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 8 TeV with the ATLAS detector at the LHC

Daniele Madaffari

▶ To cite this version:

Daniele Madaffari. 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 8 TeV with the ATLAS detector at the LHC. High Energy Physics - Experiment [hep-ex]. Aix Marseille Université, 2015. English. NNT : . tel-01326964

HAL Id: tel-01326964
https://hal.science/tel-01326964v1

Submitted on 15 Jun 2016

UNIVERSITÉ D'AIX-MARSEILLE
ECOLE DOCTORALE PHYSIQUE ET SCIENCES DE LA MATIÈRE

Discipline : Physique et Sciences de la Matière
Spécialité : Physique des Particules et Astroparticules

Centre de Physique des Particules de Marseille (CPPM) UMR 7346

Thèse présentée en vue d'obtenir le grade universitaire de docteur

Par

Daniele Madaffari

Search for the decay $H \rightarrow b\bar{b}$ 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

Soutenue le 18/12/2015 devant le jury composé de:

Freya Blekman	IIHE, Vrije Universiteit Brussel	Rapporteur
Bruno Mansoulié	IRFU-CEA Saclay	Rapporteur
Eric Kajfasz	CPPM, Marseille	Examinateur
Dominique Pallin	LPC, Clermont-Ferrand	Examinateur
Lorenzo Feligioni	CPPM, Marseille	Co-directeur de thèse
Mossadek Talby	CPPM, Marseille	Directeur de thèse

Contents

Introduction

The description of the elementary particles and their interactions is provided by the Standard Model (SM) [1, 2, 3] of particle physics. The SM was built sixty years ago, has been widely tested experimentally since then. Until 2012 only one of the predicted particles remained to be observed, the Brout-Englert-Higgs boson [4, 5, 6, 7], hereafter called Higgs boson. This particle underlies the spontaneous symmetry breaking of the electroweak interaction and is thus responsible for the acquisition of mass of the elementary particles.

A new particle, likely to be the Higgs boson, was discovered in 2012 by the ATLAS and CMS experiments [8, 9] through the analysis of pp collisions produced by the Large Hadron Collider (LHC) located at CERN.

The SM is a great success of modern physics, though it is widely known that this is not a complete theory. Indeed a series of experimental observations at the fundamental scale, such as the non-zero masses of the neutrinos, and at the cosmological scale, such as the existence of dark matter and dark energy, are not accounted for by the SM.

The study of the properties of this recently discovered boson is fundamental to understand its nature. In particular, the intensity of its Yukawa coupling to the top quark (Y_{top}) can provide hints on where the physics beyond the standard model can manifest itself [10]. Indirect constraints on Y_{top} come from the study of the processes where the interaction of the Higgs boson with other elementary particles is mediated by fermion loops, such as the gluon-gluon fusion production and the decay into a pair of photons. A direct measurement of Y_{top} allows to provide an independent constraint and gives insight on the presence of new particles inside such loops. This direct measurement is possible at LHC via the search of the associated production of a Higgs boson with a top-antitop quark pair ($t\bar{t}H$). Several attempts have been performed to measure $t\bar{t}H$ exploiting different decays of both Higgs boson and top quarks [11, 12, 13, 14, 15], but this observation has been out of reach so far.

In this doctoral thesis it is presented the first attempt ever performed to measure the cross section of $t\bar{t}H$ ($H \to b\bar{b}$) in the topology where the top and antitop quarks decay hadronically with the ATLAS detector. The experimental signature for this signal consists of eight hadronic jets, four of which are originated from b-quarks. The signal events are hidden by background events from $t\bar{t}$ + jets process, which has very similar signature with the signal and cross section ~ 100 times bigger. In particular the production of $t\bar{t} + b\bar{b}$ is an irreducible background. A great challenge is posed by the multijet production that has a cross section $\sim 10^{+6}$ times bigger. An ad-hoc data-driven technique, called $\mathrm{TRF}_{\mathrm{MJ}}$ method, has been developed to estimate this background.

The fully hadronic analysis uses 20.3 fb^{-1} of pp collisions collected at $\sqrt{s} = 8$ TeV with the ATLAS detector during Run 1 of the LHC. Data are collected using a multi-jet trigger. The search uses events with at least five energetic jets and uses a Boosted Decision Tree (BDT) to discriminate the signal from the otherwise overwhelming backgrounds. To increase the signal efficiency, dedicated trigger scale factors have been evaluated to use all signal events collected, even in kinematical regions where data and simulations behave differently.

After fitting the BDT output discriminant to the data, for a Higgs Boson mass of 125 GeV, a 95% CL upper limit of 6.5 times the SM cross section is observed, while the expected limit is 5.4. A $t\bar{t}H$ signal strength of 1.6 ± 2.6 times the SM value is obtained.

The document is concluded by a discussion about the future prospects for the fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$) analysis in Run 2 where the analysis benefits from increase in signal cross section, from the increase of center of mass energy. Moreover, larger integrated luminosity delivered by the LHC and upgrades in the ATLAS trigger system and at the detector level, by the addition of the Insertable B-Layer (IBL) [16], can further benefit the analysis.

Chapter 1

Theoretical framework

In the last decades a long series of discoveries have extended our knowledge of the properties and physics laws of elementary particles. Last of these is the discovery of a new particle [8, 9] in 2012, that that very much resembles the Brout-Englert-Higgs[1] boson predicted by the Standard Model (SM) of particle physics. To assess if this newly discovered particle is the one predicted by the SM, the precise measurement of its properties is necessary.

The first part of this chapter is intended to describe briefly the SM of particle physics and the role played by the Higgs boson.[2] The second part is devoted to a summary of the measurements in the Higgs sector that have been performed as of today and a comparison with SM theoretical expectations. Particular attention is devoted to the Yukawa coupling of the Higgs boson with the top quark and what a precision measurement of this parameter can unveil on the existence of new physics beyond the SM.

1.1 Standard model of particle physics

The SM precisely describes the nature of the elementary particles and their mutual interaction. It has been postulated in 1961 thanks to the pioneer work of S. L. Glashow, S. Weinberg and A. Salam [1, 2, 3] and has been developed and tested extensively since then.

According to the SM, the elementary particles can be classified in two categories based on their spin: *fermions* with spin $\frac{1}{2}$ and *bosons* with spin zero or one. A schema showing the elementary particles described in the SM and their main characteristics is shown in Figure 1.1.

1.1.1 Fermions

Fermions are the fundamental bricks of the known matter in our universe. Elementary fermions can be divided in two categories: *quarks* and *leptons*, as shown in

---

[1] Called hereafter Higgs boson.

[2] For a more complete and pedagogical description of the SM the reader is referred to [17, 18].

