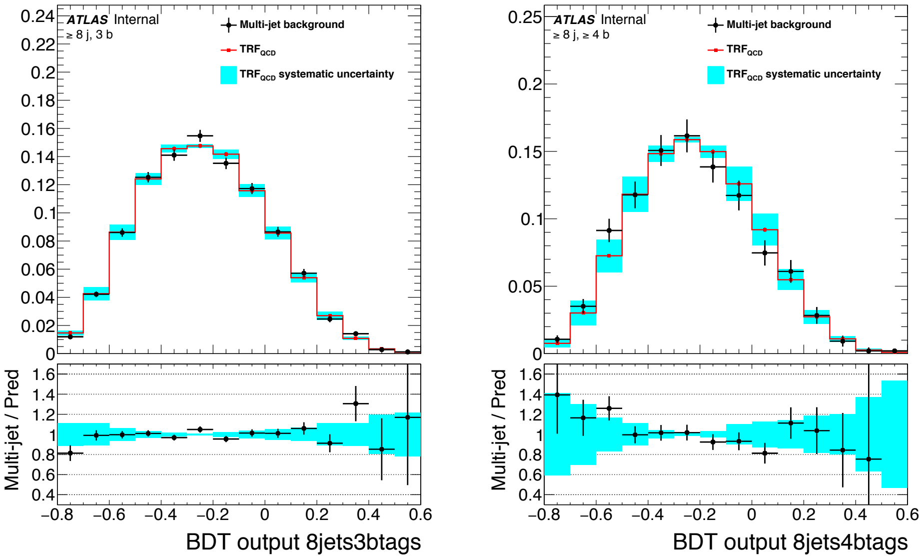


Figure 5.16: Shape systematic variations induced by different multijet parametrisations on the BDT output. Both figures represent the variations in the ≥ 8 jets for exactly 3 and ≥ 4 b-tagged jet regions, respectively [123].

5.7.3 Systematics for sub-leading backgrounds

The single top background represents a minor fraction ($\leq 1\%$) of the total background in the analysis regions. For this reason a less refined treatment of the uncertainties related to this small background has been adopted. An overall uncertainty of 7% is assumed for the theoretical cross sections [115, 116], computed from the theoretical uncertainty on the t- and Wt-channel production.

An uncertainty of 30% in the total cross section is assumed for $t\bar{t}V$ [121, 122]. Studies have been performed at truth level to test the modelling of this background by changing the parameters of the MadGraph+Pythia 6 generator like ISR, FSR, QCD emission scale, renormalisation and factorisation scale [46]. The variation of the QCD emission scale produces the largest shape differences, of the order of 10%. Thus all $t\bar{t}V$ samples are reweighted to account for QCD scale variations.

5.7.4 Multijet-related systematics

The multijet background is estimated using data in regions with exactly two b-tagged jets after subtraction of contributions from other events using MC simulation. All systematic uncertainties mentioned above are fully propagated to the data-driven multijet background estimation and treated in a correlated manner.

To estimate the uncertainties associated with the multijet background, the $\mathrm{TRF}_{\mathrm{MJ}}$ parameterisation is varied as a function of different sets of variables, which are sensitive to the amount and the mechanism of heavy-flavour production.

Figure 5.16 shows the shape variation induced on the BDT classifier by all the parametrisations in the ≥ 8 jets regions, for both 3 and ≥ 4 b-tagged jets.

It should be noted that the normalisation of the multijet background is evaluated independently in each of the six analysis regions through the fitting procedure, as motivated and described in Section 5.8.

Therefore the systematic uncertainties considered for the multijet background involve shape variations, only.

5.7.5 Signal systematics

The modelling uncertainties of $t\bar{t}H$ events are obtained by varying the factorisation and renormalisation scales in the PowHel generator. The nominal static scale of the process ($\mu_r = \mu_f = m_t + m_H/2$) is scaled by a factor two. A sample with a dynamic function $\mu_r = \mu_f = (m_t^{\mathrm{T}} m_{\bar{t}}^{\mathrm{T}} m_H^{\mathrm{T}})^{1/3}$ is also exploited since this functional form is the one used in the Pythia generator. Static scales are in principle more robust against divergences, whereas dynamic scales perform better at energies higher than the production threshold, allowing a better description of the process in that phase space. The nominal $t\bar{t}H$ (PowHel) is reweighed to reproduce the observed variations, as function of the top and $t\bar{t}H$ p_{T}.

The uncertainties related to the choice of the PDF are evaluated using the recommendations of the PDF4LHC [153]: the envelope of the differences among three PDF sets (MSTW2008, CTEQ6.6 and NNPDF2.0) and their internal uncertainties are computed.

The systematic uncertainties from the parton shower and fragmentation models are evaluated using PowHel+Herwig++ samples. Corrections to the PowHel+Herwig++ sample are considered in order to match the Higgs branching fraction in the NLO calculations used for the PowHel+Pythia 8 sample.

The uncertainty due to the choice of generator is evaluated by comparing PowHel+Pythia 8 with MadGraph5_aMC@NLO +Herwig++. The kinematic distributions of the $t\bar{t}$, $t\bar{t}H$ and Higgs boson systems show the largest differences, especially for the p_{T} distributions of the Higgs boson. Therefore a reweighing of this observables is used as estimation for this systematic uncertainty.

5.8 Results

The BDT discriminants for each of the six analysis regions are combined as inputs to a test statistic to search for the presence of a signal. A maximum likelihood fit, as described in Chapter 4, is performed: it allows the impact of systematic uncertainties on the search sensitivity to be reduced, by taking advantage of the highly populated background-dominated control regions included in the likelihood fit.

To obtain the final result, a simultaneous fit to the data is performed on the distributions of the discriminants in six regions: four analysis regions and two control regions. The fit is performed under the signal-plus-background hypothesis to obtain the best value of the signal strength μ, which is a free parameter in the fit. The normalisation of each component of the background and μ are determined simultaneously from the fit. Contributions from $t\bar{t}$+jets, $t\bar{t}V$ and single top backgrounds are constrained by the uncertainties of the respective theoretical calculations, the uncertainty in the luminosity, and experimental data. The multijet background normalisations, instead, are free parameters in the fit and are independent in each region.

First, a blind fit is performed in order to determine whether the nuisance parameters are consistently describing the uncertainties in all the independent regions. The blinding procedure is the same adopted for the validation of the input variables and described in Section 5.6.4: these regions are excluded in order to test the description of the systematic uncertainties with a sufficient amount of data. This is a necessary step for the validation of the multijet background predictions with the $\mathrm{TRF}_{\mathrm{MJ}}$ method. Once all closure tests behave as expected and the robustness of the analysis is proven, the complete dataset is used and final results are obtained.

5.8.1 Expected performance of the fit

The expected performance of the fit has been studied using the so-called Asimov dataset, described in Section 4.3.5. As the expected nuisance parameters corresponding to systematic uncertainties are all centred on zero and the normalisation scale factors are all centred around 1. The majority of the parameters, especially the ones related to the detector performance, are not constrained with respect to their prior uncertainties. The constraints on the parameters related to the shape of the multijet background is expected, given that this process represents by far the largest contribution in each analysis region and the considered variations have been defined in a conservative way. The normalisation of the multijet background in each region can be known from the fit to a few percent level, as a result of the good separation achieved by the BDT discriminants. Finally, a minor constraint of the $t\bar{t} + b\bar{b}$ background is also expected given its relatively large contribution to the analysis regions with more than seven jets.

For what concerns the analysis sensitivity, the leading sources of systematic uncertainties are the ones which show a high level of correlation with the signal strength. These are the uncertainty on the normalisation of the $t\bar{t} + b\bar{b}$ background as well as its shape, and the uncertainty on the multijet process normalisation. This is also confirmed by the ranking plot on Figure 5.17: the parameters with the largest impact on the extracted signal strength are the normalisation of the multijet background in some of the analysis regions, followed by the uncertainty on the $t\bar{t} + b\bar{b}$ process. This effect is calculated by fixing the corresponding nuisance parameter at $\hat{\theta} \pm \sigma_\theta$ and performing the fit again. Here $\hat{\theta}$ is the fitted value of the nuisance parameter and σ_θ is its post-fit uncertainty. The difference between the default and the modified μ, $\Delta\mu$, represents the effect on μ of this particular systematic uncertainty.

In the presence of a signal compatible with the one predicted by the SM, the expected error on the signal strength is 2.79 in case of all sources of systematics are considered, and 0.92 in case only the data statistical component and the template statistics are taken into account.

The expected signal significance is 0.37, in case of a SM signal, which corresponds to a p-value of 34%, while the expected (median, for the background-only hypothesis) 95% CL upper limits are 5.42 and 1.73 when all systematic uncertainties are included or not (“statistics only”), respectively.

5.8.2 Fit to data in the analysis regions

The yields in the different analysis regions considered in the analysis after the fit (post-fit) are summarised in Table 5.7. In each region, the variation of background and signal events with respect to the pre-fit values, shown in Table 5.3, are modest and, in particular, the fitted multijet background component is well constrained by the fit within an uncertainty of 8%.

The results of the final fit under the signal-plus-background hypothesis is shown in Figures 5.18 and 5.19. Figures 5.20 and 5.21 show the BDT output distributions for data and the predictions in each analysis region, both before (left) and after (right panels) the fit to data. As it can be seen, the fit significantly improves the agreement, especially in regions with 4 b-tags, mainly thanks to the adjustment of the multijet normalisation scale factors. The relative uncertainties decrease significantly in all regions due to the constraints provided by the data, exploiting the correlations between the uncertainties in the different analysis regions. Appendix B presents the pre- and post-fit plots of the BDT input variables in all the analysis regions they are used in.

The signal strength in the all-hadronic $t\bar{t}H$ decay mode, for $m_H = 125$ GeV, is measured to be:

$$\mu(m_H = 125\ \text{GeV}) = 1.6 \pm 2.6. \tag{5.5}$$

The observed (expected) significance of the signal is 0.6 (0.4) standard deviations, corresponding to an observed (expected) p-value of 27% (34%). The observed and expected limits are summarised in

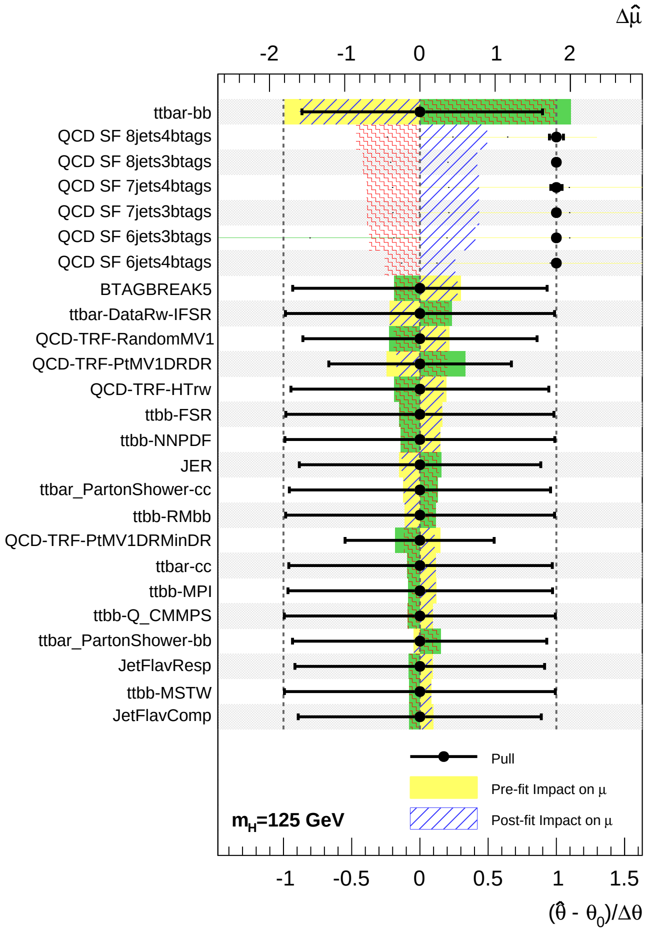


Figure 5.17: Ranking of nuisance parameters obtained from the fit to the Asimov dataset, corresponding to the sources of systematic uncertainty with the largest impact on the fitted signal strength μ. The points, which are drawn conforming to the scale of the bottom axis, show the deviation of each of the fitted nuisance parameters $\hat{\theta}$ from θ_0, which is the nominal value of that nuisance parameter, in units of the pre-fit standard deviation $\Delta\theta$. The plain yellow area represents the pre-fit impact on μ and the hashed blue area its post-fit impact. The error bars show the post-fit uncertainties σ_θ, which have a size close to one if the data do not provide any further constraint on that uncertainty. Conversely, an error bar for σ_θ smaller than one indicates a reduction with respect to the original uncertainty. The nuisance parameters are sorted according to their post-fit impact $\Delta\theta$ (top horizontal scale). Multijet scale factors (QCD SF) show the fitted values and uncertainties of the normalisation parameters that are freely floating in the fit. These normalisation parameters have a pre-fit value of unity [123].

Table 5.7: Event yields from simulated backgrounds and the signal, as well as measured events in each of the analysis regions after the fit. The quoted uncertainties include statistical and systematic effects. The sum of all contributions may slightly differ from the total value due to rounding. The *tH* background is not shown as fewer than 1.5 events in each region are predicted.

	6j, 3b	6j, ≥4b	7j, 3b	7j, ≥4b	≥8j, 3b	≥8j, ≥4b
Multijet	15940 ± 320	1423 ± 66	12060 ± 350	1233 ± 78	10020 ± 490	1280 ± 100
$t\bar{t}$+light	1750 ± 270	55 ± 13	1650 ± 340	54 ± 15	1550 ± 450	54 ± 21
$t\bar{t} + c\bar{c}$	350 ± 170	22 ± 11	490 ± 240	28 ± 14	750 ± 360	66 ± 33
$t\bar{t} + b\bar{b}$	230 ± 120	31 ± 17	350 ± 190	63 ± 34	560 ± 320	139 ± 75
$t\bar{t} + V$	15.0 ± 6.2	1.9 ± 1.5	23.3 ± 8.9	3.6 ± 2.2	43 ± 15	8.7 ± 4.2
Single top	184 ± 59	6.7 ± 3.6	153 ± 52	9.4 ± 4.4	123 ± 48	11.8 ± 5.8
Total background	18470 ± 320	1539 ± 58	14720 ± 320	1391 ± 69	13030 ± 340	1561 ± 63
$t\bar{t}H$ (m_H=125 GeV)	23.4 ± 6.3	5.6 ± 2.8	39.1 ± 8.9	11.9 ± 4.5	71 ± 15	28.8 ± 8.5
Data events	18508	1545	14741	1402	13131	1587

Table 5.8. A $t\bar{t}H$ signal 6.4 times larger than predicted by the SM is excluded at 95% CL. A signal 5.4 times larger than the signal of a SM Higgs boson is expected to be excluded for the background-only hypothesis.

Figure 5.22 summarises the post-fit event yields for data, total background and signal expectations as a function of $\log_{10}(S/B)$. The signal is normalised to the fitted value of the signal strength ($\mu = 1.6$). The

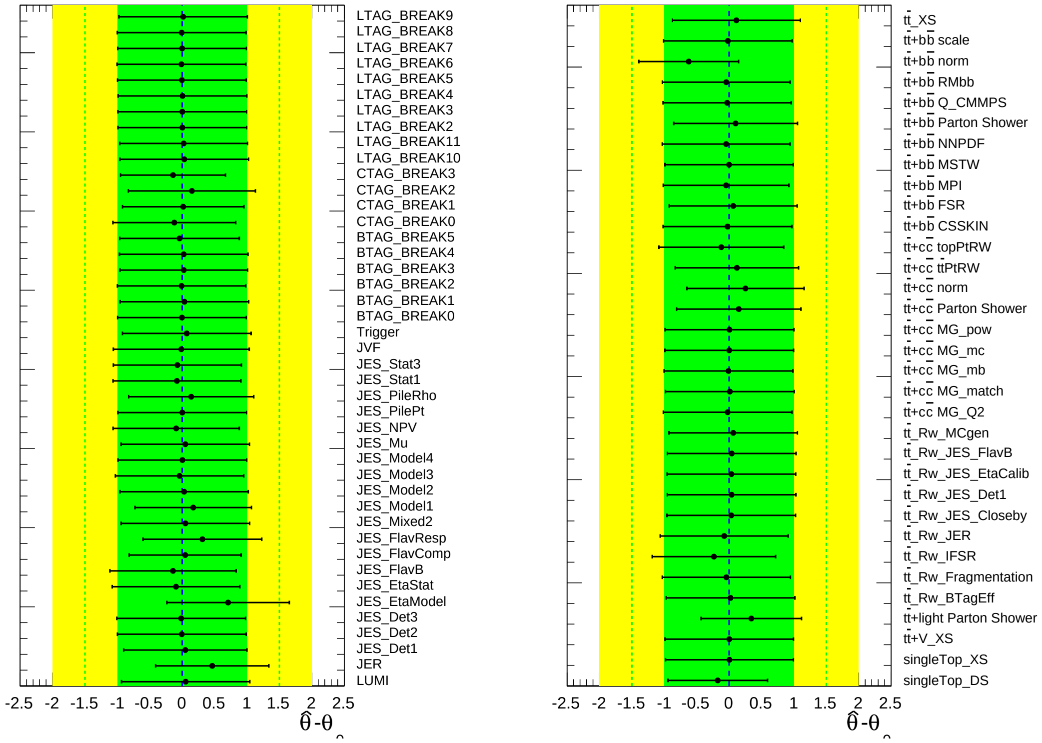


Figure 5.18: Fitted nuisance parameters and normalisation scale factors to the data under the signal-plus-background hypothesis (assuming $m_H = 125$ GeV), for (a) detector-related uncertainties and (b) uncertainties related to simulated background modelling and normalisation [123].

Table 5.8: Observed and expected upper limits at 95% CL on $\sigma(t\bar{t}H)$ relative to the SM prediction assuming $m_H = 125$ GeV, for the background-only hypothesis. Confidence intervals around the expected limits under the background-only hypothesis are also provided, denoted by $\pm 1\sigma$ and $\pm 2\sigma$, respectively. The expected (median) upper limit at 95% CL assuming the SM prediction for $\sigma(t\bar{t}H)$ is shown in the last column.

	Observed	Expected if $\mu = 0$					Expected if $\mu = 1$
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Median
Upper limit on μ at 95%	6.4	2.9	3.9	5.4	7.5	10.1	6.4

effect of a signal strength 6.4 times larger than predicted by the SM is also illustrated.

Figures 5.23 show a summary of the 95% CL upper limits of $\sigma(t\bar{t}H)$ relative to the SM prediction and the observed signal strength of this analysis compared with the other $t\bar{t}H(H \to b\bar{b})$ channels.

Figure 5.24 shows the effect of the major systematic uncertainties on the fitted value of μ and the constraints provided by the data. The ranking, from top to bottom, is determined by the post-fit impact on μ. This is also shown in Table 5.9.

The largest systematic effect arises from the uncertainty in the normalisation of the irreducible $t\bar{t} + b\bar{b}$ background. The $t\bar{t} + b\bar{b}$ background normalisation is smaller by 30% in the fit than in the prediction, resulting in a decrease of the observed $t\bar{t} + b\bar{b}$ yield with respect to the Powheg+Pythia prediction. The second largest effect comes from the multijet background normalisation. The data-driven method focuses on modelling the shape of the multijet background, while the normalisation is constrained by the regions dominated by the multijet background. The uncertainty in the normalisation parameters amounts to a few

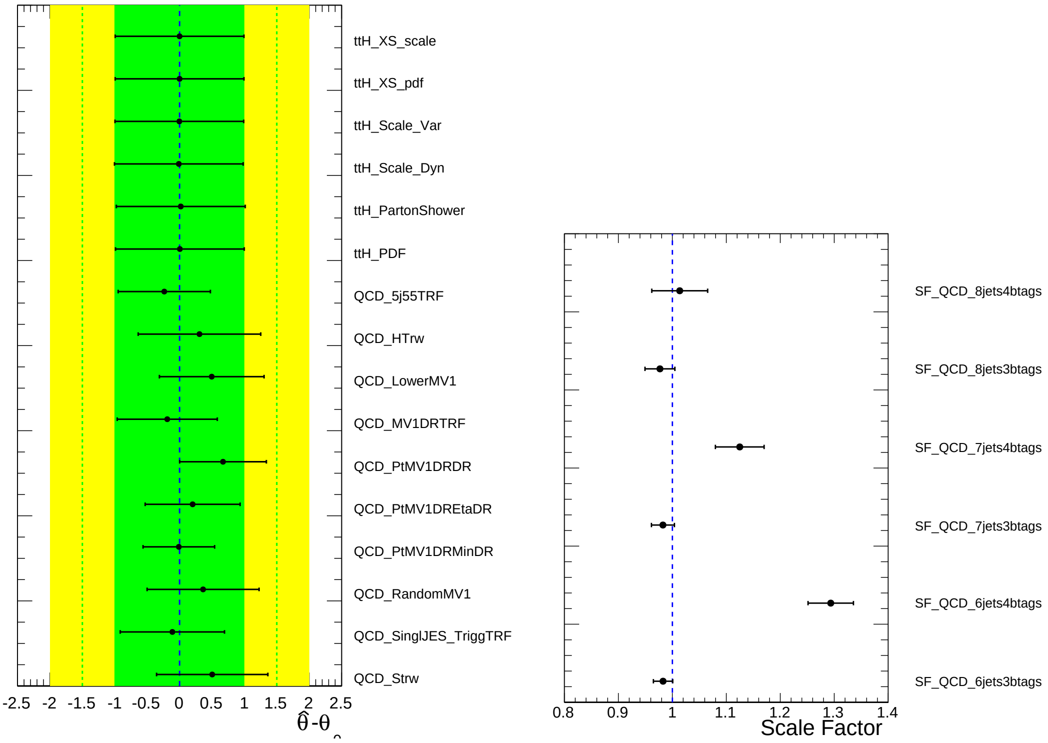
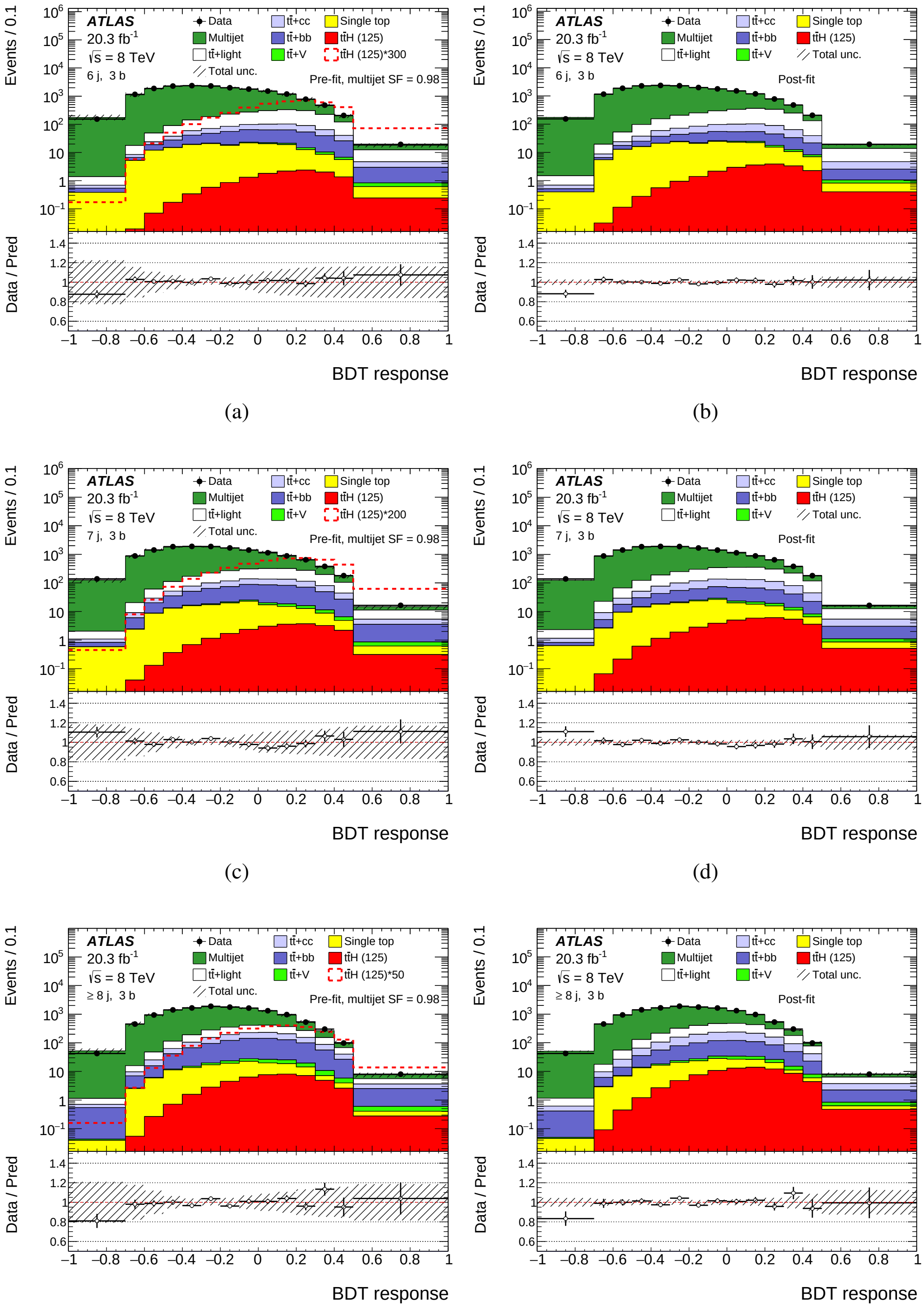


Figure 5.19: Fitted nuisance parameters and normalisation scale factors to the data under the signal-plus-background hypothesis (assuming $m_H = 125$ GeV), for (a) uncertainties related to the multijet background and (b) multijet normalisation scale factors (SF_QCD). The fitted values are within the a priori expected variations of the various TRF parametrisations [123].

Table 5.9: Effect of the different sources of systematic uncertainties on the signal strength μ, expressed in terms of percentage of the fitted value of μ sorted according to their post-fit effect [104].

Source of systematic uncertainty	$\pm1\sigma$ post-fit impact on μ
$t\bar{t}$ normalisation	108%
Multijet normalisation	71%
Multijet shape	60%
Main contributions from $t\bar{t}$ modelling	34%–41%
Flavour tagging	31%
Jet energy scale	27%
Signal modelling	22%
Luminosity+trigger+JVF+JER	18%

percent and the values from each region are consistent with the variations applied to these parameters to account for systematic uncertainties. Two of the multijet background shape uncertainties are ranked fourth and fifth, and their pulls are slightly positive. Other important uncertainties include b-tagging and JES. Uncertainties arising from jet energy resolution, jet vertex fraction, jet reconstruction and JES that affect primarily low-p_{T} jets, as well as the $t\bar{t}$ + light background modelling uncertainties, do not have a significant impact on the result.



(a) (b)

(c) (d)

(e) (f)

Figure 5.20: Comparison between data and prediction for the BDT discriminant in the, from top to bottom, (6–8 j, 3b) regions before (left) and after (right) the fit. The fit is performed under the signal-plus-background hypothesis. Pre-fit plots show an overlay of the multijet distribution normalised to data for illustration purposes only. The bottom panels display the ratios of data to the total prediction. The hashed areas represent the total uncertainty in the background predictions, containing the statistical and systematic contribution. The binning is the same as that used in the fit [104].

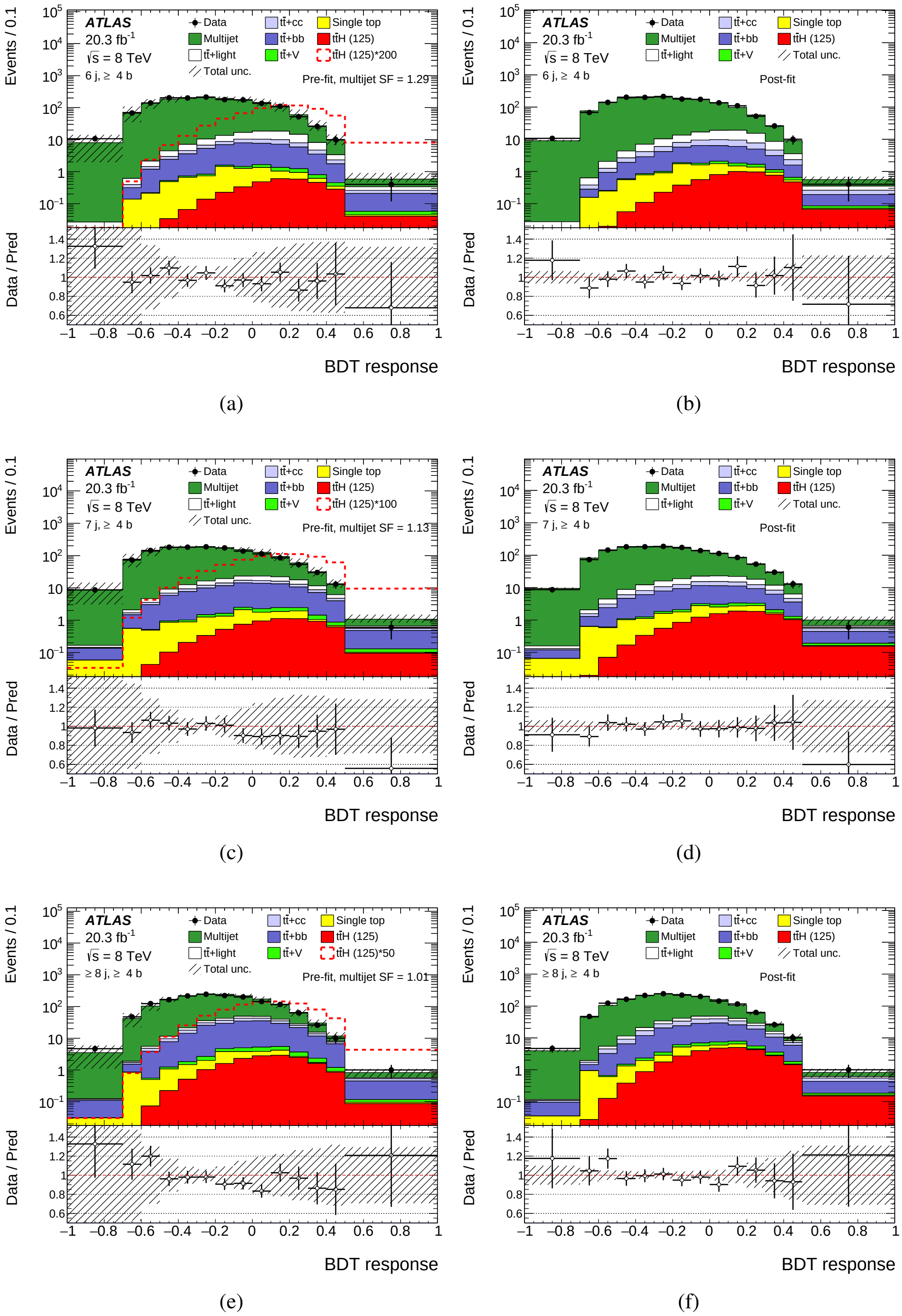


Figure 5.21: Comparison between data and prediction for the BDT discriminant in the, from top to bottom, (6–8j, $\geq 4b$) regions before (left) and after (right) the fit. The fit is performed under the signal-plus-background hypothesis. Pre-fit plots show an overlay of the multijet distribution normalised to data for illustration purposes only. The bottom panels display the ratios of data to the total prediction. The hashed areas represent the total uncertainty in the background predictions, containing the statistical and systematic contribution. The binning is the same as that used in the fit [104].

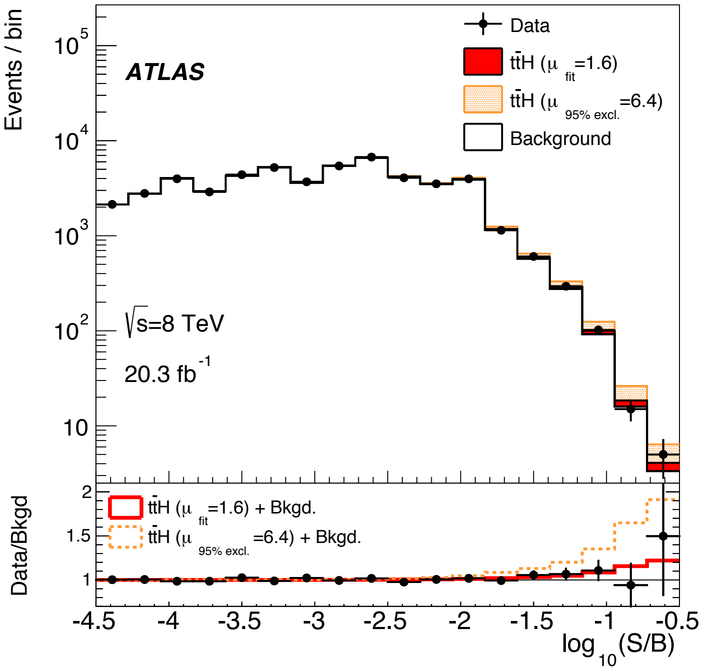
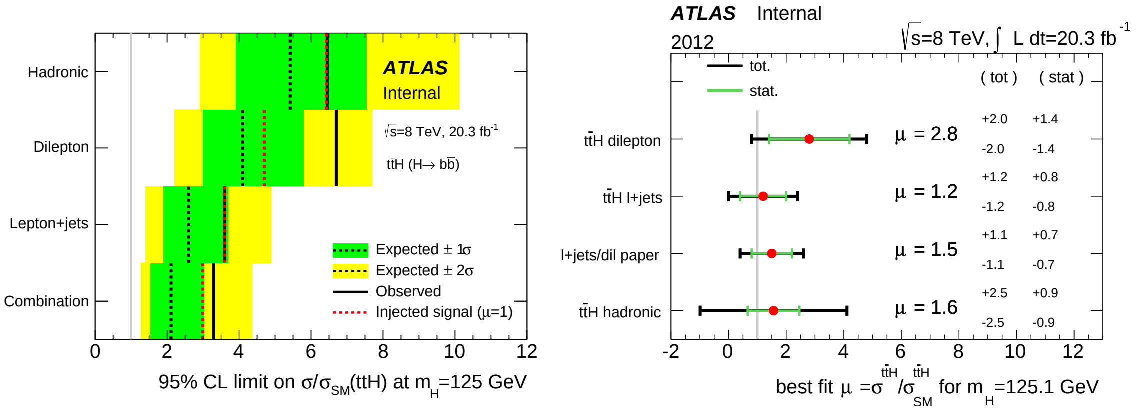


Figure 5.22: Event yields as a function of $\log_{10}(S/B)$, where S (expected signal yield) and B (expected background yield) are taken from the corresponding BDT discriminant bin. Events from all fitted regions are included. The predicted background is obtained from the global signal-plus-background fit. The $t\bar{t}H$ signal is shown both for the best-fit value ($\mu = 1.6$) and for the upper limit at 95% CL ($\mu = 6.4$) [104].



(a) (b)

Figure 5.23: Summary of (a) the 95% CL upper limit on $\sigma(t\bar{t}H)$ relative to the SM prediction and (b) the signal strength measurement for the $t\bar{t}H(b\bar{b})$ channels [104].

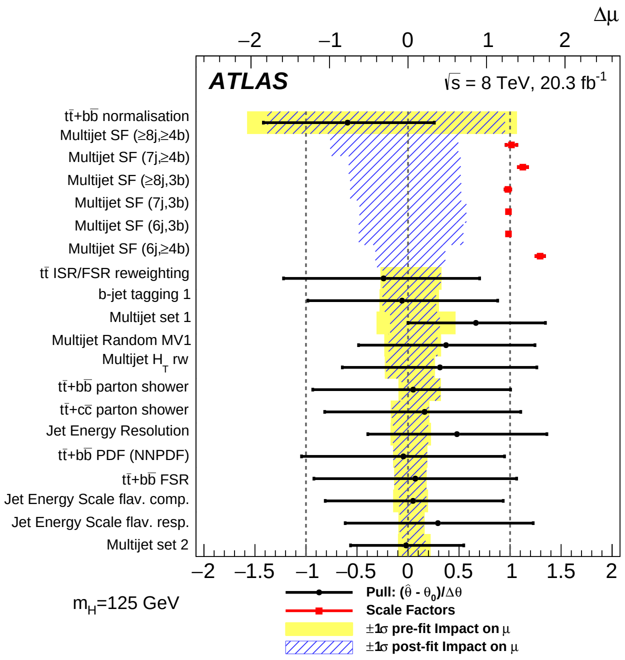


Figure 5.24: Ranking of nuisance parameters obtained from the fit to data [104].

CHAPTER 6

Search for the $t\bar{t}H$ vertex via the three leptons final state

The description of the search of the Higgs boson production in association with a top anti-top quark pair in the multilepton channel is provided in this Chapter. The multilepton $t\bar{t}H$ search described here requires exactly three light leptons in the final state, allowing to have a good compromise between signal purity and branching ratio. The multilepton channel has a larger sensitivity to the top Yukawa coupling compared to the fully hadronic but the analysis follows an analogous strategy. A remark in this context is in order: this is the first multilepton $t\bar{t}H$ search that exploits a MVA technique to assess the top Yukawa coupling, employing data collected by ATLAS at LHC. The search uses data corresponding to an integrated luminosity of 36.5 fb^{-1}at a centre-of-mass energy of 13 TeV, recorded in Run 2.

After a brief introduction in Section 6.1 about the advantage of such multilepton signature, Section 6.2 contains a description of the dataset and the MC simulations used. The definition of the physics objects, the event selection and the background composition are studied in Sections 6.3, 6.4 and 6.5, respectively: the main source of background is coming from "non-prompt" leptons, namely, leptons coming from semileptonic hadron decay or jets misidentified as leptons (also called "fakes"); it results to be the most challenging background to model for several three leptons analyses.

Section 6.6 introduces the BDT-based analysis: the set of most discriminating variables is determined according to an optimisation procedure to attain the best BDT classifier, which is eventually deployed for the profile likelihood fit. The full set of systematic uncertainties considered in the likelihood is presented in Section 6.7. In Section 6.8, the top Yukawa coupling is estimated and the 95% C.L. upper limit on the signal strength μ is provided. Finally, some prospects at higher integrated luminosities are presented and expected sensitivity estimates are given in Section 6.9.

6.1 The multilepton channel

As already described in Section 2.7, the measurement of the top Yukawa coupling y_t is a key parameter of the Standard Model. The combination of ATLAS and CMS results in Run 1 did not give a conclusive answer concerning possible anomalous deviations of this parameter from the SM prediction: the combined fit result yielded $\mu_{t\bar{t}H} = \sigma/\sigma_{\text{SM}} = 2.3^{+0.7}_{-0.6}$, where the excess was primarily driven by multileptonic final states [50]. Therefore, the multilepton measurement in Run 2 could play a crucial role to finally prove or disprove a possible anomalous excess of y_t.

Multilepton signatures are characterised by the presence of multiple leptons and of jets. The excellent performance of the ATLAS detector in reconstructing and triggering light leptons allow to have a clear signature and easily reject several background processes, unlike in the case of the fully hadronic channel.

The multilepton $t\bar{t}H$ channel has a lower branching ratio than the fully hadronic channel: it mostly targets the $H \to WW^*$, ZZ^* and $\tau\tau$ decay modes, while the $t\bar{t}H$ fully hadronic channel benefits from the large branching fractions of $H \to b\bar{b}$. As a result, the multilepton analysis has a good signal purity, to the detriment of the number of collected events available for this search. Albeit substantial dissimilarities in terms of selection and background composition, multilepton and fully hadronic searches follow the same strategy benchmarks, described in Chapter 4.

As the centre-of-mass energy changed from 8 to 13 TeV in 2015, the cross section for $t\bar{t}H$ production raised by a factor of ~ 3.9 [32, 34]: higher-energy searches can, therefore, reach comparable sensitivity to existing results with significantly less integrated luminosity. Moreover, towards Run 2, the performances of the pixel tracker, the muon detectors and calorimeters have been improved; more accurate Monte Carlo simulations have been adopted and new calibrations of physics objects have been performed, leading to substantial changes in the treatment of systematics related to the detector performance and calibration. Finally, improvements in the level of accuracy of calculations for some physical processes like the $t\bar{t}H$, $t\bar{t}V$ ($V = W, Z$) and diboson have been reached recently [34], giving a better precision on the predicted cross sections in comparison with 8 TeV analyses, like the fully hadronic presented in Chapter 5. In the following Sections it will be mentioned whenever the cross section precision or the treatment of simulations and systematic uncertainties changed between the two LHC Runs.

6.2 Dataset and Monte Carlo event generation

6.2.1 Dataset

The dataset used in this analysis has been recorded by the ATLAS experiment during 2015 and 2016 at $\sqrt{s} = 13$ TeV. The peak instantaneous luminosity achieved during that period was $13.7{\times}10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, at a bunch crossing every 25 ns. The total luminosity delivered by the LHC was 43.1 fb^{-1}, while the total luminosity recorded by ATLAS was 39.9 fb^{-1}.

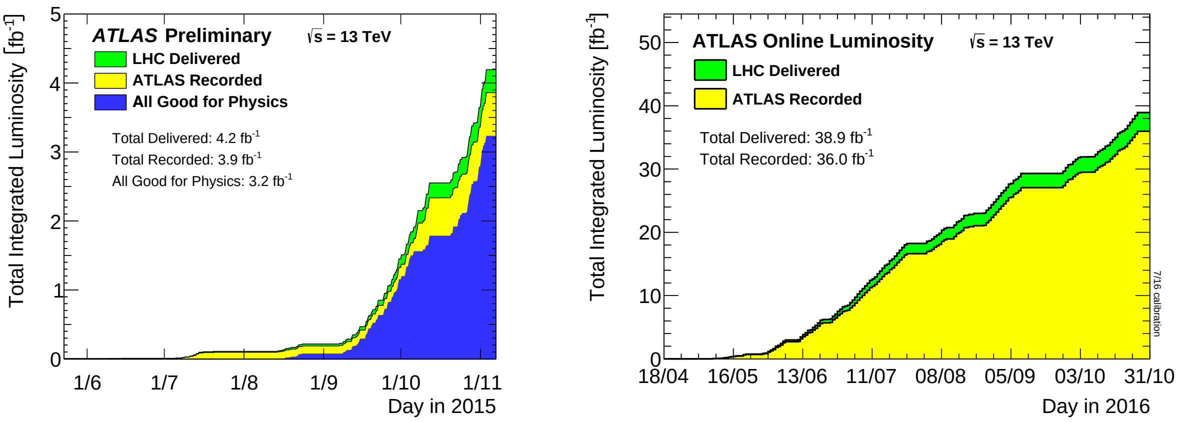
The data with IBL on, and verifying data quality cuts (GRL) is used in this analysis. The events passing the GRL filter, shown in blue in Figure 6.1, are considered good quality events (“good for physics”) and used in ATLAS analyses, corresponding to a total integrated luminosity of 36.5 fb^{-1}(3.2 fb^{-1} for 2015, and 33.2 fb^{-1} for 2016), with the current calibration.

6.2.2 Signal and background modelling

The dominant background contribution to the multilepton $t\bar{t}H$ signal comes from different sources: diboson processes, $t\bar{t}$+jets, and the associated production of a vector boson with a $t\bar{t}$ pair, $t\bar{t}V$. A small fraction comes from the production of a single top quark, Z+jet and rarer processes ($t\bar{t}\,t\bar{t}$, $t\bar{t}WW$, tZ and tH). All contributions are estimated using MC simulations: a detailed description of the method adopted to control the normalisations of the dominant backgrounds is given in Section 6.5.

The event generator programs and configurations used for simulating the signal and background processes are shown in Table 6.1. In addition, detailed descriptions of the generator configurations may be found in References [155–158].

The production of $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$ is simulated with a NLO QCD matrix-element computed by MadGraph5_aMC@NLO, matched to the Pythia 8 parton shower generator. In the case of $t\bar{t}Z$, the inclusive $t\bar{t}\ell^+\ell^-$ matrix-element is computed, including off-shell Z and γ^* contributions with $m_{\ell^+\ell^-} >$ 5 GeV. For studies of systematic variations, samples with variations of the QCD factorisation and



(a) (b)

Figure 6.1: Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in (a) 2015 and (b) 2016 [61].

renormalisation scales, different parton shower or matrix-element generator are used. More details are given in Section 6.7.4.

The Higgs boson production cross section, branching ratios and their uncertainties are taken from the CERN Yellow Report 4, prepared by the LHC Higgs Cross Section Working Group [33–35, 113]. The overall $t\bar{t}H$ production cross section, 507.1 fb, is computed at NLO in QCD and electroweak couplings [159–165] as compiled in References [32, 34] and has uncertainties of $^{+5.8\%}_{-9.2\%}$ from the QCD renormalisation/factorisation scale choice and 3.6% from PDF uncertainties (including α_s uncertainties), as shown in Table 6.2.

The cross sections for $t\bar{t}V$ production, including the process $pp \to t\bar{t}\ell^+\ell^-$ over the full Z/γ^* mass spectrum, are computed at NLO in QCD and electroweak couplings using the configurations of References [85, 165]. These have QCD scale uncertainties of $\approx$ 12% and PDF+α_s uncertainties of 3–4%, and total cross sections of 123.7 fb and 566.2 fb for $t\bar{t}\ell^+\ell^-$ (with $m_{\ell^+\ell^-} > 5$ GeV) and $t\bar{t}W^\pm$, respectively.

$t\bar{t}$+jets background events are generated with Powheg v2.0 and interfaced with Pythia 6 for the parton showering and fragmentation. The Perugia 2012 tune with the CTEQ6L PDF set is used for the underlying event description. A filter requiring at least two leptons is included in the sample production, in order to obtain a large amount of simulated events in the interesting region. Analogously to the fully hadronic analysis, the $t\bar{t}$+jets sample is normalised to the theoretical calculation performed at NNLO in QCD, as calculated with top++2.0 and includes resummation of NNLL soft-gluon terms [11–16]. Table 6.2 shows the accounted uncertainties from PDF variation (4.2%) and from QCD renormalisation and factorisation scale choices ($^{+2.4\%}_{-3.5\%}$). The total $t\bar{t}$ cross section is 831.76 pb at $\sqrt{s} = 13$ TeV.

The Powheg+Pythia 6 simulation is also used to model other top backgrounds such as single top t-channel, s-channel and Wt. They are normalised to the approximate NNLO theoretical cross sections [115, 116] using the MSTW2008 NNLO PDF set [117, 118]. Overlaps between the $t\bar{t}$ and Wt final states are removed [119] by the diagram removal method [120].

A $t\bar{t} + \gamma$ sample is used in addition to the $t\bar{t}$+jets to improve the modelling of the $t\bar{t}$ background with a hard scattered photon. The sample is generated with MadGraph5_aMC@NLO interfaced to the Pythia 8 parton shower: the matrix-element is generated at LO including the decays of the top quarks. The photons can be emitted either in the production or in the decay stage. The renormalisation and factorisation scales are the same of the $t\bar{t}$+jets sample. In order to avoid infrared and collinear singularities, a set of kinematic requirements are applied ($p_{\mathrm{T}}(\gamma) > 10$ GeV, $|\eta(\gamma)| < 5$ and ΔR between the photon and any charged particle below 0.2). The cross section is normalised to the NLO reference calculation [166]. The partial overlap between the $t\bar{t}$+jets sample and $t\bar{t} + \gamma$ is removed by subtracting from the $t\bar{t}$+jets process the component with a hard photon ($p_{\mathrm{T}} > 15$ GeV) produced at the matrix-element level; then, $t\bar{t} + \gamma$ is incorporated in the $t\bar{t}$+jets background.

Diboson processes are generated with Sherpa 2.1 at LO, using the CT10 PDF set. For the fully leptonic diboson decays (including τ leptons) the matrix-elements consider the production with up to three additional partons. A generator level cut of 5 GeV on the transverse momentum of the two highest p_{T} leptons is also imposed for all samples and any opposite-sign dilepton invariant mass is required to be larger than 5 GeV. Diboson processes can be grouped into 4ℓ, $3\ell\nu$, $2\ell2\nu$ and $2\ell qq$, where $\ell^+\ell^-$ refers to the full Z/γ^* mass spectrum. Out of these, only the 4ℓ and $3\ell\nu$ processes were considered for the generation of samples, in order to increase the number of MC events in the signal region. The sample is not reweighted to match higher order predictions: it is scaled by 0.91, correcting the α_{QED} and electroweak parameters difference between two generators, Powheg being better than Sherpa in predicting data. The remaining difference is due to the multijet merging where the more complete calculation in Sherpa describes the process better than Powheg: Sherpa is more suitable for final states with high jets activity, the reason why it is used as baseline.

The Z+jet background events are simulated using the Sherpa 2.2.1 generator, with matrix-elements calculated for $\ell^+\ell^-$ with 0, 1 or 2 additional partons at NLO and $\ell^+\ell^-$ with 3 or 4 additional partons at LO accuracy. These matrix-elements are merged with the Sherpa parton shower using the MEPS@NLO prescription [167]. The NNPDF 3.0 NLO PDF set is used in conjunction with a dedicated parton shower tuning developed in Sherpa. The sample is reweighted to match the cross section predictions at NNLO: a global 5% uncertainty is assigned to the total Z+jet inclusive cross section. Alternative Z+jet samples have also been generated with either MadGraph+Pythia 8, Powheg+Pythia 8 or Sherpa 2.1, and are used as backup for background modelling validation and to compute systematic uncertainties (shower, hadronisation, generator variations).

Rare backgrounds, such as the associated production of single top quark and Higgs boson, $tHqb$ and tWH (referred to as tH), are also simulated and the cross sections are computed using the MadGraph5_aMC@NLO generator [85] at NLO in QCD.

Table 6.2 summarises the cross section values and their uncertainties for the signal and for the background processes considered at the state-of-the-art precision.

6.2.3 Common treatment of MC samples

Additional minimum-bias pp collisions (pileup) generated with Pythia 8 (MSTW2008LO PDF set [117] and A2 tune [175]) were overlaid to model the effects of both in- and out-of-time pileup, from additional pp collisions in the same and nearby bunch crossings. The pileup distribution is reweighted to reflect the luminosity profile of the recorded data. Leading-logarithm photon emission, either modelled by the parton shower generator or by PHOTOS [131], is included in all generated events. Samples using Pythia 6 and Pythia 8 have heavy flavour hadron decays modelled by EvtGen 1.2.0 [136]. All samples are then passed through the full Geant 4 [134] simulation of the ATLAS detector.

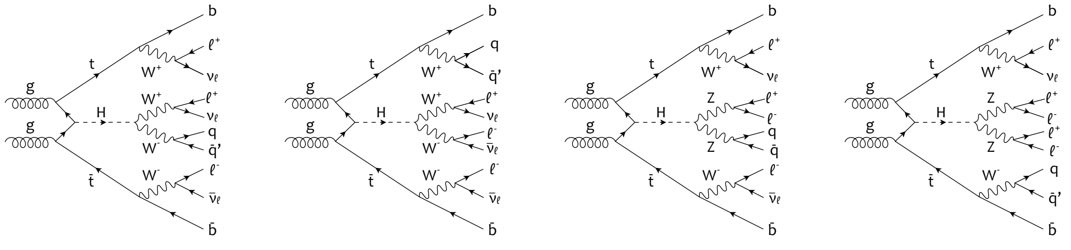
All simulated events were processed using the same reconstruction algorithms and analysis chain as the data. Simulated events are corrected so that the object reconstruction and identification efficiencies, energy scales and energy resolutions match those determined from data, as described in Chapter 3.

Table 6.1: Configurations used for the event generation of signal and background processes. If only one parton distribution function (PDF) is shown, the same one is used for both the matrix-element (ME) and parton shower generators; if two are shown, the first is used for the matrix-element calculation and the second for the parton shower. “V” refers to the production of an electroweak boson (W or Z/γ^*). “Tune” refers to the underlying-event tune of the parton-shower generator.

Process	ME Generator	Parton Shower	PDF	Tune
$t\bar{t}H$	MadGraph5_aMC@NLO [85]	Pythia 8 [110]	NNPDF 3.0 NLO [168]/ NNPDF 2.3 LO [170]	A14 [169]
$tHqb$	MadGraph5_aMC@NLO	Pythia 8	CT10 [109]/ NNPDF 2.3 LO	A14
tHW	MadGraph5_aMC@NLO	Herwig++ [139]	CT10/CTEQ6L1 [137, 138]	UE-EE-5 [171]
$t\bar{t}W$	MadGraph5_aMC@NLO	Pythia 8	NNPDF 3.0 NLO/ NNPDF 2.3 LO	A14
$t\bar{t}(Z/\gamma^*)$	MadGraph5_aMC@NLO	Pythia 8	NNPDF 3.0 NLO/ NNPDF 2.3 LO	A14
$t(Z/\gamma^*)$	MadGraph5_aMC@NLO	Pythia 6 [172]	CTEQ6L1	Perugia2012 [141]
$tW(Z/\gamma^*)$	MadGraph5_aMC@NLO	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}\,t\bar{t}$	MadGraph5_aMC@NLO	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}WW$	MadGraph5_aMC@NLO	Pythia 8	NNPDF 2.3 LO	A14
$t\bar{t}$	Powheg -BOX [140]	Pythia 6	CT10/CTEQ6L1	Perugia2012
$t\bar{t}+\gamma$	MadGraph5_aMC@NLO	Pythia 8	NNPDF 2.3 LO	A14
s-, t-channel, Wt single top	Powheg -BOX [120, 142]	Pythia 6	CT10/CTEQ6L1	Perugia2012
VV, $qqVV$, VVV	Sherpa 2.1.1 [89]	Sherpa	CT10	Sherpa default
$Z \to \ell^+\ell^-$	Sherpa 2.2	Sherpa	NNPDF 3.0 NLO	Sherpa default
$W \to \ell\nu$	Sherpa 2.1.1	Sherpa	CT10	Sherpa default

6.3 Object and event preselection

The $t\bar{t}H \to 3\ell$ final state is composed of exactly three leptons, originating from both top and Higgs decay products, jets from (u, d, s)-quarks or gluons (namely, light jets) and jets from c- or b-quarks (heavy-flavour jets). Several three-lepton configurations are thinkable, according to the Higgs (WW, ZZ, $\tau\tau$) and top-quark decay modes. Feynman diagrams in Figure 6.2 illustrates some possible final states.



(a) (b) (c) (d)

Figure 6.2: Feynman diagrams for some multilepton configurations of the $t\bar{t}H$ process: (a-b) the $H \to WW^*$ and (c-d) the $H \to ZZ^*$ decay modes. In (d) 5 leptons are produced, but one lepton can be an hadronically decaying τ and a second one can fail the reconstruction due to detector acceptance or reconstruction inefficiencies.

Table 6.2: Production cross sections for signal $t\bar{t}H$, at $m_H = 125$ GeV, and various simulated background processes at $\sqrt{s} = 13$ TeV. The uncertainties quoted are variations of the renormalisation and factorisation scales (first uncertainty) and uncertainties in the parton distribution functions (second uncertainty). The $t\bar{t}$ uncertainties refer instead to the renormalisation and factorisation scales (first) and uncertainties in the parton distribution functions (second) and top mass uncertainty (third). The $t\bar{t}H$ cross section used is reported in the CERN Yellow Report [34] at NLO QCD with the zero-width-approximation. All $t\bar{t}V$ processes are calculated in the five-flavour scheme and numbers are taken from Reference [85]. The tZ and $t\bar{t}WW$ processes are calculated in the four-flavour scheme and numbers are taken from Reference [85].

Process	σ [pb]	State-of-the-art precision
$t\bar{t}H$	$0.509^{+5.7\%+8.8\%}_{-9.3\%-8.8\%}$	NLO
$tHqb$	$0.0743^{+6.5\%\ +3.6\%}_{-14.9\%-3.6\%}$	NLO
tHW	$0.022^{+20\%+20\%}_{-20\%-20\%}$	NLO
$t\bar{t}W$	$0.566^{+11.2\%\ +1.7\%}_{-10.6\%\ -1.3\%}$	NLO
$t\bar{t}(Z/\gamma^*)$	$0.124^{+9.7\%\ \ +1.9\%}_{-11.1\%\ -2.2\%}$	NLO
$tbjZ$ t-channel	$0.240^{+7.7\%\ +0.9\%}_{-7.9\%\ -0.9\%}$	NLO
$t\bar{t}\,t\bar{t}$	$0.009^{+30.8\%\ +5.5\%}_{-25.6\%\ -5.9\%}$	NLO
$t\bar{t}WW$	$0.010^{+10.9\%\ +2.1\%}_{-11.8\%\ -2.1\%}$	NLO
$t\bar{t}$ inclusive	$831.8^{+2.4\%\ +4.2\%\ +2.7\%}_{-3.5\%\ -4.2\%\ -2.8\%}$	NNLO+NNLL
single top Wt-channel	$71.7^{+2.5\%\ +4.7\%}_{-2.5\%\ -4.7\%}$	aNNLO
single top t-channel	$217^{+3.1\%\ +2.8\%}_{-2.1\%\ -2.8\%}$	aNNLO
single top s-channel	$10.3^{+2.8\%\ +2.6\%}_{-2.3\%\ -2.6\%}$	aNNLO
VV, $qqVV$ (lep-filtered)	117.1 [173, 174]	NLO
VVV	0.015 [173, 174]	NLO
$Z \to \ell^+\ell^-$	$6300^{+5.0\%}_{-5.0\%}$	NNLO

In order to ensure orthogonality with other $t\bar{t}H$ multilepton analyses ($2\ell + 0\tau_{\text{had}}$, $2\ell + 1\tau_{\text{had}}$ and 4ℓ), all multilepton channels share a common jet, lepton and overall event preselection. After the channel categorisation a tighter object selection is applied, according to the analysis needs.

Events are required to pass single or dilepton triggers: single lepton triggers achieve full efficiency for isolated e (μ) with $p_{\text{T}} > 25$ (21) GeV [176], while the dilepton triggers for e (μ) with p_{T} that is 1 GeV above the lowest trigger threshold. Overall the trigger efficiency is > 94–99% efficient for events passing final signal region selections.

At least one reconstructed primary vertex is required, consistent with the luminous region of the beams in the transverse plane. If more than one vertex is found, the primary vertex is chosen as the vertex with highest $\sum p_{\text{T}}^2$ of associated tracks [71].

Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter that are associated with charged particle tracks reconstructed in the inner detector [177]. Only candidates with $p_{\text{T}} > 10$ GeV are considered. They are required to satisfy $|\eta_{\text{cluster}}| < 2.47$. Candidates in the transition region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are rejected. A multivariate likelihood discriminant combining shower shape and track information is used to distinguish real electrons from hadronic showers. Isolation variables are used to reduce the background from non-prompt electrons produced in hadronic decays. Calorimetric isolation uses the sum of transverse energies of calorimeter clusters within a cone of $\Delta R \equiv \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.2$ of the electron candidate. Track isolation uses the sum of transverse momenta of tracks within a cone of

Table 6.3: Tight and loose light lepton definitions.

	Loose		Tight	
	e	μ	e	μ
Track isolation	99% eff.	99% eff.	$< 0.06 \times p_{\mathrm{T}}$	$< 0.06 \times p_{\mathrm{T}}$
Calorimeter isolation	99% eff.	99% eff.	$< 0.06 \times p_{\mathrm{T}}$	99% eff.
Identification working point	Loose	Loose	Tight	Loose
Transverse impact parameter $\|d_0\|/\sigma_{d_0}$	< 10	< 10	< 5	< 3
z impact parameter $\|\Delta z_0 \sin\theta_\ell\|$	< 2 mm	< 2 mm	< 0.5 mm	< 0.5 mm

$\Delta R = \min(0.2, 10\ \mathrm{GeV}/p_{\mathrm{T}}(e))$. For the object preselection, a loose electron discriminant working point is used, and an isolation selection tuned to be 99% efficient for prompt electrons in both calorimetric and tracking variables is chosen. To further reduce the non-prompt electron contribution, the track is required to be consistent with originating from the primary vertex: requirements are imposed on the transverse impact parameter significance and the longitudinal impact parameter, as shown in Table 6.3.

Muon candidates are reconstructed by combining inner detector tracks with track segments or full tracks in the muon spectrometer [77]. In the region $|\eta| < 0.1$, muon candidates are also reconstructed from inner detector tracks matched to isolated energy deposits in the calorimeters consistent with the passage of a minimum-ionizing particle. Candidates are required to satisfy $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.5$. Calorimetric and track isolation are defined for muon candidates similarly as for electron candidates, except that the track isolation uses a larger cone size at low p_{T} ($\Delta R = \min(0.3, 10\ \mathrm{GeV}/p_{\mathrm{T}}(\mu))$). The impact parameter requirement for preselected muon candidates is the same as for loose electrons (Table 6.3).

Jets are reconstructed from calibrated topological clusters built from energy deposits in the calorimeters [178], using the anti-k_t algorithm with a radius parameter $R = 0.4$ [78, 145]. Jets with potential energy contributions from noise or detector effects are removed from consideration [179], and only jets satisfying $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$ are used in this analysis. For jets with $p_{\mathrm{T}} < 60$ GeV and $|\eta| < 2.4$, a jet-track association algorithm, called jet-vertex-tagger (JVT) [180], is used to confirm that the jet originates from the selected primary vertex, in order to reduce the impact of pileup collisions [181].

Jets containing b-hadrons are identified (b-tagged) via a multivariate discriminant combining information from the impact parameters of displaced tracks with topological properties of secondary and tertiary decay vertices reconstructed within the jet: it corresponds to an improved version of the multivariate tagger used for Run 1 [182]. Improvements are due to the addition of the IBL, which results in an extra pixel layer closer to the beampipe, and due to several improvements to the tracking and b-tagging algorithms. The working point used for this search corresponds to approximately 70% efficiency for b-jets in the tracking coverage arising from top-quark decay. The expected rejection factors against light and c-jets are 380 and 12 [183], respectively.

Hadronically decaying τ lepton candidates (τ_{had}) are reconstructed from clusters in the calorimeters and associated inner detector tracks [184]. The candidates are required to have either one or three associated tracks, with a total charge of ±1. Candidates with $|\eta| < 2.5$, excluding the electromagnetic calorimeter transition region $1.37 < |\eta| < 1.52$, are considered. A BDT discriminant using calorimeter and tracking-based variables is used to identify true τ_{had} candidates and reject generic jet backgrounds. The chosen working point has an efficiency of 55% (40%) for true one- (three-)prong τ_{had} decays. Electrons which are reconstructed as one-prong τ_{had} candidates are removed using a sliding cut on the electron likelihood ID variable; the rejection factor for electrons is $\sim$ 30–100 depending on η.

In order to avoid double counting objects and remove leptons likely originating from hadron decays, ambiguities are resolved following an Overlap Removal procedure summarised in Table 6.4. This algorithm is applied to the preselected objects and all channel-specific further quality requirements on the leptons and jets start with the surviving candidates after this procedure.

Events are then organised into various multilepton channels ($2\ell + 0\tau_{\mathrm{had}}$, $2\ell + 1\tau_{\mathrm{had}}$, 3ℓ and 4ℓ) based on the object multiplicity after the preselection. No events are therefore shared between channels.

Keep	Remove	Cone size (ΔR)
μ	e	0.1
e	e	0.1, keep the highest p_{T} one
e	jet	0.3
jet	μ	$\min(0.4 + 10[\ \mathrm{GeV}]/p_{\mathrm{T}}(\mu)[\ \mathrm{GeV}])$
e	τ	0.2
μ	τ	0.2
τ	jet	0.3

Table 6.4: Summary of overlap removal between electrons, muons, taus and jets.

6.4 Three lepton selection: definition of signal and control regions

The three lepton analysis requires exactly three loose leptons at preselection level, with a total charge of ±1. Background processes arising from the addition of a fake or non-prompt lepton to an opposite-sign dilepton pair (such as dileptonic $t\bar{t}$ or Z+jet productions) would easily mimic the signal process. The additional lepton would have the same sign of one real lepton; consequently, the two same-sign (SS) leptons are always required to pass tight selections, in terms of isolation, identification and impact parameter, as shown in Table 6.3. Additionally, they are required to have a transverse momentum above 20 GeV.

The lepton with a different charge among the three, hereafter named opposite-sign (OS) lepton, is very rarely non-prompt: thus, it has to fulfill much looser prerequisites on the transverse impact parameter significance and the longitudinal impact parameter (Table 6.3). Leptons are ordered, by convention, according to the sign and distance to the opposite-sign lepton: the first lepton is defined as the opposite-sign, the second as the closest same-sign and the third as the remaining one.

At least one e or μ candidate in the event is required to match one of the triggers and satisfy the corresponding p_{T} threshold: 25 GeV for an electron and 21 GeV for a muon if it fires a 2015 single lepton trigger, and 25 GeV for both flavours, instead, in the case of a 2016 single lepton trigger,; concerning dilepton triggers the p_{T} threshold that the lepton has to pass is 1 GeV more than the lowest triggering p_{T}.

To reject the $t\bar{t}Z$ background, all opposite-sign same-flavour (OSSF) lepton pairs in the event must be outside the Z mass resonance peak[1] by 10 GeV. To remove potential backgrounds with light dilepton resonances, the $\ell^+\ell^-$ invariant mass is also required to be larger than 12 GeV. Moreover, to remove potential backgrounds with Z decays to $\ell\ell\gamma \rightarrow \ell\ell\ell'(\ell')$, where one lepton has very low momentum and is not reconstructed, the three-lepton invariant mass must satisfy $|m_{3\ell} - 91.2\ \mathrm{GeV}| > 10$ GeV.
Finally, at least three jets are required in the event in order to further suppress $t\bar{t}$ and diboson backgrounds.

In Table 6.5 the signal region (SR) and the control region (CR) selections are summarised, together with some validation regions (VR) used to validate the modelling of the main backgrounds. The control regions will be used in the profile likelihood fit to improve the knowledge on the normalisation of some

---

[1] The Z mass resonance is taken to be 91.2 GeV, as quoted by the Particle Data Group [7].

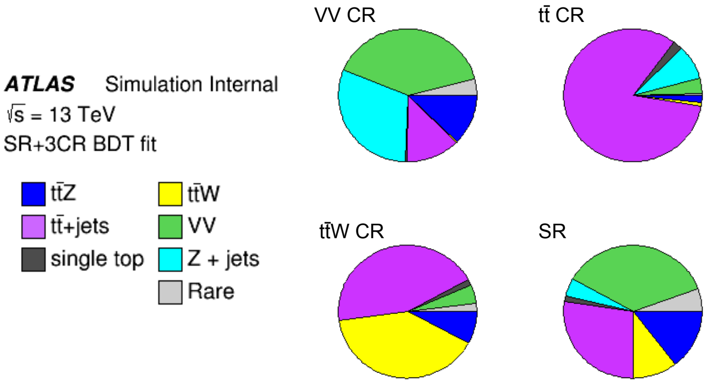


Figure 6.3: Pie-chart of the predicted signal and background events simulated by MC after selection for the signal and three control regions.

backgrounds, as described in Section 6.8: by inverting the tight requirements on the two same-sign leptons, the $t\bar{t}$ control region ($t\bar{t}$ CR) gets enriched of backgrounds with ≤ 2 real leptons (reducible background), whereas reversing the Z mass veto in the signal region enriches the VV control region (VV CR) of diboson events; keeping the signal region selection but requiring exactly two b-tagged jets produces a third control region ($t\bar{t}W$ CR), highly pure in $t\bar{t}W$ processes.

The b-tagging of jets is performed by a new MVA algorithm, named “MV2c10”, which superseded MV1 used in Run 1.

A description about the composition and treatment of the reducible and irreducible background sources is given in the next Section.

The sensitivity of the 3ℓ signal region to different Higgs boson decay modes is shown in Table 6.6: being mostly sensitive to the WW decay will be one of the main features to define the pseudo matrix-element discriminator.

The event yields in the different regions prior to the fit (“pre-fit”) are summarised in Table 6.7. The relative proportions of the background yields obtained in each topology are shown in Figure 6.3.

6.5 Backgrounds

The backgrounds characterising the multilepton channels can be categorised into:

- irreducible background, namely processes where all selected leptons are produced in decays of electroweak bosons (prompt leptons);
- reducible background, as processes in which at least one lepton arises from another source. In this case, the leptons arise from hadron decays (non-prompt), detector interactions (charge misreconstruction or fake), or improper reconstruction of other particle species (fake).

6.5.1 Irreducible backgrounds

The irreducible background corresponds to background processes with at least three prompt leptons in the final state, thus with a signature very similar to the signal. It is estimated using MC simulation: diboson production, in particular WZ, is the largest background with prompt leptons. $t\bar{t}V$ production ($t\bar{t}Z$ and $t\bar{t}W$) instead is the subleading source. Rare processes (tZ, tWZ, $t\bar{t}t\bar{t}$, $t\bar{t}W^+W^-$) also contribute to a

Table 6.5: Selections for the signal region (SR), control region (CR) and validation regions (VR). The variable $H_{\mathrm{T, jets}}$ is the scalar sum of the transverse momenta for the considered jets. Same-flavour, opposite-charge lepton pairs are referred to as OSSF pairs. Trigger-matched leptons correspond to an object reconstructed by the trigger, and must have $p_{\mathrm{T}} > 25$ GeV (21 GeV for muons in 2015 data); for dilepton triggers the p_{T} threshold is 1 GeV more than the lowest triggering p_{T}. In all regions at least one selected light lepton is required to be trigger-matched.

Region	Selection
3ℓ SR	3 light leptons; total charge ±1 2 SS leptons must be tight and have $p_{\mathrm{T}} > 20$ GeV OS lepton must satisfy tight impact parameter selection $\lvert M(\ell^+\ell^-) - 91.2\ \mathrm{GeV}\rvert > 10$ GeV for all OSSF pairs $N_{\mathrm{jets}} \geq 3$
VV CR	3 light leptons; total charge ±1 2 SS leptons must have $p_{\mathrm{T}} > 20$ GeV OS lepton must satisfy tight impact parameter selection $\lvert M(\ell^+\ell^-) - 91.2\ \mathrm{GeV}\rvert < 10$ GeV for at least one OSSF pair $N_{\mathrm{jets}} \geq 3$
$t\bar{t}$ CR	3 light leptons; total charge ±1 2 SS leptons must have $p_{\mathrm{T}} > 20$ GeV OS lepton must satisfy tight impact parameter selection $\lvert M(\ell^+\ell^-) - 91.2\ \mathrm{GeV}\rvert > 10$ GeV for all OSSF pairs at least 1 SS lepton must be non-tight $N_{\mathrm{jets}} \geq 3$
$t\bar{t}W$ CR	3 light leptons; total charge ±1 2 SS leptons must be tight and have $p_{\mathrm{T}} > 20$ GeV OS lepton must satisfy tight impact parameter selection $\lvert M(\ell^+\ell^-) - 91.2\ \mathrm{GeV}\rvert > 10$ GeV for all OSSF pairs $N_{\mathrm{jets}} = 2$ and $N_{b\mathrm{-jets}} = 2$
Tight $t\bar{t}Z$ VR	3ℓ lepton selection at least one $\ell^+\ell^-$ pair with $\lvert M(\ell^+\ell^-) - 91.2\ \mathrm{GeV}\rvert < 10$ GeV $N_{\mathrm{jets}} \geq 4$ and $N_{b\mathrm{-jets}} \geq 2$
Loose $t\bar{t}Z$ VR	3ℓ lepton selection at least one $\ell^+\ell^-$ pair with $\lvert M(\ell^+\ell^-) - 91.2\ \mathrm{GeV}\rvert < 10$ GeV $N_{\mathrm{jets}} \geq 4$ and $N_{b\mathrm{-jets}} \geq 1$, or $N_{\mathrm{jets}} = 3$ and $N_{b\mathrm{-jets}} \geq 2$
WZ + heavy flavour VR	3ℓ lepton selection at least one $\ell^+\ell^-$ pair with $\lvert M(\ell^+\ell^-) - 91.2\ \mathrm{GeV}\rvert < 10$ GeV $N_{\mathrm{jets}} \geq 1$ and $N_{b\mathrm{-jets}} \geq 1$

Table 6.6: Fraction of the expected $t\bar{t}H$ signal arising from different Higgs boson decay modes in the signal region. The decays contributing to the “other” column are dominantly $H \to \mu\mu$ and $H \to b\bar{b}$. The acceptance times efficiency includes Higgs boson and top-quark branching fractions, detector acceptance, and reconstruction and selection efficiency, and is computed relative to inclusive $t\bar{t}H$ production.

	Higgs boson decay mode				$A \times \varepsilon$
Category	WW^*	$\tau\tau$	ZZ^*	Other	$(\times 10^{-4})$
3ℓ	74%	20%	4%	2%	9.2

Table 6.7: Event yields from simulated backgrounds and the signal in the control (CR) and signal regions (SR) before the fit procedure (pre-fit). The quoted uncertainties include statistical and systematic effects. The sum of all contributions may slightly differ from the total value due to rounding. The $t\bar{t}+\gamma$ yields are included in the $t\bar{t}$+jets background after the overlap removal between the two samples.

	VV CR	$t\bar{t}$ CR	$t\bar{t}W$ CR	3ℓ SR
$t\bar{t}W$	5.77 ± 0.39	13.58 ± 0.81	7.4 ± 1.8	31.6 ± 1.6
$t\bar{t}(Z/\gamma^*)$	234 ± 38	24.8 ± 3.8	1.46 ± 0.74	44.9 ± 5.7
Diboson	750 ± 150	54 ± 10	0.86 ± 0.69	111 ± 21
$t\bar{t}$	232 ± 71	1190 ± 150	8.1 ± 5.9	86 ± 23
Z+jet	580 ± 100	119 ± 29	0.024 ± 0.027	13.7 ± 4.1
single top	8.7 ± 3.0	32.9 ± 3.7	0.25 ± 0.25	3.9 ± 1.2
Rare	74 ± 24	7.4 ± 1.5	0.35 ± 0.12	17.6 ± 3.7
Total background	1880 ± 210	1440 ± 160	18.5 ± 6.7	305 ± 35
$t\bar{t}H$ (SM)	8.36 ± 0.87	17.5 ± 1.7	0.87 ± 0.12	25.1 ± 2.4

Table 6.8: Expected and observed event yields in the validation regions (VR). The quoted uncertainties include all systematic uncertainties. "Purity" indicates the fraction of events in the VR expected to arise from the targeted process ($t\bar{t}Z$ for the first two VRs and WZ for the third).

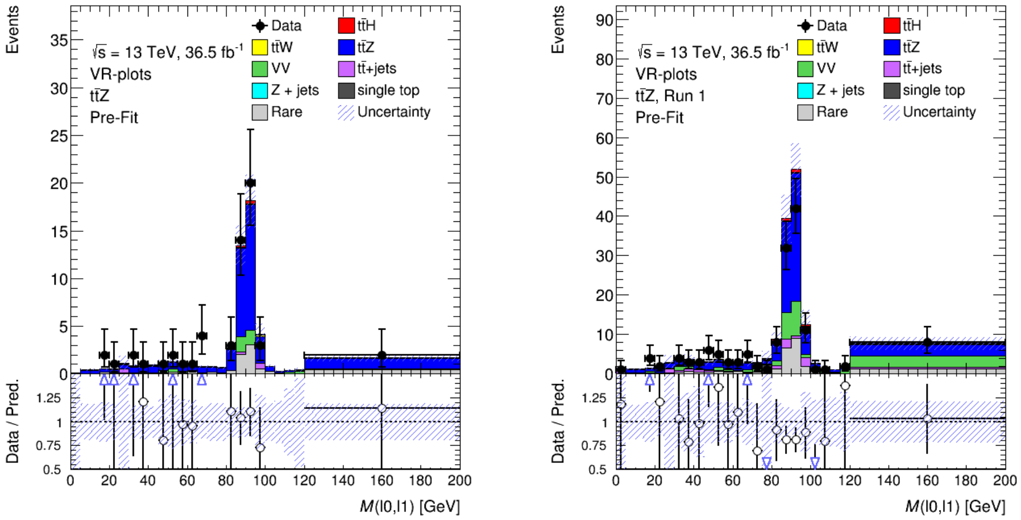
VR	Purity	Expected	Data
Tight $t\bar{t}Z$	70%	54 ± 8	57
Loose $t\bar{t}Z$	64%	168 ± 4	147
WZ + 1 b-tag	25%	455 ± 64	432

lesser extent (≈ 5%). For the purposes of this analysis, the associated production of a single top quark and a Higgs boson is considered as a background, since it contributes negligibly to the SM cross section.

The major backgrounds are studied in control regions that are enhanced in specific processes and disjoint from the signal regions: these allow to control their normalisations. Additional validation regions are also defined to study the dependence on the jet multiplicity of specific backgrounds, and to validate normalisations. The validation regions are summarised in Table 6.5 and a comparison of expected and observed yields is shown in Table 6.8.

A validation region for $t\bar{t}Z$ is defined by inverting the Z veto of the 3ℓ signal region, and in addition tightening the jet selection by requiring ≥ 4 jets of which ≥ 2 are b-tagged. To gain additional events for comparison, a selection with higher acceptance in multiplicity of jets is also used; the latter includes a larger fraction of WZ events. Figure 6.4 shows the invariant mass plots for these VRs: a perfect agreement between data and simulation is observed for the $t\bar{t}Z$ background.

With the present limited amount of data, it is not possible to constrain the cross section for the production of WZ or other diboson processes in association with c- and b-quark with high accuracy. To probe these processes, trilepton events with a Z candidate and a single b-tagged jet were studied as a validation region. This region is expected to contain similar numbers of mistagged WZ + light quark/gluon jet, $WZ+c$, and $WZ+b\bar{b}$ events, with the fraction of $WZ+b\bar{b}$ increasing with jet multiplicity. The jet multiplicity in these events is modelled well by simulation as shown in Figure 6.5. The good agreement indicates that the heavy flavour versus light flavour content of these events is reasonably simulated.



(a)

(b)

Figure 6.4: Invariant mass of the first 2 leptons for a) tight and b) loose $t\bar{t}Z$ validation regions. The leptons are labelled in the same way as for the 3ℓ signal region. Events away from the Z peak are those satisfying the Z selection with the first and third lepton.

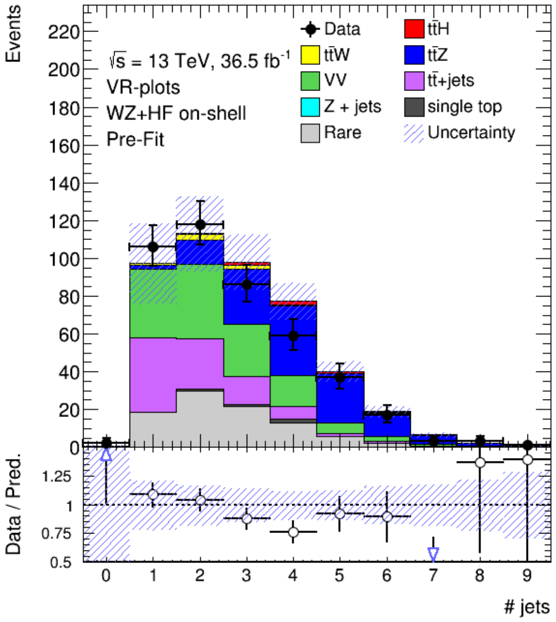


Figure 6.5: Jet multiplicity in the WZ + one b-tag validation region.

6.5.2 Reducible backgrounds

The reducible background mostly arises from processes with at least one non-prompt or fake lepton. Monte Carlo studies indicate that, with the chosen lepton identification requirements, the vast majority of e and μ candidates in signal and validation regions corresponds to real e and μ leptons, but originated from decays of hadrons or photon conversions in detector material, and not from jets misidentified as leptons. In addition to the non-prompt source, the trident process $e^{\pm} \to \gamma^{(*)}e^{\pm} \to e^{\mp}e^{\pm}e^{\pm}$ can result in a high-p_{T} electron with opposite charge to the original prompt electron (charge misreconstruction); anyhow events with at least one charge misidentified lepton in the 3ℓ final state have a negligible impact ($\lesssim 0.01\%$ of the total background in the SR are produced).

Simulation studies also show that the dominant source of non-prompt leptons in the signal region is decays of heavy flavour hadrons, primarily in $t\bar{t}$ events, followed by the Z+jet and single top (mostly Wt production mode) processes. Relative systematic uncertainties are considered, due to theoretical uncertainties on cross sections and variations of PDF, showering, generator, renormalisation scale and factorisation scale; they are summarised in Section 6.7.

In the 3ℓ analysis performed at 3.2 fb^{-1} [185], the reducible background was estimated using a data-driven technique, which determined a transfer factor from anti-tight to tight lepton in the $2\ell 0\tau_{\mathrm{had}}$ channel; the transfer factor was then applied to an anti-tight 3ℓ side-band region in order to extrapolate the non-prompt contribution to the signal region. This oversimplified method and the choice of the $2\ell 0\tau_{\mathrm{had}}$ channel to determine the transfer factor were primarily driven by the lack of data in other 3ℓ control regions to estimate the rate. Moreover, the analysis strategy was cut-based and the signal region was tighter in terms of multiplicities of jets and b-tagged jets, in order to reach the highest signal purity.

The BDT-based analysis presented in this thesis, instead, uses a different definition of the signal region, more relaxed in terms of multiplicity of jets, and employs several control regions; this approach allows to use simulation to model the shape of the reducible background and to control the normalisation through a dedicated control region. The diboson-enriched region is motivated by the need to control the non-prompt component of the WZ process, where one non-prompt lepton is present; thus, it can be considered as a non-prompt control region, like the $t\bar{t}$ region. The $t\bar{t}W$ enriched region, instead, is introduced because of the inconclusive checks on the normalisation of such background: the cross section for $t\bar{t}W$ production rises more slowly with energy than those for $t\bar{t}Z$ or $t\bar{t}H$, and as a result the statistical power of the current data sample for validating predictions of $t\bar{t}W$ production is weak, unlike the $t\bar{t}Z$ background. Indeed, the dedicated measurement of $t\bar{t}W$ production with 13 TeV data in ATLAS [186] finds a cross section consistent with the Standard Model prediction with 56% uncertainties, dominantly statistical.

6.6 Analysis method

Analogously to the analysis method followed in the fully hadronic search and described in Section 5.6, the TMVA framework [90] is used to train a Gradient BDT classifier in order to separate the $t\bar{t}H$ signal from the total background in the signal region.

Then a hypothesis test, based on the profile likelihood method described in Chapter 4, is performed exploiting the resulting BDT. The details about the training and the validation of the BDT classifier, and about the fit model are given in Sections 6.6.3, 6.6.4 and 6.8, respectively.

6.6.1 Variables entering the BDT

A set of 14 variables is considered for optimizing the BDT. The input variables include global event variables, such as $n_{\mathrm{jets}} + 10 \cdot n_{b\text{-jets}}$, $H_{\mathrm{T}}^{\mathrm{jets}}$ and $H_{\mathrm{T}}^{\mathrm{lep}}$, as well as event-shape variables, like the average mass of reconstructed leptonic W bosons. Other variables are calculated from pairs of objects: the invariant masses of lepton pairs or physical features (i.e., p_{T}, mass and ΔR) of the lepton and b-tagged jet pair with the largest vector sum p_{T}.

Finally, a Pseudo Matrix-Element discriminator, $\mathcal{D}$, is used; it returns the probability of an event to be a signal candidate, compared to the probability of being a background candidate. Details about the definition of such discriminator are provided in the next Section.

6.6.2 The Pseudo Matrix-Element discriminator

A Pseudo Matrix-Element discriminator $\mathcal{D}$ is built through an approximated matrix-element approach, which aims at identifying signal-like events by partially reconstructing resonances and exploiting some peculiar kinematic informations via the so-called "truth-matching" method: it determines the origin of a reconstructed object, at simulation level, by matching it to the closest true particle (fermion or boson before hadronisation and parton showering) within $\Delta R < 0.3$.

The discriminant between signal and background is defined as

$$\mathcal{D}(\mathbf{x}) = \frac{P^{\mathrm{sig}}(\mathbf{x})}{P^{\mathrm{bkg}}(\mathbf{x})} \tag{6.1}$$

where $P^{\mathrm{sig}}(\mathbf{x})$ and $P^{\mathrm{bkg}}(\mathbf{x})$ represent the probability density functions (p.d.f.s) of a given event under the signal hypothesis ($t\bar{t}H$) and under the background hypothesis ($t\bar{t}X$), respectively: the background hypothesis ($t\bar{t}X$) refers to the sum of the $t\bar{t}$-like processes ($t\bar{t}Z$, $t\bar{t}W$ and $t\bar{t}$+jets) in the signal region. The discriminator $\mathcal{D}$ intends essentially to discriminate those processes, which exhibit a very signal-like behaviour. Both p.d.f.s are functions of $\mathbf{x}$, representing the four-momentum vectors of all final-state particles at the reconstruction level, leptons (ℓ), jets and b-jets (j and b), selected in the analysis.

Since the signal and the most of backgrounds result from $t\bar{t}$ decay, there are few experimental handles available to discriminate between them. The most prominent features are the different resonances present in the decay (i.e. the Higgs boson in the case of $t\bar{t}H$, W and Z for $t\bar{t}W$ and $t\bar{t}Z$, no resonance for $t\bar{t}$+jets), and the different flavour content of the jets forming those and the top-quark resonances. This is one of the main ingredients in the construction of $P^{\mathrm{sig}}(\mathbf{x})$ and $P^{\mathrm{bkg}}(\mathbf{x})$ in this analysis, so that $\mathbf{x}$ is extended to include not only the four-momenta of jets, $\mathbf{p}_{\mathrm{jet}}$, but also the value of their multivariate b-tagging discriminant w_{jet}, *i.e.*, $\mathbf{x} \equiv \mathbf{p}_{\ell_i}, (\mathbf{p}_{j_k}, w_{j_k})$ ($i = 1, ..., N_{\mathrm{leptons}}$ and $k = 1, ..., N_{\mathrm{jets}}$). There is also some angular information from the different spins of the daughter resonances (Higgs and W boson) that is exploited.

The calculation of $P^{\mathrm{sig}}(\mathbf{x})$ and $P^{\mathrm{bkg}}(\mathbf{x})$ is discussed in the following two Sections. The level of separation achieved between signal and background with the resulting discriminant $\mathcal{D}$ is illustrated in Section 6.6.4.

Signal probability

The construction of $P^{\mathrm{sig}}(\mathbf{x})$ is described step by step to illustrate the method. As shown in Table 6.6 the 3ℓ final state is characterised by many possible Higgs decay modes: the simultaneous description of all those states would be difficult and a pure matrix-element method would become troublesome in terms of high dimensionality of the integration and computing time. However, many of those decay modes contribute very little to the signal event p.d.f., and therefore a first approximation can be introduced by considering only the $H \to WW$ decay mode ($\sim$ 80% of summary event p.d.f.).

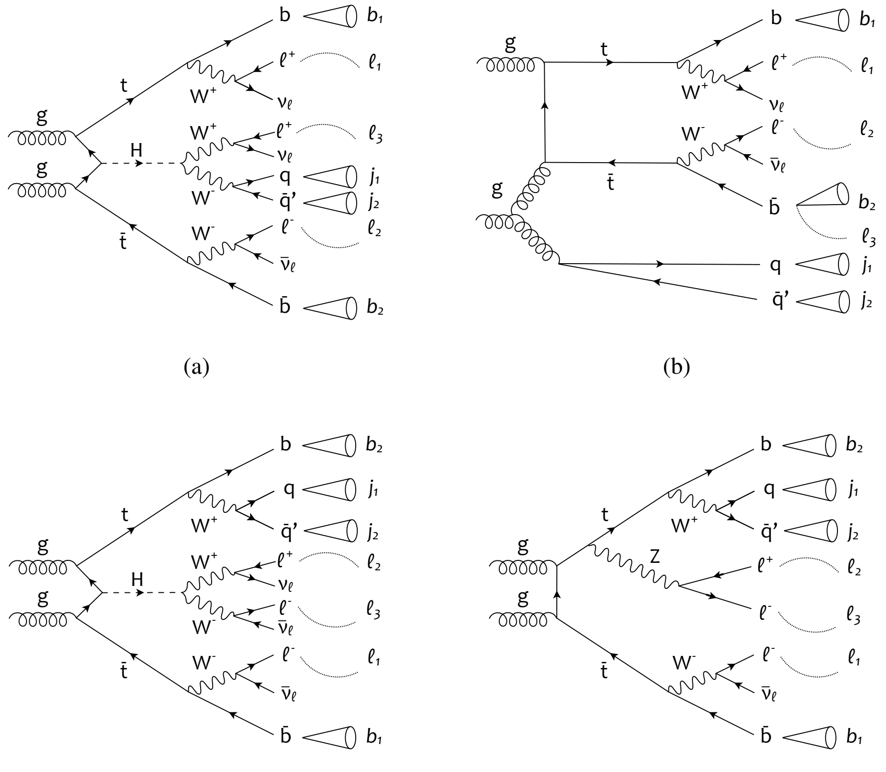
A second approximation comes from MC studies on the truth origin of leptons in $t\bar{t}H$ events; the most likely categories are found to be:

- two leptons from the top-quark decay and one from the Higgs boson decay (labelled as $t2\ell$)
- one leptons from the top-quark decay and two from the Higgs boson decay (labelled as $h2\ell$).

Thus two possible signal final states can be indicated in the following way (Figures 6.6(a) and 6.6(c)):

$$t(W(\ell\nu)\,b)\cdot t(W(\ell\nu)\,b)\cdot H(W(\ell\nu)\,W(jj)) \quad \text{for } t2\ell$$
$$t(W(\ell\nu)\,b)\cdot t(W(jj)\,b)\cdot H(W(\ell\nu)\,W(\ell\nu)) \quad \text{for } h2\ell\,,$$

where the adopted formalism outlines the reconstructed objects present in the final state, shown in the parentheses, and their origin from the $t\bar{t}H$ process. According to these two most likely final states for $t\bar{t}H$ in the signal region, two different signal probabilities can be constructed.



(a) (b)

(c) (d)

Figure 6.6: Representative Feynman diagrams illustrating the partonic configurations and parton–jet assignments considered in the construction of the signal and the background probabilities in the 2 categories used in the definition of the final discriminant, (top) $t2\ell$ and (bottom) $h2\ell$ for signal and background. The diagram (b) represents the $t\bar{t}$+jets background while the diagram (d) refers to $t\bar{t}Z$.

A probability $P^{\mathrm{sig}}(\mathbf{x})$ is then defined as the product of the normalised p.d.f.s (mass, p_{T}, distance, ...) of each reconstructed resonance in the event. The p.d.f.s are determined by employing "templates", which are constructed from simulated signal events using the reconstructed lepton and/or jets corresponding to the correct parton–object assignment, determined by truth-matching. These templates are constructed as unit-normalised one-dimensional histograms.

In the $t2\ell$ category, the p.d.f. for the semileptonic top-quark resonance is calculated by considering two templates: the distance between a lepton and a b-jet from a top quark in the transverse plane, $\Delta\Phi_{\ell b}$, and the "corrected mass", $M_{\ell b\,\mathrm{corr}}$. The p.d.f. for the Higgs boson resonance is calculated, instead, by considering the invariant mass of two light-jets, M_{jj}, originating from an hadronically decaying W boson from an Higgs boson, and a corrected mass, $M_{\ell jj} \ominus M_{jj}$, of the two previous light-jets and a lepton originating from a Higgs boson.

Analogously, in the $h2\ell$ category, the p.d.f. for the semileptonic top-quark resonance is calculated by employing the same templates described above, while the p.d.f. for the hadronic top-quark resonance is calculated by considering the invariant mass of two light-jets, M_{jj}, originating from an hadronically decaying W boson from a top quark, and a corrected mass, $M_{bjj} \ominus M_{jj}$, of the two previous light-jets and a b-jet originating from an hadronically decaying top quark. Finally, the p.d.f. for the Higgs boson resonance uses the distance in the transverse plane between two leptons, $\Delta\Phi_{\ell\ell}$, originating from two leptonically decaying W bosons from an Higgs boson, and their corrected mass, $M_{\ell\ell\,\mathrm{corr}}$.

The templates used in the signal p.d.f.s are shown in Figures 6.8 and 6.9: they are determined in a region very close to the signal one, using the Powheg +Herwig++ signal samples.

Since only one-dimensional templates are employed, a more effective $P^{\mathrm{sig}}(\mathbf{x})$ can be built by reducing the correlation among templates. The invariant mass, $M_{\ell jj}$, of one lepton and two jets originating from a Higgs boson is strongly correlated to the mass of the two jets, M_{jj}: therefore their difference in quadrature, $M_{\ell jj} \ominus M_{jj}$, is used as template instead of $M_{\ell jj}$. For the same reason the $M_{bjj} \ominus M_{jj}$ template is used to describe the hadronically decaying top quark. The corrected mass for the semileptonic top-quark decay, $M_{\ell b\,\mathrm{corr}}$, is used to fully decorrelate the (ℓ, b) mass and distance in the transverse plane:

$$\begin{aligned} m^2_{\ell b}(\Delta\Phi_{\ell b}) &= m^2_\ell + m^2_b + 2[E_\ell E_b - \vec{p}_{/\!/,\ell} \cdot \vec{p}_{/\!/,b} - p_{T,\ell}p_{T,b}\cos\Delta\Phi_{\ell b}] \\ \implies\ M_{\ell b\,\mathrm{corr}} &\equiv m^2_{\ell b} + 2\,p_{T,\ell}p_{T,b}\cos\Delta\Phi_{\ell b} \neq f(\Delta\Phi_{\ell b}) \end{aligned} \tag{6.2}$$

Finally, the expression for P^{sig}, just making use of the above kinematic information, denoted by $P^{\mathrm{sig}}_{\mathrm{kin}}$, is:

$$P^{\mathrm{sig}}_{\mathrm{kin}}(\mathbf{x}) = \begin{cases} P^{\mathrm{sig}}(\Delta\Phi_{\ell_1 b_1}, M_{\ell_1 b_1\,\mathrm{corr}}) \cdot P^{\mathrm{sig}}(\Delta\Phi_{\ell_2 b_2}, M_{\ell_2 b_2\,\mathrm{corr}}) \cdot P^{\mathrm{sig}}(M_{j_1 j_2}, M_{\ell_3 j_1 j_2} \ominus M_{j_1 j_2}) & \text{for } t2\ell \\ P^{\mathrm{sig}}(\Delta\Phi_{\ell_1 b_1}, M_{\ell_1 b_1\,\mathrm{corr}}) \cdot P^{\mathrm{sig}}(M_{j_1 j_2}, M_{b_1 j_1 j_2} \ominus M_{j_1 j_2}) \cdot P^{\mathrm{sig}}(\Delta\Phi_{\ell_2 \ell_3}, M_{\ell_2 \ell_3\,\mathrm{corr}}) & \text{for } h2\ell \end{cases} \tag{6.3}$$

In practice, the partonic origin of the jets is not known on data, and thus it is necessary to evaluate $P^{\mathrm{sig}}(\mathbf{x})$ by averaging over the N_p possible parton–object assignments, which dilutes the kinematic information. The b-tagging information can be used in addition to improve the assignment, suppressing those combinations which are inconsistent with the correct parton flavours, as follows:

$$P^{\mathrm{sig}}(\mathbf{x}) = \frac{\sum_{k=1}^{N_p} P_{\mathrm{btag}}(\mathbf{x}^k) \cdot P^{\mathrm{sig}}_{\mathrm{kin}}(\mathbf{x}^k)}{\sum_{k=1}^{N_p} P_{\mathrm{btag}}(\mathbf{x}^k)} \tag{6.4}$$

where $P^{\mathrm{sig}}_{\mathrm{kin}}(\mathbf{x})$ is given by equation 6.3 and $P_{\mathrm{btag}}(\mathbf{x})$ is defined as:

$$P_{\mathrm{btag}}(\mathbf{x}) = P_b(w_{j_1}) \cdot P_b(w_{j_2}) \cdot P_{W\to q'\bar{q}}(w_{j_3}, w_{j_4}) \tag{6.5}$$

where

$$P_{W\to q'\bar{q}}(w_{j_3}, w_{j_4}) = \tfrac{1}{4}[2 \cdot P_q(w_{j_3}) \cdot P_q(w_{j_4}) + P_c(w_{j_3}) \cdot P_q(w_{j_4}) + P_q(w_{j_3}) \cdot P_c(w_{j_4})] \tag{6.6}$$

with j_i ($i = 1, ..., 4$) representing the parton–jet assignment being evaluated, and $P_f(w_{j_i})$ denoting the probability that the i^{th} jet, characterised by its four-momentum $\mathbf{p}_{j_i}$ and b-tagging weight value w_{j_i}, originates from a parton with flavour f (b, c, or q; q for light parton). P_{btag} corresponds to the probability density function to have 2 jets (j_1,j_2) originated from 2 b-quarks and 2 additional light-jets (j_3,j_4) from a hadronic W decay.