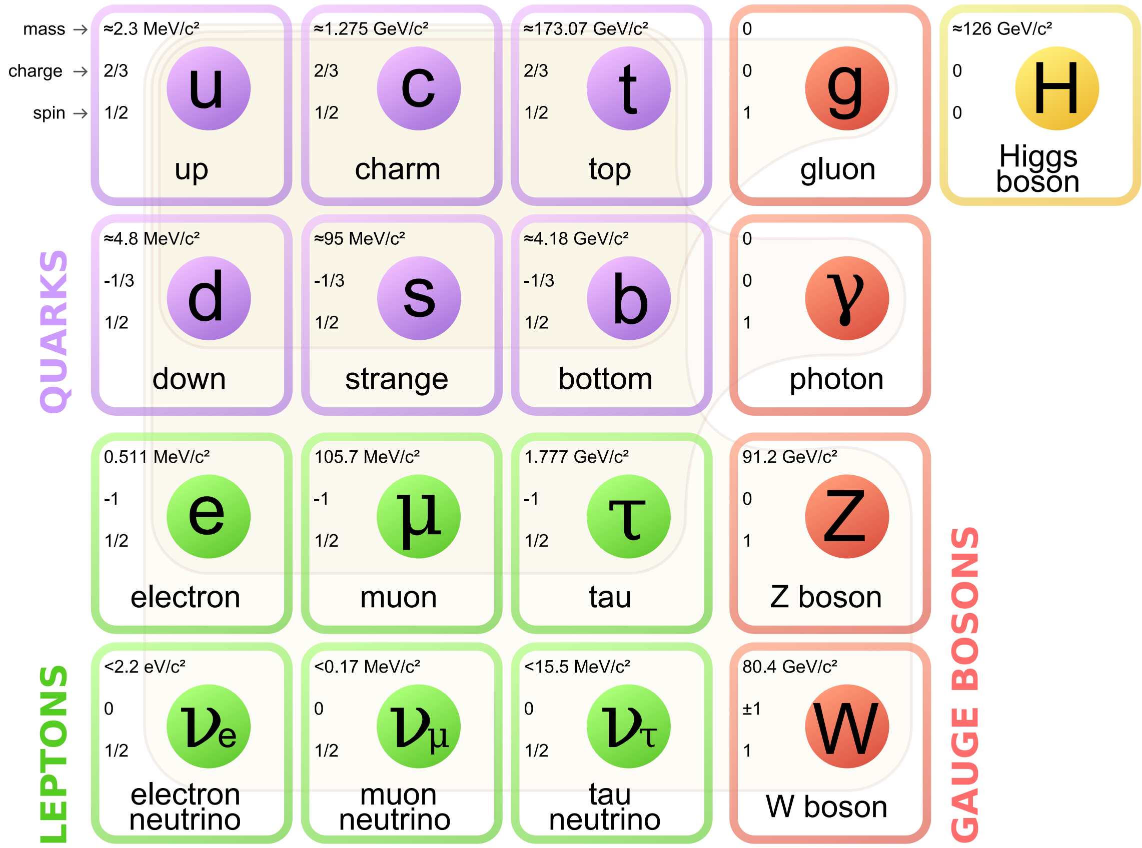


Figure 1.1: Elementary particles described in the Standard Model.

Figure 1.1. Leptons are subdivided in two categories: electrically charged leptons ($e^{\pm}, \mu^{\pm}, \tau^{\pm}$) sensitive to weak and electromagnetic interactions and electrically neutral leptons (ν_e, ν_μ, ν_τ) sensitive to weak interaction only. Quarks carry, in addition to a fractional electric charge, a color charge and are sensitive, beside the weak and electromagnetic interaction, to the strong interaction as well.

In the SM quarks and leptons are grouped into three weak isospin doublets or *generations*. For quarks each generation contains one up-type quark, with electric charge $+\frac{2}{3}e$, and one down-type quark, with electric charge $-\frac{1}{3}e$, with e it is meant the absolute value of the electric charge of the electron. The first generation contains the *up* (u) and *down* (d) quarks, the second generation contains the quarks *charm* (c) and *strange* (s), the third generation contains the quarks *top* (t) and *bottom* (b). Each quark can carry one of the three color charges: red, blue and green. Only colorless states[3] like hadrons (baryons and mesons) can be observed as free states.

For leptons, the first generation contains the electron (e^-) and the electron-neutrino (ν_e), the second generation contains the muon (μ^-) and the muon-neutrino (ν_μ) and the third generation contains the tau (τ^-) and the tau-neutrino (ν_τ).

1.1.2 Bosons

Five bosons are considered by the SM. Four of them have spin 1 and are also called “gauge bosons”. They are the mediators of the fundamental interactions between spin $\frac{1}{2}$ elementary particles and appear in the SM as the consequence of the invariance of the Lagrangian describing free fermions with respect to local gauge transformations.

The gauge bosons are:

- The photon (γ), the mediator of the electromagnetic interaction. It interacts with all the particles that have an electric charge different from zero. It has no electric charge and no mass. As a consequence of the latter the electromagnetic interaction is a long range interaction.

- The two $W^{\pm}$ bosons and the Z boson, which mediate the weak interaction. Both the W and the Z bosons have a relatively large mass. As a consequence, the weak interaction is a short range interaction. The $W^{\pm}$ and Z bosons interact not only with quarks and leptons but also among themselves.

- The gluon (g), which mediates the strong interaction. Gluons are mass-less and have no electric charge, carry color charge and can interact strongly not only with the quarks but also among themselves. Eight type of gluons exist with different color charge combinations.

The fifth boson described in the SM has spin 0 and it is known as the Higgs boson. It results from the spontaneous symmetry breaking of the electroweak symmetry ($SU(2)_L \otimes U(1)_Y \to U(1)_{\text{em}}$).

---

[3] For example objects with the combination of the three different colors or with the combination of the same color and anti-color.

1.1.3 Gauge symmetries and interactions

In the SM interactions between fundamental particles stem from the invariance of the Lagrangian of free fermions with respect to local gauge symmetries (transformations).[4]

A simple example: electromagnetism

The Lagrangian associated to the field that describes a free electron is:

$$\mathcal{L} = i\overline{\psi}\gamma_\mu \partial_\mu \psi - m\overline{\psi}\psi, \tag{1.1}$$

where each of the four component of the bi-spinors ψ and $\overline{\psi}$ are considered as independent complex fields. This Lagrangian is invariant under the phase transformation of the field:

$$\psi(x) \to e^{i\alpha}\psi \tag{1.2}$$

where α is a real constant. This transformation belongs to the group of unitary transformations $U(1)$. This type of symmetry is called global "gauge" [5] symmetry.

If one applies to the fermion field ψ a real phase (gauge) transformation, in which α is no longer a constant but a function depending on space-time coordinates:

$$\psi(x) \to e^{i\alpha(x)}\psi \tag{1.3}$$

the Lagrangian is not anymore invariant. This happens because the derivative term of Equation 1.1 is not invariant under this transformation:

$$\partial_\mu \psi \to e^{i\alpha(x)}\partial_\mu \psi + e^{i\alpha(x)}\psi \partial_\mu \alpha. \tag{1.4}$$

To recover the invariance of the Lagrangian under local gauge transformation one needs to replace in the Lagrangian the derivative four-vector ∂_μ by a new type of derivative, called "covariant derivative" D_μ, that transform like the field ψ under the local phase transformation.

$$D_\mu \psi \to e^{i\alpha(x)} D_\mu \psi \tag{1.5}$$

It is then necessary to introduce a vector field A_μ that transform in a way to cancel the symmetry violating term of Equation 1.4. With these requirements the covariant derivative is defined as:

$$D_\mu \psi \equiv \partial_\mu - ieA_\mu \tag{1.6}$$

with A_μ transforming as:

$$A_\mu \to A_\mu - \frac{1}{e}\partial_\mu \alpha(x) \tag{1.7}$$

---

[4] For a more complete description of how local gauge symmetries are related to fields' interaction, the reader is referred to [17, 19, 20].

[5] Historical name for "phase".

Introducing the covariant derivative in the Lagrangian defined in Equation 1.1 one obtains:

$$\begin{aligned}\mathcal{L} &= i\overline{\psi}\gamma^{\mu}D_{\mu}\psi - m\overline{\psi}\psi \\ &= \overline{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi + e\overline{\psi}\gamma^{\mu}A_{\mu}\psi \qquad (1.8)\end{aligned}$$

The result of requiring invariance of the Lagrangian under a local phase transformation is the introduction of a vector field A_{μ} (called gauge field). It has to be noted from Equation 1.8 that A_{μ} couples to the electron with a coupling constant $-e$. To allow this new field propagate it is necessary to introduce in the Lagrangian a kinetic term for it, and this term has to be invariant under the gauge transformation of Equation 1.7. In this way one obtains the Lagrangian :

$$\mathcal{L} = \overline{\psi}\left(i\gamma_{\mu}\partial_{\mu} - m\right)\psi + e\overline{\psi}\gamma^{\mu}A_{\mu}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \qquad (1.9)$$

Where:

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} \qquad (1.10)$$

is the electromagnetic tensor and $A_{\mu\,(\nu)}$ is the electromagnetic potential four-vector.

It is important to notice that a mass term of the type $\frac{1}{2}m^2A_{\mu}A^{\mu}$ is forbidden since it violates the gauge invariance. As a consequence A_{μ} is a massless gauge field and is identified to the photon.

Electroweak interaction

The electromagnetic and weak interactions are described in a unified way in the SM of electroweak (EW) interactions which describes the interactions among quarks and leptons mediated by four vector bosons, meaning γ, W^+, W^- and Z. This EW theory has been proposed for the first time in 1961 by S. Glashow [1].

To understand how EW theory is built one needs to start with a Lagrangian describing a weak isospin doublet ψ of free spinorial massless fields:

$$\mathcal{L} = i\overline{\psi}\gamma^{\mu}\partial_{\mu}\psi \qquad (1.11)$$

This Lagrangian is not invariant under local special unitary transformations of dimension 2 ($SU(2)$):

$$\psi \equiv \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \quad \rightarrow \quad \psi' \equiv \begin{pmatrix}\psi'_1\\ \psi'_2\end{pmatrix} = e^{i\vec{\theta}(x)\cdot\frac{\vec{\sigma}}{2}} \begin{pmatrix}\psi_1\\ \psi_2\end{pmatrix} \qquad (1.12)$$

Where (ψ_1, ψ_2) and (ψ'_1, ψ'_2) are the components of the weak isospin doublet ψ before and after the gauge transformation. $\vec{\sigma}$ is the vector of the Pauli matrices σ^1, σ^2 and σ^3, the generators of the $SU(2)$ transformation. $\vec{\theta}(x)$ is a three dimensional vector of arbitrary functions.