The calibration of the b-tagging algorithm is provided for fixed thresholds on the multivariate b-tagging discriminant variable, corresponding to different average b-tagging efficiencies in $t\bar{t}$ events of 60%, 70%, and 80%, also referred to as “operating points” (OP). The corresponding thresholds are denoted by $w_{\text{cut}}^{\text{OP}}$, with OP = 60%, 70%, or 80%. Parameterisations of the b-tagging efficiencies for different jet flavours as functions of jet p_{T} and η are available for each of these operating points, $\varepsilon_f^{\text{OP}}(p_{\text{T}}, \eta)$, which can be used to compute P_f as follows: if the jet b-tagging weight lies within the thresholds for operating points OP$_1$ and OP$_2$, $w_{\text{cut}}^{\text{OP}_1} < w_j \leq w_{\text{cut}}^{\text{OP}_2}$, then $P_f = \varepsilon_f^{\text{OP}_1} - \varepsilon_f^{\text{OP}_2}$; alternatively, if the jet b-tagging weight is below (above) the threshold corresponding to the 80% (60%) operating point, $P_f = 1 - \varepsilon_f^{80\%}$ ($P_f = \varepsilon_f^{60\%}$).

The b-tagging maps for b-jets employed to compute the probabilities $P(\mathbf{x})$ are shown in Figures 6.7. Maps use for c- and light-jets are shown in Appendix C.

Background probability

Since no b-tagged jets are required, the diboson background is one of the the major processes contaminating the signal region. However, it does not result as the most difficult process to discriminate from signal. Subdominant processes like $t\bar{t}$+jets and $t\bar{t}V$, instead, exhibit a very signal-like behaviour: thanks to the presence of the $t\bar{t}$ system with additional jets or weak bosons, they are the most difficult background to distinguish against $t\bar{t}H$. Thus, the background p.d.f. is simplified by considering only the dominant $t\bar{t} + X$ sources (≈ 65% of the total background), in order to focus on the separation between $t\bar{t}H$ and $t\bar{t}X$.

The strategy is to split the background contributions into 2 categories ($t2\ell$ and $h2\ell$) corresponding to the following final states, similarly to what is done for the signal categories:

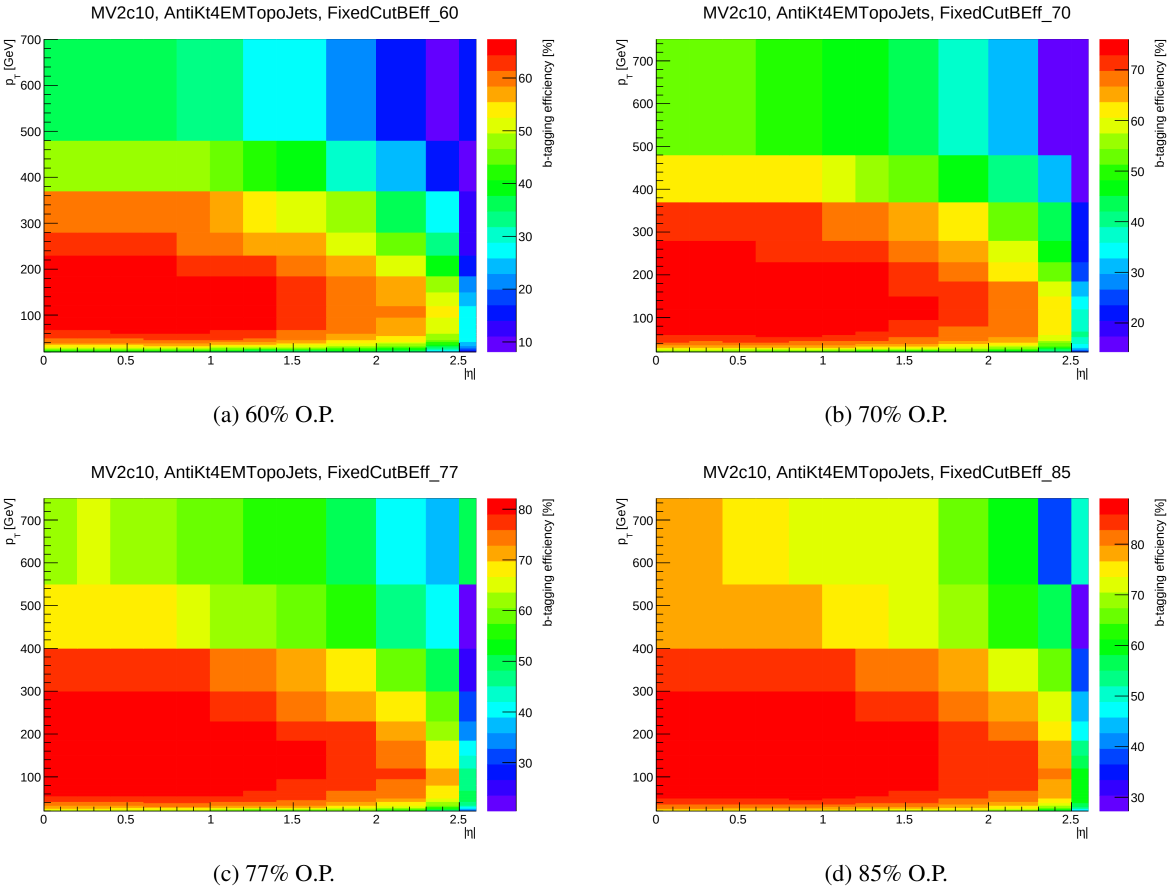
$$\begin{aligned} &t(W(\ell\nu)\,b) \cdot t(W(\ell\nu)\,b) \cdot X(\ell\, jj) \quad \text{for } t2\ell \\ &t(W(\ell\nu)\,b) \cdot t(W(jj)\,b) \cdot X(\ell\ell) \quad \text{for } h2\ell\,. \end{aligned}$$

The calculation of P^{bkg} follows a similar approach to P^{sig}. Of course P^{bkg} differs from P^{sig} because of the absence of the Higgs resonance (replaced by X): $t\bar{t}W$ and $t\bar{t}Z$ are characterised by a bosonic resonance, while $t\bar{t}$+jets not. Thus the $X(\ell\, jj)$ and $X(\ell\ell)$ p.d.f.s result to be a combination of $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}$+jets histograms, calculated in a more relaxed region than the signal one, in order to construct more reliable templates, exploiting all the available MC statistics. Then, the total background template takes into account the expected background proportions in the signal region.

The expression for P^{bkg}, only making use of kinematic informations, denoted by $P_{\text{kin}}^{\text{bkg}}$, is:

$$P_{\text{kin}}^{\text{bkg}}(\mathbf{x}) = \begin{cases} P^{\text{bkg}}(\Delta\Phi_{\ell_1 b_1}, M_{\ell_1 b_1\,\text{corr}}) \cdot P^{\text{bkg}}(\Delta\Phi_{\ell_2 b_2}, M_{\ell_2 b_2\,\text{corr}}) \cdot P^{\text{bkg}}(M_{j_1 j_2}, M_{\ell_3 j_1 j_2} \ominus M_{j_1 j_2}) & \text{for } t2\ell \\ P^{\text{bkg}}(\Delta\Phi_{\ell_1 b_1}, M_{\ell_1 b_1\,\text{corr}}) \cdot P^{\text{bkg}}(M_{j_1 j_2}, M_{b_1 j_1 j_2} \ominus M_{j_1 j_2}) \cdot P^{\text{bkg}}(\Delta\Phi_{\ell_2 \ell_3}, M_{\ell_2 \ell_3\,\text{corr}}) & \text{for } h2\ell \end{cases} \tag{6.7}$$

where the templates employed in the top-quark p.d.f.s are the same signal ones.



(a) 60% O.P.

(b) 70% O.P.

(c) 77% O.P.

(d) 85% O.P.

Figure 6.7: b-tagging maps for a true b-jet used to calculate the pseudo matrix-element discriminator. The 2-D maps are provided for four different fixed O.P., 3 jet flavours (*i.e.*, from b-, c- and light-jet sources, see also Appendix C) and are parametrised in (p_{T},η) of the jet.

Besides the top-quark templates, two additional templates are considered in the p.d.f. of the $t2\ell$ category: the invariant mass of the hadronic dijet, M_{jj}, and the corrected mass of the (ℓ, j, j) triplet, $M_{\ell jj} \ominus M_{jj}$, originated from the "pseudo" resonance X. Two templates for the "pseudo" resonance X ($\Delta\Phi_{\ell\ell}$ and the corrected mass $M_{\ell\ell\,\mathrm{corr}}$) are employed in the $h2\ell$ category, instead. Figures 6.8 and 6.9 show the distributions of the background templates employed in the construction of the probability P^{bkg}. It should be noted that the top-related templates are similar for the signal and the individual backgrounds, proving that the $t\bar{t}$ system behaves exactly the same for the different $t\bar{t}H/t\bar{t}X$ processes and motivates the choice of employing in P^{bkg} the signal top-related templates, which have a much larger number of simulated events.

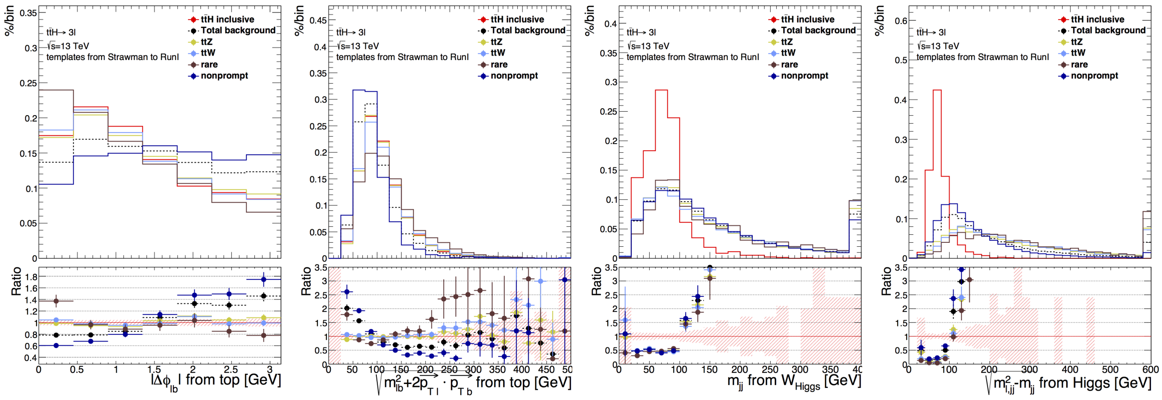
As in Equation 6.4, the expression for $P^{\mathrm{bkg}}(\mathbf{x})$ becomes:

$$P^{\mathrm{bkg}}(\mathbf{x}) = \frac{\sum_{k=1}^{N_p} P_{\mathrm{btag}}(\mathbf{x}^k) \cdot P_{\mathrm{kin}}^{\mathrm{bkg}}(\mathbf{x}^k)}{\sum_{k=1}^{N_p} P_{\mathrm{btag}}(\mathbf{x}^k)} \tag{6.8}$$

where $P_{\mathrm{btag}}(\mathbf{x})$ is computed as discussed in Section 6.6.2.

Final discriminant

The final discriminant $\mathcal{D}$ is computed for each event as given in Equation 6.1, using the definitions for P^{sig} and P^{bkg} given in Equations 6.4 and 6.8, respectively. Since two discriminants $\mathcal{D}$ have been defined for two different categories ($t2\ell$ and $h2\ell$), the maximal among them is chosen as final discriminant: it results to be the best way to maximise the signal versus background discrimination. Finally, the discriminant is used as an input variable to the BDT and results to be very discriminating. The discriminator shows to be also quite performing against VV, in spite of the fact that diboson backgrounds were not considered in the computation of P^{bkg}.



(a) (b) (c) (d)

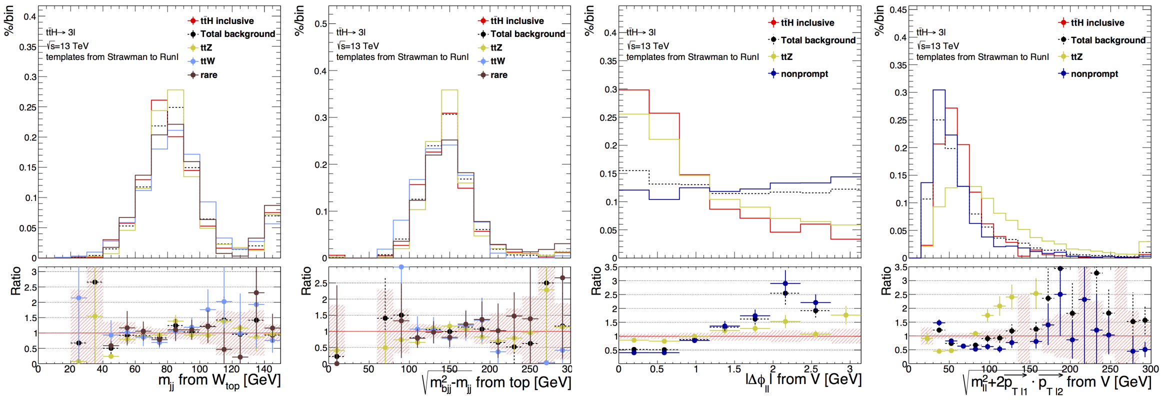
Figure 6.8: Signal and background templates used to compute the Pseudo-Matrix Element discriminator in the $t2\ell$ category: the red and dashed-black lines show the signal and total background distributions for (a-b) the lepton b-jet from top ($\Delta\Phi_{\ell b}$ and $M_{\ell b\,\text{corr}}$) and (c-d) the lepton jets from Higgs or X (M_{jj} and $M_{\ell jj} \ominus M_{jj}$). The templates are extracted in a more relaxed region than the signal one; the $t\bar{t}X$ background templates are renormalised to the respective yield in the signal region.

6.6.3 BDT optimisation

For the sake of optimisation the sum of all backgrounds except for the Z+jet is considered as background: because of the exiguous amount of events of the Monte Carlo sample, the Z+jet shape is modelled with the $t\bar{t}$ background. The signal is modelled by Monte Carlo simulation using the MadGraph5_aMC@NLO +Pythia 8 samples. In this context the systematic uncertainty on the background yield is not taken into account (mostly coming from theoretical uncertainties on diboson and $t\bar{t}$ backgrounds). Both the signal and the background samples are randomly split in half, for training and testing purposes.

Using the initial set of 14 discriminating variables described in Table 6.9, different multivariate methods provided by TMVA (*i.e.*, Boosted Decision Trees, Neural Networks, one- and multi-dimensional Likelihoods, Nearest-Neighbour methods, Fisher discriminants) have been tried in the signal region and their discriminating power have been studied, in terms of signal efficiency vs. background rejection through the ROC integral. The Gradient BDT method resulted to be the most performing for this analysis: its parameters have been customised, as well, in order to achieve even better performance, with a further gain of 10%.

All possible sets of k variables are considered, where k varies from 2 (the smallest set) to 14: the total number of combinations is easily deduced to be $\sum_{k=2}^{N}\binom{N}{k}$, where $N = 14$ is the total number of variables



(a) (b) (c) (d)

Figure 6.9: Signal and background templates used to compute the Pseudo-Matrix Element discriminator in the $h2\ell$ category: the red and dashed-black lines show the signal and total background distributions for (a-b) the jets from top (M_{jj} and $M_{bjj} \ominus M_{jj}$) and (c-d) the two leptons from Higgs or X ($\Delta\Phi_{\ell\ell}$ and $M_{\ell\ell\,\mathrm{corr}}$). The templates are extracted in a more relaxed region than the signal one; the $t\bar{t}X$ background templates are renormalised to the respective yield in the signal region.

and k the number of variables per combination.

$$\sum_{k=2}^{N} \binom{N}{k} = \sum_{k=2}^{N} \frac{N!}{(N-k)! \cdot k!} = 2^N - N - 1 = 16369.$$

For each combination a BDT is trained and tested and the ROC integral computed. Finally the best combination of input variables is kept, which results to be a set of the seven "check-marked" variables listed and described in Table 6.9.

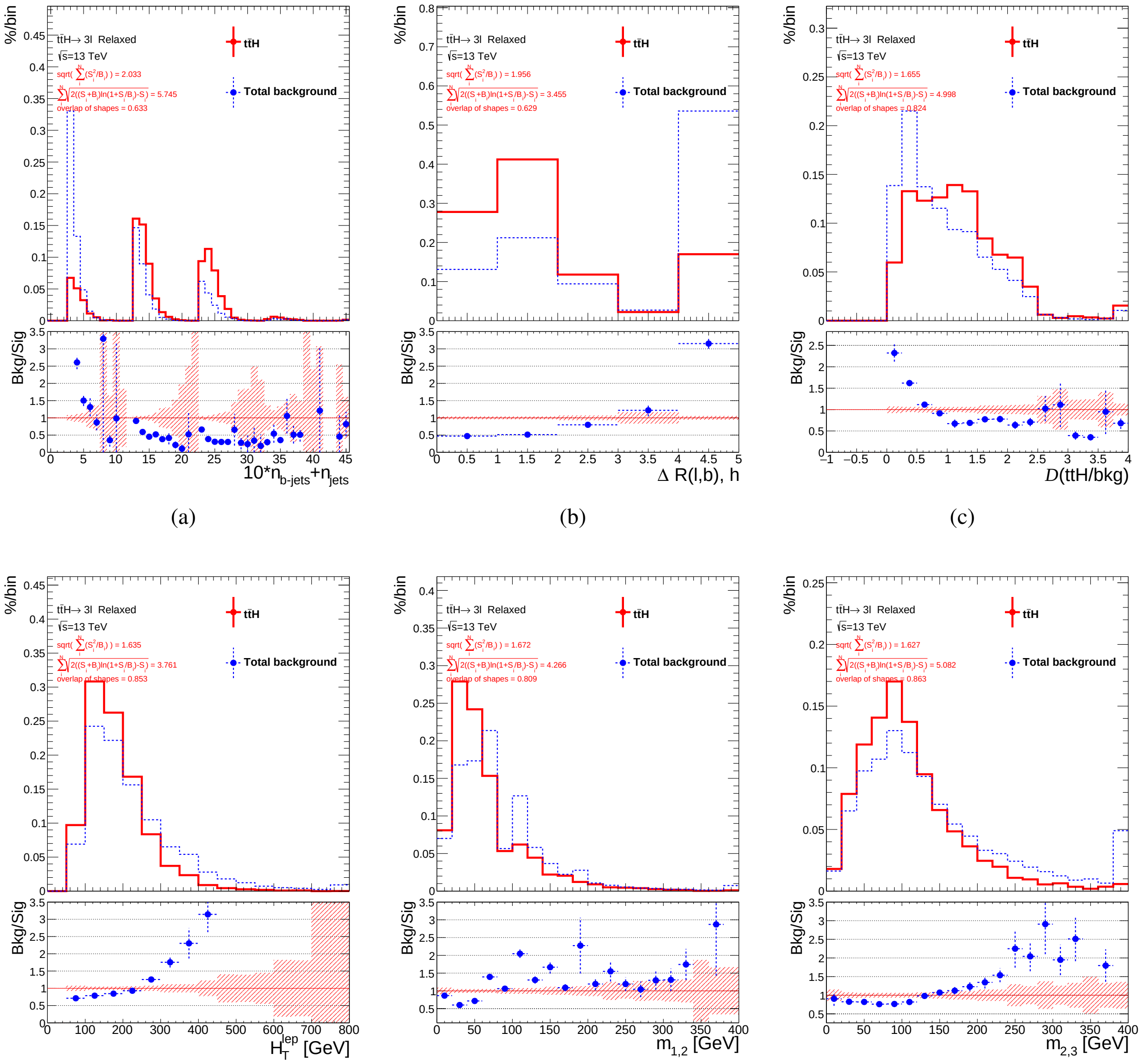
Figures 6.10 and 6.11 show the distributions of the seven variables entering the BDT as well as the BDT distribution: a good separation between signal and background is obtained. In order to check whether the training procedure suffers of possible overtraining the ROC curves obtained using the training samples and the test samples are compared; no effect of overtraining has been observed.

6.6.4 Validation of the variables entering the BDT

The modelling of the variables entering the BDT classifier is validated by comparing the expected distributions with data in both signal regions and control regions. The high amount of events in the $t\bar{t}$ and VV control regions can be used to check the modelling and the pre-fit normalisation of such backgrounds. More details about the fit and how control regions are exploited to infer the normalisation of such backgrounds are provided in Section 6.8.

In the $t\bar{t}$ and VV control regions the signal is almost absent in those regions and thus data can be studied even before unblinding. For the signal and the $t\bar{t}W$ control regions, instead, an "anti-BDT" cut needs to be applied, in order to exclude bins where the expected S/B ratio exceeds 10%, such to remove signal contamination. The blinding thresholds are shown in Table 6.10.

Figure 6.12 show the data/MC comparison after unblinding, for the signal region. Additional plots can be found in Appendix D. The uncertainties discussed in Section 6.7 are considered in the comparisons. No significant shape disagreement is observed: all the observed differences are within the total uncertainty



(a) (b) (c)

(d) (e) (f)

Figure 6.10: Distributions of the six variables after optimisation for simulated signal (solid red line) and total background (dashed blue line) events.

Table 6.9: List of variables used in the BDT optimisation. The check marks indicate variables chosen by the optimisation procedure.

Variable	Definition	
$n_{\text{jets}} + 10 \cdot n_{b\text{-jets}}$	Multiplicity of jets multiplied by 10 times the multiplicity of b-jets	✓
$p_{\text{T}}^{\text{max}}(\ell, b)$	p_{T} of the lepton and b-tagged jet pair with the largest vector sum p_{T}	
$m(\ell, b)^{p_{\text{T}}^{\text{max}}}$	Mass of the lepton and b-tagged jet pair with the largest vector sum p_{T}	
$\Delta R(\ell, b)^{p_{\text{T}}^{\text{max}}}$	ΔR between lepton and b-tagged jet pair with the largest vector sum p_{T}	✓
$m(OS\,\ell)^{p_{\text{T}}^{\text{max}}}$	Mass of the two opposite-sign leptons, left after defining $p_{\text{T}}^{\text{max}}(\ell, b)$ (-1 if two leptons are same sign)	
$\mathcal{D}$	Pseudo Matrix-Element ($\mathcal{P.M.E.}$) discriminator, the ratio of event probabilities under the signal and background hypotheses	✓
$E_{\text{T}}^{\text{miss}}$	Missing transverse energy of the event	✓
$H_{\text{T}}^{\text{jets}}$	Scalar sum of jet p_{T}	✓
$H_{\text{T}}^{\text{lep}}$	Scalar sum of lepton p_{T}	
$m_{(}\ell_0, \ell_1)$	Invariant mass of the first and second lepton pair	✓
$m_{(}\ell_0, \ell_2)$	Invariant mass of the first and third lepton pair	
$m_{(}\ell_1, \ell_2)$	Invariant mass of the second and third lepton pair	✓
m_{leptons}	Invariant mass of the 3 leptons system	
$< m_W >$	averaged leptonic W mass, constructed by assuming that $t\bar{t}H$ events contain three neutrinos with a p_{T} exactly equal to $E_{\text{T}}^{\text{miss}}/3$ and direction of flight opposite to the "all visible objects" Lorentz vector (very harsh approximation)	

Analysis region	anti-BDT cut
$t\bar{t}$ CR	not blind
VV CR	not blind
$t\bar{t}W$ CR	0.16
SR	-0.16

Table 6.10: This table shows for each region the value of the BDT discriminant below which less than 10% of the signal events are expected. The values are used for blinding.

band.

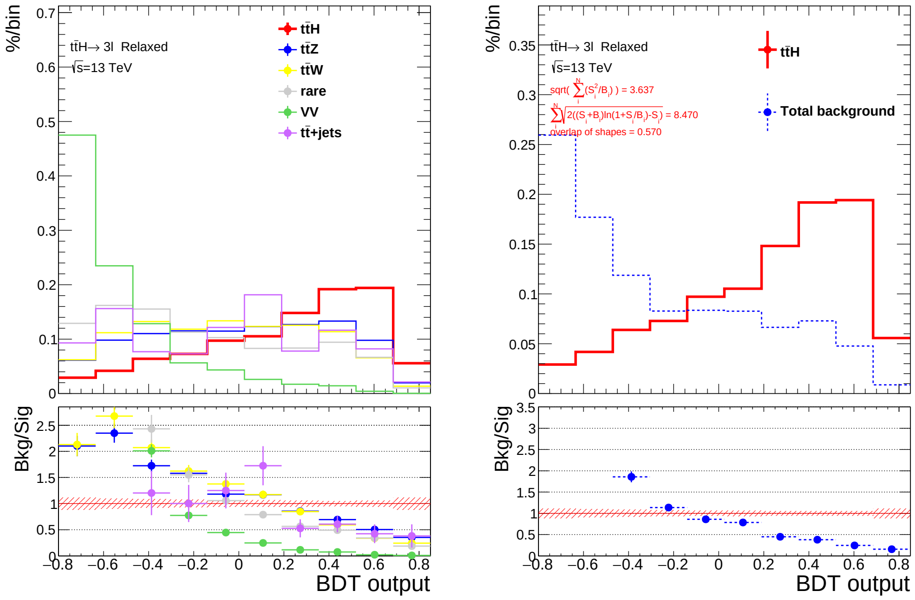
All the validation plots are shown before any fit procedure ("pre-fit"); the fit procedure, described in Section 6.8, will further improve the knowledge of the background normalisation. After the result of the fit ("post-fit"), performed on the final discriminants, the level of agreement for each variable is also verified and found to be satisfying. Post-fit distributions can also be found in Section 6.8.

Finally, the distributions of the BDT classifiers in the control and signal regions are illustrated in Figure 6.16 with the same binning used in the fit in the VV, $t\bar{t}$, $t\bar{t}W$ control regions and signal regions, respectively. A good level of agreement is found for the BDT distributions in all regions.

As already mentioned above, the binning is a very important parameter of the fit procedure and is determined as compromise between the best BDT separation, the amount of events that are available in data and in the MC samples and the shape of the most important systematics.

6.7 Systematic uncertainties

The systematic uncertainties considered in this analysis can be grouped into five main categories, as summarised in Table 6.11. Each systematic uncertainty is represented by an independent parameter,



(a)

(b)

Figure 6.11: Response of the BDT algorithm for simulated signal (red line), $t\bar{t}Z$ background (blue), $t\bar{t}W$ (yellow), $t\bar{t}$+jets (pink), diboson (green) and rare (grey) events on the left, simulated signal (solid red line) and total background (dashed blue) events on the right. The binning is the same as the one used in the fit.

referred to as a nuisance parameter: the effect of the various systematics result on either yield variation and/or shape distortion, as described in Chapter 4.

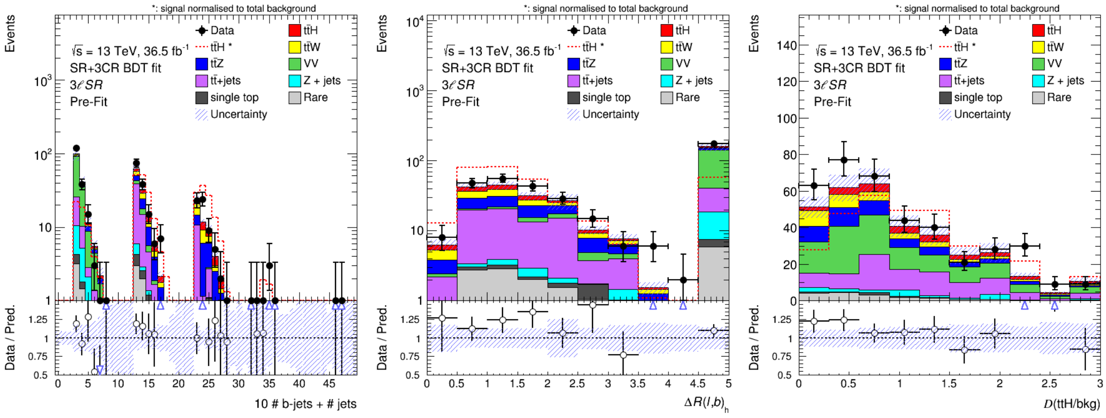
The uncertainties in the integrated luminosity, triggers, reconstruction of the physics objects, and the signal and background MC models follow the recommended treatment. In total, 96 fit parameters are considered. The determination and treatment of the systematic uncertainties are detailed in this Section. Their impact on the fitted signal strength is summarised in Figure 6.18 in Section 6.8.

6.7.1 Detector-related systematics

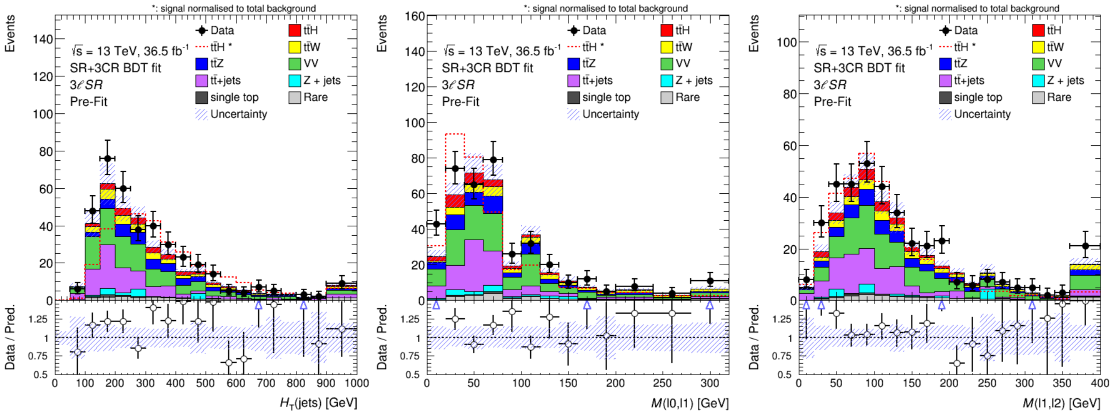
The preliminary uncertainty on the combined 2015+2016 integrated luminosity is 2.9%, derived following a methodology as that detailed in Reference [187] and [105], from a preliminary calibration of the luminosity scale using x–y beam-separation scans performed in August 2015 and May 2016.

Several quantities used in this analysis are subject to experimental systematic uncertainties. Each systematic effect has been evaluated individually using the given uncertainties on event-by-event basis. These uncertainties are related to the trigger efficiency, leptons reconstruction and identification, jet calibration, b-tagging efficiencies and the global event activity. The experimental systematic treatments are evaluated by ATLAS performance groups and are used in the multilepton analysis either as an overall event re-weighting or as a rescaling of the object energy and momentum.

Uncertainties associated with the lepton selection arise from the imperfect knowledge of the trigger, reconstruction, identification and isolation efficiencies, and lepton momentum scale and resolution [72, 77, 188]. The reconstruction and identification efficiency of electrons and muons, as well as the efficiency of the triggers used to record the events, differ between data and simulation. Scale factors and their uncertainties are derived using a tag-and-probe method applied to electrons and muons from Z and



(a) (b) (c)



(d) (e) (f)

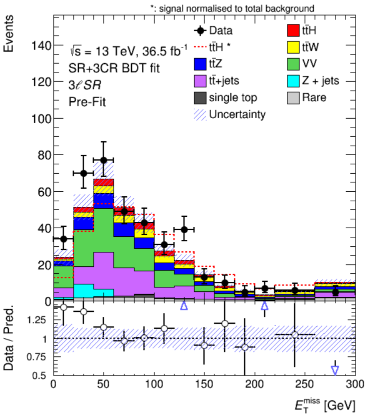


Figure 6.12: Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the signal region before the fitting procedure ("pre-fit"). The uncertainty band includes the statistical and systematic contribution.

Table 6.11: Sources of systematic uncertainty considered in the analysis grouped in five categories. "N" denotes uncertainties affecting only the normalisation for the relevant processes and regions, whereas "S" denotes uncertainties which are considered to affect only the shape of normalised distributions. "SN" denotes uncertainties affecting both shape and normalisation. Some sources of systematic uncertainty are split into several components. The number of components is also reported.

Systematic uncertainty source	Type	Number of components
Luminosity	N	1
Trigger	SN	3
Physics Objects		
Lepton reconstruction	SN	3
Lepton identification	SN	4
Lepton isolation efficiency	SN	3
Lepton momentum scale	SN	2
Lepton momentum resolution	SN	2
Jet energy scale	SN	21
Jet vertex fraction	SN	1
Jet energy resolution	SN	1
b-tagging efficiency	SN	7
c-tagging efficiency	SN	4
Light-jet tagging efficiency	SN	12
Background MC Model		
$t\bar{t}$ hard scattering	S	1
$t\bar{t}$ fragmentation	S	1
$t\bar{t}$ radiation	S	1
$t\bar{t}$ NLO generator	S	1
VV hard scattering	S	1
VV fragmentation	S	1
Z+jet cross section	N	1
Z+jet LO generator	SN	1
$t\bar{t}Z$ hard scattering	SN	1
$t\bar{t}Z$ NLO generator	SN	1
$t\bar{t}Z$ scale	N	2
$t\bar{t}Z$ scale acceptance	N	6
$t\bar{t}W$ hard scattering	S	1
$t\bar{t}W$ NLO generator	S	1
$t\bar{t}W$ scale	N	2
$t\bar{t}W$ scale acceptance	N	6
single top cross section	N	1
Signal Model		
$t\bar{t}H$ fragmentation	SN	1
$t\bar{t}H$ scale	N	2

W bosons and J/ψ particles [177, 189]. For trigger systematics each lepton in the event is weighted according to $\mathrm{SF}_{\mathrm{trig}}(p_{\mathrm{T}}, \eta)$, the uncertainty of which is propagated to the shape and normalisation of the

BDT classifier distribution.

The accuracy of lepton momentum scale and resolution in simulation is checked using reconstructed distributions of the $Z \rightarrow \ell^+\ell^-$ and $J/\psi \rightarrow \ell^+\ell^-$ masses. In the case of electrons, E/p studies using $W \rightarrow e\nu$ events are also used. Small discrepancies are observed between data and simulation, and corrections for the lepton energy scale and resolution in the latter are taken into account. In the case of muons, momentum scale and resolution corrections are only applied to the simulation. Uncertainties on both the momentum scale and resolutions in the muon spectrometer and the tracking systems are considered, and varied separately.

The uncertainties in physics objects are also related to the reconstruction and b-tagging of jets. The jet energy resolution (JER) and the jet energy scale (JES) uncertainties are derived combining the information from test-beam data and simulation [147]. The JES uncertainties are split into 21 uncorrelated components: the largest of these uncertainties is due to the jet flavour composition. The systematic associated to JVT association algorithm requires a particular treatment; this systematic error is determined from the variation of the corresponding cut and data to MC comparisons and ranges from 2% to 1% per jet for p_{T} from 20 to 60 GeV.

The systematic uncertainties related to the b-tagging are modelled with six independent parameters, while four parameters model the c-tagging uncertainty [79]: they are eigenvalues obtained by diagonalising the matrix which parameterises the tagging efficiency as a function of p_{T}. The light-jet tagging systematic uncertainties are parametrised as twelve parameters, depending on p_{T} and η of the jet [152]. The per-jet b-tagging uncertainties are 3%–5%, about 10% for c-tagging and 20% for light jet tagging.

6.7.2 Systematic uncertainties on the $t\bar{t}$+jets background

Since the $t\bar{t}$ background is the main source of non-prompt background, which could be not well modelled by simulations, its normalisation is fully determined through the simultaneous fit described in Section 6.8. Hence, all the $t\bar{t}$-related systematic uncertainties due to variations of the hard scattering generation, the fragmentation algorithm and the radiation modelling affect only the shape of the BDT classifier.

The radiation uncertainties relate to the way additional parton radiation is modelled and are determined through two alternative Powheg +Pythia 8 samples (high/low additional radiation). Variations due to the choice of the hard scattering generator are determined comparing Powheg +Pythia 8, the default $t\bar{t}$ sample of the analysis, with a aMC@NLO +Powheg 8 one. Finally, parton shower variations are calculated by comparing two aMC@NLO samples with different fragmentation schemes (Herwig++/Powheg 8) and applying the difference to the default $t\bar{t}$. Moreover, the shape differences between the default $t\bar{t}$ and a NLO one, Sherpa 2.2.1, is determined and included as additional systematic uncertainty. Basically, the largest uncertainties in the $t\bar{t}$ background description arise from radiation modelling.

6.7.3 Systematic uncertainties on the diboson background

Analogously to the $t\bar{t}$+jets background, normalisation systematics are not applied to the diboson background, because its normalisation is determined through the fit procedure. The systematic uncertainties included, thus, affect only the shape of the BDT classifier and arise from the choice of the hard scattering generation or the fragmentation algorithm.

The default VV sample is Sherpa 2.2.1: the differences between the default diboson and a LO one, MadGraph +Herwig++, is determined and included as systematic uncertainty.

The largest contribution of this process to the signal and control regions includes contributions from $WZ + b\bar{b}$, $WZ + c/c\bar{c}$, and WZ+light jet production. The WZ+light jet mistag contribution is validated in the VV control region: its normalisation factor is fully compatible with the cross section predicted by the

Sherpa simulation. The jet multiplicity in WZ events where no b-tags were found, shown in Figure D.1 pre-fit and in Figure D.2 post-fit, matches Monte Carlo within errors up to 5 jets. $WZ + b\bar{b}$ production process is validated instead in the single b-tagged jet bins of the same $10 \times \#b$-jets + #jets distribution.

6.7.4 Systematic uncertainties on $t\bar{t} + V$

Systematic uncertainties on the cross section and shape for the $t\bar{t}Z/\gamma^*$ backgrounds are obtained from theoretical calculations and Monte Carlo event simulation. Variations of hard process renormalisation/factorisation scale and PDF uncertainties are considered, as are uncertainties in the A14 parton shower tune for the $t\bar{t}Z$ samples. The renormalisation and factorisation scales were varied between 2μ and $\mu/2$ around the nominal μ. The A14 tune [169] optimises 10 parameters corresponding to multiparton interactions, initial and final state radiation. These variations were reduced to a subset of tune variations, of which one pair is sensitive to underlying event effects, another pair mainly for jet structure effects, and three pairs for different aspects of extra jet production.

Cross-section uncertainties from QCD scale and PDF are taken from Tables 6.2 for the $t\bar{t}Z$ process. Acceptance variations from QCD scale variation, from eigenvalues, from diagonalisation of effects including PDF+α_s, modelling, matching, parton shower and underlying events have been evaluated: they are considered as normalisation only variations and their contribution is up to 2-3%.

Finally, the systematic uncertainty due to the variation of the hard scattering generation is considered and determined comparing aMC@NLO +Pythia 8 with the default aMC@NLO +Pythia 8. In addition, the differences between the default NLO generator and an alternative one Sherpa 2.2 are also considered.

Analogous treatment of hard scattering generation and NLO generator systematics are considered for the $t\bar{t}W$ background. However, they are affecting only the shape of the process, since its normalisation will be derived by the fit in the dedicated control region.

6.7.5 Z+jet -related systematics

The default Z+jet sample is Sherpa 2.2.1: the differences between the default diboson and a LO one, MadGraph +Herwig++, is determined and included as systematic uncertainty. An additional and conservative systematic uncertainty on the normalisation is included ($\sim$ 50%) to cover possible misestimate of the cross section.

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

The systematic uncertainties on the $t\bar{t}H$ signal process come from the theoretical cross section uncertainty of Table 6.2 and the shower and fragmentation models. The uncertainty from the parton shower and fragmentation models are evaluated comparing the nominal $t\bar{t}H$ (aMC@NLO +Pythia 8) with aMC@NLO +Herwig++.

6.7.7 Rare SM production and systematics

Rare SM category consists of four main processes: tZ, tH, $t\bar{t}t\bar{t}$, and $t\bar{t}WW$. Table 6.2 shows the NLO cross sections for tZ, $t\bar{t}t\bar{t}$, and $t\bar{t}WW$, together with their fractional QCD scale and PDF uncertainties, calculated with MadGraph5_aMC@NLO [85]. Theory uncertainties on tH cross section at 13 TeV are not yet available. A conservative overall normalisation uncertainty of 50% is assigned to these background processes in the fit. Due to their small contribution in the Signal Region, theory uncertainties are considered as normalisation only.

6.8 Results

The binned distributions of the BDT output discriminants for each region are combined as inputs to a test statistic to search for the presence of a signal. The analysis uses the profile-likelihood method, described in Chapter 4, which allows to reduce the impact of systematic uncertainties on the search sensitivity by exploiting the highly populated background-dominated control regions that are included in the fit.

The fit is performed under the signal-plus-background hypothesis to obtain the value of the signal strength μ, as a free parameter in the fit. Contributions from $t\bar{t}$+jets, diboson and $t\bar{t}V$ backgrounds are constrained simultaneously by the fit in the signal region and the three dedicated control regions, each enriched of the relevant process, and their normalisations are inferred. Production of $t\bar{t}H$ is assumed to be SM-like except for an overall scaling of the cross section, while single top-Higgs boson associated production is fixed to its SM rate.

First the expected performance of the fit is studied using the Asimov dataset (see also Section 4.3.5), then a partially blind fit is performed in order to test the description of the systematic uncertainties with a sufficient amount of data. The blinding procedure is the same adopted for the validation of the input variables and described in Section 6.6.4. The blind fit is a step for a final validation of the main backgrounds. Once all the control regions show a good modelling and the robustness of the analysis in extracting each normalisation is proven, the complete dataset is used and results are obtained.

6.8.1 Expected performance of the fit

The expected performance of the fit has been studied using the so-called Asimov dataset, described in Section 4.3.5. As expected, the nuisance parameters corresponding to systematic uncertainties are all centred on zero and the normalisation scale factors are all centred around 1. The majority of the parameters, especially the ones related to the detector performance, are not constrained with respect to their prior uncertainties. The constraints on the parameters related to the shape of the diboson background can be expected given the large amount of events in the dedicated control region and the conservative way in which considered variations have been defined; as a result, the normalisation of the diboson background can be known to a few percent level. The shape and normalisation of the $t\bar{t}$+jets source are also well constrained, given the large amount of Asimov data in the dedicated control region. On the other hand, a minor constraint of the $t\bar{t}W$ background is expected because of the low purity of the process and the low number of events in the corresponding control region.

For what concerns the analysis sensitivity, the leading source of systematic uncertainties are those that show a high level of correlation with the signal strength. These are the uncertainty on the normalisation of the $t\bar{t}$+jets background, as well as its shape, and the uncertainty on the $t\bar{t}W$ process normalisation. This is also confirmed by the ranking plot in Fig. 6.13; the parameter with the largest impact on the extracted signal strength is, by far, the normalisation of the $t\bar{t}W$ background, followed by the uncertainties on the $t\bar{t}Z$ process and the pileup modelling. The impact of a nuisance parameter on the fit $\mu_{t\bar{t}H}$ is calculated by fixing the corresponding nuisance parameter at $\hat{\theta} \pm \sigma_{\hat{\theta}}$ and performing the fit again. Here $\hat{\theta}$ is the fitted value of the nuisance parameter and $\sigma_{\hat{\theta}}$ is its post-fit uncertainty. The difference between the default and the modified μ, $\Delta\mu$, represents the effect on μ of this particular systematic uncertainty.

In the presence of a signal compatible with the SM prediction, the expected uncertainty is 0.75 in case all sources of systematics are considered, and 0.66 in case only the data statistical constitution and the template statistics are taken into account.component The expected signal significance is 1.41σ, in case of a SM signal, which corresponds to a p-value of 16%, while the expected (median, for the background-only hypothesis) 95% CL upper limits are 1.7 and 1.2, when all systematic uncertainties are included or not (“statistics only”), respectively.

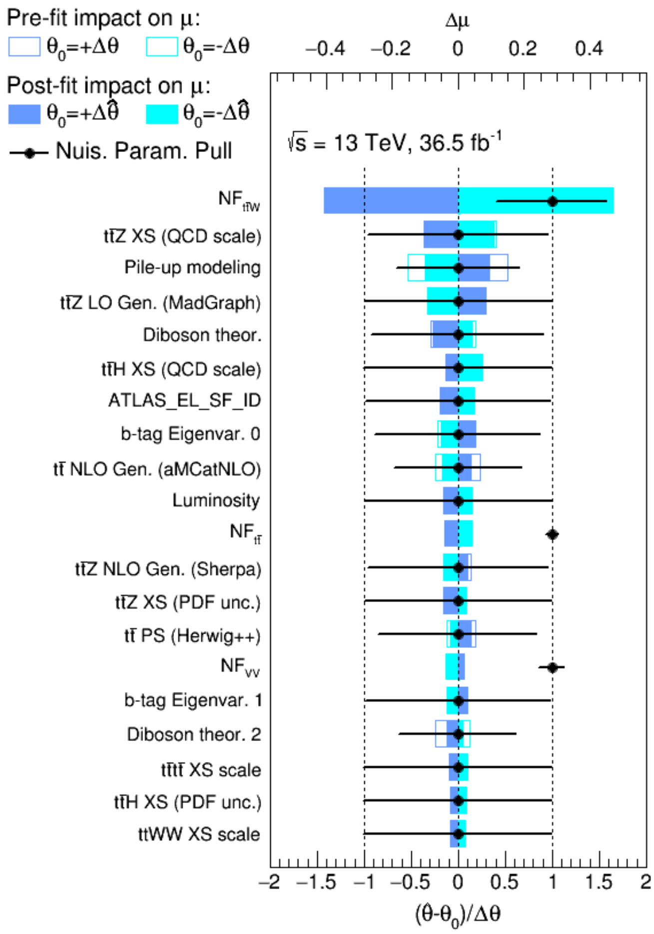


Figure 6.13: Ranking of nuisance parameters obtained from the fit to the Asimov dataset, corresponding to the sources of systematic uncertainty with the largest impact on the fitted signal strength μ. The points, which are drawn conforming to the scale of the bottom axis, show the deviation of each of the fitted nuisance parameters $\hat{\theta}$ from θ_0, which is the nominal value of that nuisance parameter, in units of the pre-fit standard deviation $\Delta\theta$. The plain blue area represents the pre-fit impact on μ and the blue line its post-fit impact. The error bars show the post-fit uncertainties σ_θ, which have size close to one if the data do not provide any further constraint on that uncertainty. Conversely, an error bar for σ_θ smaller than one indicates a reduction with respect to the original uncertainty. The nuisance parameters are sorted according to their post-fit impact $\Delta\theta$ (top horizontal scale). The $t\bar{t}$+jets, VV and $t\bar{t}W$ scale factors (NF) show the fitted values and uncertainties of the normalisation parameters that are freely floating in the fit. These normalisation parameters have a pre-fit value of one.

Table 6.12: Event yields from simulated backgrounds and the signal as well as measured events in the control (CR) and signal regions (SR) used for the fit (post-fit). The quoted uncertainties include statistical and systematical effects. The sum of all contributions may slightly differ from the total value due to rounding. The $t\bar{t}+\gamma$ yields are included in the $t\bar{t}$+jets background after the overlap removal between the two samples.