For the Lagrangian given in Equation 1.11 to be invariant under such a transformation it is necessary to replace the derivative four-vector ∂_{μ} by an appropriate covariant derivative, similar to the one defined in Equation 1.5.

$$\partial_{\mu} \rightarrow D_{\mu} = \partial_{\mu} + ig_wW^a_{\mu}\frac{\sigma^a}{2} \qquad (1.13)$$

Where g_W is a coupling constant and W^a_μ (with $a = 1, 2, 3$) is a set of three massless gauge fields which transform under $SU(2)$ as:

$$W^a_\mu(x) \to W'^a_\mu(x) = W^a_\mu(x) - \frac{1}{g_W}\partial_\mu\theta^a(x) - \epsilon^{abc}\theta^b W^c_\mu(x) \tag{1.14}$$

where ϵ^{abc} is the Levi-Civita tensor.

The Lagrangian, including this covariant derivative and the kinetic term for the three massless gauge bosons, becomes:

$$\mathcal{L} = i\overline{\psi}\gamma^\mu\partial_\mu\psi - g_w W^a_\mu\overline{\psi}\gamma^\mu\frac{\sigma^a}{2}\psi - \frac{1}{4}W^a_{\mu\nu}W^{a,\mu\nu} \tag{1.15}$$

where

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g_W\epsilon^{abc}W^b_\mu W^c_\nu \tag{1.16}$$

with $a = 1, 2, 3$, is the gauge field tensor. The second term of the Lagrangian describes the interaction of the spinorial field ψ with the gauge fields W^a_μ. The third term contains terms describing the self interaction between W^a_μ bosons.

As already anticipated fermions are grouped into weak isospin doublets ($I_w = \frac{1}{2}$) to account for the experimental evidence that all elementary fermions are sensitive to weak interactions and, due to parity-violation, that only their left-handed chiral components participate to this interaction:

$$f_L = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}_L \equiv \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L$$

with $I^3_w = +\frac{1}{2}$ for the f_{1L} component and $I^3_w = -\frac{1}{2}$ for the f_{2L} component. As the right-handed component of elementary fermions do not participate to charge current weak interaction, they are considered as weak isospin singlet states ($I_w = 0$):

$$f_R \equiv e_R, \quad \mu_R, \quad \tau_R, \quad u_R, \quad c_R, \quad t_R, \quad d_R, \quad s_R, \quad b_R \tag{1.17}$$

The theory built in this way do not include the fact that photons can interact with both left-handed and right-handed fermions. To incorporate the electromagnetic interaction one needs to require a new invariance of the Lagrangian with respect to another local gauge symmetry $U(1)_Y$:

$$\psi(x) \to \psi'(x) = e^{i\alpha(x)Y_w}\psi(x) \tag{1.18}$$

Where the operator Y_W is the generator of the hypercharge symmetry. This is the analogous of the hypercharge defined in the Gell-Mann-Nishijima formula:

$$Q = I^3_W + \frac{1}{2}Y_W \tag{1.19}$$

where Q is the electric charge and I^3_W is the third component of the weak isospin.

With this new local gauge transformation requirement the covariant derivative becomes:

$$\partial_\mu \to D_\mu = \partial_\mu + ig_w W_\mu^a \frac{\sigma^a}{2} + ig' \frac{Y_w}{2} B_\mu \tag{1.20}$$

where B_μ[6] is a new gauge field (similar to the A_μ field in QED) associated to the hypercharge Y_W and g' is a new coupling constant (equivalent of e in QED).

The gauge symmetry group associated with the two local gauge transformations:

$$\psi \to e^{i\vec{\theta}(x)\cdot\frac{\vec{\sigma}}{2}}\psi \qquad \text{and} \qquad \psi \to e^{i\alpha(x)Y_W}\psi \tag{1.21}$$

is:

$$SU(2)_L \otimes U(1)_Y \tag{1.22}$$

where $SU(2)_L$ is the weak isospin gauge symmetry group involving only left-handed fermion components and $U(1)_Y$ the weak hypercharge gauge symmetry group involving both the left-handed and right-handed components of the fermions.

The gauge fields associated to the generators of the two symmetry groups are:

$$\text{for} \quad SU(2)_L \ : \ W_\mu^1, \ \ W_\mu^2, \ \ W_\mu^3 \quad \text{and for} \quad U(1)_Y \ : \ B_\mu \tag{1.23}$$

These are massless gauge bosons since any mass term for these bosons in the Lagrangian spoils the invariance under local gauge transformations.

Strong interaction

Strong interactions between quarks and gluons are described by the Quantum Chromo Dynamics (QCD) theory. The gauge symmetry group that generates the interaction is $SU(3)$. The Lagrangian to be considered in this case is of the form:

$$\mathcal{L}_0 = \overline{q}_j \left(i\gamma^\mu \partial_\mu - m\right) q_j \tag{1.24}$$

where $q_j (j = 1, 2, 3)$ is one of the three color spinorial quark fields. The gauge local transformation for which the Lagrangian is required to be invariant is:

$$q(x) \to q'(x) \equiv e^{i\alpha_a(x)T_a} q(x) \tag{1.25}$$

where the transformation is shown only for one color field, since it is the same for all color fields. T_a are a set of eight linearly independent and traceless 3×3 matrices called Gell-Mann matrices, $\alpha_a(x)$ are eight scalar functions with $a = 1, \ldots, 8$. Following the same logic used for the derivation of Equation 1.15 it is possible to obtain the local gauge invariant QCD Lagrangian:

$$\mathcal{L} = \overline{q}\left(i\gamma^\mu \partial_\mu - m\right) q - g(\overline{q}\gamma^\mu T_a q) G_\mu^a - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \tag{1.26}$$

---

[6] The gauge field B_μ transforms under $U(1)_Y$ as:

$$B_\mu(x) \to B'_\mu(x) = B_\mu(x) - \frac{1}{g'}\partial_\mu \alpha(x)$$

with:

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f_{abc} G^b_\mu G^c_\nu \tag{1.27}$$

where G^a_μ, with $a = 1, \ldots, 8$, represents the eight gluon fields, f_{abc} are real constants called *structure constants of the group* and g is a coupling constant.

As for the photon, the gluons are required to be massless to preserve the gauge invariance of the Lagrangian. It is important to notice that the QCD Lagrangian contains terms describing a self-interaction among the gluons, with vertices with three and four gluons. This is consequence of the fact that the gluons carry color charge.

1.1.4 Spontaneous symmetry breaking

This formulation of the electroweak interaction is not yet complete, since it is not able to include the masses of the fermions and of the observed $W^\pm$ and Z bosons. To account for it, it is necessary to include in the EW theory a mechanism that generates $W^\pm$ and Z bosons masses. This mechanism is called Spontaneous Symmetry Breaking (SSB) mechanism.

SSB takes place when a physical system is described by a Lagrangian that is invariant under certain transformation while its fundamental state is not. A typical example of this phenomenon can be found in ferromagnetism [21]. From the quantum-mechanical point of view a ferromagnet is a crystalline lattice of spin 1/2 magnetic dipoles, with spin-spin forces that tends to align the neighboring spins. The Hamiltonian of such system is invariant with respect to space rotation. The ground state of this system has all the spins aligned in one direction, meaning that the symmetry of the Hamiltonian is broken and only a symmetry with respect to rotation around the axis of the spins is preserved. It is important to notice that the symmetry is not actually broken but just "hidden" by the ground state.

To understand how SSB can work into the SM and to search for such "hidden symmetries", it is better to start from a simple example. Let's consider a complex scalar field $\phi(x) = (\phi_1 + i\phi_2)/\sqrt{2}$ associated to a potential V:

$$V(\phi, \phi^\dagger) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \tag{1.28}$$

where μ^2 and λ are free parameters of the potential. The Lagrangian describing the system is:

$$\mathcal{L} = \partial^\mu \phi^\dagger \partial_\mu \phi - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 \tag{1.29}$$

The Lagrangian $\mathcal{L}$ is invariant under the $U(1)$ global symmetry group transformation:

$$\phi(x) \to \phi'(x) = e^{i\alpha} \phi(x) \tag{1.30}$$

where α is a constant. To have the potential $V(\phi, \phi^\dagger)$ bounded from the bottom it is necessary that $\lambda > 0$. The configuration of the minimum of $V(\phi, \phi^\dagger)$ is determined by the sign of the μ^2 parameter resulting into two possible solutions for the fundamental state of the scalar field $\phi(x)$. For $\mu^2 > 0$ the potential has a parabolic shape with a trivial minimum $\phi_{\text{min}} = 0$. When $\mu^2 < 0$ the potential has a “Mexican hat” shape with a circle of minima in the (ϕ_1, ϕ_2) plane, as can be seen in Figure 1.2. The equation of the circle of minima is:

$$\phi_1^2 + \phi_2^2 = v^2 \quad \text{where} \quad v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{1.31}$$

Choosing a particular solution among the infinite number of possible minima brings to a SSB.

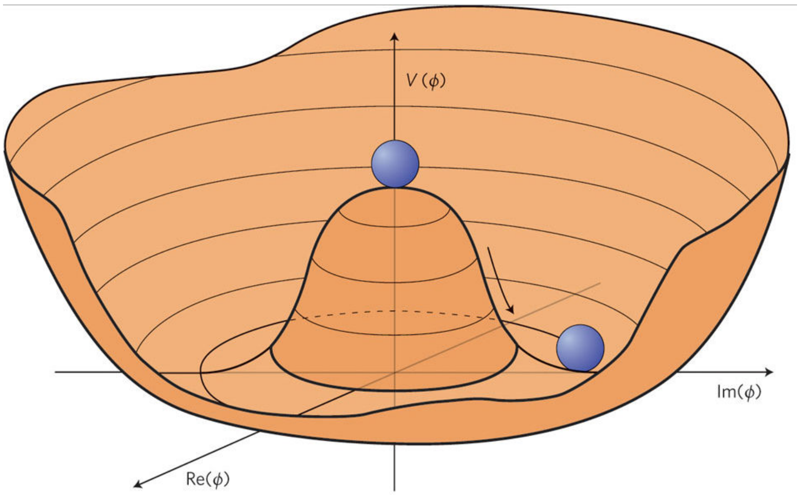


Figure 1.2: Shape of the scalar potential for $\mu^2 < 0$ and $\lambda > 0$. The blue sphere represents the minimum of the potential $V(\phi)$ of the scalar field ϕ.

Choosing the minimum that is real and positive leads to a vacuum expectation value (vev) of the field $\phi(x)$:

$$\langle 0|\phi(x)|0\rangle = \frac{v}{\sqrt{2}} \tag{1.32}$$

This corresponds to the magnitude of the field $\phi(x)$ at the minimum of the potential. It is very easy to see that this vev is not invariant under the phase transformation defined in Equation 1.30:

$$\langle 0|\phi'(x)|0\rangle = e^{i\alpha}\langle 0|\phi(x)|0\rangle = e^{i\alpha}\frac{v}{\sqrt{2}} \neq \frac{v}{\sqrt{2}} \tag{1.33}$$

Goldstone bosons

In order to have a deeper understanding of the effect of the SSB on the Lagrangian 1.29, the scalar field is rewritten as a fluctuation around the minimum of the potential:

$$\phi(x) = \frac{1}{\sqrt{2}}\left[v + H(x)\right] e^{i\frac{G(x)}{v}} \tag{1.34}$$

where $H(x)$ and $G(x)$ are two real scalar fields.

With this parametrization the kinetic term of the Lagrangian becomes:

$$\mathcal{L}_{\text{cin}} = \partial^\mu \phi^\dagger \partial_\mu \phi = \frac{1}{2}\partial^\mu H \partial_\mu H + \frac{1}{2}\left(1+\frac{H}{v}\right)^2 \partial^\mu G \partial_\mu G \tag{1.35}$$

and the potential $V(\phi, \phi^\dagger)$ is given by:

$$\begin{aligned} V(\phi,\phi^\dagger) &= \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger\phi)^2 \\ &= \lambda\left(\phi^\dagger\phi - \frac{v^2}{2}\right)^2 - \frac{\lambda}{4}v^4 \\ &= \lambda v^2 H^2 + \lambda v H^3 + \frac{\lambda}{4}H^4 - \frac{\lambda}{4}v^4 \end{aligned} \tag{1.36}$$

The total Lagrangian is:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{\text{cin}} - V(\phi,\phi^\dagger) \\ &= \frac{1}{2}\partial^\mu H \partial_\mu H - \frac{1}{2}m_H^2 H^2 + \frac{1}{2}\partial^\mu G \partial_\mu G \\ &+ \left(\frac{1}{v}H + \frac{1}{2v^2}H^2\right)\partial^\mu G \partial_\mu G - \lambda v H^3 - \frac{\lambda}{4}H^4 + \frac{\lambda}{4}v^4 \end{aligned} \tag{1.37}$$

As we can see, the two real scalar fields H and G, used to parametrize the complex scalar field ϕ, have different masses. One, $H(x)$, has a mass equal to $M_H = v\sqrt{2\lambda} = \sqrt{-2\mu^2}$, and the other, $G(x)$, is massless. In summary, SSB results in the presence of a massless scalar particle, known as Nambu-Goldstone boson. The presence of such massless particle is due to the fact that along the angular direction the potential is flat and so there is no “resistance” to excitations. The existence of such a massless scalar field is a general result stated in the Goldstone theorem [22].

1.1.5 The Brout-Englert-Higgs (BEH) mechanism

The application of the SSB to the electroweak interaction has been introduced in 1964 by several theoreticians: P. Higgs, F. Englert, R. Brout, G. S. Guralnik, C. R. Hegen, and T. W. B. Kibble [4, 5, 6, 7]. The mechanism predicts the existence of a massive scalar particle called the Higgs boson which infers masses to the weak gauge bosons.

The BEH mechanism assumes that anywhere in the vacuum there is a complex scalar field, and that the Lagrangian of this field is invariant under $SU(2)_L \otimes U(1)_Y$ local gauge symmetry. To fulfill this requirement the scalar field has to be a doublet composed of complex scalar fields[7]:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1^+ + i\phi_2^+ \\ \phi_1^0 + i\phi_2^0 \end{pmatrix} \tag{1.38}$$

---

[7] The upper component ($I_W^3 = +1/2$) has a positive electric charge and the lower component ($I_W^3 = -1/2$) is neutral.

The most general scalar field Lagrangian that can be written in this case is:

$$\mathcal{L}_{\text{Higgs}} = D^{\mu}\Phi^{\dagger}D_{\mu}\Phi - \mu^2\Phi^{\dagger}\Phi - \lambda(\Phi^{\dagger}\Phi)^2 \tag{1.39}$$

where μ^2 and λ are two free real parameters and D_μ is the covariant derivative that ensures the invariance of the kinetic term under the local gauge transformation of the symmetry group $SU(2)_L \otimes U(1)_Y$. As seen in Section 1.1.4, in the case where $\mu^2 < 0$ and $\lambda > 0$ the $SU(2)_L \otimes U(1)_Y$ symmetry group is spontaneously broken. The minimum of the potential $V(\Phi, \Phi^\dagger)$ occurs for:

$$\Phi^{\dagger}\Phi = \frac{v^2}{2}, \quad \text{with} \quad v = \sqrt{-\frac{\mu^2}{\lambda}}$$

A schematic representation of the scalar potential can be seen in Figure 1.2

There is freedom in the selection of the minimum of the scalar field potential. With any choice the SSB leads to a vacuum state that is not invariant under the $SU(2)_L \otimes U(1)_Y$ symmetry group. An appropriate choice that allows to keep the photon massless is to choose the minimum of the Higgs potential along the neutral component. In this case the fundamental state of the scalar field doublet is written:

$$|\langle 0|\Phi|0\rangle| \equiv \Phi_{\text{min}} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \tag{1.40}$$

One can rewrite the scalar field as a fluctuation around the minimum of the potential.

$$\Phi(x) = e^{i\frac{G^a(x)\sigma^a}{2v}} \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix} \tag{1.41}$$

where $G^a(x)$, $(a = 1, 2, 3)$ are the three becoming Nambu-Goldstone bosons corresponding to the three broken generators of $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{\text{em}}$ and $H(x)$ is the Higgs boson. Since the Lagrangian is invariant under gauge transformation $\Phi(x) \rightarrow e^{i\theta^a(x)}\Phi(x)$, where $\theta^a(x)$ are arbitrary functions of space-time coordinates, it is possible to choose $\theta^a(x) = -G^a(x)/v$ $(a = 1, 2, 3)$. This is called unitary gauge. With this choice the three Goldstone bosons are eliminated and the scalar field $\Phi(x)$ becomes:

$$\Phi(x) = \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix} \tag{1.42}$$