	VV CR	$t\bar{t}$ CR	$t\bar{t}W$ CR	3ℓ SR
$t\bar{t}W$	17.6 ± 4.6	41 ± 11	22.2 ± 7.7	96 ± 25
$t\bar{t}(Z/\gamma^*)$	240 ± 35	25.1 ± 3.7	1.54 ± 0.73	45.6 ± 5.4
Diboson	770 ± 160	53 ± 11	0.63 ± 0.54	112 ± 24
$t\bar{t}$	268 ± 58	1320 ± 120	8.9 ± 5.8	89 ± 19
Z+jet	627 ± 92	130 ± 28	0.027 ± 0.025	14.8 ± 4.0
single top	8.6 ± 2.7	33.1 ± 3.5	0.24 ± 0.25	4.0 ± 1.1
Rare	74 ± 24	7.3 ± 1.4	0.36 ± 0.12	17.6 ± 3.6
Total background	2010 ± 180	1610 ± 120	33.8 ± 9.6	381 ± 29
$t\bar{t}H$ (SM)	5.7 ± 6.6	12 ± 14	0.60 ± 0.68	17 ± 19
Data	1931	1634	31	389

6.8.2 Fit to data in the analysis regions

The yields in the four analysis regions at the integrated luminosity of 36.5 fb^{-1} before and after the fit are summarised in Tables 6.7 and 6.12. In each region, the post-fit variations of background and signal events with respect to the pre-fit values are modest except for the $t\bar{t}W$ background, which increases by a factor of 2.8. However, the available data can constrain the $t\bar{t}W$ normalisation only within an uncertainty of $\sim$ 25%. This result is in agreement with the measurement of 2.3 ± 1.3, performed by ATLAS at 3.2 fb^{-1} [186].

The results of the full fit under the signal-plus-background hypothesis is shown in Figures 6.14 and 6.15.

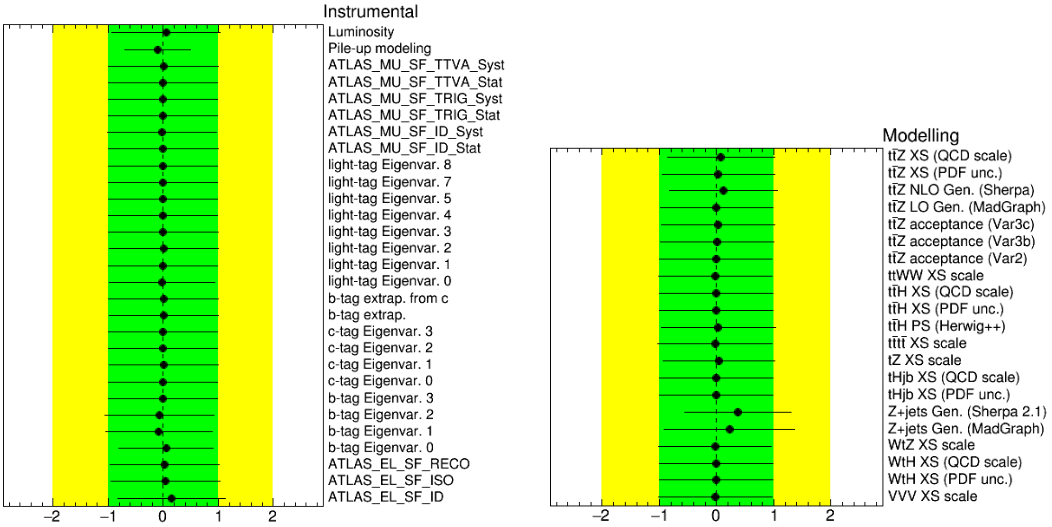


Figure 6.14: Fitted nuisance parameters and normalisation scale factors to data under the signal-plus-background hypothesis, for (a) detector-related uncertainties and (b) uncertainties related to the simulated background modelling and normalisation.

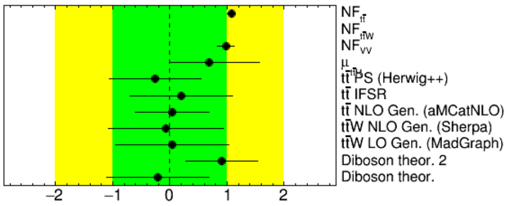


Figure 6.15: Fitted nuisance parameters and normalisation scale factors to data under the signal-plus-background hypothesis, for modelling uncertainties and the normalisation scale factors (NF) of the main sources of background.

Table 6.13: Best-fit value of the signal strength μ and 95% CL upper limits obtained using the CL$_s$ method. For the best-fit and expected-fit values, the first uncertainty is statistical and the second systematic. For the expected upper limits, the median is reported, and the uncertainties give the 68% expected range. The signal injected upper limit is the median expected upper limit that is set by this procedure in the presence of a SM $t\bar{t}H$ signal.

Luminosity (fb^{-1})	Best fit $\mu_{t\bar{t}H}$	Expected fit $\mu_{t\bar{t}H}$	Observed (expected)	Signal injected
			95% CL upper limit	95% CL upper limit
36.5	$0.68\ ^{+0.56}_{-0.56}\ ^{+0.58}_{-0.39}$	$1.0\ ^{+0.66}_{-0.63}\ ^{+0.42}_{-0.36}$	$2.3\ (1.7\ ^{+0.8}_{-0.5})$	2.5

Figure 6.16 shows the BDT output distributions for data and the predictions in each analysis region after the fit to data. The relative uncertainties decrease significantly in all regions due to the constraints provided by the data, exploiting the correlations between the uncertainties in the four analysis regions. As it can be seen the fit significantly improves the agreement especially in the $t\bar{t}W$ region, mainly due to the adjustment of the $t\bar{t}W$ normalisation scale factor.

Figure 6.18 presents the post-fit plots of the seven discriminating variables entering the final BDT and used to validate the fit result, in the signal region. A good data/MC agreement is observed and no mismodelling issues have been found. Additional plots can be found in Appendix D.

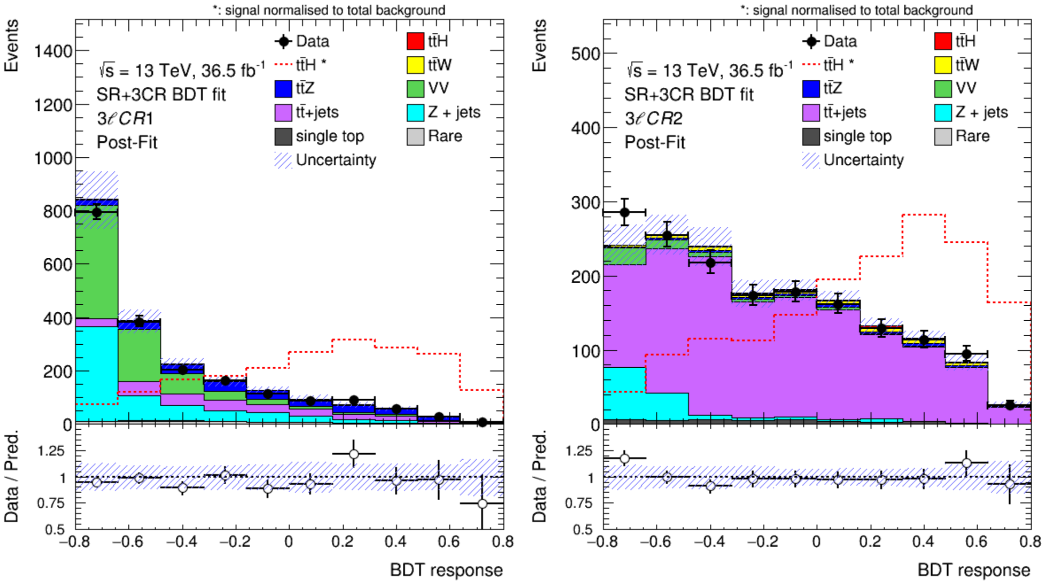
The signal strength $\mu_{t\bar{t}H}$ is measured to be $\mu_{t\bar{t}H} = 0.68^{+0.89}_{-0.68}$. The normalisation factors are measured to be:

$$\begin{aligned} \mathrm{NF}_{t\bar{t}+\mathrm{jets}} &= 1.07 \pm 0.09\,, \\ \mathrm{NF}_{t\bar{t}W} &= 2.82 \pm 0.79\,, \\ \mathrm{NF}_{VV} &= 0.98 \pm 0.16\,. \end{aligned} \tag{6.9}$$

The observed (expected) significance of the signal is 0.84 (1.41) standard deviations corresponding to an observed (expected) p-value of 40% (16%). The observed and expected limits are summarised in Table 6.13. A $t\bar{t}H$ signal 2.3 times larger than predicted by the SM is excluded at 95% CL. A signal 1.7 times larger than the signal of a SM Higgs boson is expected to be excluded.

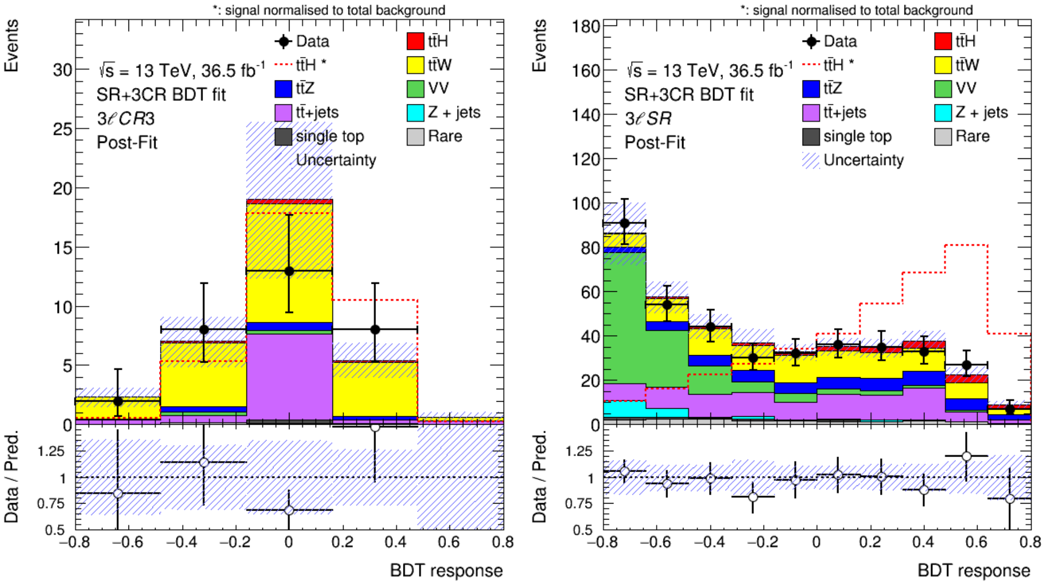
Figure 6.18 shows the effect of the major systematic uncertainties on the fitted value of μ and the constraints provided by the data. The ranking, from top to bottom, is determined by the post-fit impact on μ. This is also shown in Table 6.14 according to the different sources of systematic effects: the impact table refers only to the systematic uncertainties considered in the fit model; the effect of the normalisation factors is not included and can be directly deduced from the ranking plot.

By far the largest uncertainty on the value of $\mu_{t\bar{t}H}$ is due to the estimate of the $t\bar{t}W$ background. The $t\bar{t}W$ background cross section is larger by almost a factor 3 in the fit and shows a high impact on the fitted $\mu_{t\bar{t}H}$. The $t\bar{t}$+jets background normalisation increases by 7% respect to the prediction, with a normalisation uncertainty of 9%, and has the second largest impact on $\mu_{t\bar{t}H}$, together with the $t\bar{t}Z$ background modelling. The diboson background cross section is in very good agreement with the MC prediction and has, instead,



(a) VV CR

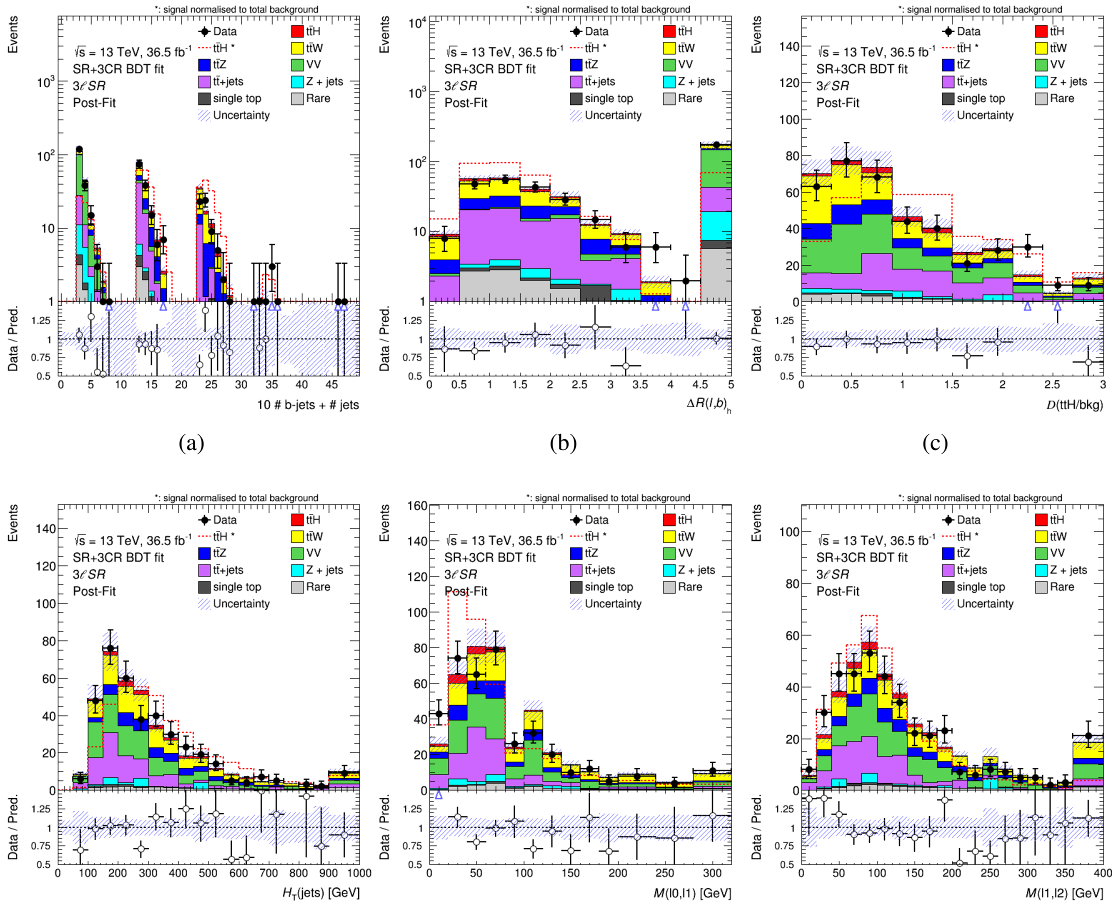
(b) $t\bar{t}$ CR



(c) $t\bar{t}W$ CR

(d) SR

Figure 6.16: Comparison between data and prediction for the BDT discriminant in the signal region and the three control regions after the fit. The fit is performed under the signal-plus-background hypothesis. The binning is the same as that used in the fit. The dashed red line shows the BDT distribution for the signal normalised to the total background yield. The bottom panels display the ratios of data to the total prediction. The hashed areas represent the total uncertainty in the background predictions.



(a) (b) (c)

(d) (e) (f)

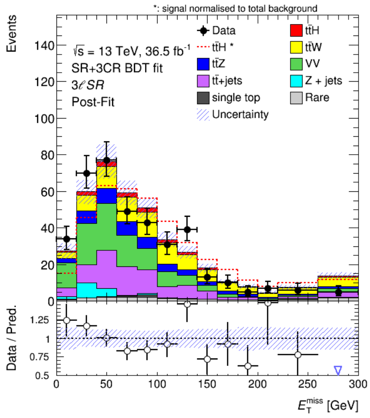


Figure 6.18 Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the signal region after the fitting procedure ("post-fit"). The uncertainty band contains the statistical and systematic contribution.

()

Table 6.14: Summary of the effects of systematic uncertainties on μ, expressed in terms of $\Delta\mu$ of the fitted value of μ and sorted according to their post-fit effect. The total systematic uncertainty is obtained from the subtraction in quadrature of the statistical uncertainty from the total uncertainty. Due to (anti)correlations between the different sources of uncertainties, the total systematic uncertainty can be different from the sum in quadrature of the individual sources.

Uncertainty Source	$\Delta\mu$	
$t\bar{t}$+jets modelling	+0.13	−0.03
diboson modelling	+0.13	−0.03
$t\bar{t}W$ modelling	+0.07	−0.05
$t\bar{t}Z$ modelling	+0.24	−0.13
$t\bar{t}H$ modelling	+0.05	−0.04
Jet-vertex association, pileup modelling	+0.23	−0.12
Jet energy scale and resolution	+0.12	−0.05
Luminosity	+0.05	−0.03
Jet flavour tagging	+0.08	−0.03
Light lepton (e, μ) and τ_{had} ID, isolation, trigger	+0.11	−0.04
Other background modelling	+0.05	−0.05
Total systematic uncertainty	+0.53	−0.12

small impact on the fitted $\mu_{t\bar{t}H}$. The value of $\mu_{t\bar{t}H}$ also depends on the assumed SM cross section and acceptance for $t\bar{t}H$ production: the uncertainties in these quantities have an impact on $\mu_{t\bar{t}H}$ comparable to the equivalent uncertainties in $t\bar{t}$+jets and VV modelling.

The most important detector-related systematic uncertainty arises from the identification efficiency of electrons. Luminosity uncertainty and uncertainties in lepton reconstruction and trigger efficiencies have very small impact.

The fit value $\mu_{t\bar{t}H} = 0.68^{+0.89}_{-0.68}$ agrees with the previous measurements performed by ATLAS at 13.2 fb^{-1}in the $t\bar{t}H$ multilepton channels and with the measured performed by CMS. The ATLAS result at 13.2 fb^{-1}in the multilepton channels is $\mu_{t\bar{t}H} = 2.5^{+1.3}_{-1.1}$ [53], showing a slight signal excess in those channels: the measurement in the final state with exactly three leptons is, instead, $\mu_{t\bar{t}H} = 0.5 \pm 1.7$, which also agrees with the measurement performed in this thesis.

The $t\bar{t}H$ combined value measured by ATLAS is $\mu_{t\bar{t}H} = 1.7 \pm 0.8$ [52]. while the combined $t\bar{t}H$ signal strength measured by CMS is $\mu_{t\bar{t}H} = 2.0 \pm 0.8$ [51],

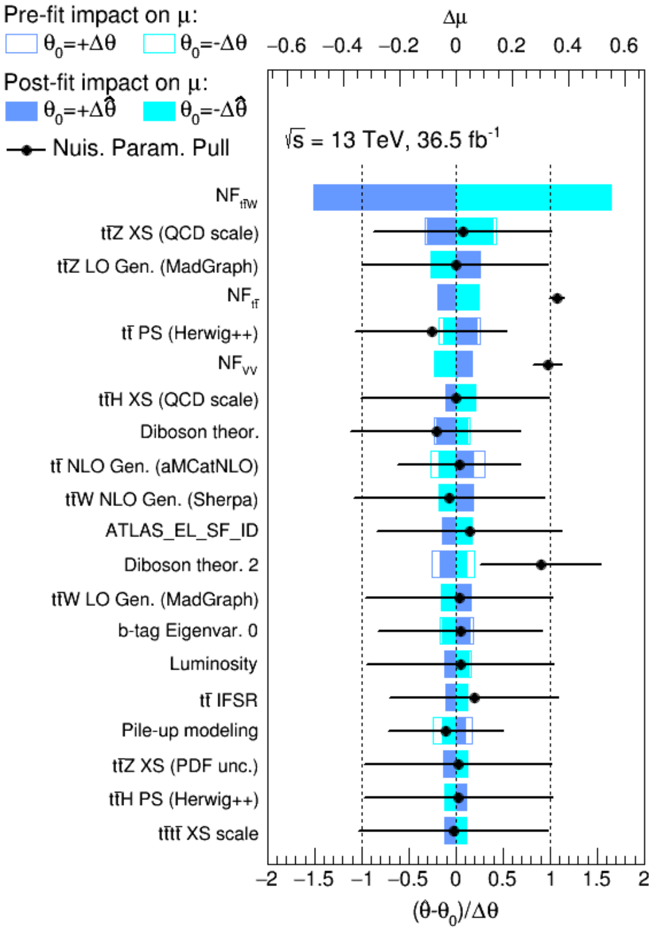


Figure 6.18: Ranking of nuisance parameters obtained from the fit to data.

6.9 Prospects at 150 fb^{-1}

A projection at higher integrated luminosity is shown in this section: the luminosity benchmark considered, according to the current plan of the LHC, is 150 $\mathrm{fb}^{-1}$and corresponds to the expected amount of data by the end of the 2018 at the same centre-of-mass energy of 13 TeV. All the systematic uncertainties are considered to be the same as the fit performed at the current luminosity of 36.5 fb^{-1}. The $t\bar{t}$+jets, $t\bar{t}W$ and diboson background normalisations are well constrained by the fit within a stat-only uncertainty of $\sim 2\%$, $\sim 20\%$ and $\sim 3\%$ respectively, compared to the $\sim 3\%$, $\sim 50\%$ and $\sim 5\%$ at 36.5 fb^{-1}.

The expected fit value and 95% CL upper limit on $\mu_{t\bar{t}H}$ at the current and future integrated luminosities are shown in Table 6.15. With higher luminosities the fit is able to constrain $\mu_{t\bar{t}H}$ more: at 150 fb^{-1}(~ 4 times the current luminosity) an improvement of 40% is expected in constraining the signal strength. The expected compatibility with the no-$t\bar{t}H$ hypothesis is 2.4σ.

Table 6.15: Expected fit value of the signal strength μ and expected 95% CL upper limits obtained using the CL_s method. The table shows fit results for different integrated luminosities.

Luminosity (fb^{-1})	Expected fit $\mu_{t\bar{t}H}$	Expected 95% CL upper limit	Signal injected 95% CL upper limit
36.5	$1.00\,^{+0.66}_{-0.63}\,^{+0.42}_{-0.36}$	$1.7\,^{+0.8}_{-0.5}$	2.2
150	$1.00\,^{+0.32}_{-0.31}\,^{+0.35}_{-0.31}$	$0.9\,^{+0.4}_{-0.2}$	–

About long future prospects, it is difficult to provide reliable expected sensitivities, for two main reasons. First, the analysis relies on data in order to reduce the impact of the systematic uncertainties: more data will allow to introduce new regions to better constrain the main sources of background. Secondly, a better knowledge of the background modelling, like new Monte Carlo techniques for event generation, can strongly affect the sensitivity of the results.

Therefore, in future bigger data statistics may allow to use a more complex approach, exploiting a more refined strategy. Dedicated BDTs could be also trained in the different regions in order to better resolve the specific backgrounds, new event reconstruction variables could be introduced and a more accurate pseudo matrix-element discriminator employed, with the prospect to introduce many discriminators targeting different background processes.

CHAPTER 7

Summary and conclusions

With the discovery of the Higgs boson, one main goal of the LHC program has been fulfilled and an important step towards the comprehension of nature was made. Thereafter the focus shifted to the precise tests of the Higgs boson properties: the presence of new physics processes could be revealed by possible deviations from the SM predictions.

One of the parameters that can help to infer about new processes intervening at high-energy scales (10^{15}-10^{16} GeV, depending on the model), is the Yukawa coupling between the Higgs boson and the top quark. The top Yukawa coupling influences the evolution of the effective Higgs potential with the energy: any tension between the values allowed by the Standard Model and the observation would demand for new physics to solve the inconsistency.

Unlike the gluon fusion production process and the $H \to \gamma\gamma$ decay mode, where the top Yukawa coupling enters indirectly via top-quark loops, the associated production of the Higgs boson with a top quark pair ($t\bar{t}H$) allows for a direct measurement and is the main topic of this dissertation.

The two analyses presented in this thesis describe two complementary $t\bar{t}H$ searches, in the fully hadronic and the multilepton channels, at a centre-of-mass energy of 8 and 13 TeV, respectively. The data, collected by the ATLAS detector and used for the analyses, amount to 20.3 fb^{-1} of pp collisions for the fully hadronic analysis, and 36.5 fb^{-1} for the multilepton one. Both analyses exploit multivariate techniques to provide better discrimination between the signal and background sources, and several signal and control regions to reduce the impact of background uncertainties on the analysis sensitivity. Pseudo matrix-element discriminators $\mathcal{D}$ are constructed and play a crucial role in the discrimination of the multijet background in the fully hadronic channel, and the $t\bar{t}V$ ($t\bar{t}$ plus a vector boson, W or Z) and $t\bar{t}$+jets backgrounds in the multilepton one; it identifies signal-like events by partially reconstructing resonances and exploiting the kinematic characteristics of the signal and background processes. The discriminator is employed as input variable to the multivariate classifier among other variables; the best set of variables is determined through an iterative optimisation procedure that gives the maximal signal-to-background rejection.

The 8 TeV analysis is the first fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) search ever performed and uses a novel data-driven technique to determine the multijet background contribution. The 13 TeV analysis, instead, is the first multilepton $t\bar{t}H$ search which adopts a multivariate technique to assess the top Yukawa coupling.

In the $t\bar{t}H$ fully hadronic search, the selected events are required to have at least six jets, of which at least three b-tagged, and are further categorised into six regions, according to their jet and b-tag multiplicities (nj, mb): two control regions ($6j$, $3b$) and ($6j$, $\geq 4b$), four signal regions ($7j$, $3b$), ($7j$, $\geq 4b$), ($\geq 8j$, $3b$) and ($\geq 8j$, $\geq 4b$). The dominant background source is the multijet process and is determined through the extrapolation of a data-driven Tag Rate Function from regions with two b-tagged jets. Six dedicated boosted decision trees (BDTs) are optimised in the six analysis regions and combined as inputs of the profile likelihood fit to search for the presence of a signal.

The best fit value for the signal strength $\mu_{t\bar{t}H}$, expressed in units of the SM expectations, in the fully hadronic $t\bar{t}H$ decay mode is measured to be 1.6 ± 2.6: it corresponds to an observed (expected) excess of 0.6 (0.4) standard deviations with respect to the background-only hypothesis. Translated into a 95% confidence level (C.L.) upper limit of the $t\bar{t}H$ cross section, $\mu_{t\bar{t}H} < 6.4$; a signal 5.4 times larger than the signal of a SM Higgs boson was expected to be excluded.

The largest impact on the sensitivity of the analysis comes from the limited knowledge of the irreducible $t\bar{t}+b\bar{b}$ background, driven by the large uncertainties considered for the normalisation and shape modelling. At the time of the analysis, the first theoretical calculations of $t\bar{t}+b\bar{b}$ background at NLO became available together with first direct measurements, but a conservative approach on the treatment of such backgrounds was adopted. Employing several control regions in the fit model helps to reduce the uncertainties on the $t\bar{t} + b\bar{b}$ background, and suggests that future measurements with a larger amount of data collected would be able to reduce the impact on the $t\bar{t}H$ measurement in the fully hadronic channel, besides possible improvements on the theoretical constraint to the $t\bar{t} + b\bar{b}$ cross section. Moreover, improvements in the estimation of the multijet background, the second-major source of uncertainty, will improve the final result other than being of great interest for further searches of new physics in fully hadronic channels.

Combining all the $t\bar{t}H$ ($H \to b\bar{b}$) searches at 8 TeV gives a signal strength of $\mu = 1.4 \pm 1.0$, and an observed (expected) significance of 2.3σ (1.5σ), close to the 3σ needed to claim the evidence of such process. The combined results demonstrate that this analysis gives a non-negligible contribution, resulting in the most stringent upper limit, $\mu_{t\bar{t}H} < 3.3$, among the $t\bar{t}H$ searches in ATLAS or CMS.

The $t\bar{t}H$ multilepton search requires exactly three light leptons with a total charge of ± 1 and at least three reconstructed jets. The $t\bar{t}Z$ and diboson contributions are suppressed by vetoing events with an opposite-sign same-flavour lepton pair with the invariant mass within 10 GeV of the Z boson mass. The background is dominated by diboson, $t\bar{t}$+jets and $t\bar{t}V$ processes. Three orthogonal control regions are defined by changing the Z mass (VV CR), the lepton quality ($t\bar{t}$+jets CR) or the jet multiplicity requirements ($t\bar{t}W$ CR). A BDT is optimised in the signal region, which is combined with the three control regions to a profile likelihood fit. The background-enriched control regions are exploited by the fit to determine the corresponding background normalisations.

The signal strength is measured to be $0.68^{+0.89}_{-0.68}$. The observed (expected) significance of the signal is 0.84 (1.41) standard deviations corresponding to an observed (expected) p-value of 40% (16%). A $t\bar{t}H$ signal 2.3 times larger than predicted by the SM is excluded at 95% CL.

By far the largest impact on the measured signal strength is given by the normalisation of the $t\bar{t}W$ background. The theoretical prediction of the $t\bar{t}W$ process is underestimated by data in the $t\bar{t}W$ control region by a factor 2.8, with a total uncertainty of 30%: the value is in agreement with the measurement of 2.3 ± 1.3, performed by ATLAS. With the current luminosity, the estimation of such background is strongly affected by the poor amount of data events populating the $t\bar{t}W$ control region. The 45 fb^{-1} expected to be delivered by ATLAS in 2017 will double the events in this region, allowing to better constrain the $t\bar{t}W$ cross section and to reduce its impact on the $t\bar{t}H$ measurement. Despite being the main source of non-prompt leptons, which are not expected to be well modelled by MC simulations in terms of shape and normalisation, the $t\bar{t}$+jets background normalisation agrees with the theoretical prediction, with a normalisation uncertainty of 9%. It has the second largest impact on $\mu_{t\bar{t}H}$, together with the $t\bar{t}Z$ background modelling, also well modelled by simulation within the systematic uncertainties. Finally, the diboson background cross section is in very good agreement with the MC prediction but has a low impact on the fitted $\mu_{t\bar{t}H}$.

In future bigger data statistics may allow to use a more complex approach in the multilepton analysis. Dedicated BDTs could be trained in the different regions for a better resolution of the specific backgrounds and more accurate pseudo matrix-element discriminators could be employed, targeting different background processes.

A fully hadronic analysis with data from Run 2 could improve the data-driven method for extracting the multijet background, employing novel *b*-jet triggers at the beginning of the chain of the analysis. The employment of an improved pseudo matrix-element discriminator may be also considered, exploiting the pseudo continuous *b*-tagging probability, which was a crucial ingredient in the multilepton channel.

About long future prospects, it is difficult to provide reliable expected sensitivities. Both analyses rely considerably on data in order to reduce the impact of the systematic uncertainties: more data will allow to introduce new regions to better constrain the main sources of background. In addition, a better knowledge of the background modelling, like new data-driven methods and Monte Carlo techniques for event generation, can strongly affect the sensitivity of the results, especially in the fully hadronic analysis, where the systematic uncertainties dominate the total uncertainty on the signal strength $\mu_{t\bar{t}H}$.

Bibliography

[1] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579 (cit. on pp. 3, 5).

[2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264 (cit. on pp. 3, 5).

[3] A. Salam, "Weak and Electromagnetic Interactions", *Elementary particle theory, Relativistic groups and analyticity*, Proceedings of the Eighth Nobel Symposium, (Aspenäsgarden, Lerum, 19th–25th May 1968), ed. by N. Svartholm, Almquist & Wiksell, 1968 367 (cit. on pp. 3, 5).

[4] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508 (cit. on pp. 3, 5).

[5] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) 1156 (cit. on pp. 3, 5).

[6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321 (cit. on pp. 3, 5).

[7] C. Patrignani et al., *Review of Particle Physics*, Chin. Phys. **C40** (2016) 100001 (cit. on pp. 4, 8, 12, 106).

[8] CDF Collaboration, F. Abe et al., *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74** (1995) 2626, arXiv: `hep-ex/9503002 [hep-ex]` (cit. on p. 12).

[9] DØ Collaboration, V. M. Abazov et al., *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632, arXiv: `hep-ex/9503003 [hep-ex]` (cit. on p. 12).

[10] *First combination of Tevatron and LHC measurements of the top-quark mass*, ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416, FERMILAB-TM-2582-E (2014), arXiv: `1403.4427 [hep-ex]` (cit. on p. 12).

[11] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612, arXiv: `1111.5869 [hep-ph]` (cit. on pp. 12, 66, 101).

[12] P. Bärnreuther et al., *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001, arXiv: `1204.5201 [hep-ph]` (cit. on pp. 12, 66, 101).

[13] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054, arXiv: `1207.0236 [hep-ph]` (cit. on pp. 12, 66, 101).

[14] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080, arXiv: `1210.6832 [hep-ph]` (cit. on pp. 12, 66, 101).

[15] M. Czakon, P. Fiedler and A. Mitov,
The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$,
Phys. Rev. Lett. **110** (2013) 252004, arXiv: `1303.6254 [hep-ph]` (cit. on pp. 12, 66, 101).

[16] M. Czakon and A. Mitov,
Top++: a program for the calculation of the top-pair cross-section at hadron colliders,
185 (2014) 2930, arXiv: `1112.5675 [hep-ph]` (cit. on pp. 12, 66, 101).

[17] LHCTopWG - LHC Top Physics Working Group, 2016,
URL: `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWG` (cit. on p. 13).

[18] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
arXiv: `1207.7214 [hep-ex]` (cit. on pp. 14, 17, 18).

[19] CMS Collaboration,
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,
Phys. Lett. B **716** (2012) 30, arXiv: `1207.7235v1 [hep-ex]` (cit. on pp. 14, 17).

[20] J. R. Ellis and G. L. Fogli,
New bounds on m(t) and first bounds on M(H) from precision electroweak data,
Phys. Lett. **B249** (1990) 543 (cit. on pp. 14, 15).

[21] J. R. Ellis, G. L. Fogli and E. Lisi, *Bounds on M(H) from electroweak radiative corrections*,
Phys. Lett. **B274** (1992) 456 (cit. on pp. 14, 15).

[22] J. R. Ellis, G. L. Fogli and E. Lisi,
Indirect bounds on the Higgs boson mass from precision electroweak data,
Phys. Lett. **B318** (1993) 148 (cit. on pp. 14, 15).

[23] T. Hambye and K. Riesselmann, *Matching conditions and Higgs mass upper bounds revisited*,
Phys. Rev. **D55** (1997) 7255, arXiv: `hep-ph/9610272 [hep-ph]` (cit. on pp. 14, 15).

[24] R. Barate et al., *Search for the standard model Higgs boson at LEP*, Phys. Lett. **B565** (2003) 61,
arXiv: `hep-ex/0306033 [hep-ex]` (cit. on p. 14).

[25] Tevatron New Physics Higgs Working Group, CDF, DØ, "Updated Combination of CDF and DØ Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data", 2012,
arXiv: `1207.0449 [hep-ex]` (cit. on p. 14).

[26] H. Flächer et al.,
Revisiting the Global Electroweak Fit of the Standard Model and Beyond with Gfitter,
Eur. Phys. J. **C60** (2009) 543, [Erratum: Eur. Phys. J. C 71 (2011) 1718],
arXiv: `0811.0009 [hep-ph]` (cit. on p. 14).

[27] M. Baak et al., *Updated Status of the Global Electroweak Fit and Constraints on New Physics*,
Eur. Phys. J. **C72** (2012) 2003, arXiv: `1107.0975 [hep-ph]` (cit. on p. 14).

[28] F. Bezrukov et al., *Higgs Boson Mass and New Physics*, JHEP **10** (2012) 140, [,275(2012)],
arXiv: `1205.2893 [hep-ph]` (cit. on p. 16).

[29] M. Shaposhnikov, *Cosmology: theory*, PoS **EPS-HEP2013** (2013) 155,
arXiv: `1311.4979 [hep-ph]` (cit. on pp. 16, 17).

[30] D. Buttazzo et al., *Investigating the near-criticality of the Higgs boson*, JHEP **12** (2013) 089,
arXiv: `1307.3536 [hep-ph]` (cit. on pp. 16, 17, 22, 23).

[31] ATLAS Collaboration, *Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data*, (2013), ATLAS-CONF-2013-013, URL: `https://cds.cern.ch/record/1523699` (cit. on p. 18).

[32] LHC Higgs Cross Section Working Group, 2016, URL: `https://twiki.cern.ch/twiki/bin/view/LHCPhysics` (cit. on pp. 19, 100, 101).

[33] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, FERMILAB-FN-1025-T (CERN, Geneva, 2016), arXiv: `1610.07922 [hep-ph]` (cit. on pp. 19, 101).

[34] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, CERN-2013-004 (CERN, Geneva, 2013), arXiv: `1307.1347 [hep-ph]` (cit. on pp. 19, 100, 101, 104).

[35] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, CERN-2012-002 (CERN, Geneva, 2012), arXiv: `1201.3084 [hep-ph]` (cit. on pp. 19, 101).

[36] A. Bredenstein et al., *Radiative corrections to the semileptonic and hadronic Higgs-boson decays H —> W W / Z Z —> 4 fermions*, JHEP **02** (2007) 080, arXiv: `hep-ph/0611234 [hep-ph]` (cit. on p. 19).

[37] B. Odom et al., *New Measurement of the Electron Magnetic Moment Using a One-Electron Quantum Cyclotron*, Phys. Rev. Lett. **97** (3 2006) 030801, URL: `http://link.aps.org/doi/10.1103/PhysRevLett.97.030801` (cit. on p. 20).

[38] J. Wess and B. Zumino, *Supergauge Transformations in Four-Dimensions*, Nucl. Phys. **B70** (1974) 39 (cit. on p. 21).

[39] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1, arXiv: `1106.0034 [hep-ph]` (cit. on p. 21).

[40] 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) 191803, arXiv: `1503.07589 [hep-ex]` (cit. on p. 21).

[41] M. Baak et al., *The global electroweak fit at NNLO and prospects for the LHC and ILC*, Eur. Phys. J. **C74** (2014) 3046, arXiv: `1407.3792 [hep-ph]` (cit. on p. 21).

[42] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, J. Exp. Theor. Phys. **120** (2015) 335, [Zh. Eksp. Teor. Fiz.147,389(2015)], arXiv: `1411.1923 [hep-ph]` (cit. on pp. 21–23).

[43] A. Kobakhidze and A. Spencer-Smith, *Electroweak Vacuum (In)Stability in an Inflationary Universe*, Phys. Lett. **B722** (2013) 130, arXiv: `1301.2846 [hep-ph]` (cit. on p. 24).

[44] K. Enqvist, T. Meriniemi and S. Nurmi, *Higgs Dynamics during Inflation*, JCAP **1407** (2014) 025, arXiv: `1404.3699 [hep-ph]` (cit. on p. 24).

[45] ATLAS Collaboration, *Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector*, Phys. Lett. B **749** (2015) 519, arXiv: `1506.05988 [hep-ex]` (cit. on p. 24).

[46] ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) 349, arXiv: `1503.05066 [hep-ex]` (cit. on pp. 24, 66, 67, 88).

[47] ATLAS Collaboration, *Search for $H \rightarrow \gamma\gamma$ produced in association with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs boson using data taken at 7 TeV and 8 TeV with the ATLAS detector*, Phys. Lett. B **740** (2015) 222, arXiv: `1409.3122 [hep-ex]` (cit. on p. 24).

[48] CMS Collaboration,
Search for the associated production of the Higgs boson with a top-quark pair, JHEP **1409** (2014) 087, arXiv: `1408.1682 [hep-ex]` (cit. on p. 24).

[49] CMS Collaboration, *Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method*, Eur. Phys. J. **C75** (2015) 251, arXiv: `1502.02485 [hep-ex]` (cit. on p. 24).

[50] ATLAS and CMS Collaborations,
Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV, (2015), ATLAS-CONF-2015-044, CMS-PAS-HIG-15-002 (cit. on pp. 24, 99).

[51] *Search for associated production of Higgs bosons and top quarks in multilepton final states at* $\sqrt{s} = 13$ TeV, CMS-PAS-HIG-16-022 (2016), URL: `https://cds.cern.ch/record/2205282` (cit. on pp. 24, 132).

[52] *Combination of the searches for Higgs boson production in association with top quarks in the $\gamma\gamma$, multilepton, and $b\bar{b}$ decay channels at $\sqrt{s}$=13 TeV with the ATLAS Detector*, (2016), ATLAS-CONF-2016-068, URL: `http://cds.cern.ch/record/2206211` (cit. on pp. 24, 132).

[53] *Search for the Associated Production of a Higgs Boson and a Top Quark Pair in Multilepton Final States with the ATLAS Detector*, (2016), ATLAS-CONF-2016-058, URL: `http://cds.cern.ch/record/2206153` (cit. on pp. 24, 132).

[54] *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, (2016), ATLAS-CONF-2016-080, URL: `http://cds.cern.ch/record/2206255` (cit. on p. 24).

[55] *Measurement of fiducial, differential and production cross sections in the $H \rightarrow \gamma\gamma$ decay channel with 13.3 fb^{-1} of 13 TeV proton-proton collision data with the ATLAS detector*, (2016), ATLAS-CONF-2016-067, URL: `http://cds.cern.ch/record/2206210` (cit. on p. 24).

[56] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*, Technical Design Report ATLAS, CERN, 1999, URL: `https://cds.cern.ch/record/391176` (cit. on pp. 25, 37).

[57] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 2*, Technical Design Report ATLAS, CERN, 1999,
URL: `https://cds.cern.ch/record/391177` (cit. on p. 25).

[58] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 25, 30, 35–37, 39, 69).

[59] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001 (cit. on pp. 25, 26, 28).

[60] ATLAS Experiment @2011 CERN, ATLAS Photos,
URL: `http://www.atlas.ch/photos/index.html` (cit. on p. 27).

[61] ATLAS Collaboration, *Luminosity Public Results*, URL:
`https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`
(cit. on pp. 28, 65, 101).

[62] J.-L. Caron, “ATLAS detector in A4 format with English captions.. Détecteur ATLAS format A4 avec légende et caractéristiques en anglais.”,
AC Collection. Legacy of AC. Pictures from 1992 to 2002., 1998,
URL: `https://cds.cern.ch/record/841458` (cit. on p. 30).

[63] M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*,
CERN-LHCC-2010-013 (2010), URL: `https://cds.cern.ch/record/1291633`
(cit. on p. 33).

[64] F. Hügging and Atlas Collaboration, *The ATLAS Pixel Insertable B-layer (IBL)*,
Nucl. Instr. Meth. A **650** (2011) 45, arXiv: `1012.2742 [physics.ins-det]` (cit. on p. 33).

[65] *ATLAS muon spectrometer: Technical design report*, (1997) (cit. on pp. 36–38).

[66] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*,
CERN-EP-2016-241 (2016), arXiv: `1611.09661 [hep-ex]` (cit. on p. 39).

[67] T. Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*,
J. Phys. Conf. Ser. **119** (2008) 032014 (cit. on p. 40).

[68] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*,
Nucl. Instrum. Meth. **A262** (1987) 444 (cit. on p. 40).

[69] T. G. Cornelissen et al., *The global chi**2 track fitter in ATLAS*,
J. Phys. Conf. Ser. **119** (2008) 032013 (cit. on p. 40).

[70] G. Piacquadio, K. Prokofiev and A. Wildauer,
Primary vertex reconstruction in the ATLAS experiment at LHC,
J. Phys. Conf. Ser. **119** (2008) 032033 (cit. on p. 40).

[71] ATLAS Collaboration,
Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ *TeV*, (2015),
URL: `http://cdsweb.cern.ch/record/2037717` (cit. on pp. 40, 104).

[72] ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*,
ATL-PHYS-PUB-2011-006, URL: `https://cds.cern.ch/record/1345327`
(cit. on pp. 41, 121).

[73] ATLAS Collaboration, *Electron identification efficiency dependence on pileup*,
ATL-COM-PHYS-2012-260,
URL: `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/EGAMMA/PublicPlots/20121403/EfficiencyPileup/ATL-COM-PHYS-2012-260/` (cit. on p. 41).

[74] ATLAS Collaboration, *Electron Efficiency Measurements for 2012 and 2011 Data*, ATL-COM-PHYS-2013-1287, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/EGAMMA/PublicPlots/20130926/ATL-COM-PHYS-2013-1287/ATL-COM-PHYS-2013-1287.pdf (cit. on p. 42).

[75] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, Eur. Phys. J. **C74** (2014) 3130, arXiv: 1407.3935 [hep-ex] (cit. on p. 41).

[76] *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, (2011), ATLAS-CONF-2011-063, URL: https://cds.cern.ch/record/1345743 (cit. on p. 41).

[77] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on pp. 41–43, 105, 121).

[78] M. Cacciari, G. P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on pp. 42, 69, 105).

[79] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11** (2016) P04008, arXiv: 1512.01094 [hep-ex] (cit. on pp. 43, 70, 87, 124).

[80] P. Berta et al., *Flavour tagging pre-recommendations for Run-II*, (2016), ATL-COM-PHYS-2016-092, URL: https://cds.cern.ch/record/2128507 (cit. on p. 43).

[81] V. Dao et al., *Expected performance of the ATLAS b-tagging algorithms in Run-2*, (2015), ATL-COM-PHYS-2015-642, URL: https://cds.cern.ch/record/2032459 (cit. on p. 44).

[82] ATLAS Collaboration, *Reconstruction and Calibration of Missing Transverse Energy and Performance in Z and W events in ATLAS Proton-Proton Collisions at 7 TeV*, ATLAS-CONF-2011-080 (2011) (cit. on p. 44).

[83] M. A. Dobbs et al., "Les Houches guidebook to Monte Carlo generators for hadron collider physics", *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 26-June 3, 2003*, 2004 411, arXiv: hep-ph/0403045 [hep-ph], URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-04-183 (cit. on p. 45).

[84] J. Alwall et al., *MadGraph/MadEvent v4: the new web generation*, JHEP **09** (2007) 028, arXiv: 0706.2334 [hep-ph] (cit. on pp. 45, 67, 68).

[85] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on pp. 45, 66, 68, 101–104, 125).

[86] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on pp. 45, 65).

[87] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on pp. 45, 65).

[88] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 040, arXiv: `1002.2581 [hep-ph]` (cit. on pp. 45, 65).

[89] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv: `0811.4622 [hep-ph]` (cit. on pp. 45, 66, 103).

[90] Höcker H. et al., *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040 (cit. on pp. 47–49, 111).

[91] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389** (1997) 81, See also `http://root.cern.ch/` (cit. on p. 47).

[92] Overfitting - Machine Learning Wiki, URL: `http://mlwiki.org/index.php/Overfitting` (cit. on p. 49).

[93] H.-J. Yang, B. P. Roe and J. Zhu, *Studies of boosted decision trees for MiniBooNE particle identification*, Nucl. Instrum. Meth. **A555** (2005) 370, arXiv: `physics/0508045 [physics]` (cit. on p. 50).

[94] T. Hastie et al, *The elements of statistical learning*, Springer (2001) (cit. on p. 50).

[95] V. Barger, J. Ohnemus and R. Phillips, *Event shape criteria for single lepton top signals*, Phys. Rev. D **48** (1993) 3953, arXiv: `hep-ph/9308216` (cit. on p. 50).

[96] S. Wertz, *The Matrix Element Method at the LHC: status and prospects for Run II*, J. Phys. Conf. Ser. **762** (2016) 012053 (cit. on p. 51).

[97] R. G. L. M. Chakravarti and J. Roy, *Handbook of Methods of Applied Statistics. Volume I & II*, Journal of the American Statistical Association **63** (1968) 1047, eprint: `http://dx.doi.org/10.1080/01621459.1968.11009335`, URL: `http://dx.doi.org/10.1080/01621459.1968.11009335` (cit. on p. 53).

[98] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, arXiv: `1007.1727 [physics.data-an]` (cit. on pp. 54, 55, 57–59).

[99] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society **54** (1943) 426, ISSN: 00029947, URL: `http://www.jstor.org/stable/1990256` (cit. on p. 58).

[100] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Annals Math. Statist. **9** (1938) 60 (cit. on p. 59).

[101] K. Cranmer et al., *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016 (2012) (cit. on p. 61).

[102] A. L. Read, *Linear interpolation of histograms*, Nucl. Instrum. Meth. **A425** (1999) 357 (cit. on p. 61).

[103] R. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219, URL: `https://cds.cern.ch/record/249779` (cit. on p. 61).

[104] ATLAS Collaboration, *Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **05** (2016) 160, arXiv: `1604.03812 [hep-ex]` (cit. on pp. 63, 68, 85, 93–97).

[105] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C76** (2016) 653, arXiv: `1608.03953 [hep-ex]` (cit. on pp. 64, 85, 121).

[106] ATLAS Collaboration, *Technical aspects of the first top pair analyses: Note 8*, (2010), Support document for top paper, URL: `https://cds.cern.ch/record/1297704` (cit. on p. 64).

[107] G. Bevilacqua et al., *HELAC-NLO*, Comput. Phys. Commun. **184** (2013) 986, arXiv: `1110.1499 [hep-ph]` (cit. on p. 65).

[108] M. Garzelli et al., *Standard Model Higgs boson production in association with a top anti-top pair at NLO with parton showering*, Eur. Phys. L. **96** (2011) 11001, arXiv: `1108.0387 [hep-ph]` (cit. on pp. 65, 68).

[109] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv: `1007.2241 [hep-ph]` (cit. on pp. 65, 68, 87, 103).

[110] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: `0710.3820 [hep-ph]` (cit. on pp. 65, 68, 103).

[111] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, ATL-COM-PHYS-2012-738 (2012), https://cds.cern.ch/record/1474107 (cit. on pp. 65, 68).

[112] G. Corcella et al., *HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **01** (2001) 010 (cit. on p. 65).

[113] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, CERN-2011-002 (CERN, Geneva, 2011), arXiv: `1101.0593 [hep-ph]` (cit. on pp. 65, 101).

[114] NNPDF Collaboration, R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: `1207.1303 [hep-ph]` (cit. on pp. 65, 68).

[115] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503, arXiv: `1103.2792 [hep-ph]` (cit. on pp. 65, 88, 101).

[116] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028, arXiv: `hep-ph/1001.5034 [hep-ph]` (cit. on pp. 65, 88, 101).

[117] A. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189, arXiv: `0901.0002 [hep-ph]` (cit. on pp. 65, 87, 101, 102).

[118] A. D. Martin et al., *Uncertainties on α_s in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653, arXiv: `0905.3531 [hep-ph]` (cit. on pp. 65, 87, 101).

[119] S. Frixione et al., *Single-top production in MC@NLO*, JHEP **03** (2006) 092, arXiv: `hep-ph/0512250 [hep-ph]` (cit. on pp. 65, 101).

[120] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: `1009.2450 [hep-ph]` (cit. on pp. 65, 68, 101, 103).

[121] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^{\pm}$ production and decay at NLO*, JHEP **07** (2012) 052, arXiv: `1204.5678 [hep-ph]` (cit. on pp. 65, 88).

[122] M. V. Garzelli et al., *$t\bar{t}W^{\pm}$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, JHEP **11** (2012) 056, arXiv: `1208.2665 [hep-ph]` (cit. on pp. 65, 88).

[123] ATLAS Collaboration, *Search for the decay $H \to b\bar{b}$ of the standard model Higgs boson produced in association with hadronically decaying top quarks in pp collisions at* 8 TeV *with the ATLAS detector at the LHC*, ATL-COM-PHYS-2014-404 (2014), URL: `https://cds.cern.ch/record/1698272` (cit. on pp. 66, 70, 71, 73, 74, 84, 88, 91–93).

[124] T. Sjöstrand, S. Mrenna and P. Skands, PYTHIA *6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: `hep-ph/0603175v2 [hep-ph]` (cit. on pp. 66, 68).

[125] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section as a function of jet multiplicity and jet transverse momentum in 7 TeV proton-proton collisions with the ATLAS detector*, JHEP **2015** (2015) (cit. on p. 66).

[126] F. Cascioli et al., *NLO matching for ttbb production with massive b-quarks*, Phys. Lett. B **734** (2013) 210, arXiv: `1309.5912 [hep-ph]` (cit. on p. 66).

[127] F. Cascioli, P. Maierhöfer and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv: `1111.5206 [hep-ph]` (cit. on p. 66).

[128] F. Maltoni, G. Ridolfi and M. Ubiali, *b-initiated processes at the LHC: a reappraisal*, JHEP **07** (2012) 022, [Erratum: JHEP04,095(2013)], arXiv: `1203.6393 [hep-ph]` (cit. on p. 66).

[129] ATLAS Collaboration, *Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7\ TeV$ using the ATLAS detector*, Phys. Rev. D **90** (2014) 072004, arXiv: `1407.0371 [hep-ex]` (cit. on pp. 67, 87).

[130] J. Butterworth, J. Forshaw and M. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. **72** (1996) 637, arXiv: `hep-ph/9601371 [hep-ph]` (cit. on p. 68).

[131] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C45** (2006) 97, arXiv: `hep-ph/0506026 [hep-ph]` (cit. on pp. 68, 102).

[132] S. Jadach, J. H. Kühn and Z. Was, *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Comput. Phys. Commun. **64** (1991) 275 (cit. on p. 68).

[133] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: `1005.4568 [physics.ins-det]` (cit. on p. 68).

[134] S. Agostinelli et al., *Geant4: a simulation toolkit*, Nucl. Instrum. Meth. Phys. Res. A **506** (2003) 250, ISSN: 0168-9002 (cit. on pp. 68, 102).

[135] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, ATL-COM-PHYS-2010-838 (2010), URL: `http://cdsweb.cern.ch/record/1300517` (cit. on p. 68).

[136] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on pp. 68, 102).

[137] J. Pumplin et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP **0207** (2012) 012, arXiv: `hep-ph/0201195 [hep-ph]` (cit. on pp. 68, 103).

[138] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, arXiv: `0802.0007 [hep-ph]` (cit. on pp. 68, 103).

[139] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639, arXiv: `0803.0883 [hep-ph]` (cit. on pp. 68, 103).

[140] S. Frixione, G. Ridolfi and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **0709** (2009) 126, arXiv: `0707.3088 [hep-ph]` (cit. on pp. 68, 103).

[141] P. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D **82** (2010) 074018, arXiv: `1005.3457 [hep-ph]` (cit. on pp. 68, 103).

[142] S. Alioli et al., *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **0909** (2009) 111, arXiv: `0907.4076 [hep-ph]` (cit. on pp. 68, 103).

[143] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator* AcerMC *versions 2.0 to 3.8 with interfaces to* Pythia *6.4,* Herwig *6.5 and* Ariadne *4.1*, Comput. Phys. Commun. **184** (2013) 919, arXiv: `hep-ph/0405247 [hep-ph]` (cit. on p. 68).

[144] ATLAS Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, (2012), ATLAS-CONF-2012-020, URL: `http://cdsweb.cern.ch/record/1430034` (cit. on p. 69).

[145] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B **641** (2006) 57, arXiv: `hep-ph/0512210` (cit. on pp. 69, 105).

[146] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C72** (2012) 1896, arXiv: `1111.6097 [hep-ph]` (cit. on p. 69).

[147] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **73** (2013) 2304, arXiv: `1112.6426 [hep-ex]` (cit. on pp. 69, 87, 124).

[148] ATLAS Collaboration, *Measurement of the b-tagging efficiency in a sample of jets containing muons with 5 fb^{-1} of data from the ATLAS detector*, ATLAS-CONF-2012-043 (2012), URL: `http://cdsweb.cern.ch/record/1435197` (cit. on p. 70).

[149] ATLAS Collaboration, *The ATLAS Level-1 Calorimeter Trigger*, JINST **3** (2008) P03001 (cit. on p. 70).

[150] DØ Collaboration, V. M. Abazov et al., *Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using secondary vertex b-tagging*, Phys. Rev. D **74** (2006) 112004, arXiv: `hep-ex/0611002 [hep-ex]` (cit. on p. 71).

[151] J. Neyman and E. Pearson, Phil. Trans. R. Soc. Lond. A **231** (1 1933) (cit. on p. 76).

[152] ATLAS Collaboration, *Measurement of the mistag rate of b-tagging algorithms with 5 fb^{-1} of data collected by the ATLAS detector*, ATLAS-CONF-2012-040 (2012), URL: `http://cdsweb.cern.ch/record/1435194` (cit. on pp. 87, 124).

[153] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, (2011), arXiv: `1101.0538 [hep-ph]` (cit. on pp. 87, 89).

[154] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. **D89** (2014) 033009, arXiv: `1302.6246 [hep-ph]` (cit. on p. 87).

[155] ATLAS Collaboration, *Monte Carlo Generators for the Production of a W or Z/γ^* Boson in Association with Jets at ATLAS in Run 2*, (2015), ATL-COM-PHYS-2015-832, URL: `https://cds.cern.ch/record/2041288` (cit. on p. 100).

[156] ATLAS Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, (2015), ATL-COM-PHYS-2015-1499, URL: `https://cds.cern.ch/record/2113154` (cit. on p. 100).

[157] ATLAS Collaboration, *Simulation of top quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV*, (2015), ATL-COM-PHYS-2015-1496, URL: `https://cds.cern.ch/record/2113056` (cit. on p. 100).

[158] ATLAS Collaboration, *Modelling of the $t\bar{t}H$ and $t\bar{t}V$ ($V = W, Z$) processes for $\sqrt{s} = 13$ TeV ATLAS analyses*, (2015), ATL-COM-PHYS-2015-1510, URL: `https://cds.cern.ch/record/2113264` (cit. on p. 100).

[159] W. Beenakker et al., *Higgs radiation off top quarks at the Tevatron and the LHC*, Phys. Rev. Lett. **87** (2001) 201805, arXiv: `hep-ph/0107081 [hep-ph]` (cit. on p. 101).

[160] W. Beenakker et al., *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl. Phys. B **653** (2003) 151, arXiv: `hep-ph/0211352 [hep-ph]` (cit. on p. 101).

[161] S. Dawson et al., *Associated top quark Higgs boson production at the LHC*, Phys. Rev. D **67** (2003) 071503, arXiv: `hep-ph/0211438 [hep-ph]` (cit. on p. 101).

[162] S. Dawson et al., *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, Phys. Rev. D **68** (2003) 034022, arXiv: `hep-ph/0305087 [hep-ph]` (cit. on p. 101).

[163] Y. Zhang et al., *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider*, Phys. Lett. B **738** (2014) 1, arXiv: `1407.1110 [hep-ph]` (cit. on p. 101).

[164] S. Frixione et al., *Weak corrections to Higgs hadroproduction in association with a top-quark pair*, JHEP **09** (2014) 065, arXiv: `1407.0823 [hep-ph]` (cit. on p. 101).

[165] S. Frixione et al., *Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons*, JHEP **06** (2015) 184, arXiv: `1504.03446 [hep-ph]` (cit. on p. 101).

[166] K. Melnikov, M. Schulze and A. Scharf, *QCD corrections to top quark pair production in association with a photon at hadron colliders*, Phys. Rev. **D83** (2011) 074013, arXiv: `1102.1967 [hep-ph]` (cit. on p. 102).

[167] S. Hoeche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv: `1207.5030 [hep-ph]` (cit. on p. 102).

[168] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: `1410.8849 [hep-ph]` (cit. on p. 103).

[169] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, ATL-PHYS-PUB-2014-021 (2014), http://cds.cern.ch/record/1966419 (cit. on pp. 103, 125).

[170] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: `1207.1303 [hep-ph]` (cit. on p. 103).

[171] M. H. Seymour and A. Siodmok, *Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data*, JHEP **10** (2013) 113, arXiv: `1307.5015 [hep-ph]` (cit. on p. 103).

[172] T. Sjöstrand et al., *High-energy-physics event generation with Pythia 6.1*, Comput. Phys. Commun. **135** (2001) 238, arXiv: `hep-ph/0010017 [hep-ph]` (cit. on p. 103).

[173] J. Butterworth et al., *Single Boson and Diboson Production Cross Sections in pp Collisions at sqrts=7 TeV*, (2010), ATL-COM-PHYS-2010-695, URL: `https://cds.cern.ch/record/1287902` (cit. on p. 104).

[174] ATLAS Standard Model Working Group, *Standard Model Cross Sections for Diboson/Triboson Productions*, URL: `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SMDC14xsecs#NNLO_Cross_Sections` (cit. on p. 104).

[175] ATLAS Collaboration, *Further ATLAS tunes of* PYTHIA6 *and Pythia 8*, (2011), URL: `http://cds.cern.ch/record/1400677` (cit. on p. 102).

[176] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, (2016), ATL-DAQ-PUB-2016-001, URL: `https://cds.cern.ch/record/2136007` (cit. on p. 104).

[177] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data*, (2014), REPLACE WITH 2015 NOTE IF AVAILABLE, URL: `http://cdsweb.cern.ch/record/1706245` (cit. on pp. 104, 123).

[178] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, (2015), ATL-PHYS-PUB-2015-015, URL: `https://cds.cern.ch/record/2037613` (cit. on p. 105).

[179] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton-proton collisions with the ATLAS detector*, (2015), ATLAS-CONF-2015-029, URL: `https://cds.cern.ch/record/2037702` (cit. on p. 105).

[180] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018 (2014), URL: `https://cds.cern.ch/record/1700870` (cit. on p. 105).

[181] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions with the ATLAS detector*, Nucl. Instrum. Meth. **A824** (2016) 367, arXiv: `1510.03823 [hep-ex]` (cit. on p. 105).

[182] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, (2015), ATL-PHYS-PUB-2015-022, URL: `http://cds.cern.ch/record/2037697` (cit. on p. 105).

[183] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, (2016), ATL-PHYS-PUB-2016-012, URL: `http://cds.cern.ch/record/2160731` (cit. on p. 105).

[184] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, (2015), ATL-PHYS-PUB-2015-045, URL: `http://cds.cern.ch/record/2064383` (cit. on p. 105).

[185] ATLAS Collaboration, *Search for the associated production of a higgs boson with a top quark pair in multilepton final states with the atlas detector*, CONF-HIGG-2016-10 (2016), URL: `https://cds.cern.ch/record/2150024` (cit. on p. 111).

[186] M. Aaboud et al., *Measurement of the* $t\bar{t}Z$ *and* $t\bar{t}W$ *production cross sections in multilepton final states using 3.2* fb^{-1} *of pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 40, arXiv: `1609.01599 [hep-ex]` (cit. on pp. 111, 128).

[187] ATLAS Collaboration, *Improved luminosity determination in pp collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **73** (2013) 2518, arXiv: `1302.4393 [hep-ex]` (cit. on p. 121).

[188] ATLAS Collaboration, *Electron identification measurements in ATLAS using* $\sqrt{s} = 13$ *TeV data with 50 ns bunch spacing*, (2015), URL: `http://cdsweb.cern.ch/record/2048202` (cit. on p. 121).

[189] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton–proton collision data*, Eur. Phys. J. **C 74** (2014) 2941, arXiv: `1404.2240` (cit. on p. 123).

APPENDIX A

BDT separation in the fully hadronic channel

This Appendix contains the separation plots and overtraining checks of the boosted decision trees used in each fit region. The blue (red) histogram indicates the BDT output for signal (background) used in training, the blue (red) points indicate the BDT output distribution from the signal (background) used to test. This section contains also the correlation matrices for the set of input variables used in each fit region to evaluate the BDT response. The signal is obtained from $t\bar{t}H$ Monte Carlo samples, while the background is the sum of the data-driven multijet background ($\mathrm{TRF_{MJ}}$) and all the Monte Carlo background samples.

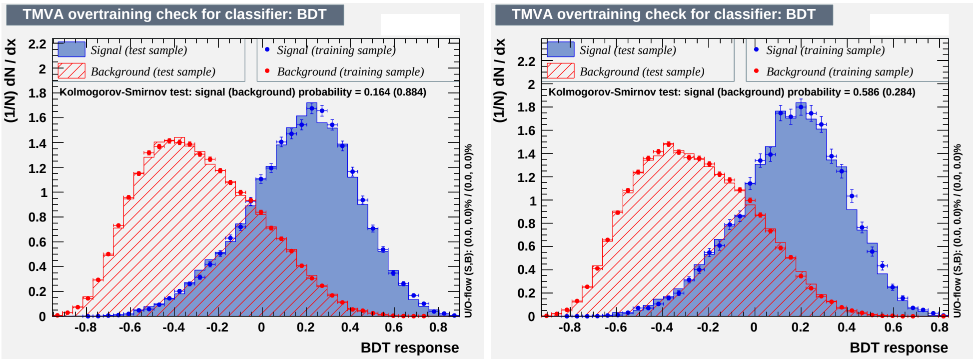


Figure A.1: The response of the boosted decision trees in the 6-jet regions: (left) exactly 6 jets, exactly 3 b-tags; (right) 6 jets, $\geq$ 4 b-tags.

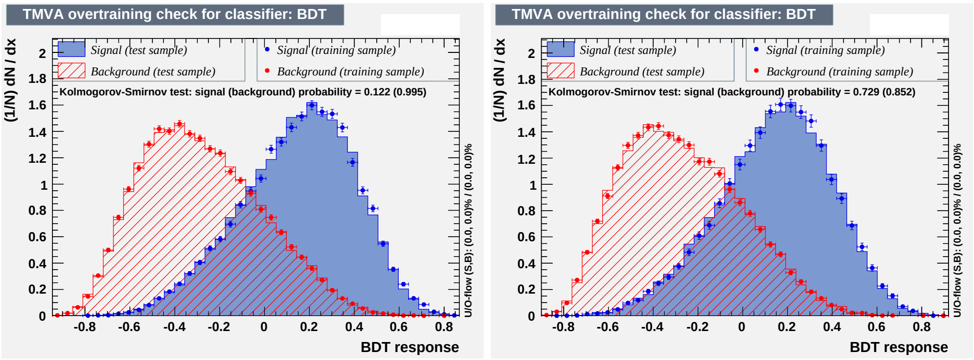
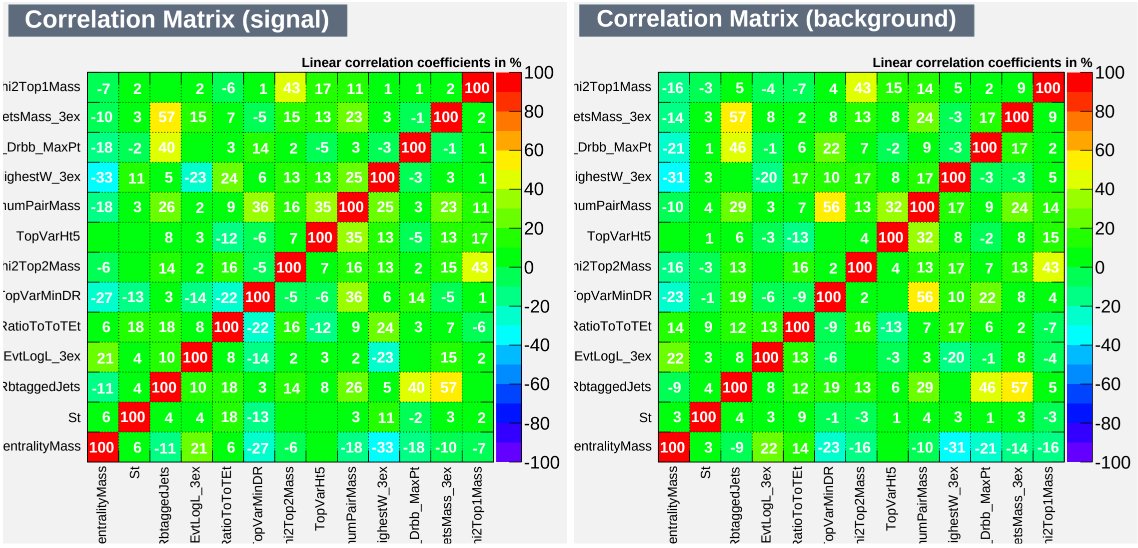


Figure A.2: The response of the boosted decision trees in the 7-jet regions: (left) exactly 7 jets, exactly 3 b-tags; (right) 7 jets, $\geq$ 4 b-tags.



Correlation Matrix (signal)

Linear correlation coefficients in %

	entralityMass	St	RbtaggedJets	EvtLogL_3ex	RatioToToTEt	opVarMinDR	hi2Top2Mass	TopVarHt5	umPairMass	ighestW_3ex	_Drbb_MaxPt	etsMass_3ex	hi2Top1Mass
hi2Top1Mass	-7	2		2	-6	1	43	17	11	1	1	2	100
etsMass_3ex	-10	3	57	15	7	-5	15	13	23	3	-1	100	2
_Drbb_MaxPt	-18	-2	40		3	14	2	-5	3	-3	100	-1	1
ighestW_3ex	-33	11	5	-23	24	6	13	13	25	100	-3	3	1
umPairMass	-18	3	26	2	9	36	16	35	100	25	3	23	11
TopVarHt5			8	3	-12	-6	7	100	35	13	-5	13	17
hi2Top2Mass	-6		14	2	16	-5	100	7	16	13	2	15	43
opVarMinDR	-27	-13	3	-14	-22	100	-5	-6	36	6	14	-5	1
RatioToToTEt	6	18	18	8	100	-22	16	-12	9	24	3	7	-6
EvtLogL_3ex	21	4	10	100	8	-14	2	3	2	-23		15	2
RbtaggedJets	-11	4	100	10	18	3	14	8	26	5	40	57	
St	6	100	4	4	18	-13			3	11	-2	3	2
entralityMass	100	6	-11	21	6	-27	-6		-18	-33	-18	-10	-7

Correlation Matrix (background)

Linear correlation coefficients in %

	entralityMass	St	RbtaggedJets	EvtLogL_3ex	RatioToToTEt	opVarMinDR	hi2Top2Mass	TopVarHt5	umPairMass	ighestW_3ex	_Drbb_MaxPt	etsMass_3ex	hi2Top1Mass
hi2Top1Mass	-16	-3	5	-4	-7	4	43	15	14	5	2	9	100
etsMass_3ex	-14	3	57	8	2	8	13	8	24	-3	17	100	9
_Drbb_MaxPt	-21	1	46	-1	6	22	7	-2	9	-3	100	17	2
ighestW_3ex	-31	3		-20	17	10	17	8	17	100	-3	-3	5
umPairMass	-10	4	29	3	7	56	13	32	100	17	9	24	14
TopVarHt5		1	6	-3	-13		4	100	32	8	-2	8	15
hi2Top2Mass	-16	-3	13		16	2	100	4	13	17	7	13	43
opVarMinDR	-23	-1	19	-6	-9	100	2		56	10	22	8	4
RatioToToTEt	14	9	12	13	100	-9	16	-13	7	17	6	2	-7
EvtLogL_3ex	22	3	8	100	13	-6		-3	3	-20	-1	8	-4
RbtaggedJets	-9	4	100	8	12	19	13	6	29		46	57	5
St	3	100	4	3	9	-1	-3	1	4	3	1	3	-3
entralityMass	100	3	-9	22	14	-23	-16		-10	-31	-21	-14	-16

Figure A.3: The correlation matrix between variables in the BDT for exactly 6 jets, exactly 3 *b*-tags;

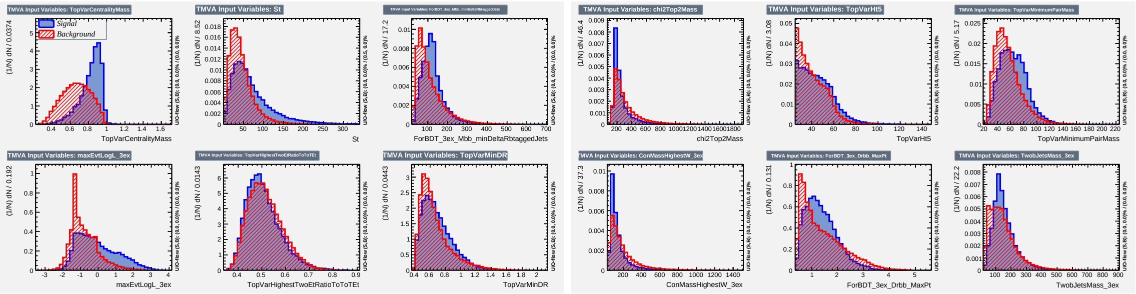
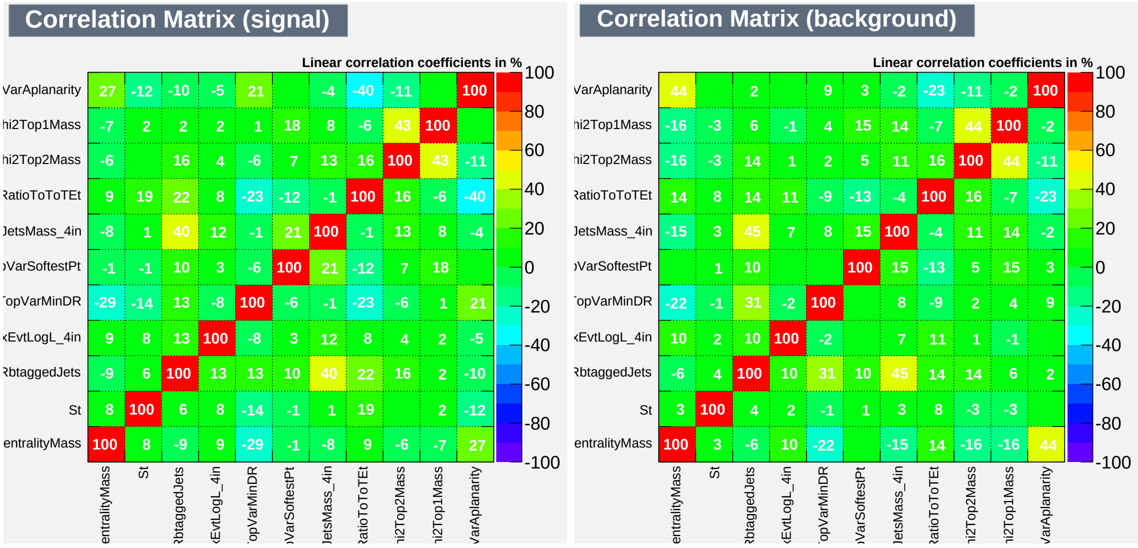


Figure A.4: The distribution of the variables entering the BDT for exactly 6 jets, exactly 3 *b*-tags; the signal is in blue, the background in red.



Correlation Matrix (signal)

	entralityMass	St	RbtaggedJets	xEvtLogL_4in	opVarMinDR	VarSoftestPt	JetsMass_4in	RatioToToTEt	hi2Top2Mass	hi2Top1Mass	VarAplanarity
VarAplanarity	27	-12	-10	-5	21		-4	-40	-11		100
hi2Top1Mass	-7	2	2	2	1	18	8	-6	43	100	
hi2Top2Mass	-6		16	4	-6	7	13	16	100	43	-11
RatioToToTEt	9	19	22	8	-23	-12	-1	100	16	-6	-40
JetsMass_4in	-8	1	40	12	-1	21	100	-1	13	8	-4
VarSoftestPt	-1	-1	10	3	-6	100	21	-12	7	18	
opVarMinDR	-29	-14	13	-8	100	-6	-1	-23	-6	1	21
xEvtLogL_4in	9	8	13	100	-8	3	12	8	4	2	-5
RbtaggedJets	-9	6	100	13	13	10	40	22	16	2	-10
St	8	100	6	8	-14	-1	1	19		2	-12
entralityMass	100	8	-9	9	-29	-1	-8	9	-6	-7	27

Correlation Matrix (background)

	entralityMass	St	RbtaggedJets	xEvtLogL_4in	opVarMinDR	VarSoftestPt	JetsMass_4in	RatioToToTEt	hi2Top2Mass	hi2Top1Mass	VarAplanarity
VarAplanarity	44		2		9	3	-2	-23	-11	-2	100
hi2Top1Mass	-16	-3	6	-1	4	15	14	-7	44	100	-2
hi2Top2Mass	-16	-3	14	1	2	5	11	16	100	44	-11
RatioToToTEt	14	8	14	11	-9	-13	-4	100	16	-7	-23
JetsMass_4in	-15	3	45	7	8	15	100	-4	11	14	-2
VarSoftestPt		1	10			100	15	-13	5	15	3
opVarMinDR	-22	-1	31	-2	100		8	-9	2	4	9
xEvtLogL_4in	10	2	10	100	-2		7	11	1	-1	
RbtaggedJets	-6	4	100	10	31	10	45	14	14	6	2
St	3	100	4	2	-1	1	3	8	-3	-3	
entralityMass	100	3	-6	10	-22		-15	14	-16	-16	44

Figure A.5: The correlation matrix between variables in the BDT for exactly 6 jets, ≥ 4b-tags;

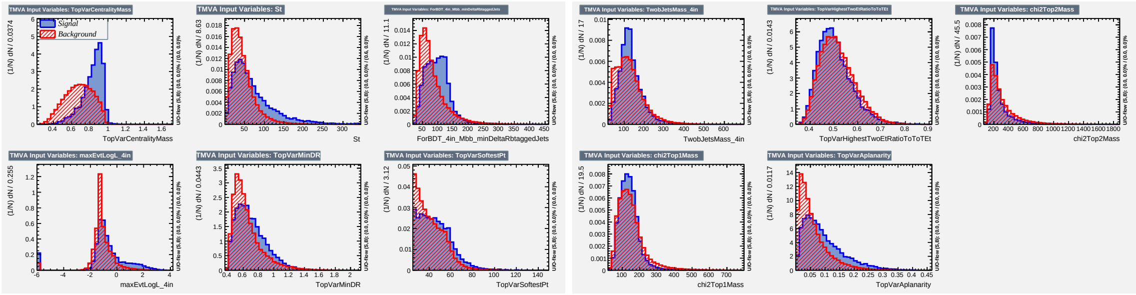
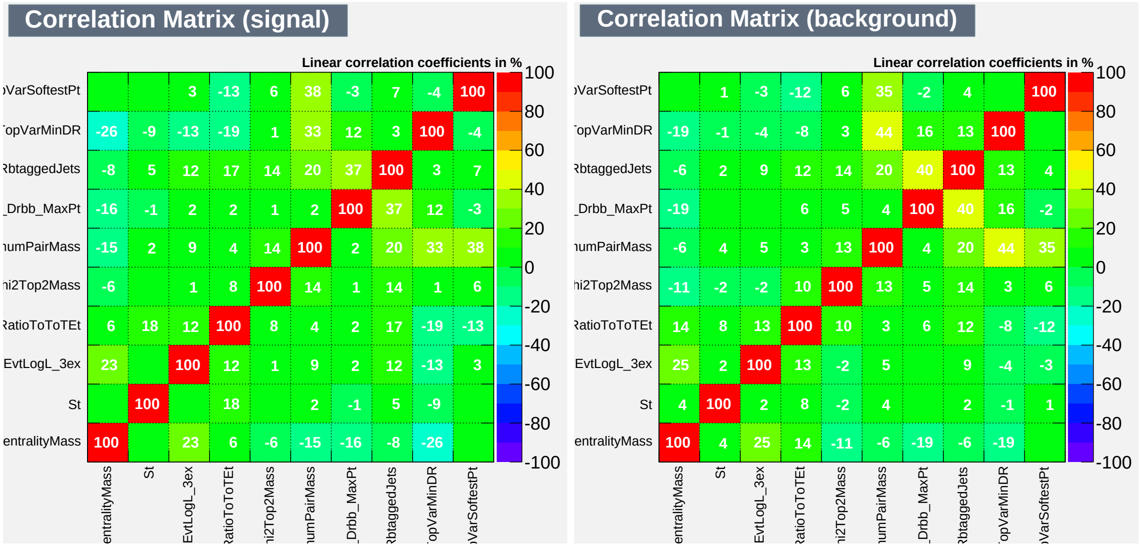


Figure A.6: The distribution of the variables entering the BDT for exactly 6 jets, ≥ 4b-tags; the signal is in blue, the background in red.



Correlation Matrix (signal)

Linear correlation coefficients in %

	entralityMass	St	EvtLogL_3ex	RatioToToTEt	hi2Top2Mass	umPairMass	_Drbb_MaxPt	RbtaggedJets	opVarMinDR	VarSoftestPt
VarSoftestPt			3	-13	6	38	-3	7	-4	100
opVarMinDR	-26	-9	-13	-19	1	33	12	3	100	-4
RbtaggedJets	-8	5	12	17	14	20	37	100	3	7
_Drbb_MaxPt	-16	-1	2	2	1	2	100	37	12	-3
umPairMass	-15	2	9	4	14	100	2	20	33	38
hi2Top2Mass	-6		1	8	100	14	1	14	1	6
RatioToToTEt	6	18	12	100	8	4	2	17	-19	-13
EvtLogL_3ex	23		100	12	1	9	2	12	-13	3
St		100		18		2	-1	5	-9	
entralityMass	100		23	6	-6	-15	-16	-8	-26	

Correlation Matrix (background)

Linear correlation coefficients in %

	entralityMass	St	EvtLogL_3ex	RatioToToTEt	hi2Top2Mass	umPairMass	_Drbb_MaxPt	RbtaggedJets	opVarMinDR	VarSoftestPt
VarSoftestPt		1	-3	-12	6	35	-2	4		100
opVarMinDR	-19	-1	-4	-8	3	44	16	13	100	
RbtaggedJets	-6	2	9	12	14	20	40	100	13	4
_Drbb_MaxPt	-19			6	5	4	100	40	16	-2
umPairMass	-6	4	5	3	13	100	4	20	44	35
hi2Top2Mass	-11	-2	-2	10	100	13	5	14	3	6
RatioToToTEt	14	8	13	100	10	3	6	12	-8	-12
EvtLogL_3ex	25	2	100	13	-2	5		9	-4	-3
St	4	100	2	8	-2	4		2	-1	1
entralityMass	100	4	25	14	-11	-6	-19	-6	-19	

Figure A.7: The correlation matrix between variables in the BDT for exactly 7 jets, exactly 3 b-tags;

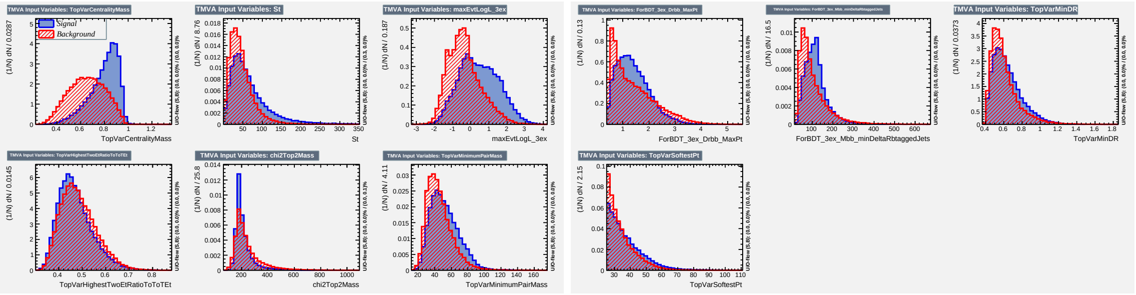
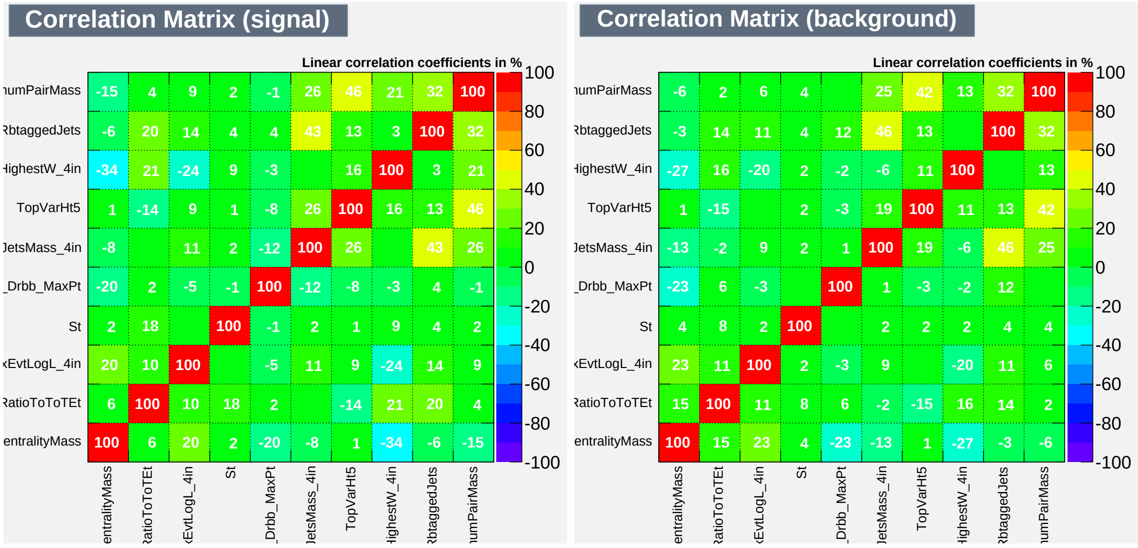


Figure A.8: The distribution of the variables entering the BDT for exactly 6 jets, exactly 3 b-tags; the signal is in blue, the background in red.



	entralityMass	RatioToToTEt	xEvtLogL_4in	St	_Drbb_MaxPt	JetsMass_4in	TopVarHt5	HighestW_4in	RbtaggedJets	numPairMass
numPairMass	-15	4	9	2	-1	26	46	21	32	100
RbtaggedJets	-6	20	14	4	4	43	13	3	100	32
HighestW_4in	-34	21	-24	9	-3		16	100	3	21
TopVarHt5	1	-14	9	1	-8	26	100	16	13	46
JetsMass_4in	-8		11	2	-12	100	26		43	26
_Drbb_MaxPt	-20	2	-5	-1	100	-12	-8	-3	4	-1
St	2	18		100	-1	2	1	9	4	2
xEvtLogL_4in	20	10	100		-5	11	9	-24	14	9
RatioToToTEt	6	100	10	18	2		-14	21	20	4
entralityMass	100	6	20	2	-20	-8	1	-34	-6	-15

	entralityMass	RatioToToTEt	xEvtLogL_4in	St	_Drbb_MaxPt	JetsMass_4in	TopVarHt5	HighestW_4in	RbtaggedJets	numPairMass
numPairMass	-6	2	6	4		25	42	13	32	100
RbtaggedJets	-3	14	11	4	12	46	13		100	32
HighestW_4in	-27	16	-20	2	-2	-6	11	100		13
TopVarHt5	1	-15		2	-3	19	100	11	13	42
JetsMass_4in	-13	-2	9	2	1	100	19	-6	46	25
_Drbb_MaxPt	-23	6	-3		100	1	-3	-2	12	
St	4	8	2	100		2	2	2	4	4
xEvtLogL_4in	23	11	100	2	-3	9		-20	11	6
RatioToToTEt	15	100	11	8	6	-2	-15	16	14	2
entralityMass	100	15	23	4	-23	-13	1	-27	-3	-6

Figure A.9: The correlation matrix between variables in the BDT for exactly 7 jets, ≥ 4b-tags;

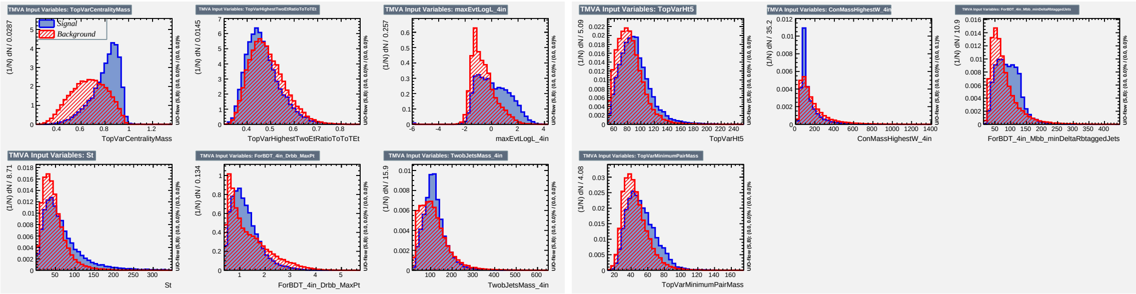


Figure A.10: The distribution of the variables entering the BDT for exactly 7 jets, ≥ 4b-tags; the signal is in blue, the background in red.

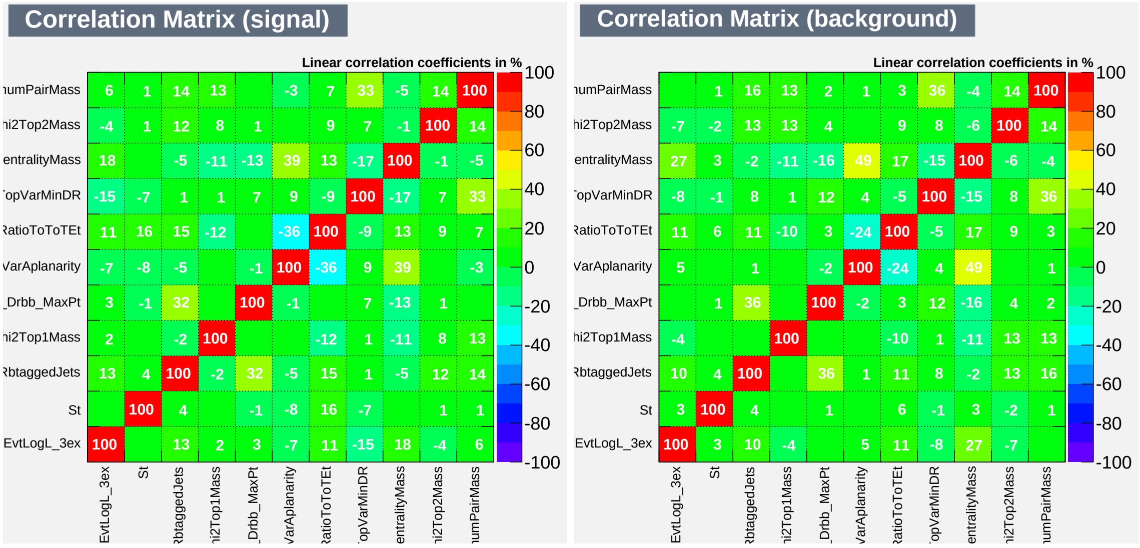


Figure A.11: The correlation matrix between variables in the BDT for ≥ 8 jets, exactly 3 b-tags;

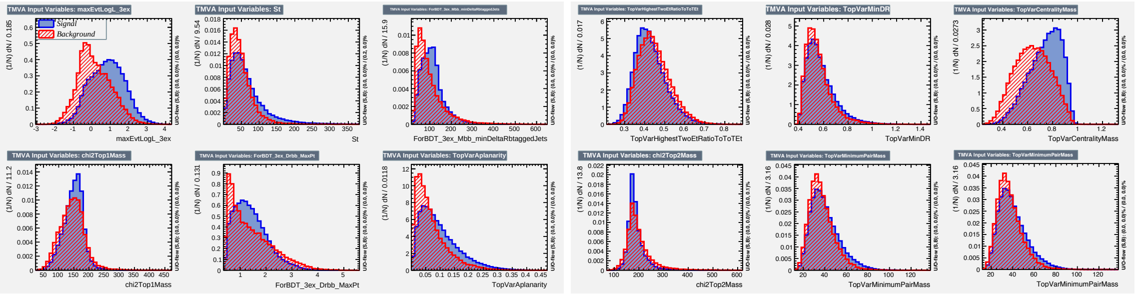
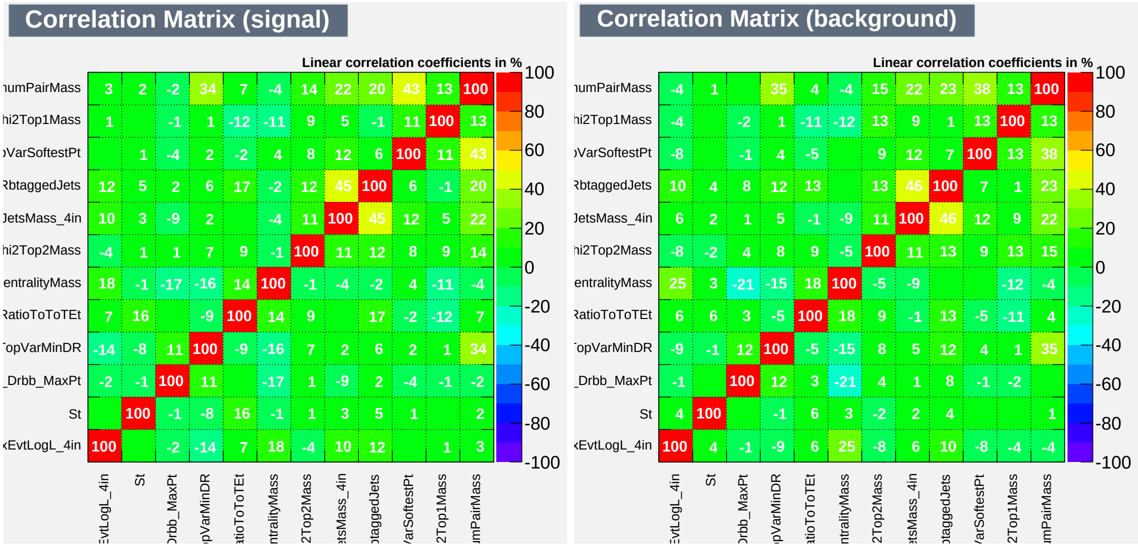


Figure A.12: The distribution of the variables entering the BDT for ≥ 8 jets, exactly 3 b-tags; the signal is in blue, the background in red.



Correlation Matrix (signal)

	EvtLogL_4in	St	Drbb_MaxPt	pVarMinDR	atioToToTEt	ntralityMass	2Top2Mass	tsMass_4in	taggedJets	VarSoftestPt	2Top1Mass	umPairMass
umPairMass	3	2	-2	34	7	-4	14	22	20	43	13	100
hi2Top1Mass	1		-1	1	-12	-11	9	5	-1	11	100	13
VarSoftestPt		1	-4	2	-2	4	8	12	6	100	11	43
RbtaggedJets	12	5	2	6	17	-2	12	45	100	6	-1	20
JetsMass_4in	10	3	-9	2		-4	11	100	45	12	5	22
hi2Top2Mass	-4	1	1	7	9	-1	100	11	12	8	9	14
entralityMass	18	-1	-17	-16	14	100	-1	-4	-2	4	-11	-4
RatioToToTEt	7	16		-9	100	14	9		17	-2	-12	7
TopVarMinDR	-14	-8	11	100	-9	-16	7	2	6	2	1	34
_Drbb_MaxPt	-2	-1	100	11		-17	1	-9	2	-4	-1	-2
St		100	-1	-8	16	-1	1	3	5	1		2
xEvtLogL_4in	100		-2	-14	7	18	-4	10	12		1	3

Correlation Matrix (background)

	EvtLogL_4in	St	Drbb_MaxPt	pVarMinDR	atioToToTEt	ntralityMass	2Top2Mass	tsMass_4in	taggedJets	VarSoftestPt	2Top1Mass	umPairMass
umPairMass	-4	1		35	4	-4	15	22	23	38	13	100
hi2Top1Mass	-4		-2	1	-11	-12	13	9	1	13	100	13
VarSoftestPt	-8		-1	4	-5		9	12	7	100	13	38
RbtaggedJets	10	4	8	12	13		13	46	100	7	1	23
JetsMass_4in	6	2	1	5	-1	-9	11	100	46	12	9	22
hi2Top2Mass	-8	-2	4	8	9	-5	100	11	13	9	13	15
entralityMass	25	3	-21	-15	18	100	-5	-9			-12	-4
RatioToToTEt	6	6	3	-5	100	18	9	-1	13	-5	-11	4
TopVarMinDR	-9	-1	12	100	-5	-15	8	5	12	4	1	35
_Drbb_MaxPt	-1		100	12	3	-21	4	1	8	-1	-2	
St	4	100		-1	6	3	-2	2	4			1
xEvtLogL_4in	100	4	-1	-9	6	25	-8	6	10	-8	-4	-4

Figure A.13: The correlation matrix between variables in the BDT for ≥ 8 jets, ≥ 4b-tags;

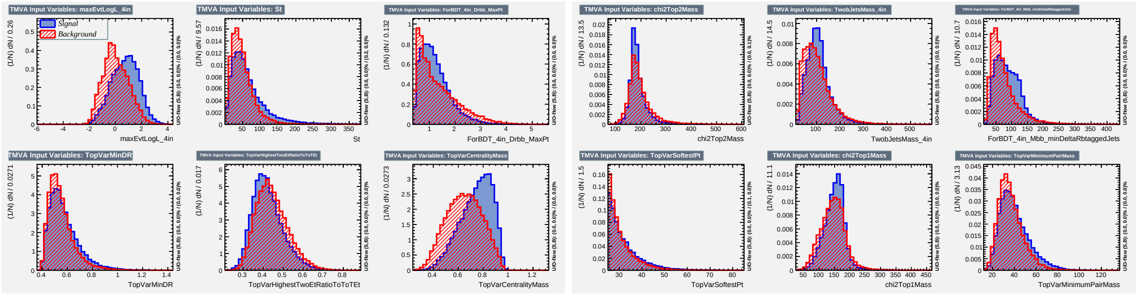


Figure A.14: The distribution of the variables entering the BDT for ≥ 8 jets, ≥ 4b-tags; the signal is in blue, the background in red.