Applying this unitary gauge to the Lagrangian 1.39 leads to[8]:

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} &= \frac{1}{2}\partial^\mu H \partial_\mu H + \frac{1}{2}\frac{g_w^2 v^2}{4\cos^2\theta_W} Z^\mu Z_\mu + \frac{g_w^2 v^2}{4} W^{+\mu} W^-_\mu \\ &+ \left(\frac{2H}{v} + \frac{H^2}{v^2}\right)\left(\frac{1}{2}\frac{g_w^2 v^2}{4\cos^2\theta_W} Z^\mu Z_\mu + \frac{g_w^2 v^2}{4} W^{+\mu} W^-_\mu\right) \\ &-\lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4 + \frac{\lambda}{4} v^4 \end{aligned} \tag{1.43}$$

The quadratic terms :

$$\frac{1}{2}\frac{g_w^2 v^2}{4\cos^2\theta_W} Z^\mu Z_\mu + \frac{g_w^2 v^2}{4} W^{+\mu} W^-_\mu \equiv \frac{1}{2} M_Z^2 Z^\mu Z_\mu + M_W^2 W^{+\mu} W^-_\mu \tag{1.44}$$

correspond to the $W^\pm$ and Z^0 bosons mass terms with:

$$M_W = \frac{g_w v}{2} \quad \text{and} \quad M_Z = \frac{g_w v}{2\cos\theta_W} \quad \Rightarrow \quad \frac{M_W}{M_Z} = \cos\theta_W \tag{1.45}$$

where g_w is the electroweak coupling constant and θ_W is the electroweak mixing angle. The ratio of the $W^\pm$ and Z^0 boson masses is usually expressed in terms of a parameter ρ, given at the tree level by:

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2\theta_w} = 1 \tag{1.46}$$

The Higgs boson mass term is given by:

$$M_H = \sqrt{-2\mu^2} = v\sqrt{2\lambda} \tag{1.47}$$

The Lagrangian $\mathcal{L}_{\text{Higgs}}$ contains also coupling terms between the Higgs boson and the gauge bosons $W^\pm_\mu$ and Z_μ, proportional to the square of their masses, and Higgs boson self coupling terms. These couplings are of the type VVH or $VVHH$ with $VV \equiv W^+W^-$ or Z^0Z^0 and self-coupling terms of types HHH and $HHHH$. There is no direct coupling between the Higgs boson and the photon.

BEH mechanism corresponds to a sort of redistribution of the scalar field $\Phi(x)$ components during the SSB. Indeed three of the four scalar field components were transformed after the SSB into the longitudinal components of the $W^\pm$ and Z^0 bosons and consequently generated their mass terms.

---

[8] The gauge fields $W^\pm$ are, in terms of W^1_μ and W^2_μ gauge fields, given by:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp i W^2_\mu\right)$$

The gauge field Z_μ is a linear superposition of the gauge field W^3_μ and B_μ:

$$Z_\mu = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W$$

where θ_W is the electroweak mixing angle called also Weinberg angle.

Yukawa interaction and fermion masses

The existence of the scalar field $\Phi(x)$ allows the presence in the Lagrangian of mass terms for fermions. This is due to the Yukawa type interaction of fermions with the scalar field $\Phi(x)$. The Lagrangian of this interaction is:

$$\mathcal{L}^{f}_{\text{Yukawa}} = -C_f \left[\overline{f}_L \Phi \, f_R + \overline{f}_R \, \Phi^\dagger f_L\right] \tag{1.48}$$

where C_f is the coupling constant of the fermions to the scalar field, called Yukawa coupling constant. Since Yukawa interaction connects left- and right-handed components of the fermions and neutrinos are only left-handed, these do not interact with the scalar field.

Leptons and quarks interacts equivalently with the scalar field, by a coupling constant that is directly proportional to the mass m_f of each particle:

$$C_f = \frac{m_f\sqrt{2}}{v} \quad ; \text{ with } \quad v = \frac{2M_W}{g_W} \tag{1.49}$$

1.1.6 Limits of the Standard Model

Despite the large amount of successful predictions the SM presents several limits, in the sense that is not capable to account for several phenomena that are observed in nature.

- SM does not include gravitational interaction.
- There is no explanation of why there are exactly three generations of fundamental particles and why these particles have the hierarchy of masses we observe.
- The Higgs field and SSB allow for the presence of a mass term in the theory, but the value of these masses are free parameters measured by experiments. The existence of many free parameters is seen as a limitation of the theory.
- Neutrinos are considered in the SM [1, 2, 3], as massless elementary fermions. However there is an experimental evidence[9] that these particles have non-zero mass [23].
- Indications from theoretical consideration exist that an extrapolation of the SM to energy scale above $\sim$ 500 GeV suffers from the so called *naturalness* problem. An example of this is coming from quantum corrections to effective-field theory calculations of the Higgs boson mass. The latter suffer of a divergence function of the virtual particles' momenta cut off scale Λ [24, 25, 26, 27, 28]:

$$\frac{\delta m_H^2}{m_H^2} = \frac{3G_F}{4\sqrt{2}\pi^2}\left(\frac{4m_t^2}{m_H^2} - \frac{2m_W^2}{m_H^2} - \frac{m_Z^2}{m_H^2} - 1\right)\Lambda^2 = \left(\frac{\Lambda}{500 \text{ GeV}}\right). \tag{1.50}$$

---

[9] Neutrino oscillations.

Higher is the value of Λ, more relevant is the divergence of ultraviolet contributions to m_H^2, which can even exceed the value of m_H. Naturalness of the Higgs boson mass leads to the so called Hierachy problem (difference between a physical parameter and its effective value like in the case of Higgs mass). Presence of new physics beyond the SM can allow for a solution to this issue.

- Cosmological observations has brought evidence of the presence in the universe of new type of matter and energy, called dark matter and dark energy, which are not accounted for in the SM.
- The asymmetry observed in the universe between matter and anti-matter is not explained by the SM.

These are only some of the motivations behind the search for physics beyond the SM and the study of the newly discovered particle likely to be the SM Higgs boson is one of the fronts where this search takes place.

1.2 Higgs physics

This section summarizes some among the most recent results.

1.2.1 Production and decay, theoretical predictions

Theoretical predictions of cross sections of different SM Higgs boson production processes at the LHC have been computed at different orders [29, 30]. The most relevant Higgs production processes and their respective cross sections for this thesis are summarised in Table 1.1 for a center of mass energy $\sqrt{s} = 8$ TeV and Higgs boson mass $M_H = 125.1$ GeV. Figure 1.3(a) shows the cross section of several SM Higgs boson production processes as function of M_H.

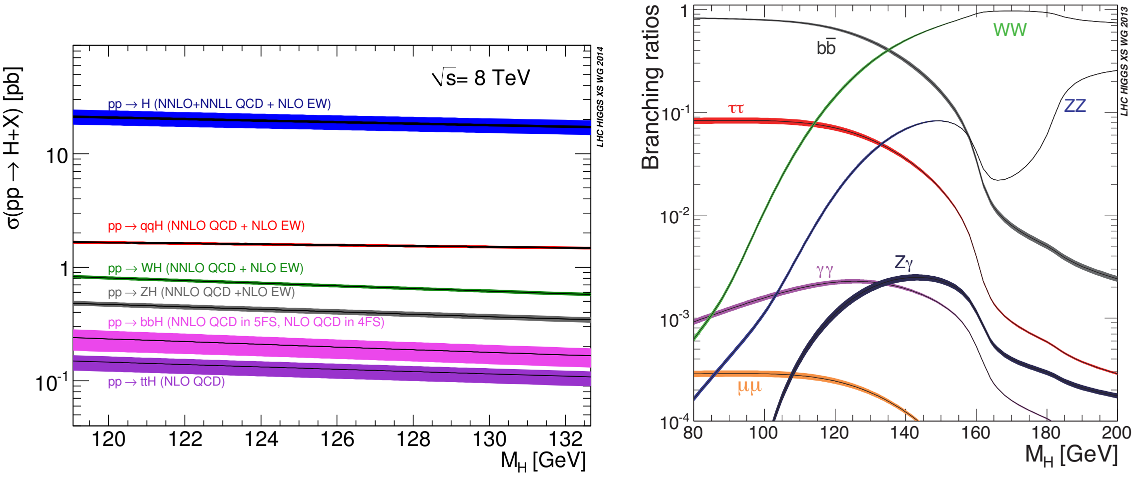
The dominant process is the gluon-gluon fusion. It is known at next-to-next-to-leading order (NNLO) in QCD, with resummation of soft-gluon contributions to next-to-next-to-leading logarithmic accuracy (NNLL). The estimated cross section is corrected for two-loop electro-weak (EW) contributions. The sub-dominant production process is the vector boson fusion (VBF). It is known at NNLO in QCD with NLO EW corrections. Associated production processes have smaller crossections with respect to the latter. Higgs boson produced together with a W boson and a Z boson have been derived at the NNLO accuracy in QCD and NLO EW corrections. Associated production of the Higgs boson with a $b\bar{b}$ pair has been computed with two approches. With the five flavour scheme (5FS), that take into account the presence in the proton of u-, d-, c- ,s- and b-quarks, or neglecting the presence of b-quarks, the four flavour scheme (4FS). The 4FS approach is designed to describe the case where $M_H/m_b \rightarrow 1$, while the 5FS approach is designed to describe the case $M_H/m_b \rightarrow \infty$. When performing the calculation the two methods are combined taking into account their different behaviours with respect to M_H. Associated production of the Higgs boson with a $t\bar{t}$ pair has been computed at NLO QCD accuracy.