APPENDIX B

Validation of the BDT approach in the fully hadronic analysis

This section contains the pre- and post-fit plots of the BDT input variables. Each histogram presents the pre-fit (post-fit) plot of the variable in each analysis region. Only variables used as input variable of the BDT in each region are presented in this Appendix.

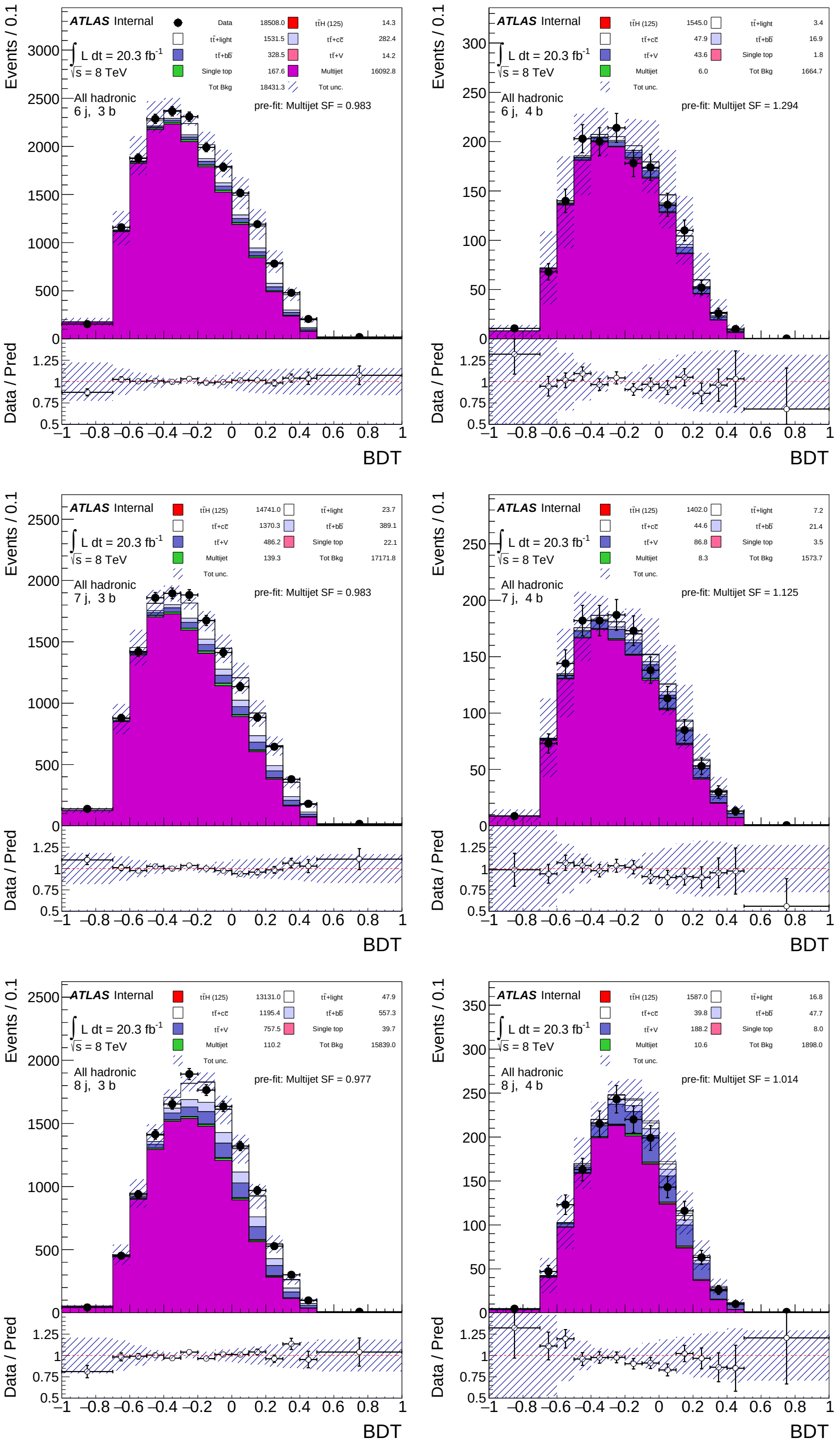


Figure B.1: Comparison between data and predictions for the final discriminant used in the analysis regions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties while the statistical uncertainty is also represented in the green band. The signal contribution is not included in the stack plot and instead displayed multiplied by 100 times (dashed red line).

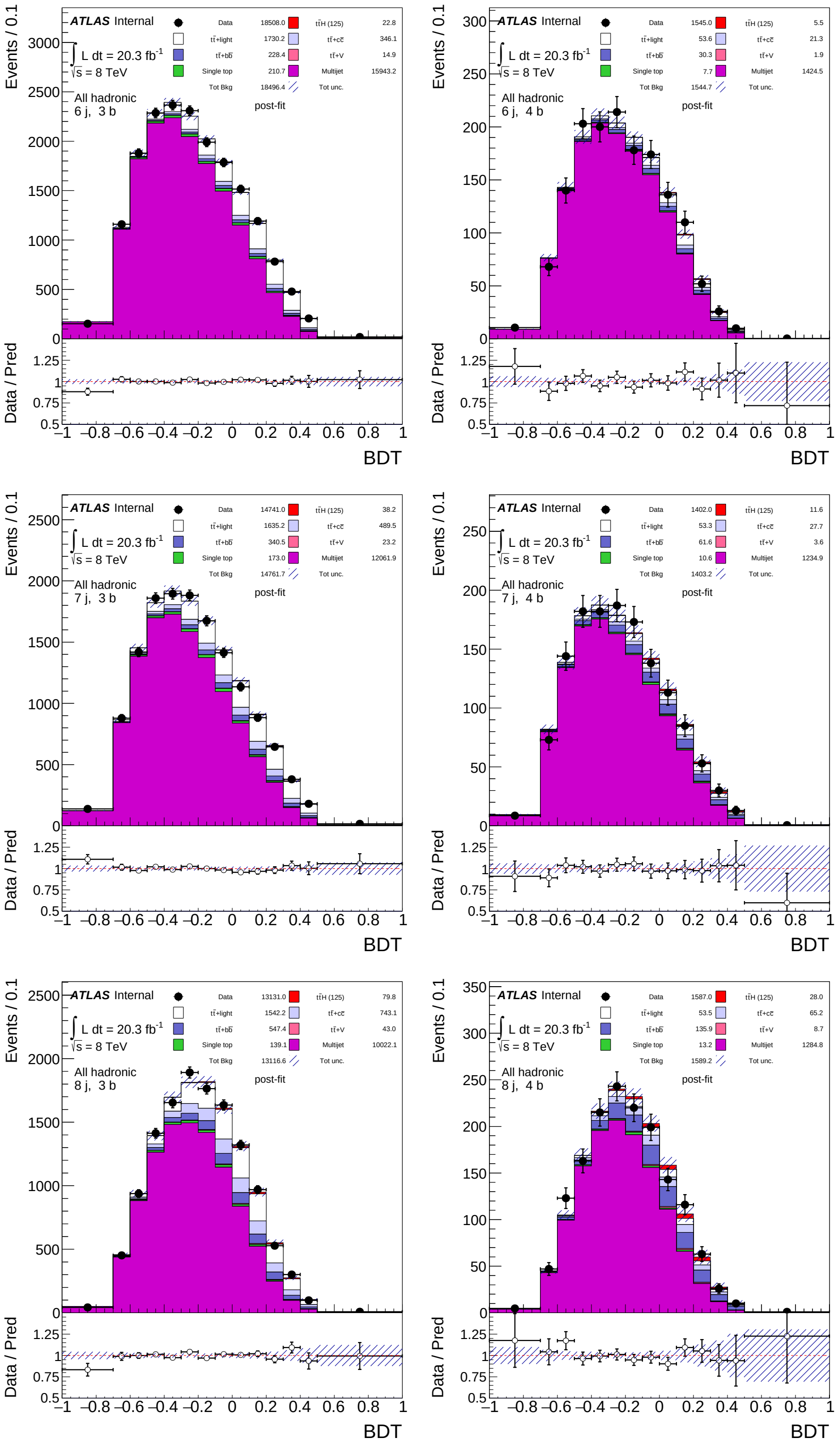


Figure B.2: Comparison between data and predictions for the final discriminant used in the analysis regions; signal and background predictions have been corrected with the result of the fit to data. The uncertainty band contains statistical and systematical uncertainties and takes into account the correlations among the nuisance parameters induced by the fit. The signal contribution is included in the stack plot.

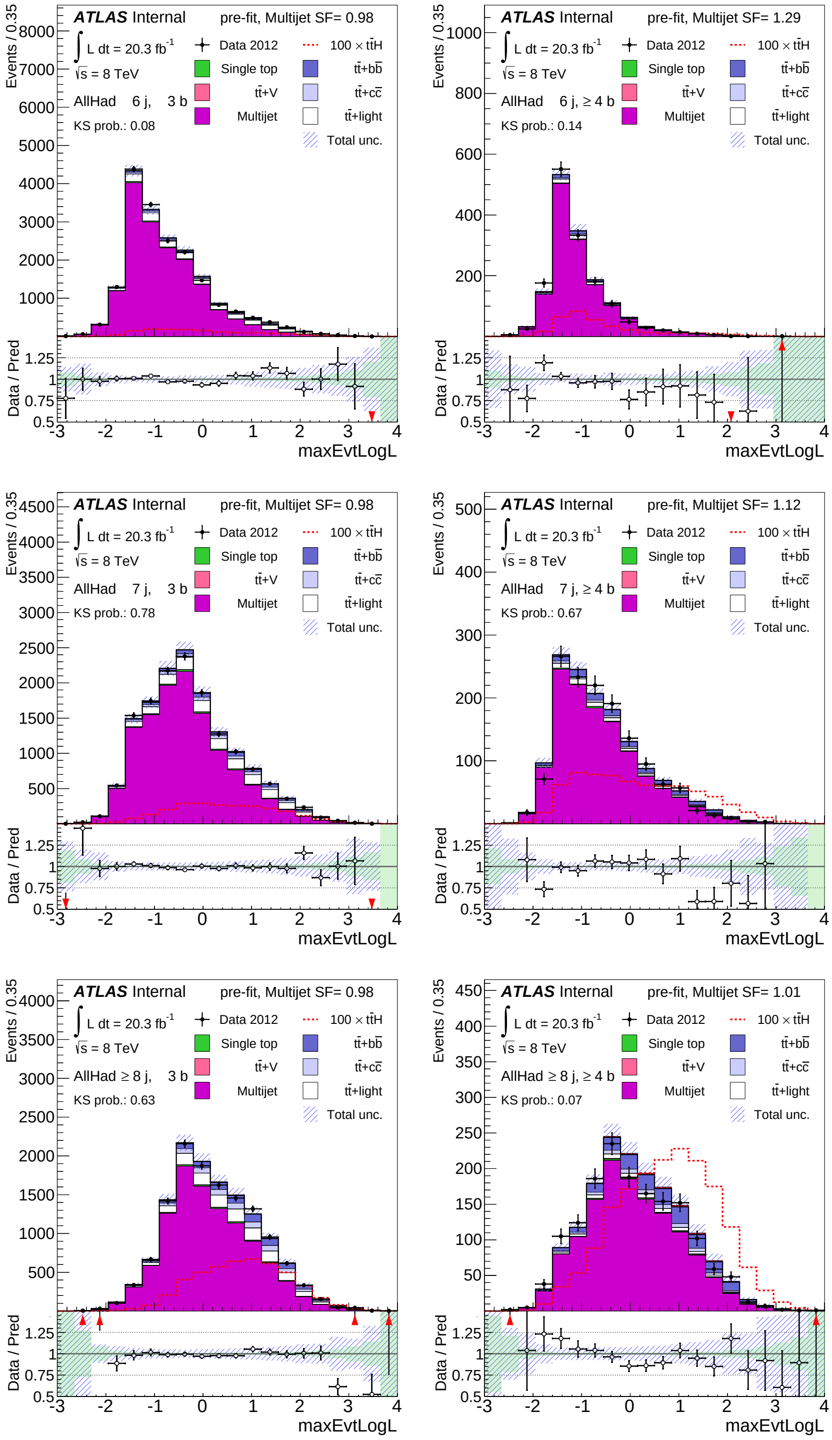


Figure B.3: $\mathcal{D}$ pre-fit.

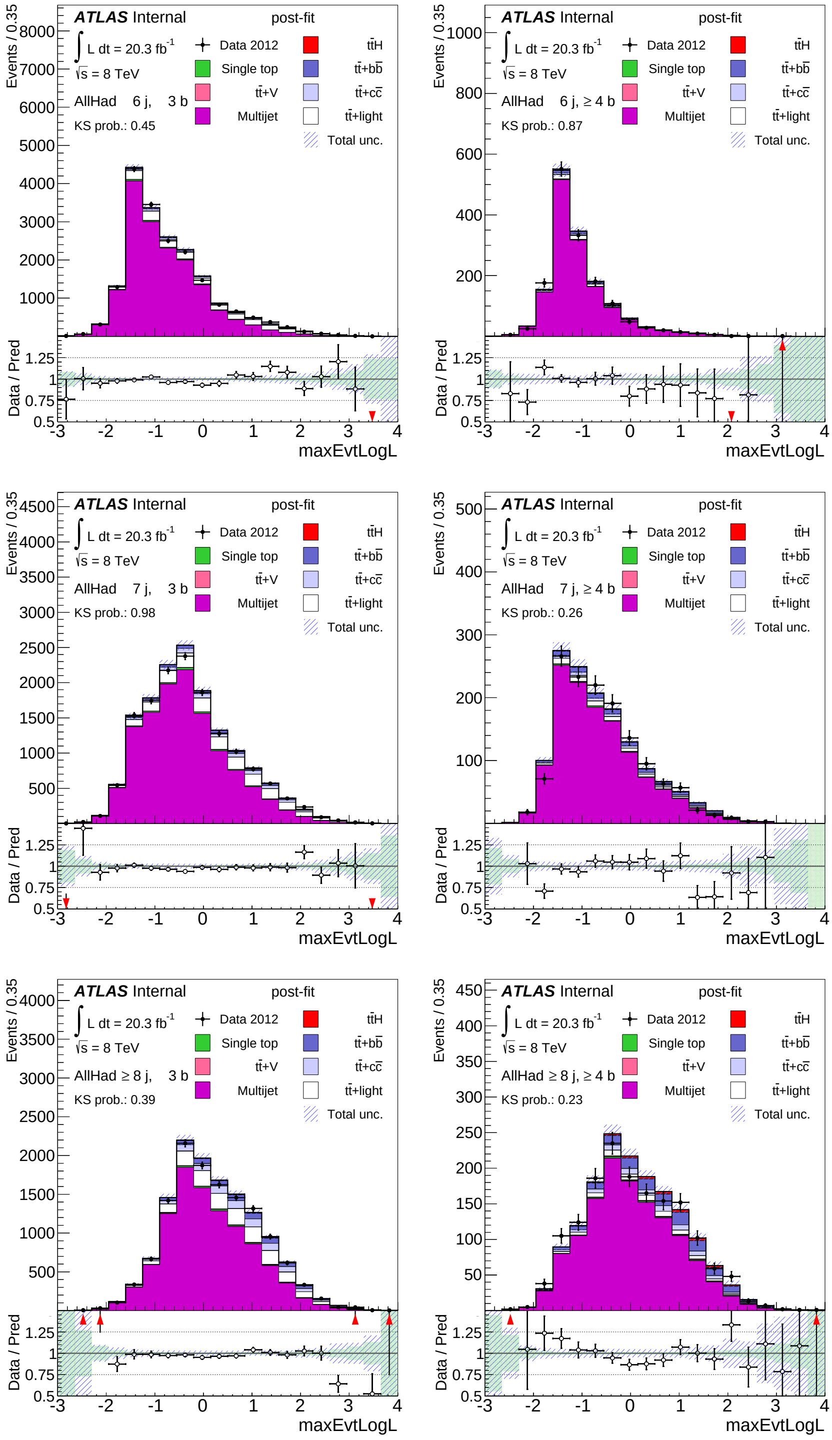


Figure B.4: $\mathcal{D}$ post-fit.

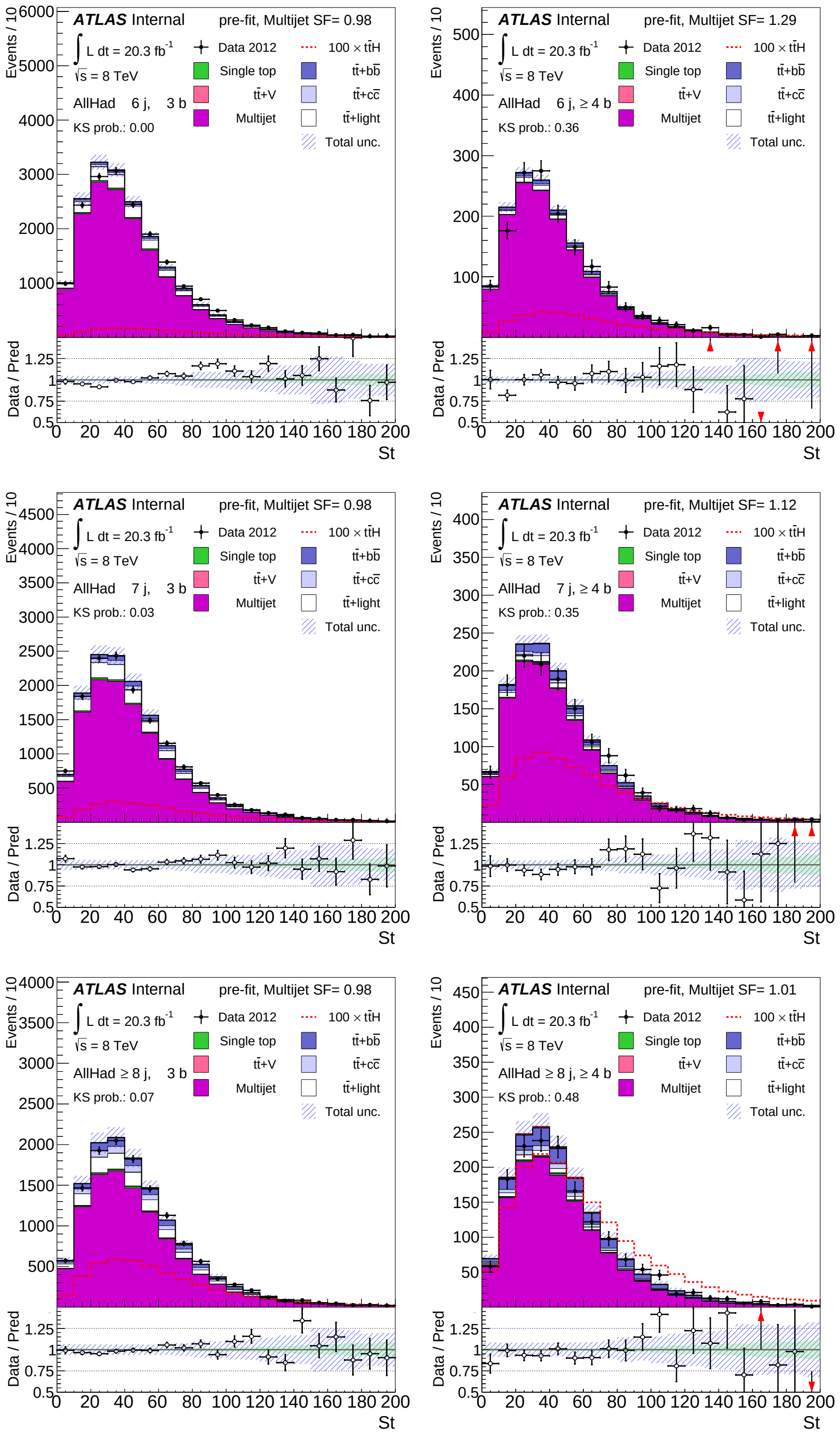


Figure B.5: S_T pre-fit.

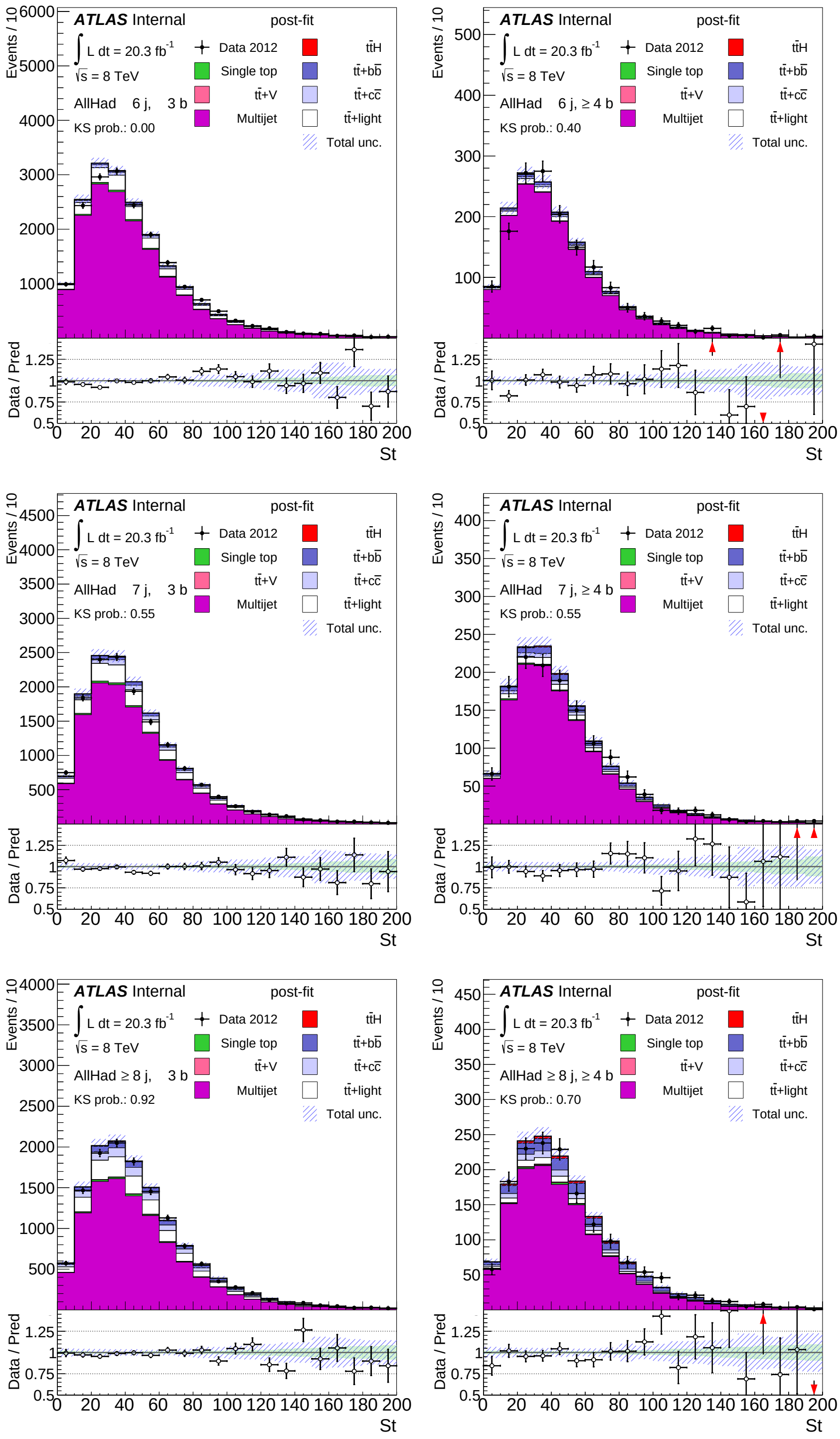


Figure B.6: S_{T} post-fit.

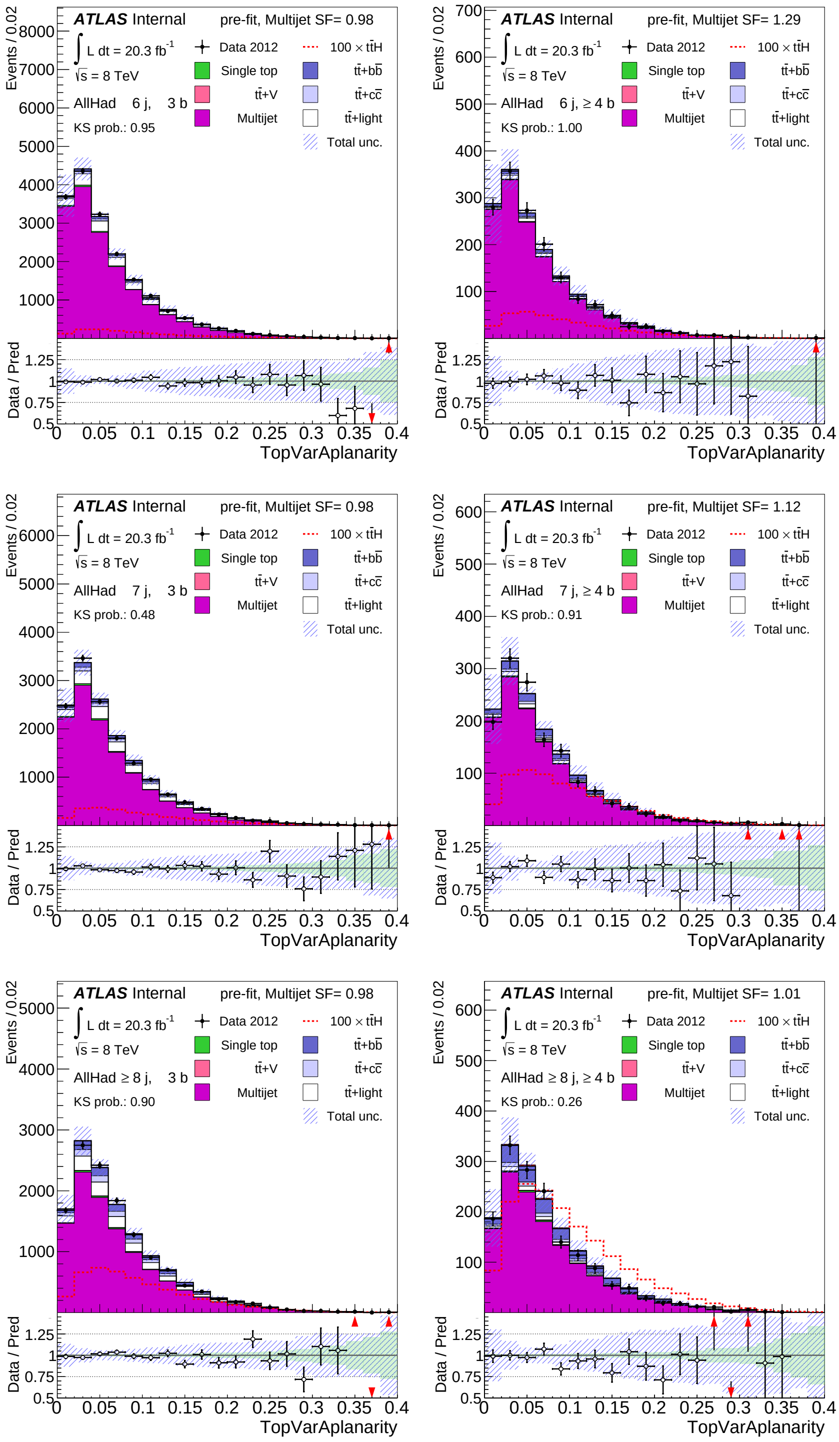


Figure B.7: Aplanarity pre-fit.

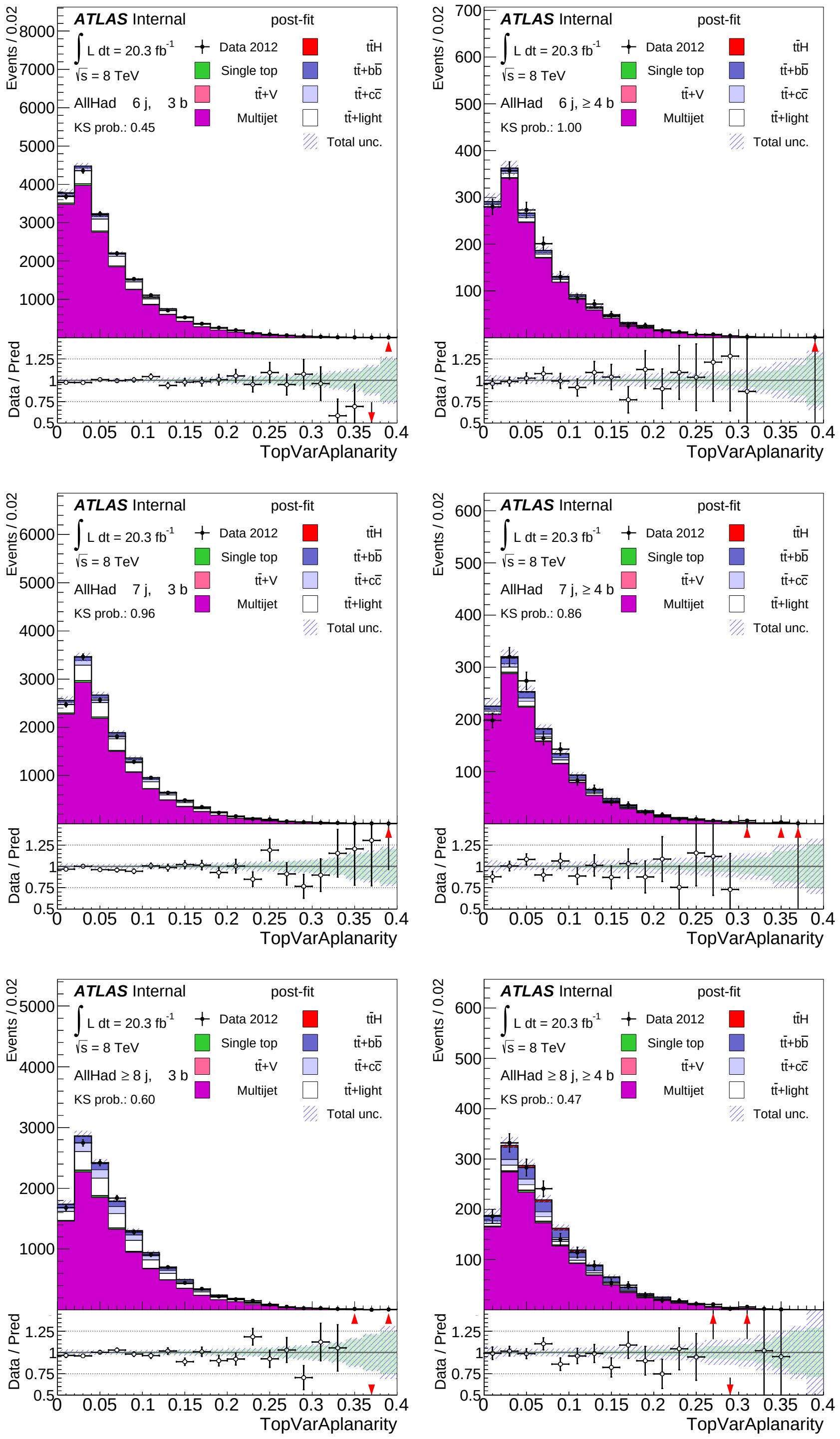


Figure B.8: Aplanarity post-fit.

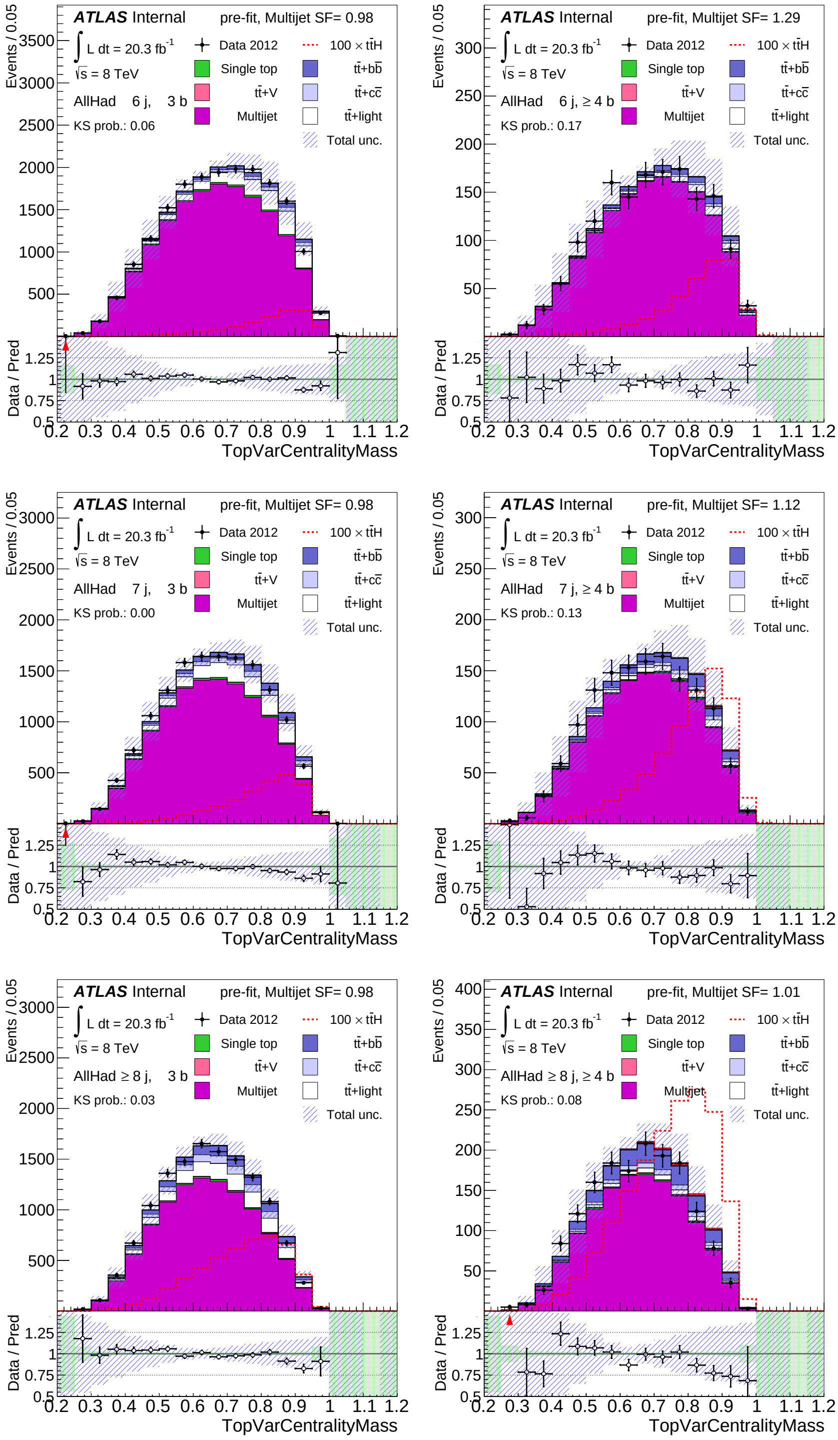


Figure B.9: Centrality$_{\text{Mass}}$ pre-fit.

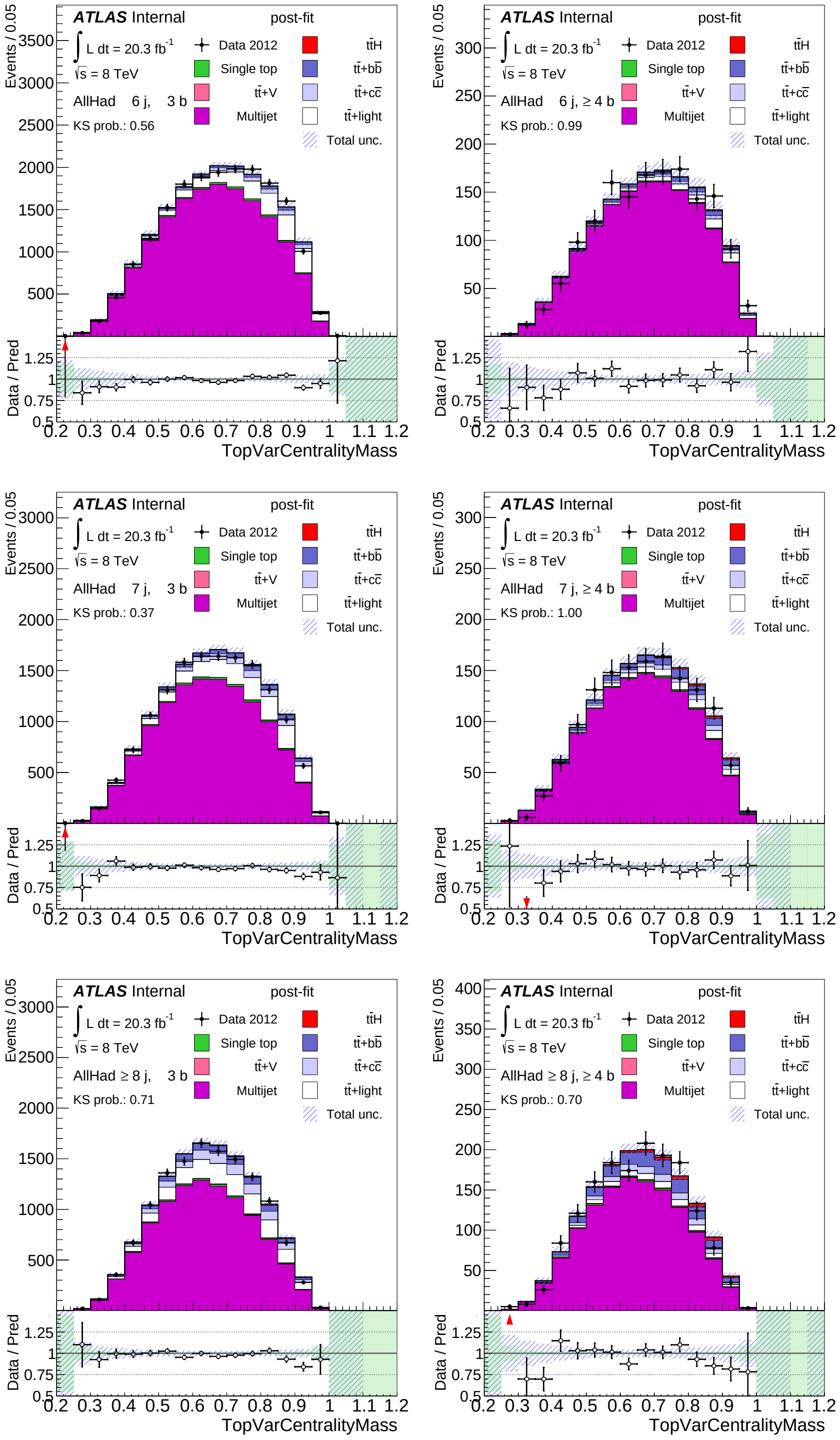


Figure B.10: $\text{Centrality}_{\text{Mass}}$ post-fit.

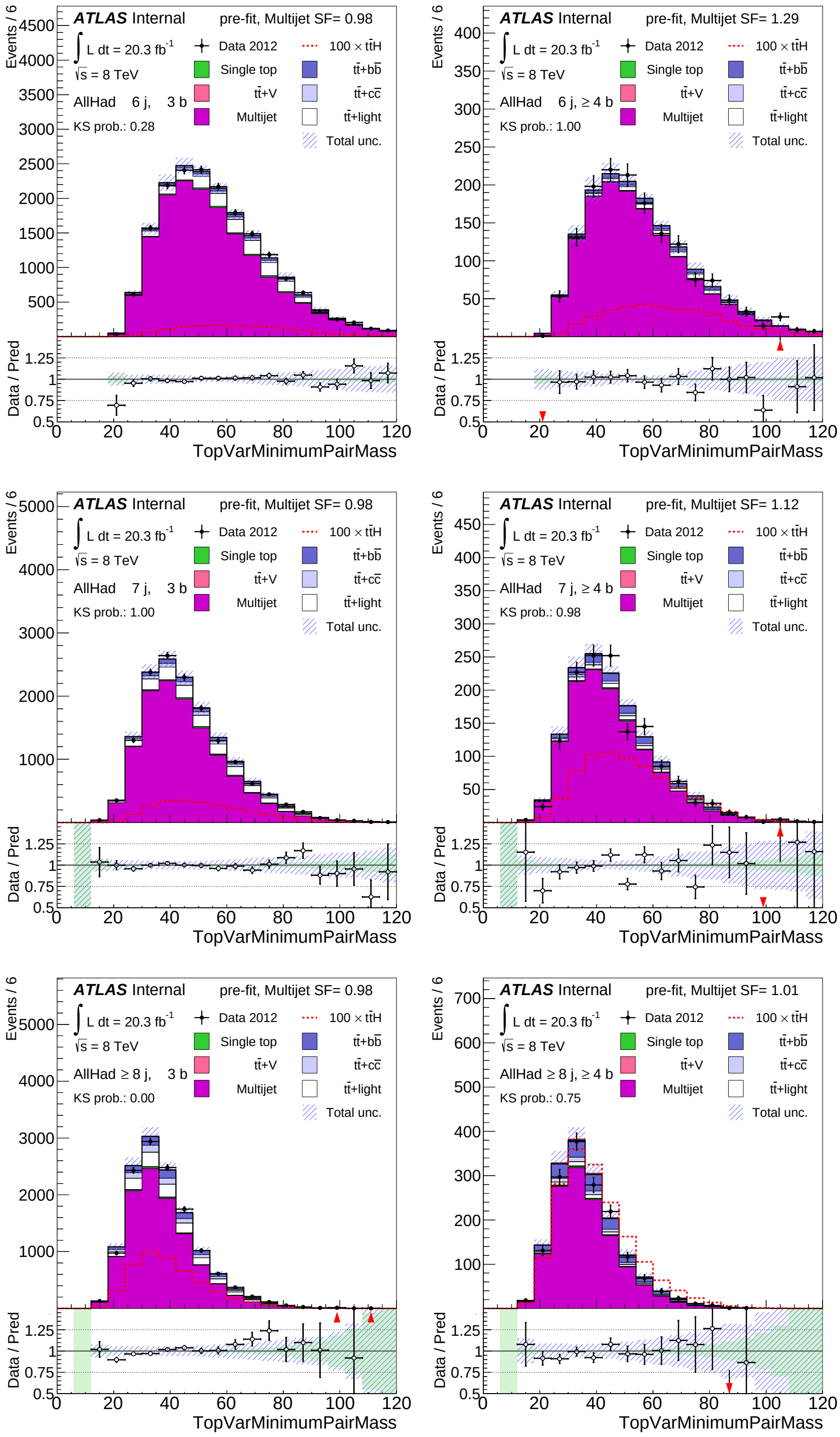


Figure B.11: $m_{jj}^{\min}$ pre-fit.

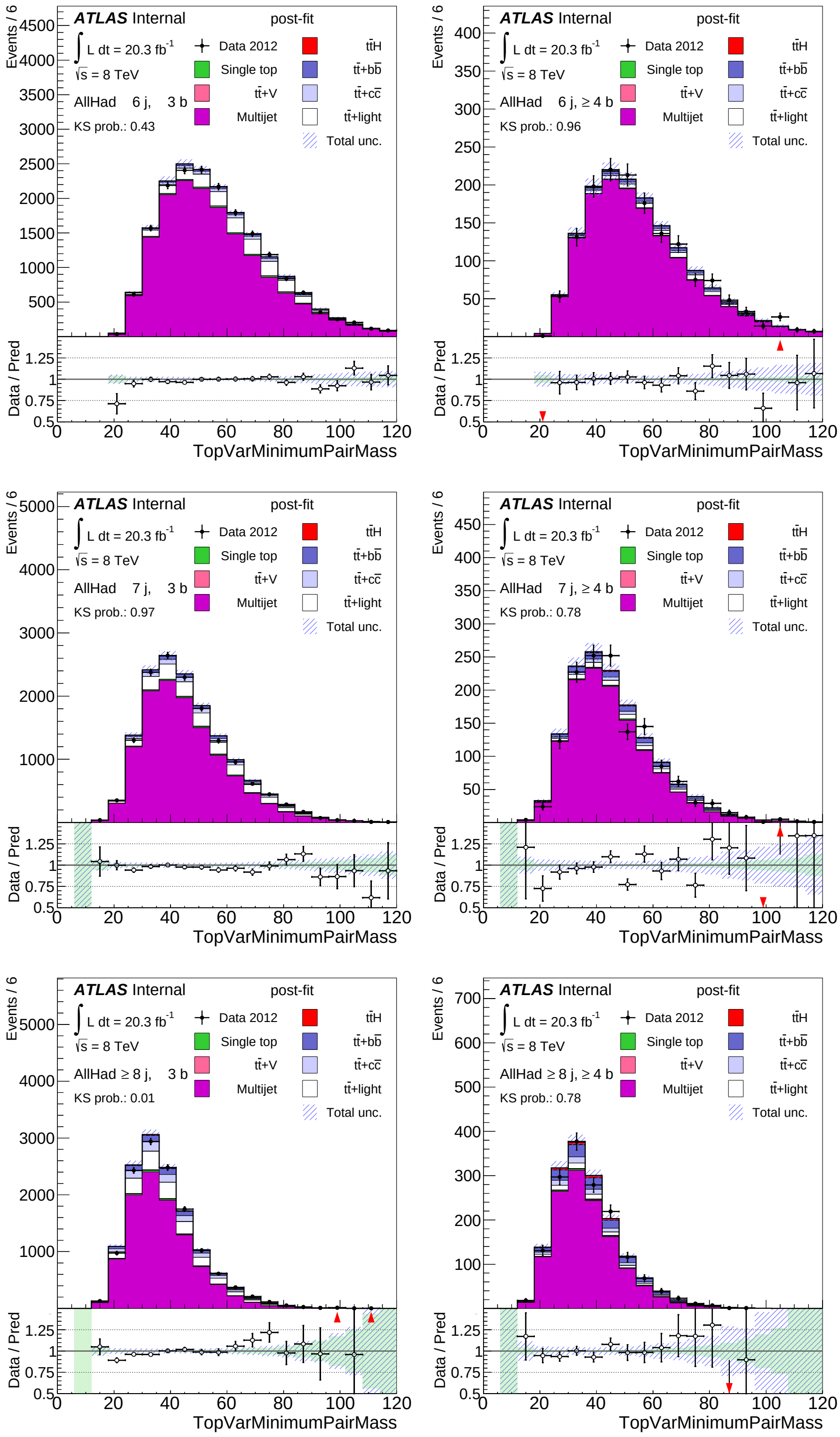


Figure B.12: $m_{jj}^{\min}$ post-fit.

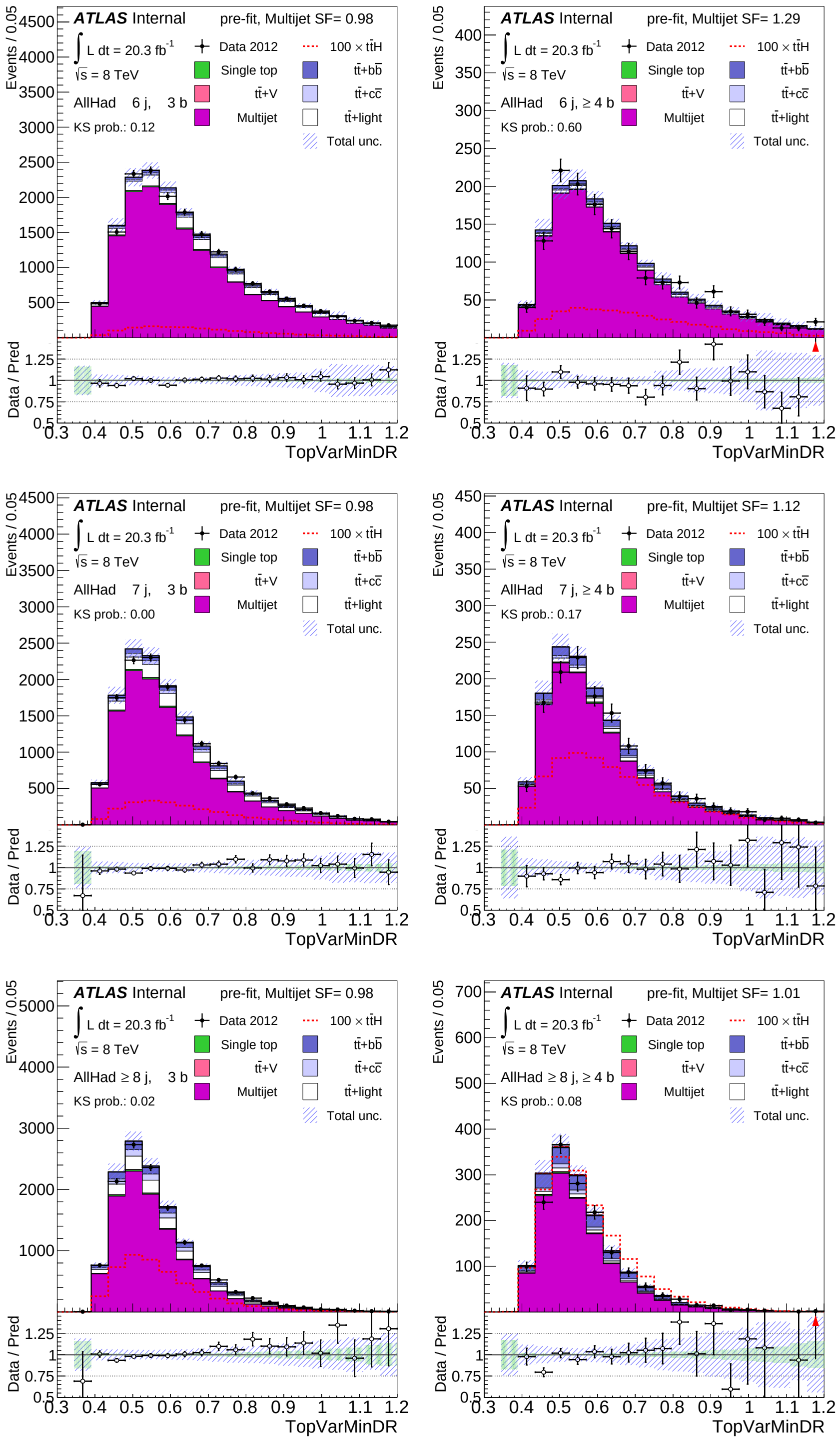


Figure B.13: ΔR^{min} pre-fit.

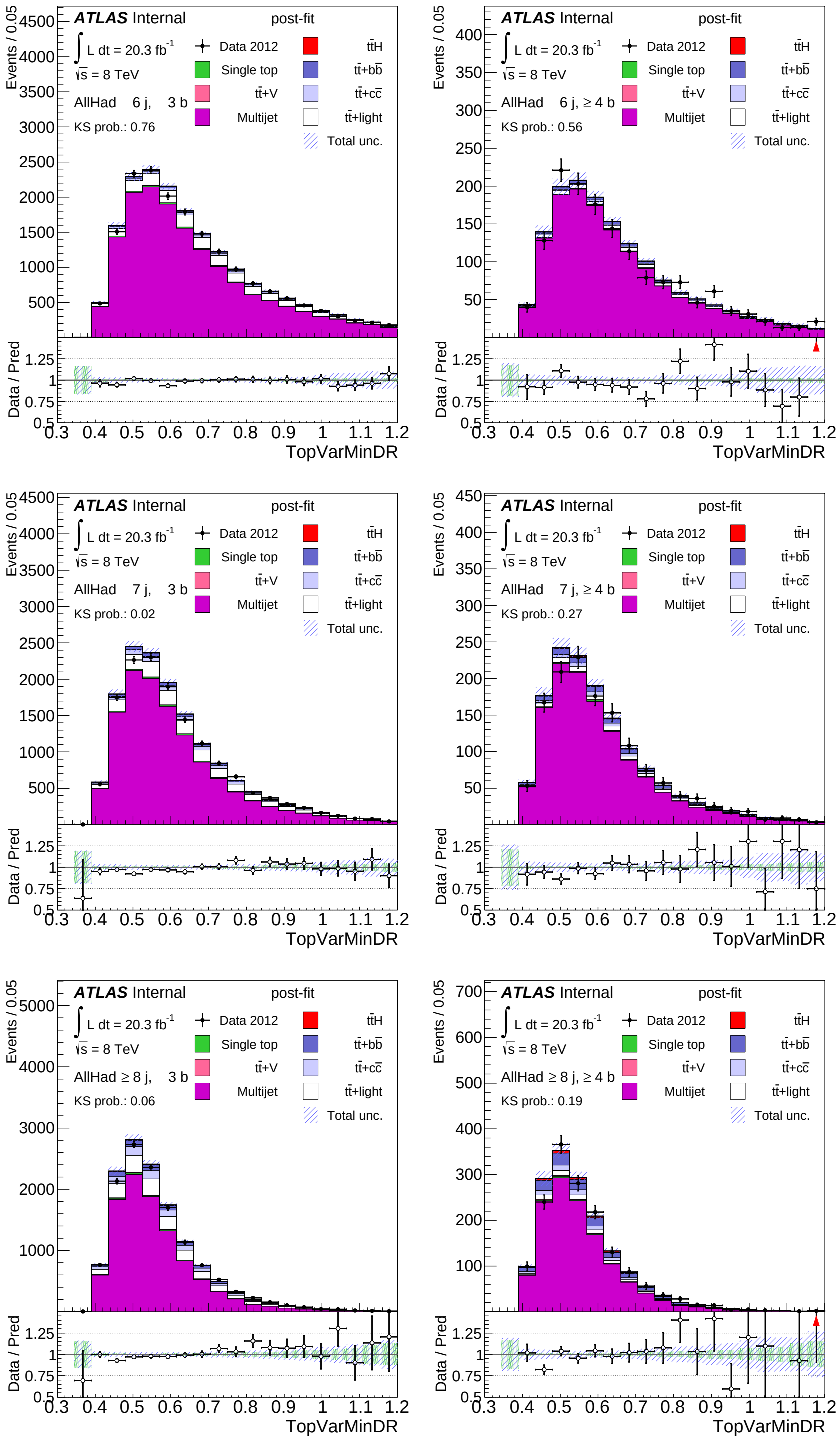


Figure B.14: ΔR^{min} post-fit.

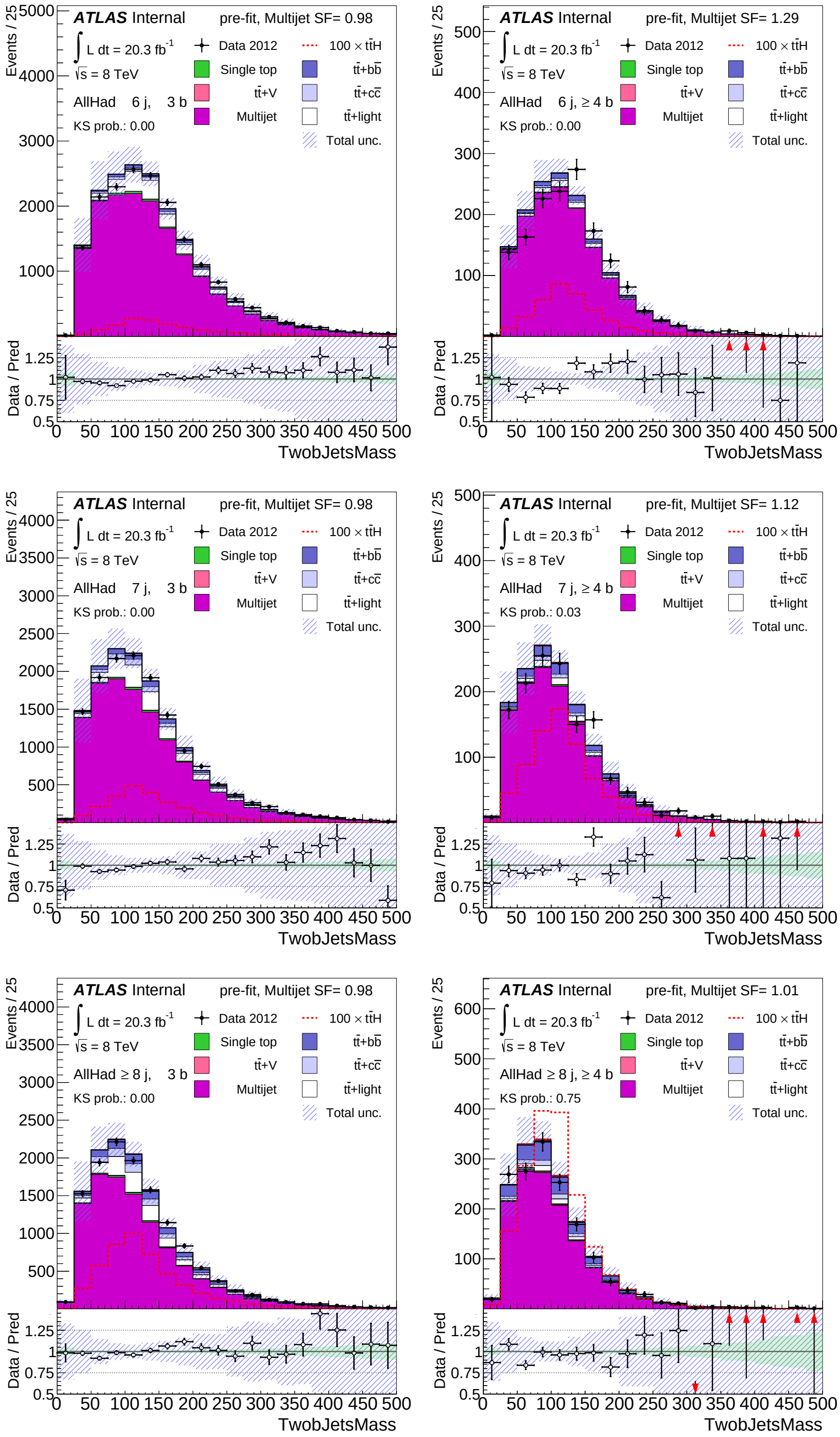


Figure B.15: $m_{bb}^{\Delta R(b,b)^{\min}}$ pre-fit.

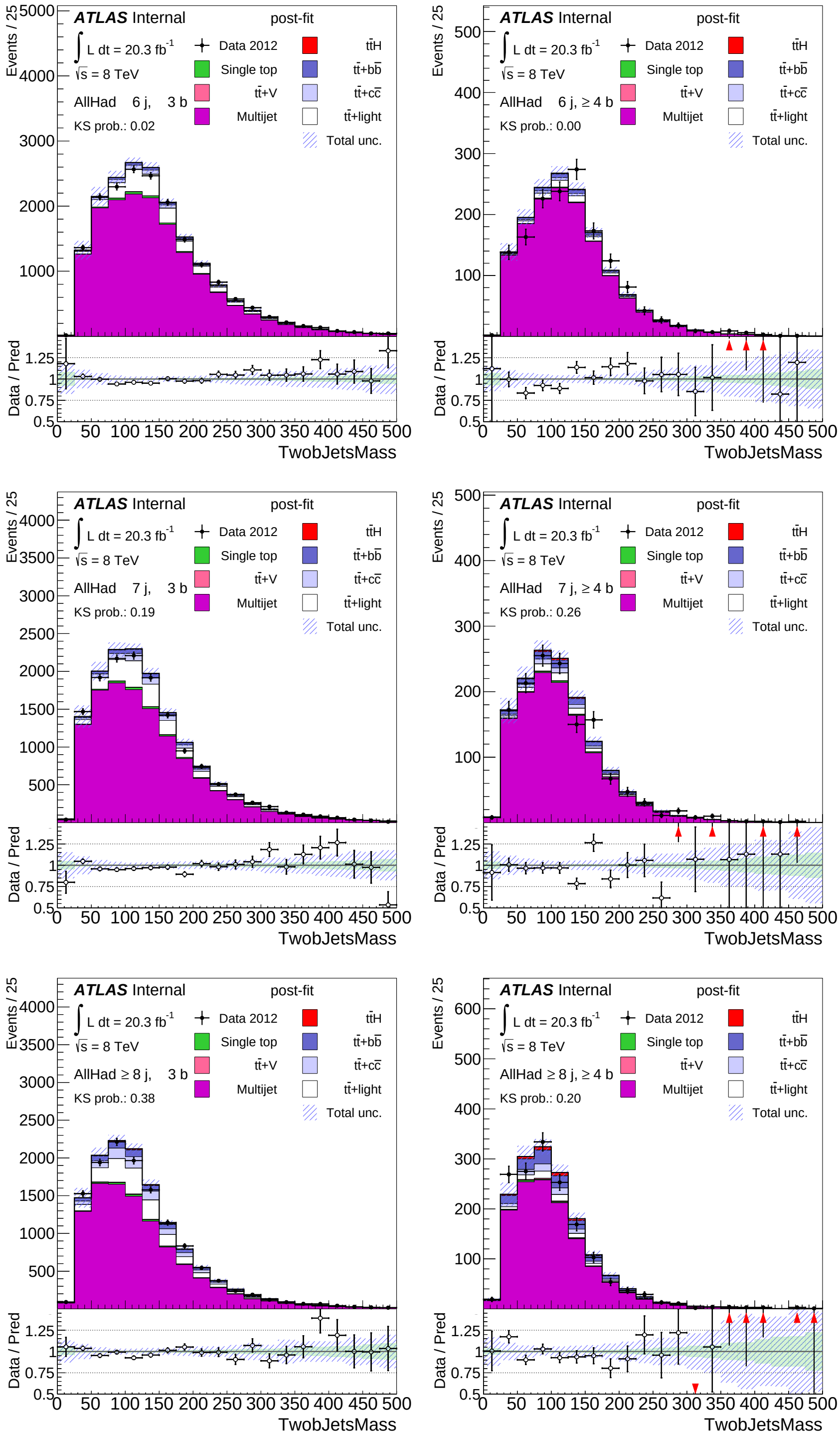


Figure B.16: $m_{bb}^{\Delta R(b,b)^{\min}}$ post-fit.

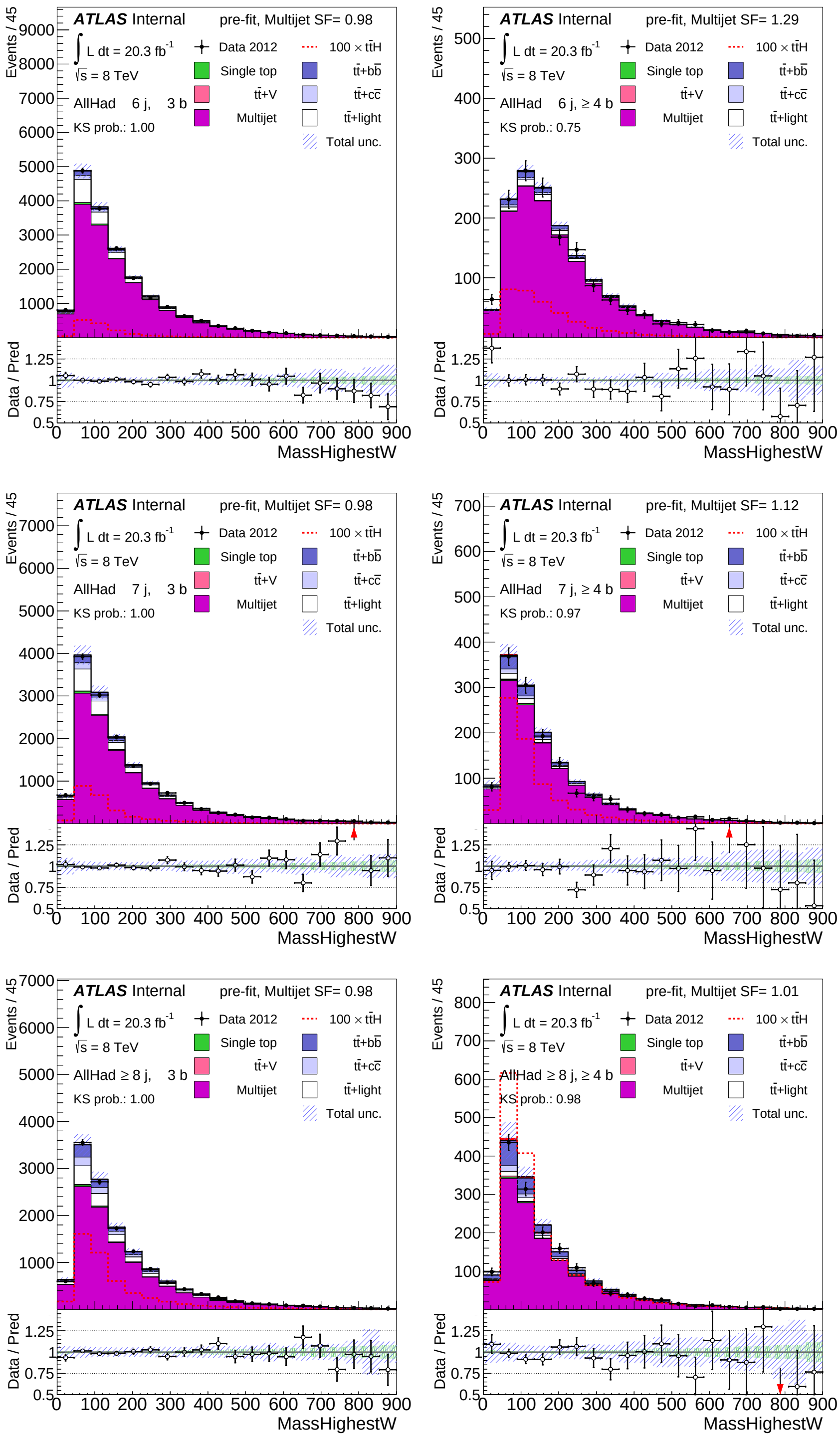


Figure B.17: $m_{2\,\mathrm{jets}}$ pre-fit.

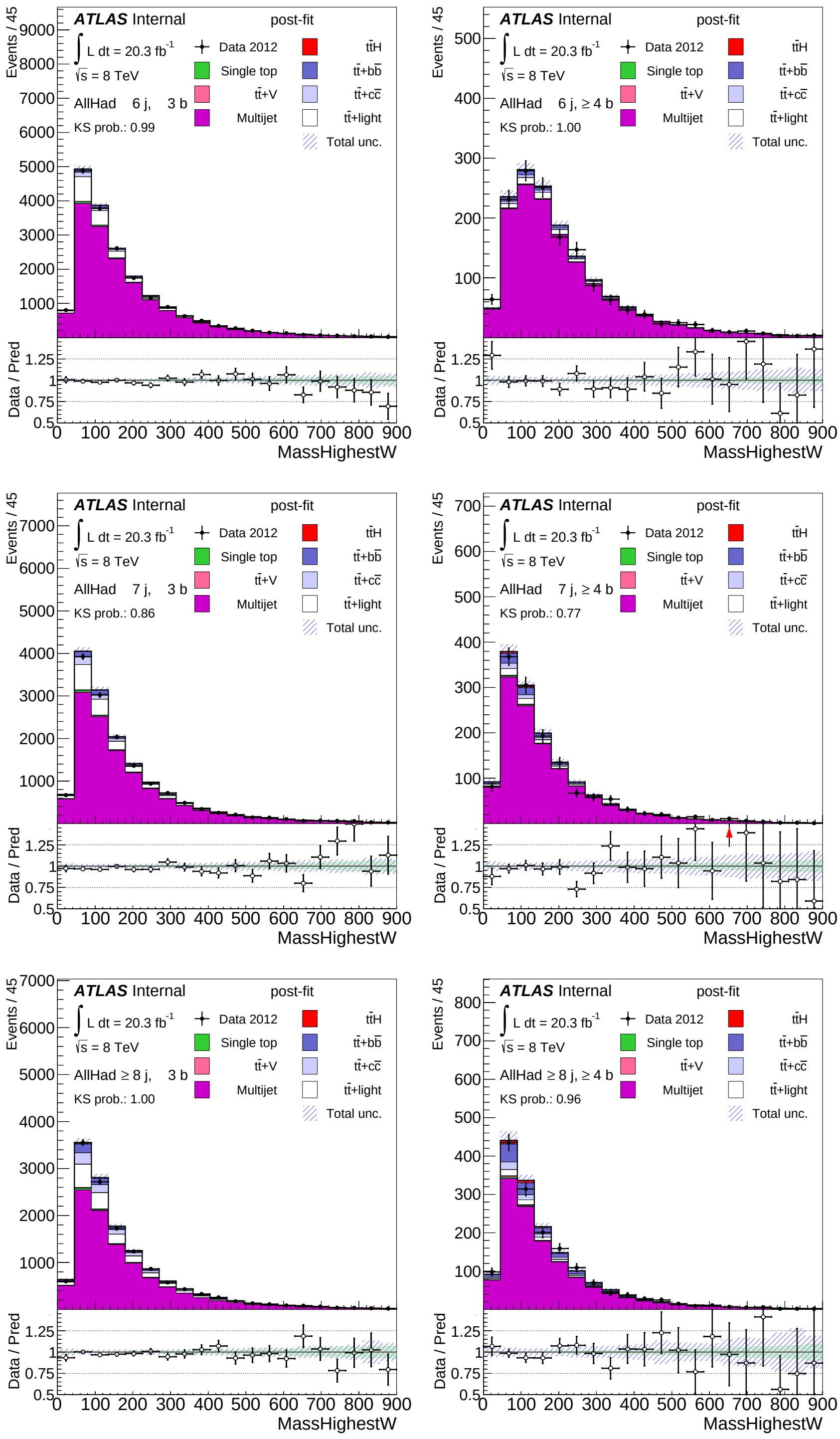


Figure B.18: $m_{2\,\mathrm{jets}}$ post-fit.

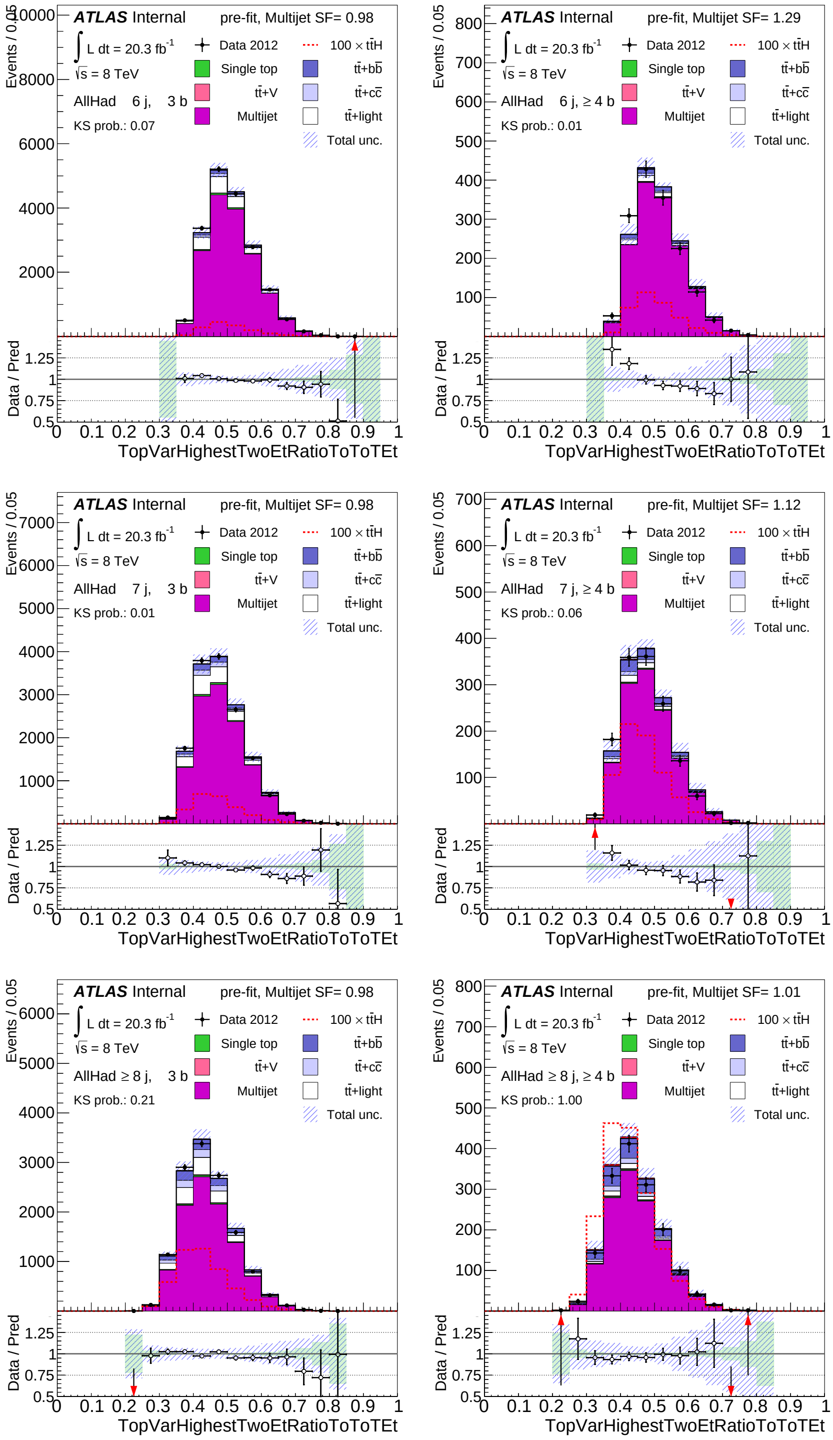


Figure B.19: $\frac{E_{\mathrm{T}\,1}+E_{\mathrm{T}\,2}}{\sum E_{\mathrm{T}}^{\mathrm{jets}}}$ pre-fit.

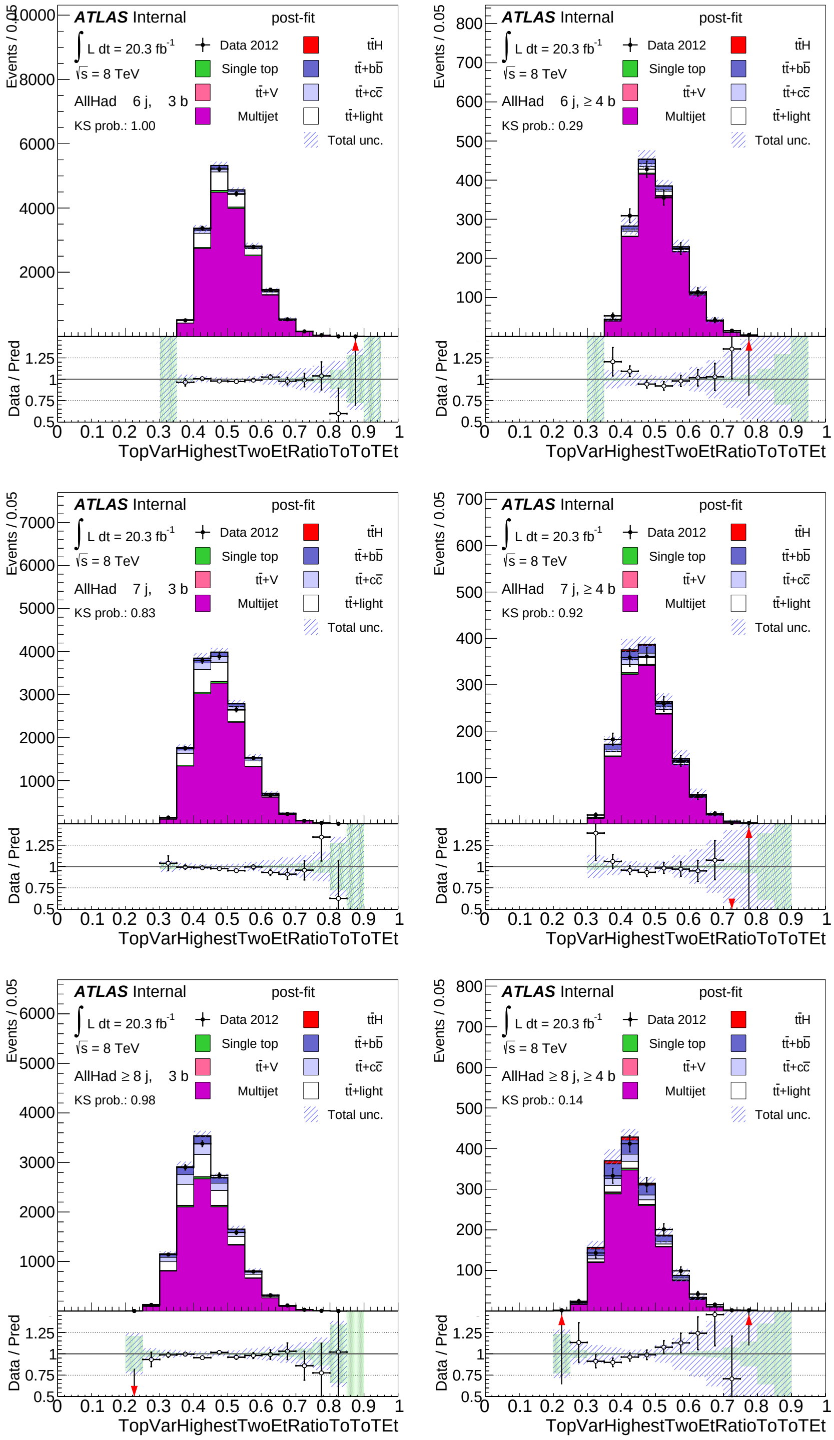


Figure B.20: $\frac{E_{\mathrm{T}\,1}+E_{\mathrm{T}\,2}}{\sum E_{\mathrm{T}}^{\mathrm{jets}}}$ post-fit.

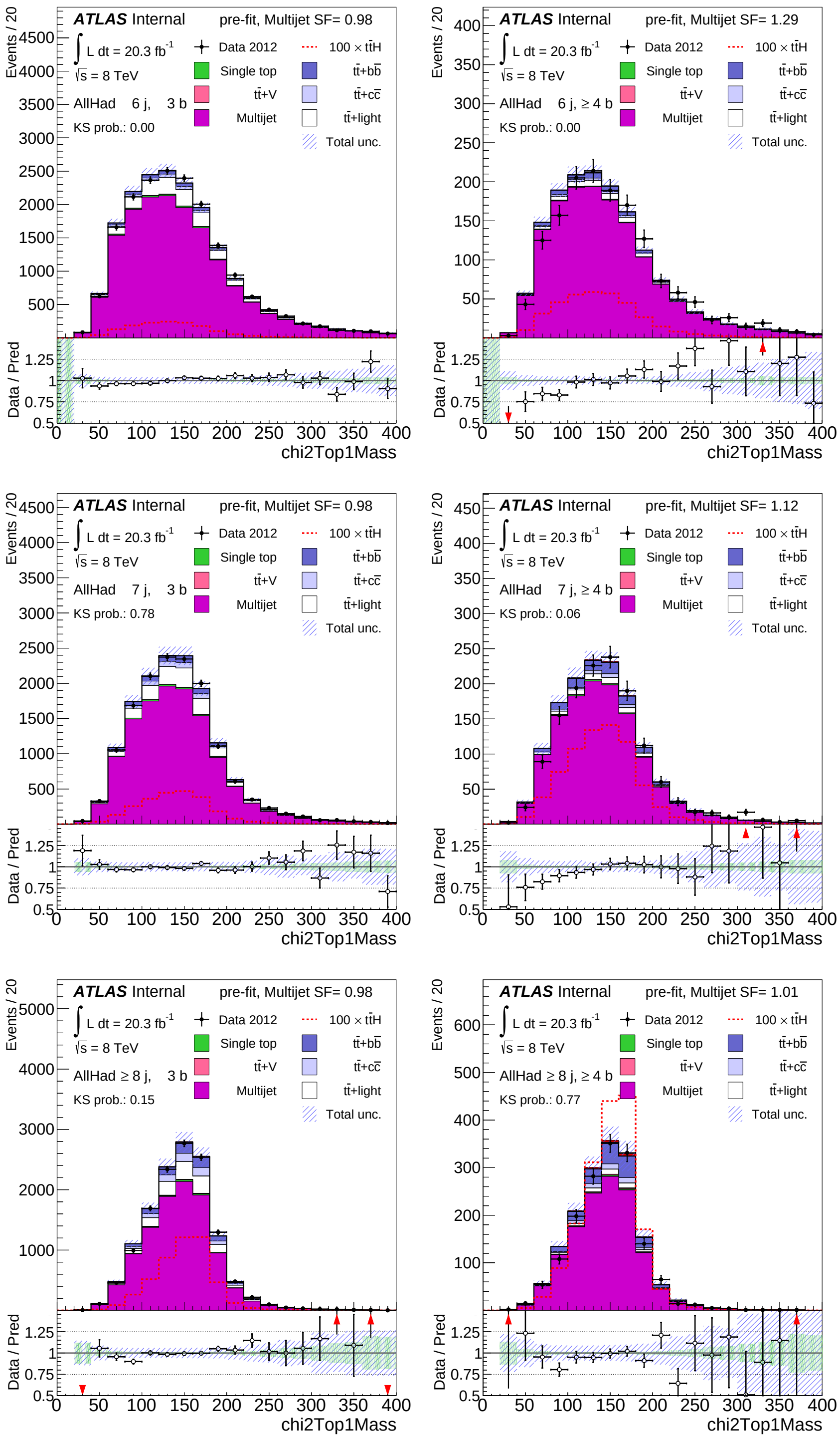


Figure B.21: $m_{\text{top},1}$ pre-fit.

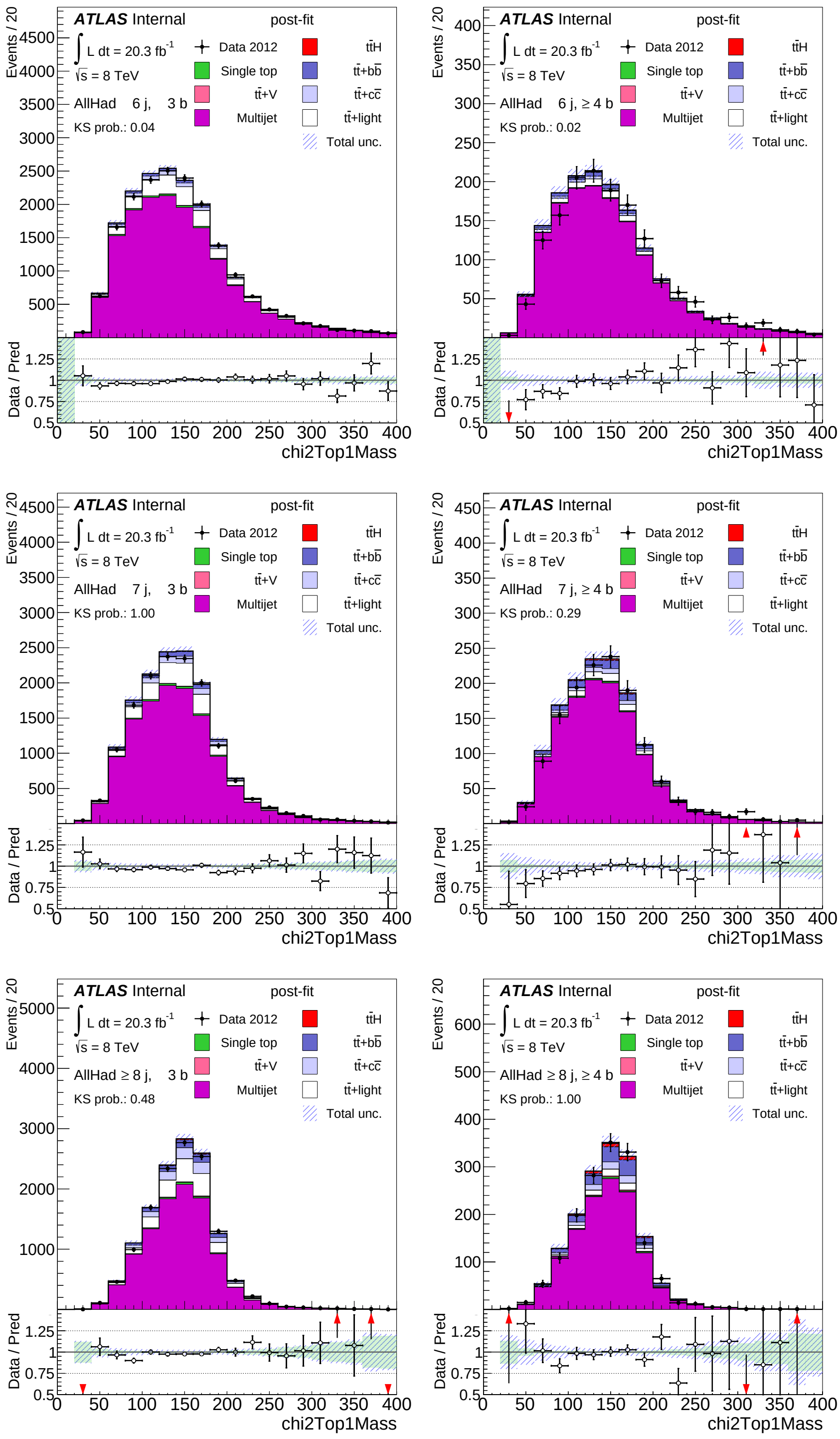


Figure B.22: $m_{\mathrm{top},1}$ post-fit.

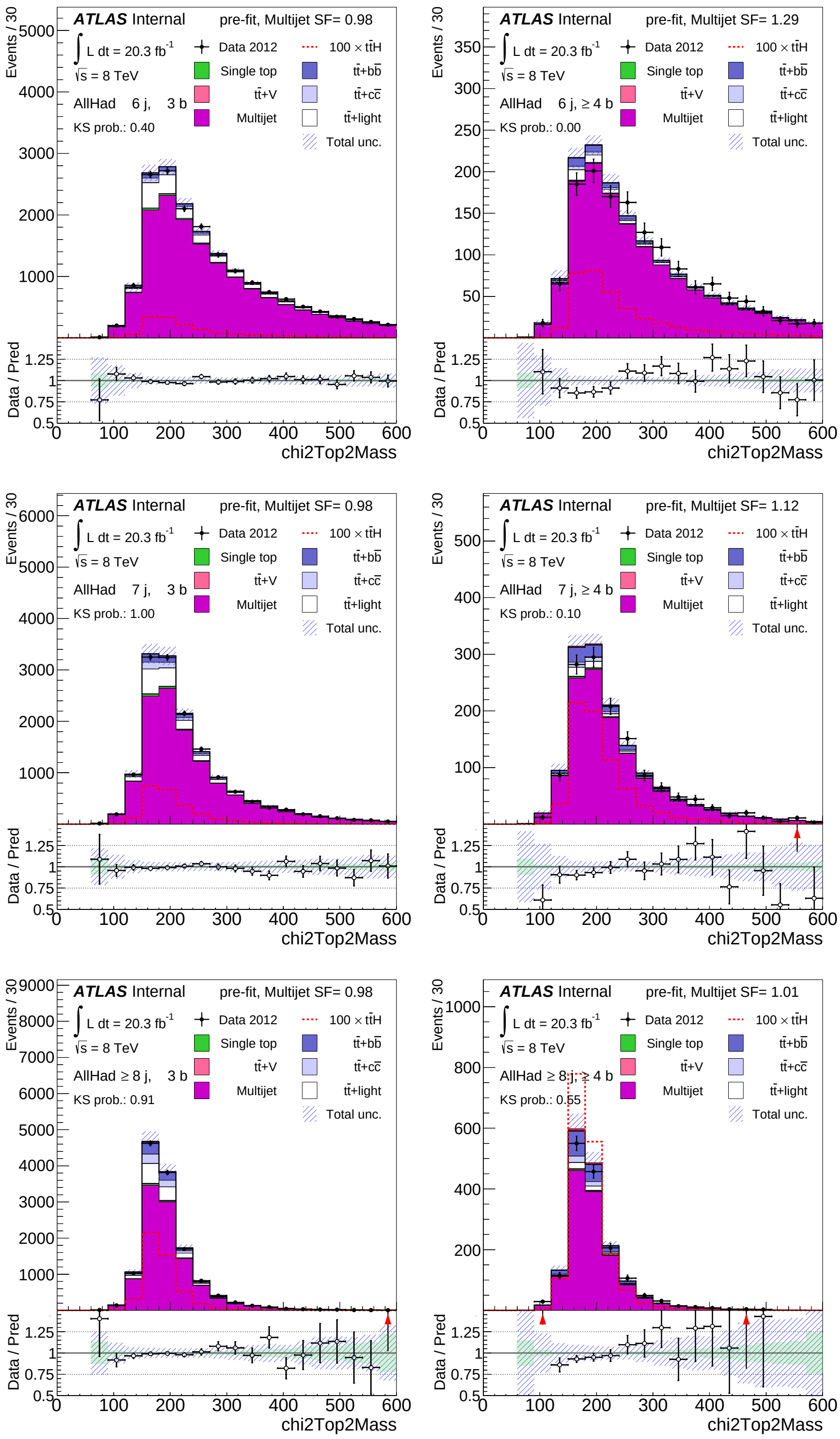


Figure B.23: $m_{\mathrm{top},2}$ pre-fit.

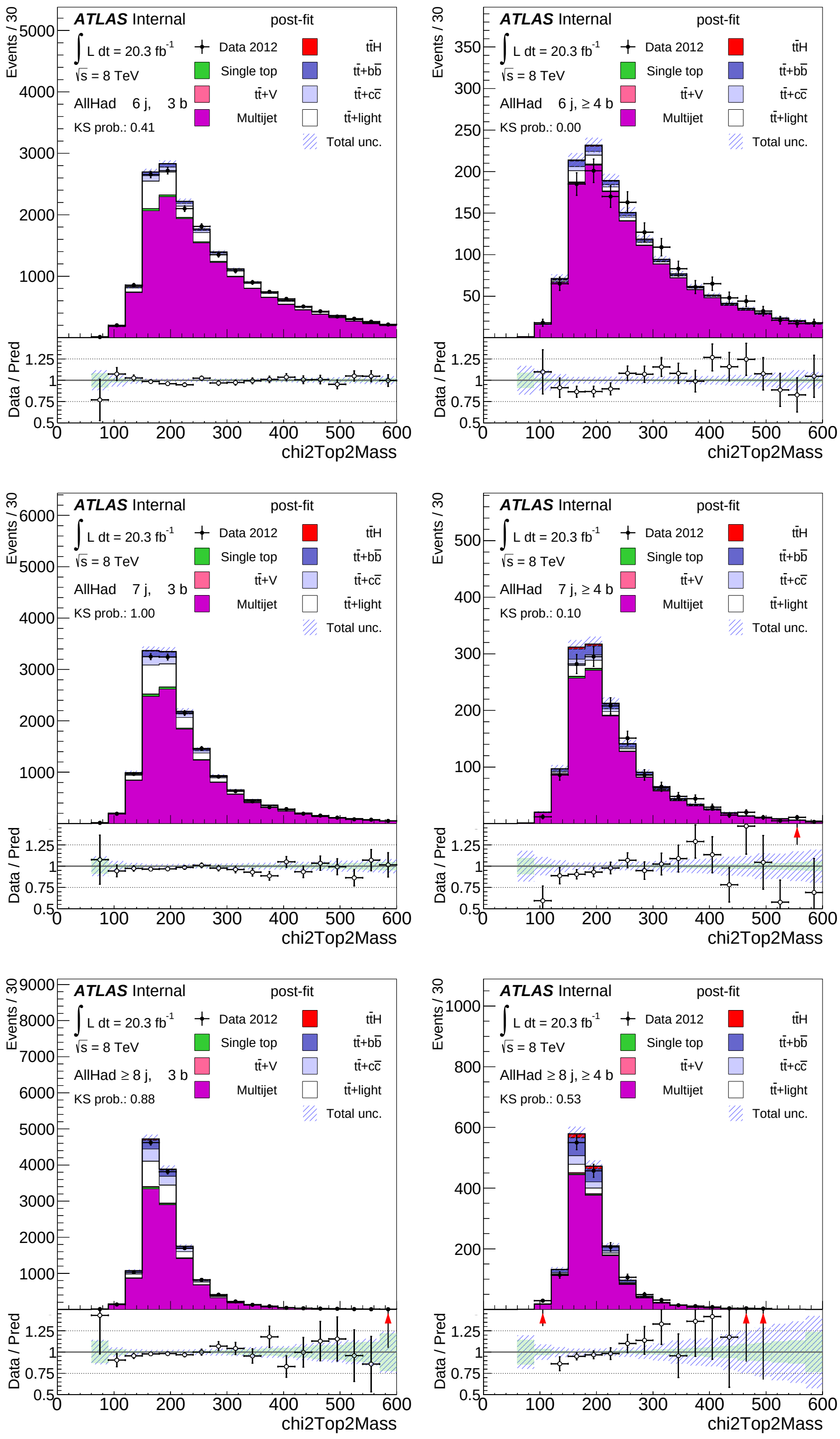


Figure B.24: $m_{\text{top},2}$ post-fit.

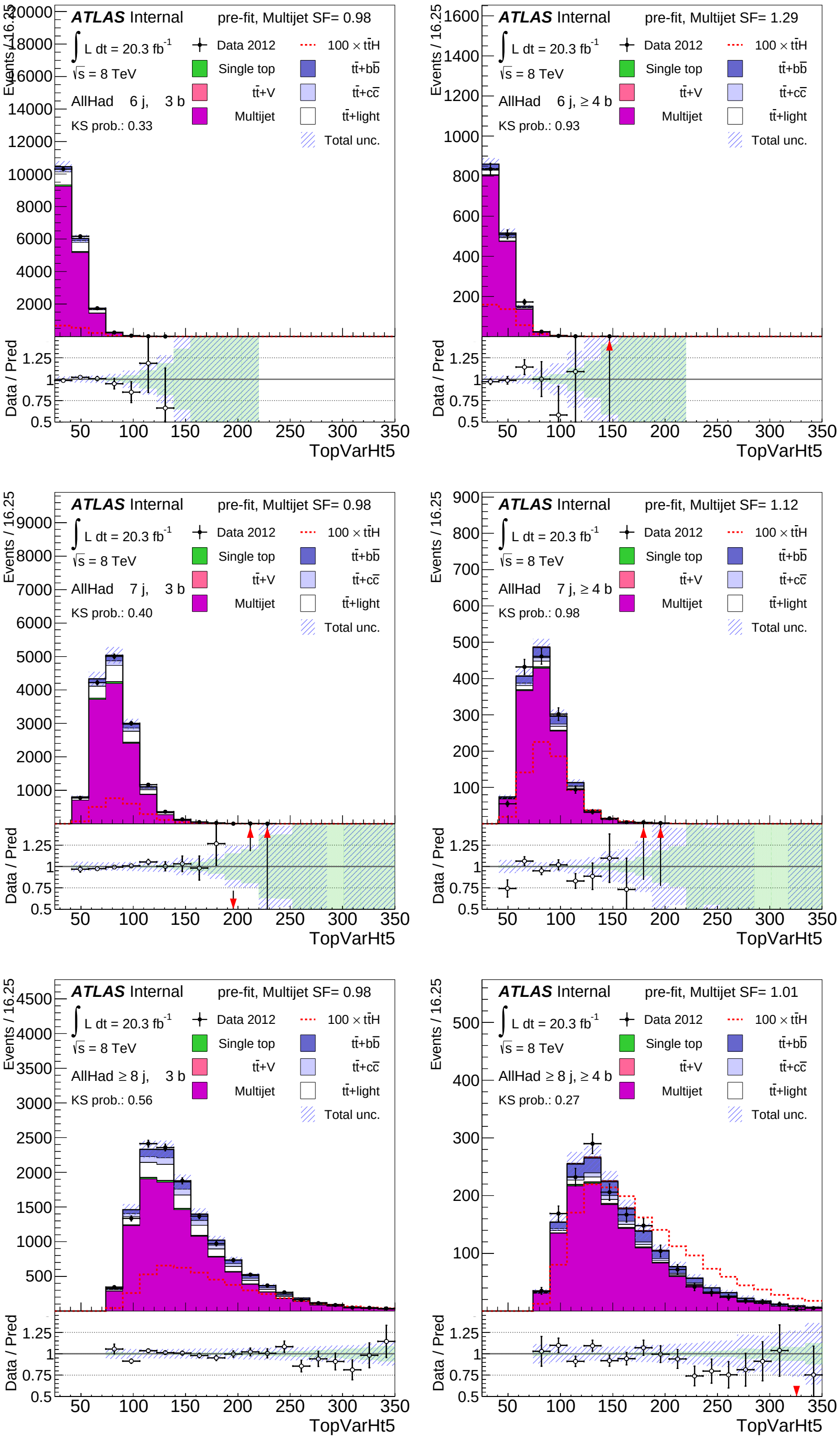


Figure B.25: $H_{\mathrm{T}\,5}$ pre-fit.