More details about the evaluation of $t\bar{t}H$ cross section and the related uncertainties are given in section 5.1.

Process	Accuracy	σ [pb]	QCD Scale [%]	PDF+α_s [%]	References
$gg \to H$	NNLO + NNLL QCD NLO EW	19.24	+7.2 −7.8	+7.5 −6.9	[31, 32, 33] [34, 35, 36, 37]
$q\bar{q}H$ (VBF)	NNLO QCD + NLO EW	1.579	±0.2	+2.6 −2.8	[38, 39]
WH	NNLO QCD + NLO EW	0.7027	±1.0	±2.3	[40, 41]
ZH	NNLO QCD + NLO EW	0.4142	±3.1	±2.5	[40, 41]
$b\bar{b}H$	NLO (4FS), NNLO (5FS)	0.2030	+10.3 −14.8	±6.2	[42, 43, 44]
$t\bar{t}H$	NLO QCD	0.1290	+3.9 −9.3	±8.1	[45, 46, 47, 48, 49]

Table 1.1: Cross section of different Higgs boson production processes at the LHC for pp collisions with $\sqrt{s} = 8$ TeV. The calculation are performed assuming a Higgs boson mass $M_H = 125.1$ GeV. The accuracy of the calculations is given for each process in the second column. The uncertainties on scale definition, PDFs choice and α_s are given in the fourth and fifth columns respectively. References are given in the last column.

Calculations have also been made to predict the Higgs boson decay Branching Ratios (BR) [30]. Figure 1.3 shows the total production cross section and decay branching ratios of the SM Higgs boson as function of its mass. For a Higgs boson mass of 125 GeV the biggest BR is to a $b\bar{b}$ pair, followed by WW, $\tau\tau$ and ZZ pairs. Non negligible are also the decay to a pair of photons or to a Z boson and a photons and to a pair of muons.



(a) (b)

Figure 1.3: Standard Model Higgs boson production cross sections at $\sqrt{s} = 8$ TeV (a) and decay branching ratios (b) as function of its mass (M_H).

1.2.2 Mass and natural width

Figure 1.3(b) shows how the Higgs boson BRs are correlated with its mass. Hence a precise measurement of it mass is fundamental to have accurate prediction to compare with experimental results. The most precise measurement of the Higgs boson mass has been performed with a joint effort between the ATLAS and CMS collaborations [50]. For this combination the $\sqrt{s} = 7$ and 8 TeV datasets collected by the two experiments are used. Corresponding, for CMS to 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and 19.5 fb^{-1} at $\sqrt{s} = 8$ TeV; for ATLAS to 4.5 fb^{-1} at $\sqrt{s} = 7$ TeV and 20.3 fb^{-1} at $\sqrt{s} = 8$ TeV. The result of the analysis of is:

$$M_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.})\ \text{GeV}. \tag{1.51}$$

Figures 1.4 and 1.5 show the contribution of the different measurements that are entering this combination.

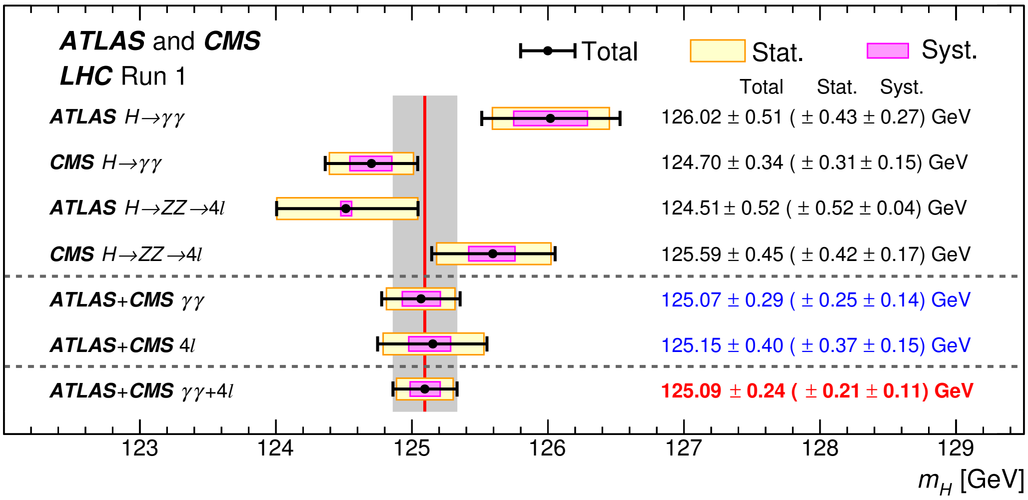


Figure 1.4: Summary of the Higgs boson mass measurements performed by the ATLAS and CMS experiments [50]. The combination of the different measurements is shown at the bottom of the plot.

In a separate publication by CMS collaboration [51] a measurement is performed on the width of the Higgs boson resonance. The width of a SM Higgs boson of mass $M_H \sim 125$ GeV is predicted to be ~ 4 MeV. This value is far smaller than the experimental mass resolution. Nevertheless it is possible to measure the observed width searching for deviation from the expected detector resolution. The decay channels used for this measurement are $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$[10], these have experimental mass resolution between 1 and 3 GeV. The result of the measurement is an observed limit at 95% CL of 1.7 GeV, with an expected limit of 2.3 GeV, summarized in Figure 1.6.

---

[10] Here and in the following ℓ refers to electrons and muons.

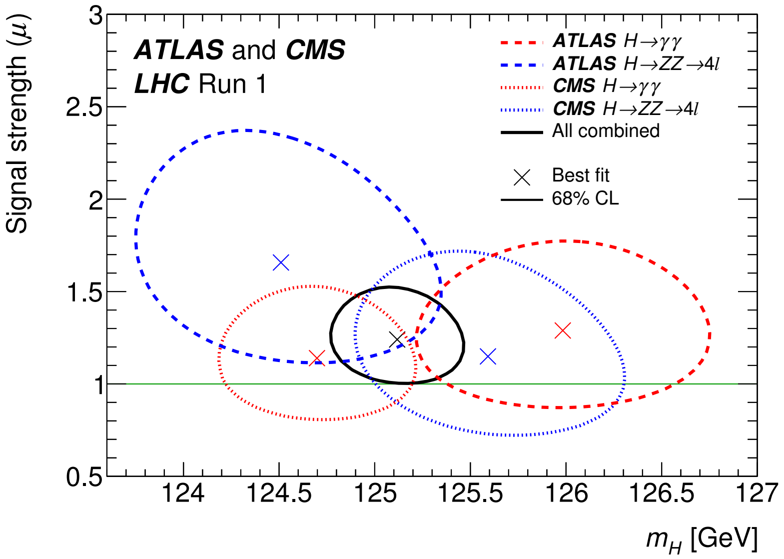


Figure 1.5: Likelihood scan of signal strength (μ) versus Higgs boson mass (m_H) The dashed and solid lines refers to the 68% CL confidence regions [50].

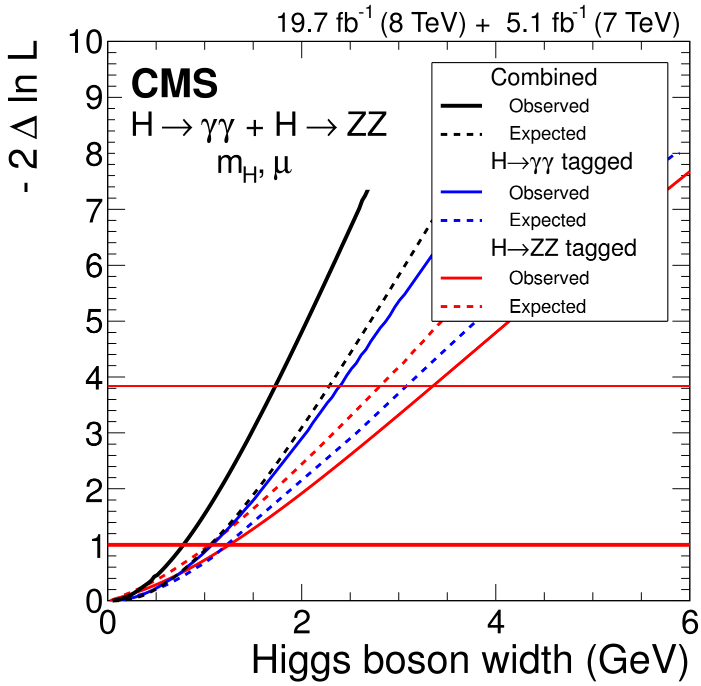


Figure 1.6: Likelihood scan of the width of the Higgs boson resonance. Observation is consistent with the SM expectation [51].

1.2.3 Electric charge

The predicted electric charge of the SM Higgs boson is zero. It is consistent with the observation of the Higgs boson candidate in final state with zero electric charge, like $\gamma\gamma$ and ZZ.

1.2.4 Spin and parity

The SM predicts the Higgs boson to have spin equal to zero and its parity to be even. these properties have been measured by the ATLAS and CMS collaborations [52, 53, 54].

CMS collaboration uses the decay channels: $H \to ZZ^*$, $Z\gamma^*$, $\gamma^*\gamma^* \to 4\ell$, $H \to WW^* \to \ell\nu\ell\nu$ and $H \to \gamma\gamma$. The results obtained are summarized in Figure 1.7 and are consistent with SM expectations.

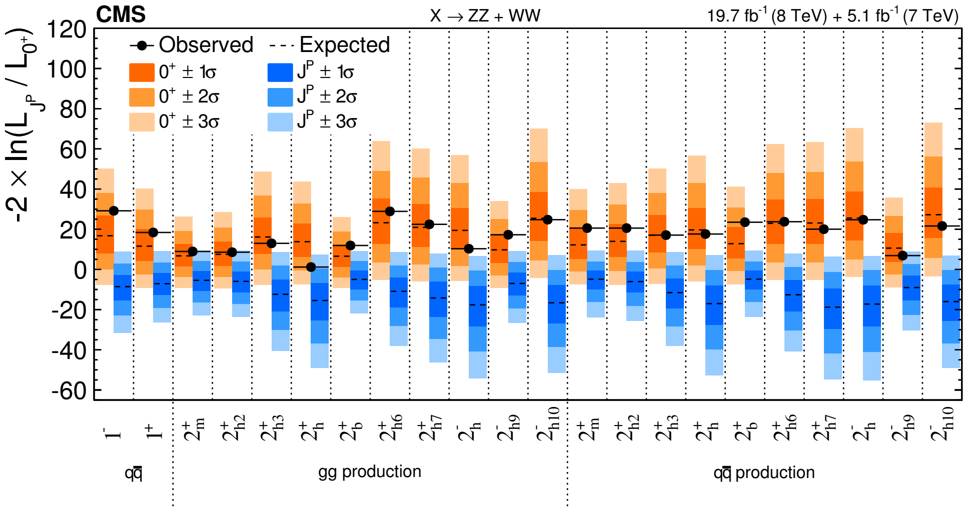


Figure 1.7: Distributions of the test statistic $q = -2\ln(\mathcal{L}_{J^P}/\mathcal{L}_{0^+})$ for the different models tested against the SM Higgs boson hypothesis. The expected median and 1,2,3 σ CL regions for the tested J^P models (blue bands) and for the SM Higgs boson hypothesis (red bands) are shown. The observed q values are indicated by the black dots [52].

ATLAS uses $H \to ZZ \to 4\ell$, $H \to WW^* \to e\nu\mu\nu$ and $H \to \gamma\gamma$ to exclude the non-SM spin hypothesis at more than 99.9% CL in favor of the SM spin-0 hypothesis.

1.2.5 Scalar couplings

To be able to measure the couplings of the Higgs boson with the other elementary particles of the SM it is necessary to perform a wide spectrum of measurement of different Higgs production processes and decays. The ATLAS and CMS collaboration have combined several analyses to extract this information [55]. The combination makes use of analyses optimized for to search for the Higgs boson decaying to $\gamma\gamma$, $ZZ(4\ell)$, $WW(\ell\nu\ell\nu)$, $\tau\tau$, $b\bar{b}$ and $\mu\mu$. The production modes studied are $t\bar{t}H$, VH, VBF and ggH. The summary of measurements of the cross-sections of these processes and of the branching ratios of the decay modes used to extract the coupling of the Higgs boson is shown in Figure 1.8. The events are categorized in different channels to maximize the sensitivity to a particular decay or production mode[11].

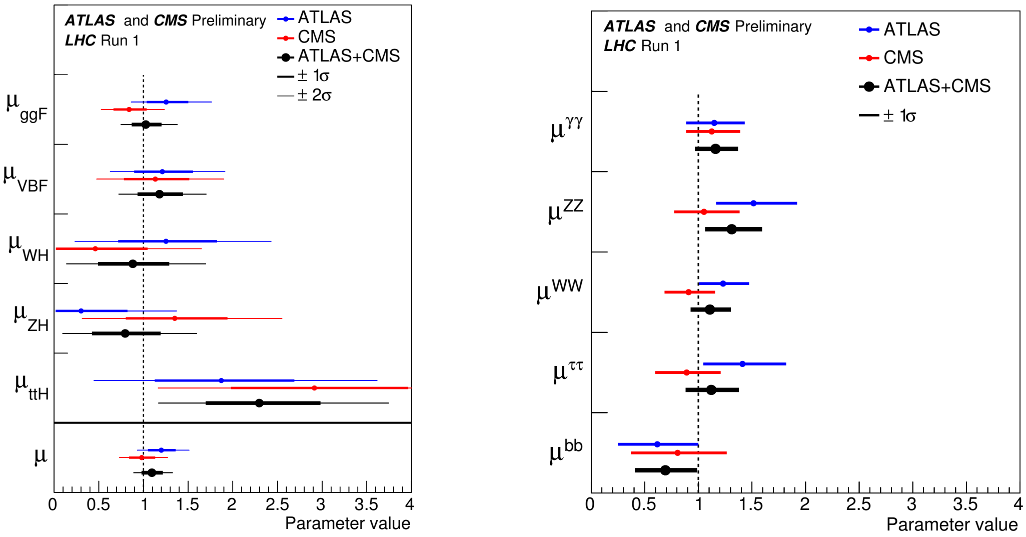


Figure 1.8: Values of the best-fit signal strength for the analyses entering in the combination. The blue and the red points indicate the values obtained by ATLAS and CMS collaboration respectively [55]. Black points represent their combination. Analyses are grouped on the left by production mode ($\mu = \sigma/\sigma_{SM}$) and on the right by decay mode ($\mu = BR/BR_{SM}$) .

Two of the production modes studied are mediated by fermions, $t\bar{t}H$ and ggH, while the others by vector bosons, VBF and VH. These are used to measure the relative strength of the Higgs boson couplings to elementary fermions and to vector bosons. The result is shown in Figure 1.9.

The measurement of the couplings is performed combining the information of both production and decay of the Higgs boson. To perform the measure it is necessary to make few assumptions. The Higgs boson candidate is assumed having $J^{PC} = 0^{++}$, as predicted by the SM. It is assumed that for this particle the narrow-width approximation can be used, allowing to consider independently decay and production processes. In this way cross section and branching ratio can be parametrized as:

$$\sigma_i \mathrm{BR}^f = \frac{\sigma_i(\vec{k}) \times \Gamma^f(\vec{k})}{\Gamma_H}, \tag{1.52}$$

where Γ_H is the total width of the Higgs boson, and Γ^f is its partial width in the final state f. These are expressed as function of a set of coupling modifiers $\vec{k}$ that

---

[11] SM is assumed true to perform such classification.