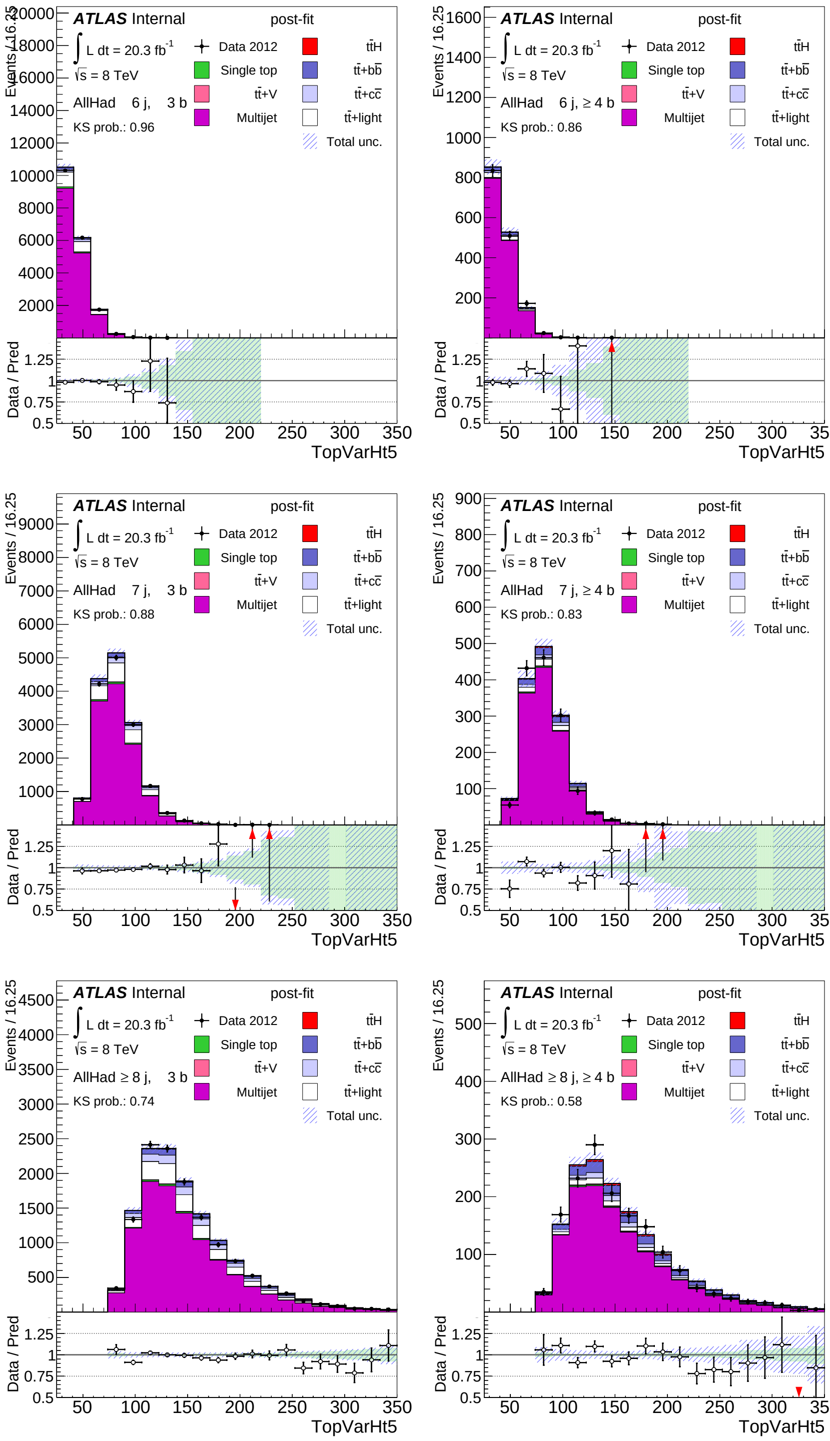


Figure B.26: $H_{\mathrm{T}5}$ post-fit.

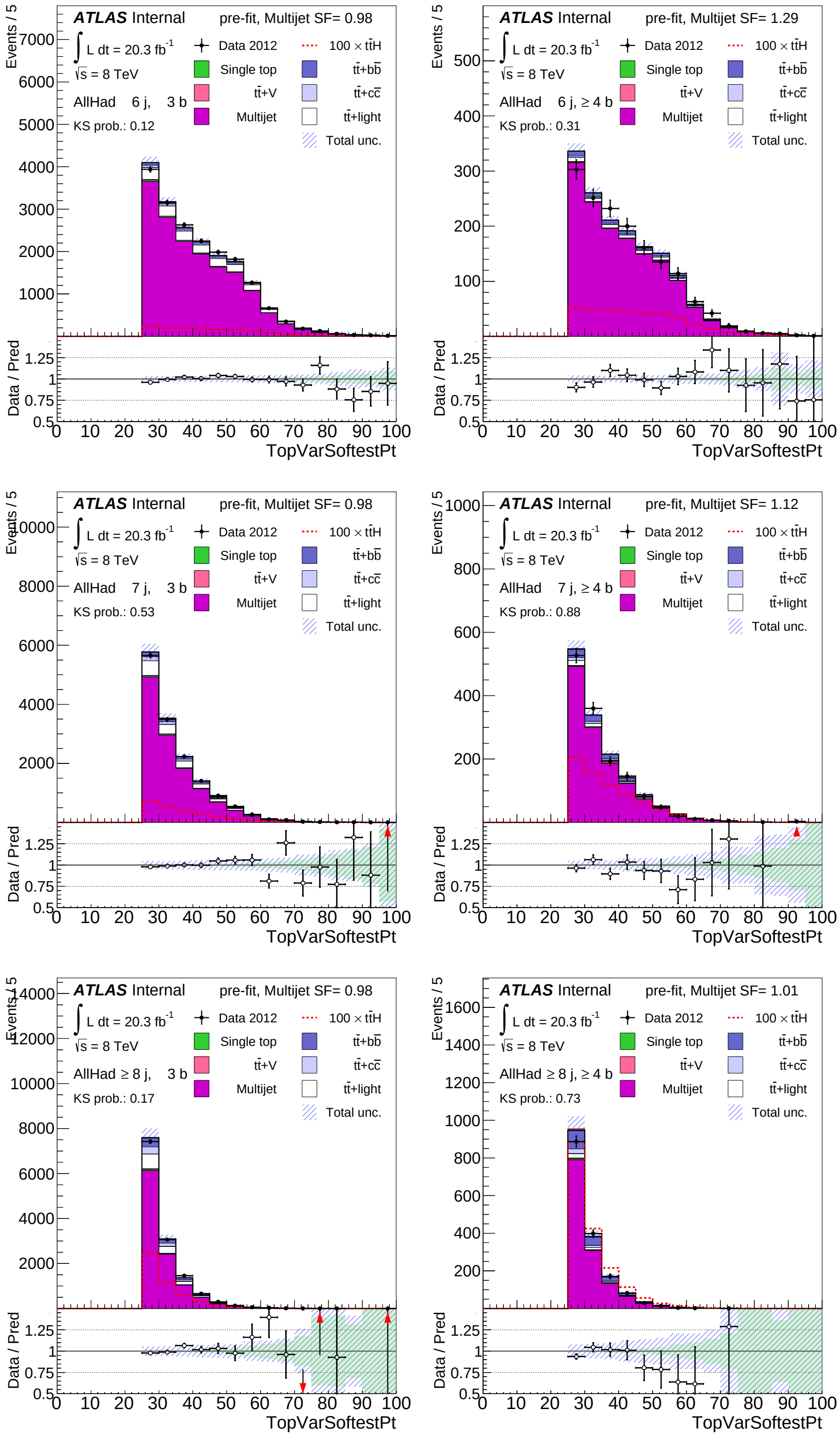


Figure B.27: $p_{\mathrm{T}}^{\text{softest jet}}$ pre-fit.

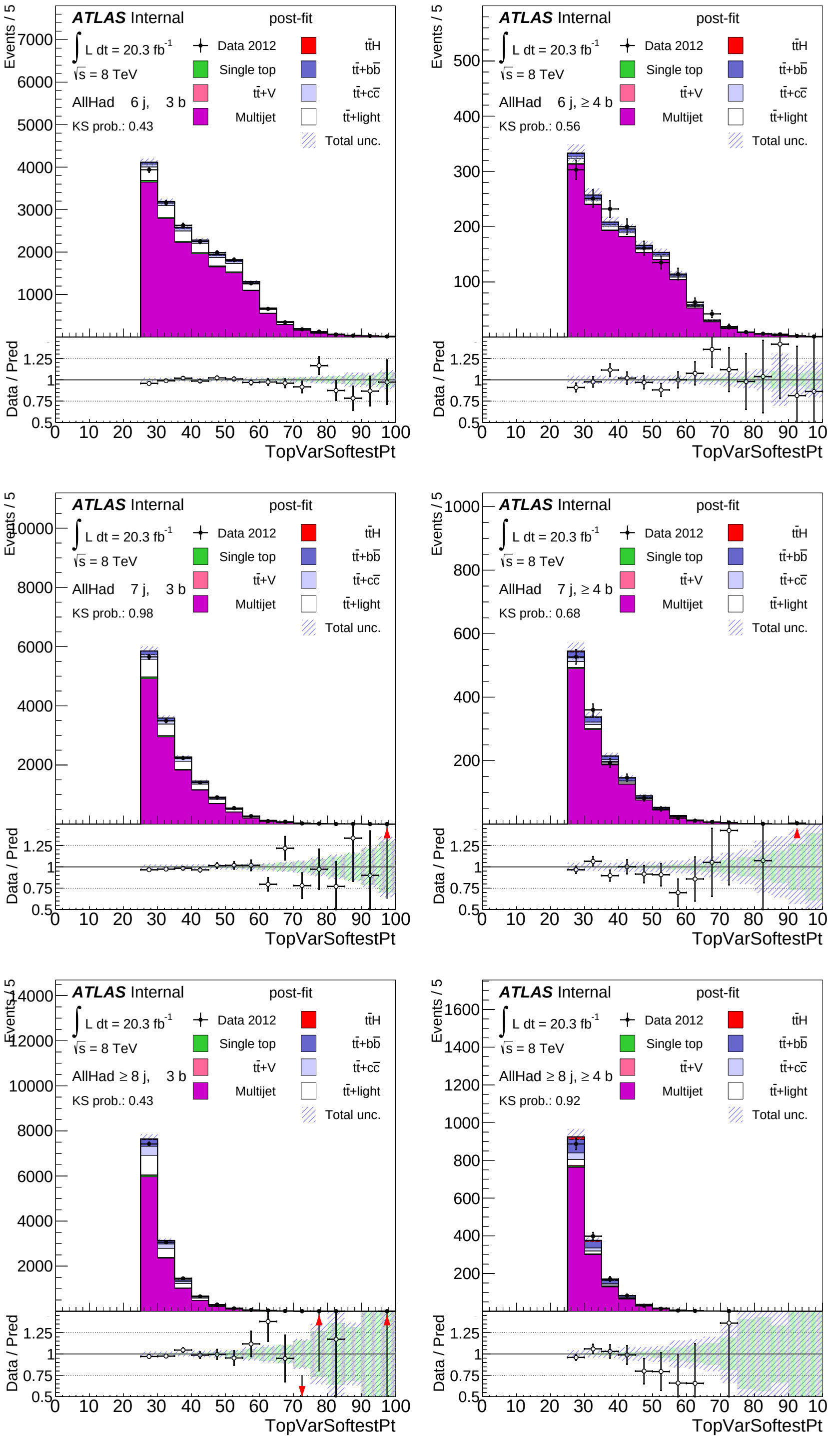
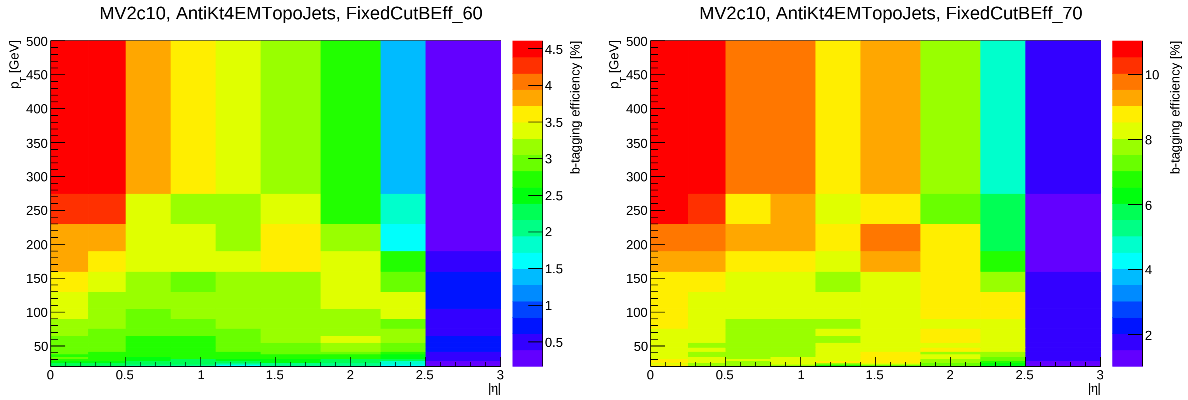


Figure B.28: $p_{\mathrm{T}}^{\text{softest jet}}$ post-fit.

APPENDIX C

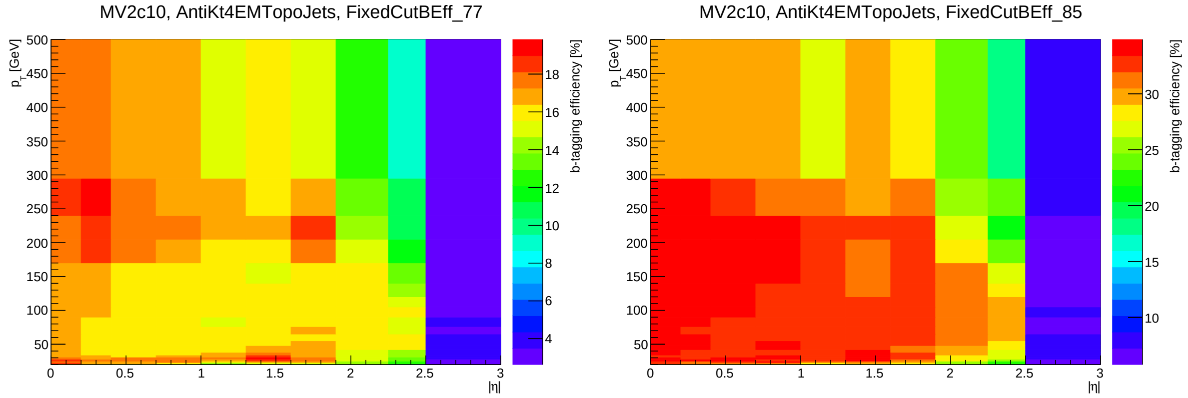
b-tagging maps employed in the Pseudo Matrix Element discriminant

This Appendix contains additional informations related to the construction og the $\mathcal{P.M.E.}$ discriminant, in the $t\bar{t}H$ multilepton analysis. The following plots show the b-tagging maps for c- and light-jets employed to compute the probabilities $P(\mathbf{x})$.



(a) 60% O.P.

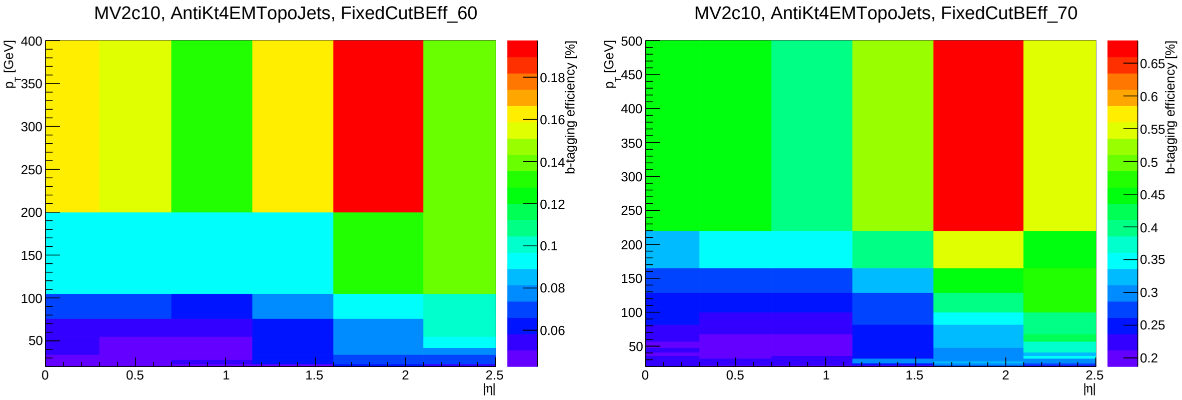
(b) 70% O.P.



(c) 77% O.P.

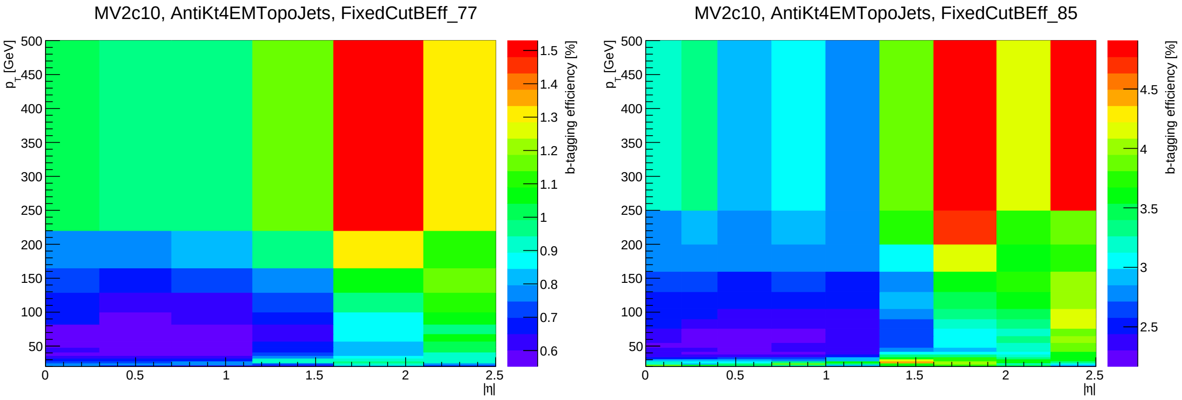
(d) 85% O.P.

Figure C.1: b-tagging maps for a true c-jet used to calculate the Pseudo-Matrix Element discriminator. The 2-D maps are provided for four different fixed O.P., 3 jet flavours (*i.e.*, from b-, c- and light-jet sources) and are parametrised in (p_T,η) of the jet.



(a) 60% O.P.

(b) 70% O.P.



(c) 77% O.P.

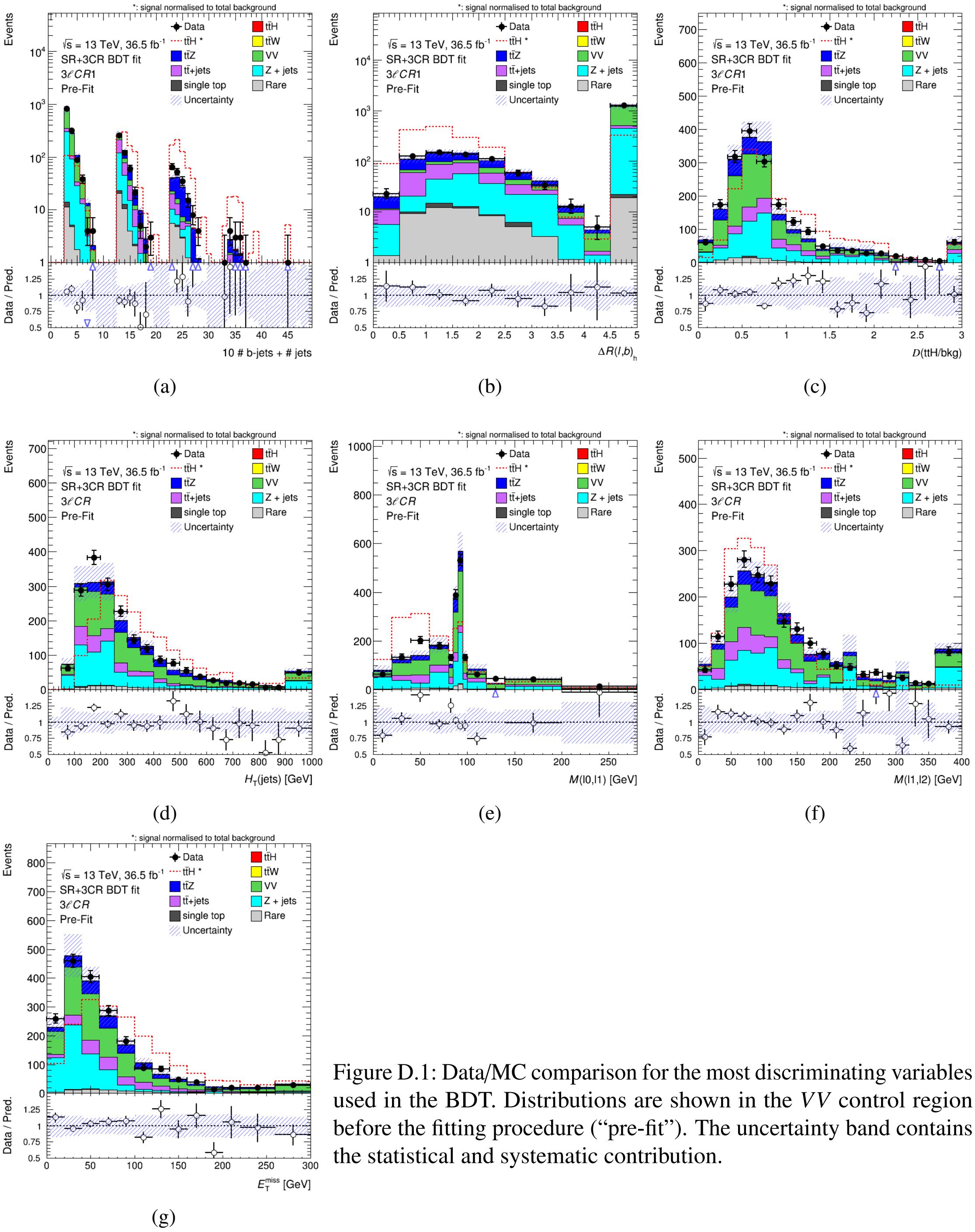
(d) 85% O.P.

Figure C.2: *b*-tagging maps for a true light-jet used to calculate the Pseudo-Matrix Element discriminator. The 2-D maps are provided for four different fixed O.P., 3 jet flavours (*i.e.*, from *b*-, *c*- and light-jet sources) and are parametrised in (p_{T},η) of the jet.

APPENDIX D

Validation of the BDT approach in the multilepton analysis

This Appendix is dedicated to the validation of the modelling of the variables entering the BDT. Each histogram presents the input variable of the BDT in one of the analysis regions, before or after the fit. The fit procedure improves level of agreement for each variable and reduce the systematic uncertainties.

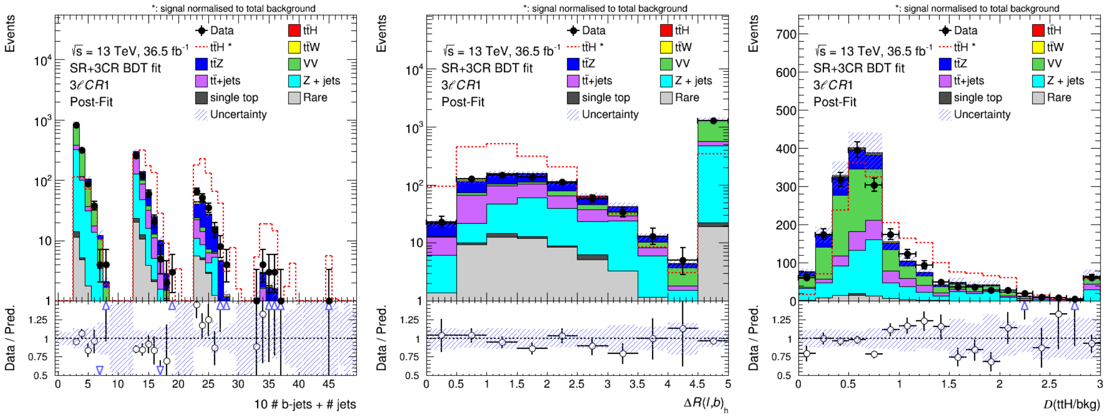


(a) (b) (c)

(d) (e) (f)

(g)

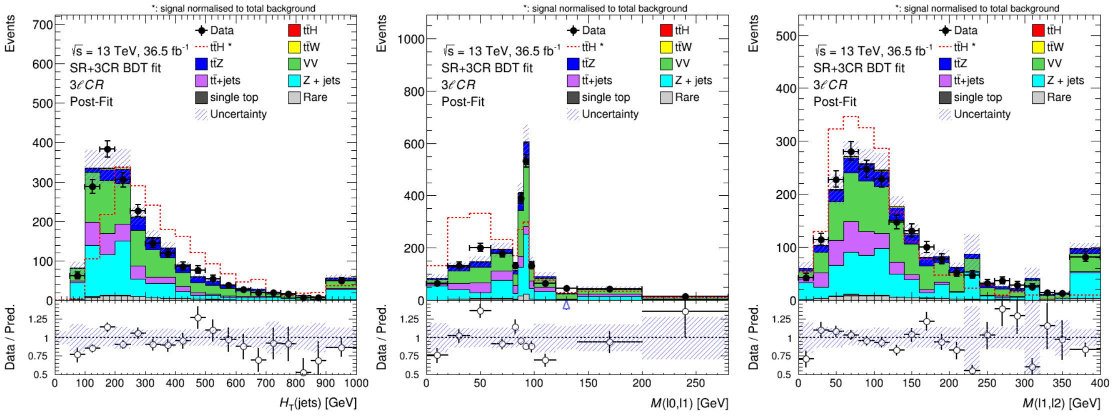
Figure D.1: Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the VV control region before the fitting procedure (“pre-fit”). The uncertainty band contains the statistical and systematic contribution.



(a)

(b)

(c)



(d)

(e)

(f)

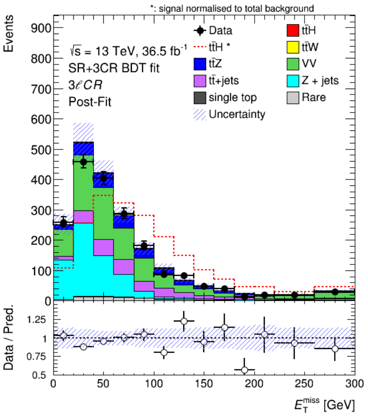
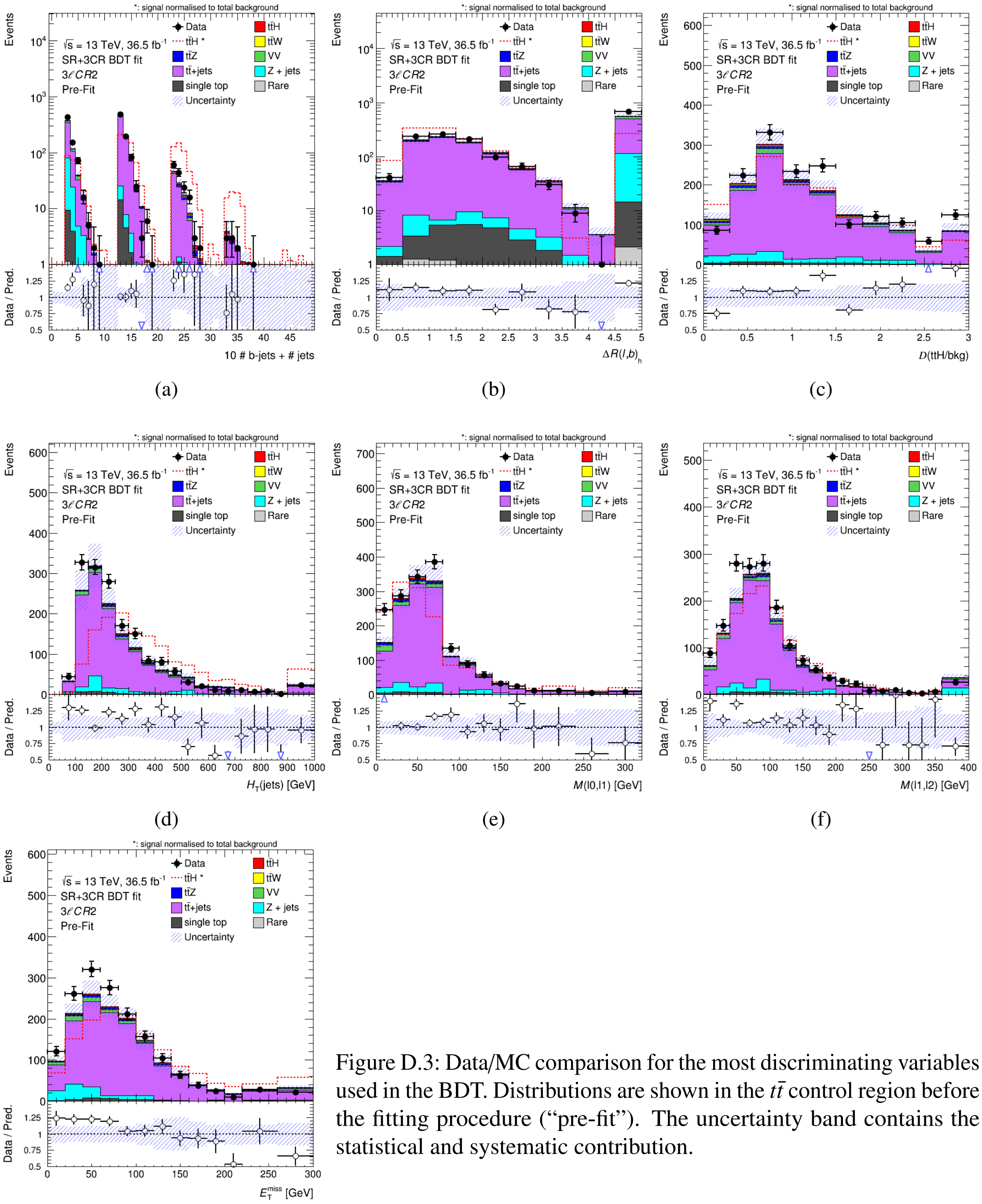


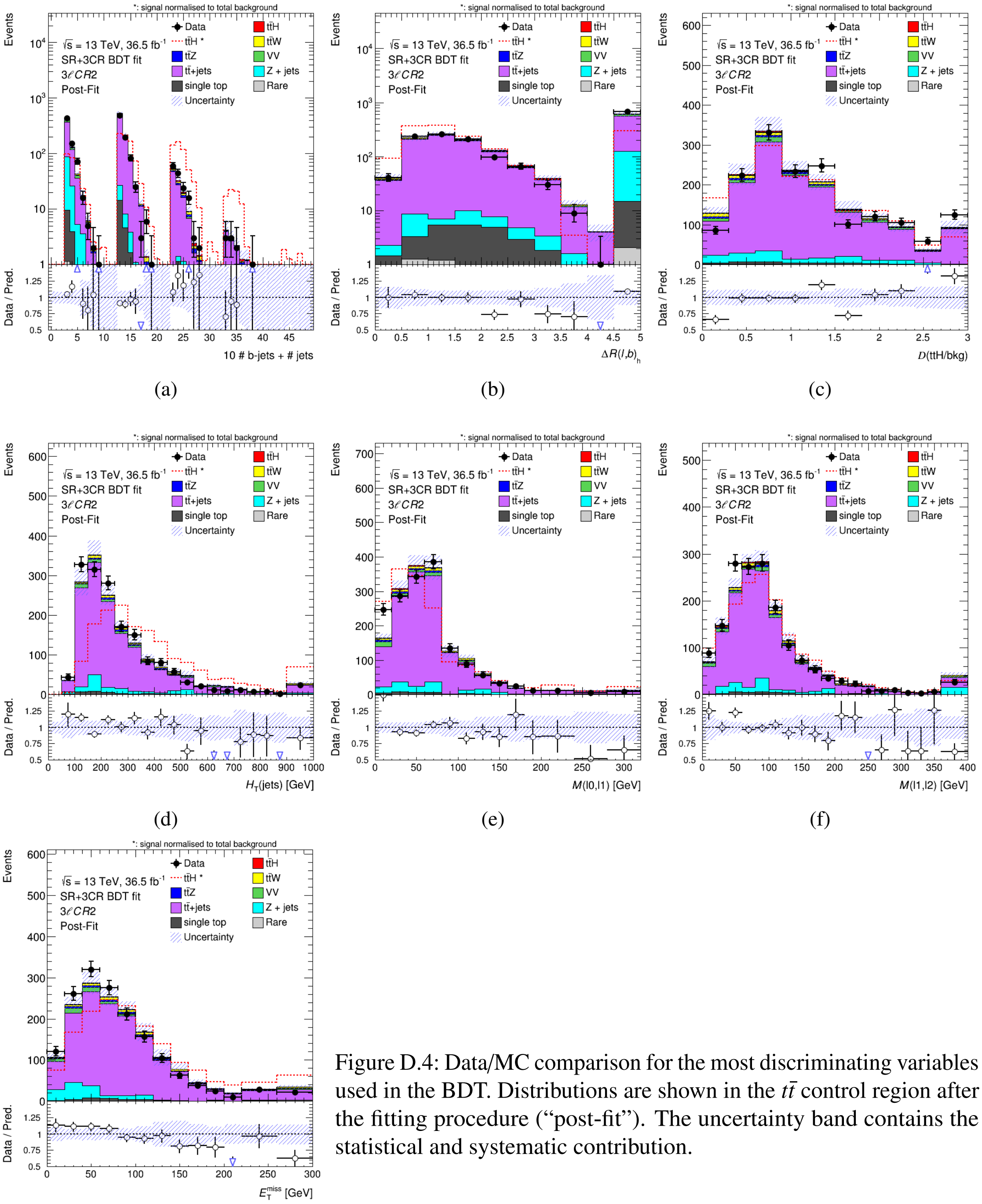
Figure D.2: Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the VV control region after the fitting procedure (“post-fit”). The uncertainty band contains the statistical and systematic contribution.



(a) (b) (c)

(d) (e) (f)

Figure D.3: Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the $t\bar{t}$ control region before the fitting procedure (“pre-fit”). The uncertainty band contains the statistical and systematic contribution.



(a)

(b)

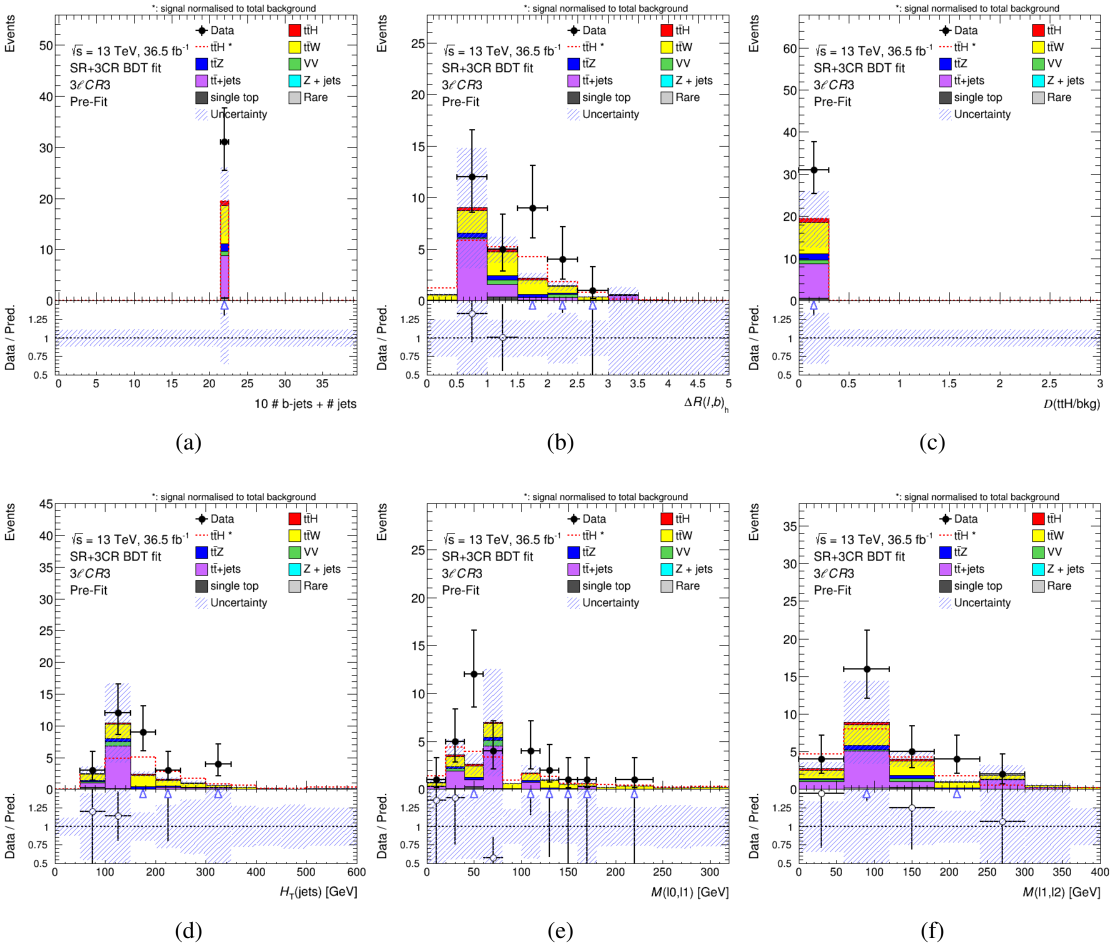
(c)

(d)

(e)

(f)

Figure D.4: Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the $t\bar{t}$ control region after the fitting procedure (“post-fit”). The uncertainty band contains the statistical and systematic contribution.



(a) (b) (c)

(d) (e) (f)

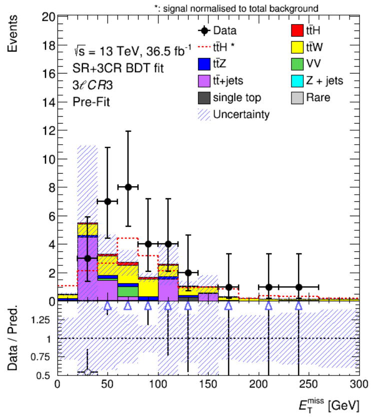
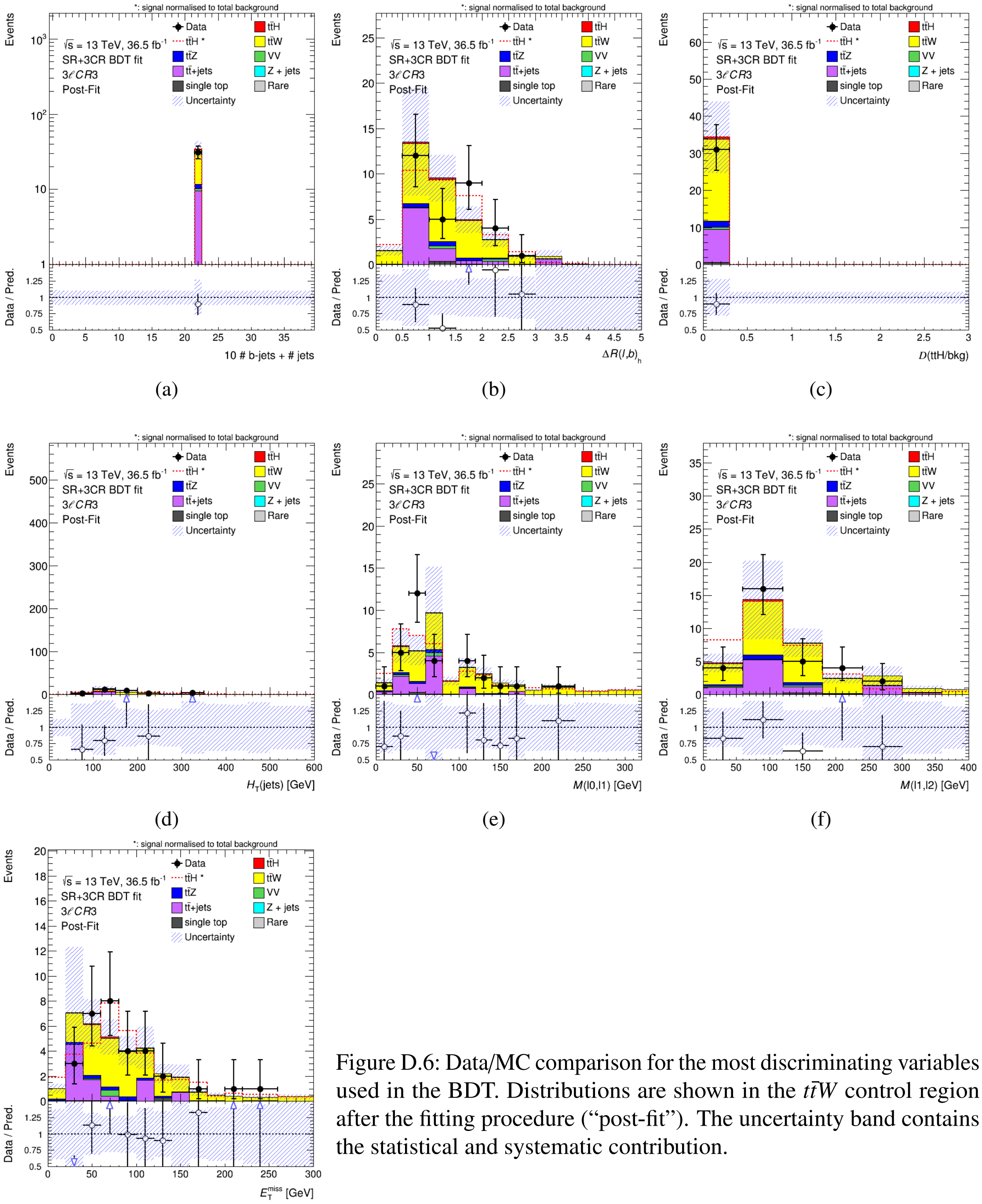


Figure D.5: Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the $t\bar{t}W$ control region before the fitting procedure (“pre-fit”). The uncertainty band contains the statistical and systematic contribution.



(a)

(b)

(c)

(d)

(e)

(f)

Figure D.6: Data/MC comparison for the most discriminating variables used in the BDT. Distributions are shown in the $t\bar{t}W$ control region after the fitting procedure (“post-fit”). The uncertainty band contains the statistical and systematic contribution.

Acknowledgements

After three challenging years, this thesis marks the end of an exciting PhD journey. Hence, the list of people to mention and thank is very long.

Firstly, I wish to express my sincere and deep gratitude to my advisor, Markus Cristinziani, for his careful guidance throughout these years. His enthusiasm, knowledge and constant support have been key ingredients for this achievement. Working with him has been a great learning opportunity for me, both as a scientist and as a person.

My sincere thanks also goes to Norbert Wermes for the many stimulating discussions and suggestions. I thank him for hosting our group and for the big opportunity I have had to work in the really nice and intellectually stimulating atmosphere of the ATLAS Bonn group.

I warmly thank Vadim Kostyukhin for his constructive criticism and assistance, for the long discussions and for his exceptional patience. His support has been essential for the writing of this thesis.

Finally, I would like to thank all (current and ex) PhD and graduate students I met during the last years in Bonn and at CERN: all of you taught me something, in one way or another. In particular, I want to thank my colleagues (in chronological order) Kaven Yau, Liza Mijović, David Hohn, Mazuza Ghneimat, Sebastian Heer, Andrea Sciandra, Carlo Gottardo, Julien Caudron and Alessandra Betti for the many useful discussions we had working together and for all the pleasant times of these years, talking about physics and more.

Publications and Proceedings

- ATLAS Collaboration, *"Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector"*, 38th International Conference on High Energy Physics (ICHEP 2016), ATLAS-CONF-2016-058.
- ATLAS Collaboration, *"Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, JHEP 05 (2016) 160, arXiv:1604.03812x.
- ATLAS Collaboration, *"Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector"*, ATL-PHYS-PUB-2014-013.
- ATLAS Collaboration, *"Measurements of the properties of the Higgs-like boson in the four lepton decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data"*, 48th Rencontres de Moriond on Electroweak Interactions and Unified Theories (Moriond 2013), ATLAS-CONF-2013-013.
- Proceeding for LHCP 2016, *"Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with hadronically decaying top quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*, edited by Proceeding of Science (PoS), PoS(LHCP2016)192.
- Proceeding for IFAE 2013, *"Measurement of the Higgs-like boson mass in the $ZZ^{(*)} \rightarrow 4\ell$ decay channel with the ATLAS detector"*, Incontri di Fisica delle Alte Energie - IFAE 2013 - Vol. 37, N. 1, 2014, pp. 272-274, edited by Societá Italiana di Fisica (SIF), 10.1393/ncc/i2014-11705-7.

Conferences, Workshops & Schools

- Talk at ATLAS-D Physics Meeting, held in Heidelberg, Germany, from October 4th to 7th, 2016, *"On the way to measure the top Yukawa coupling: search for the $t\bar{t}H$ process in multileptonic final states at 13 TeV in ATLAS"*, PDF.
- Poster at the 4th Annual Large Hadron Collider Physics Conference (LHCP2016), held in Lund, Sweden, from June 13th to 18th, 2016, *"Search for the $t\bar{t}H$ vertex via fully hadronic final state, based on "Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector""*, ATL-PHYS-SLIDE-2016-319.
- Talk at DPG 2016, Deutsche Physikalische Gesellschaft, in Hamburg, Germany, from February 29th to March 4th, 2016, *"Search for the production of the Higgs boson in association with a pair of top quarks in the 3 leptons final state at 13 TeV in ATLAS"*. Abstract.
- Talk at DPG 2015, Deutsche Physikalische Gesellschaft, in Wuppertal, Germany, from March 9th to 13th, 2015, *"Search for the decay $H \rightarrow b\bar{b}$ in association with a pair of hadronically decaying top quarks at 8 TeV in ATLAS"*. Abstract.
- Talk at SIF 2013, Societá Italiana di Fisica, in Trieste, Italy, from September 23rd to 27th, 2013, *"Hypothesis test on different Spin-CP states of the new resonance in the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ channel decay in ATLAS"*, PDF.
- Poster IFAE 2013, Incontri di Fisica delle Alte Energie, held in Cagliari, Italy, from April 3rd to 4th, 2013, organized by INFN and University of Cagliari. *"Measurement of the Higgs-like boson mass in the $ZZ^{(*)} \rightarrow 4\ell$ decay channel with the ATLAS detector"*.