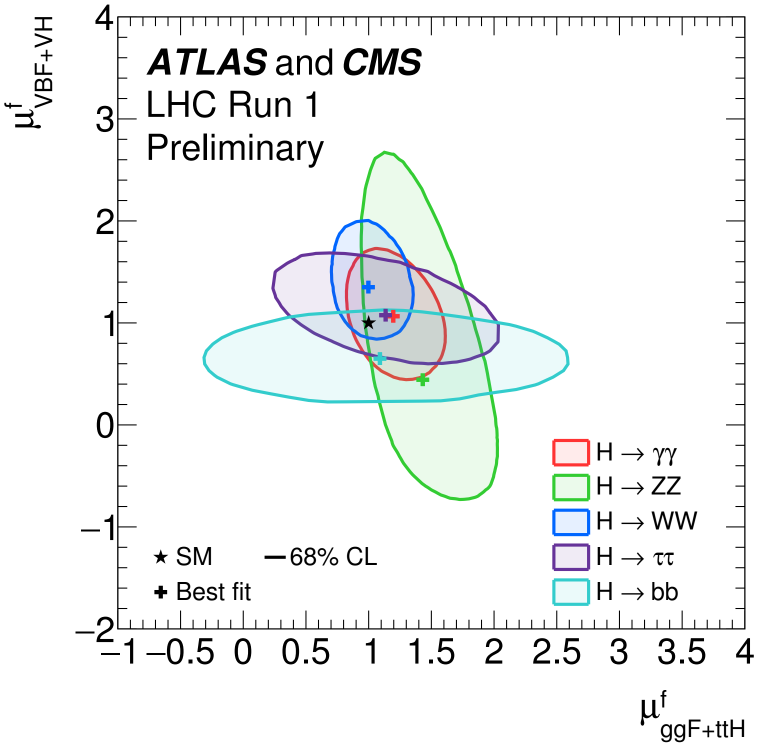


Figure 1.9: Comparison of best fit signal strength $\mu = \sigma/\sigma_{SM}$ for Higgs production modes mediated by fermions (ggH and $t\bar{t}H$) and by vector bosons (VBF and VH). Solid lines represent the contours of 68% CL confidence regions. The black star represents the SM expectation [55].

parametrize the deviation from the SM prediction. The coupling modifiers κ_i are defined as:

$$\kappa_i^2 = \sigma_i/\sigma_i^{SM} \text{ for production process } i \tag{1.53}$$
$$\kappa_i^2 = \Gamma_i/\Gamma_i^{SM} \text{ for decay process } i \tag{1.54}$$

In the SM these κ_i parameters are all equal to +1. To clarify the interpretation of the result, it is important to remember that only some SM particles have tree level interaction with the Higgs boson: W boson (κ_w), Z boson (κ_Z), bottom quark (κ_b), tau lepton (κ_τ), top quark (κ_t) and muons (κ_μ). The effective coupling modifier to gluons (κ_g) and to photons (κ_γ) are expressed as function of the tree-level coupling modifiers, since interactions with the Higgs boson is mediated by fermion and boson loops. The scaling factors for the second fermion generation are assumed to be the same as the one for the third. The coupling of the Higgs boson with electron, up- and down-quarks are neglected.

Figure 1.10 shows the result of the measured Higgs boson Yukawa couplings to fermions ($\kappa_f = \kappa_b = \kappa_t = \kappa_\tau$), as function of its couplings to vector bosons ($\kappa_V = \kappa_W = \kappa_Z$). For this test it is assumed that there is no physics beyond the SM, and both κ_f and κ_V are assumed to be positive. As a cross check, it is interesting to notice the shapes of the confidence region on Figure 1.10. Analyses with decays into vector bosons constrain κ_V more that κ_f, while it is the opposite for analysis with the Higgs boson decaying to fermions.

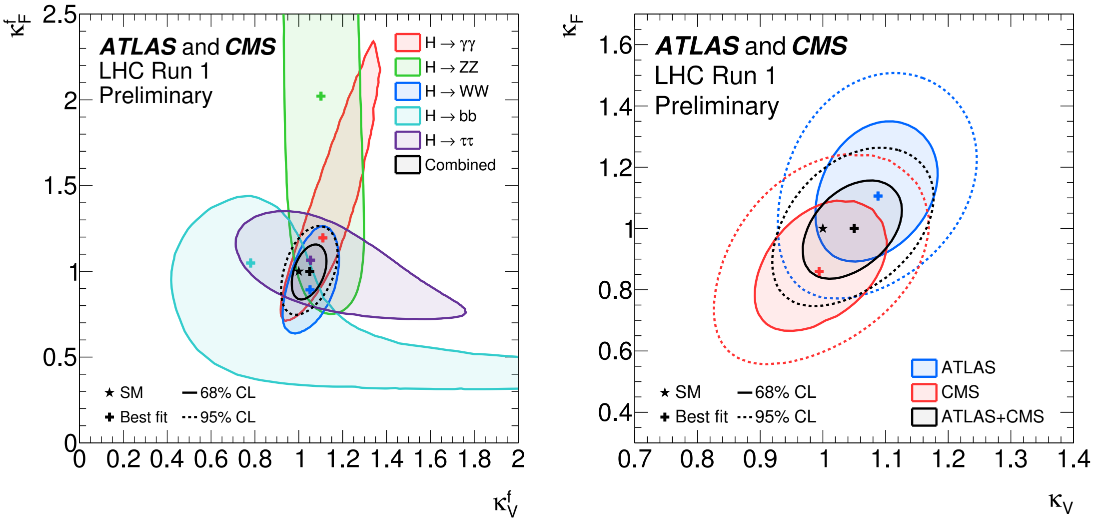
The case where κ_f and κ_V are allowed to be negative has also been considered. The measurements entering in the combination are capable to determine if κ_V and κ_f have the same sign. For this reason quadrants with positive κ_V are degenerate with respect to the one with negative κ_V, in Figure 1.11 are shown only the quadrants with positive κ_V.

Several BSM theories predict invisible or undetected Higgs boson decay. The Higgs boson width is sensitive to the presence of such decays, and it is inversely proportional to the Higgs boson production in the observed decay channel. Two scenarios compatible with the presence of new physics in the loop-mediated process of $gg \to H$ production and $H \to \gamma\gamma$ decay are considered:

- Higgs boson is assumed to not have any BSM decays ($BR_{BSM} = 0$);
- BR_{BSM} is left free, but it is assumed that $\kappa_W \leq 1$ and $\kappa_Z \leq 1$;

In these two cases each of the coupling modifiers to the SM particles involved are also obtained from the fit (i.e. are not set to SM prediction). Figure 1.12(a) shows the results of these studies.

Another test for the search for the presence of BSM particles can be performed assuming all the direct couplings of the Higgs boson to be the same as in the SM. Figure 1.12(b) shows the comparison of κ_g with κ_γ with the assumption that there is no physics BSM. A tension between the observation and the SM prediction should be observed if this assumption is wrong.



(a) (b)

Figure 1.10: Likelihood scan in the plane (κ_V, κ_f), κ_f and κ_V are assumed positive. The solid line represents the contour of the 68% CL confidence region. (a) Result of the ATLAS and CMS combination and for the individual channels. (b) Result of the ATLAS and CMS combination compared to the two experiments individual combinations [55].

After the verification that the coupling modifiers κ_g and κ_γ are compatible with the SM prediction, a new fit is performed assuming that there is no presence of BSM particles in the loops. Hence only the coupling modifiers of SM particles are considered. Figure 1.13(a) shows the result of the fit.

As stated in Section 1.1.5, the SM coupling of the Higgs boson to fermions is proportional to the fermion mass (m_f), while the coupling with vector bosons is proportional to the square of the boson mass (m_V). To check this property the coupling modifiers obtained from the fit described above are written as reduced coupling modifiers:

$$y_{V,i} = \sqrt{\kappa_{V,i}\frac{g_{V,i}}{2v}} = \sqrt{\kappa_{V,i}}\frac{m_{V,i}}{v} \tag{1.55}$$

$$y_{F,i} = \kappa_{F,i}\frac{g_{F,i}}{\sqrt{2}} = \kappa_{F,i}\frac{m_{F,i}}{v} \tag{1.56}$$

where v is the SM vev of the Higgs boson, $v = 246.22$ GeV [56], and the masses as well as the absolute Higgs boson coupling strength for vector boson (V) and for fermions (F) are denoted with $m_{V/F}$ and $g_{V/F}$ respectively. Figure 1.13(b) shows the results.

Test of the up- and down-type fermion symmetry has been performed extracting from the fit the parameters $\lambda_{du} = \kappa_l/\kappa_u$, $\lambda_{Vu} = \kappa_V/\kappa_u$, $\kappa_{uu} = \kappa_u\kappa_u/\kappa_H$. Figure 1.14(a) shows the result of the fit. Lepton and quark symmetry has been tested

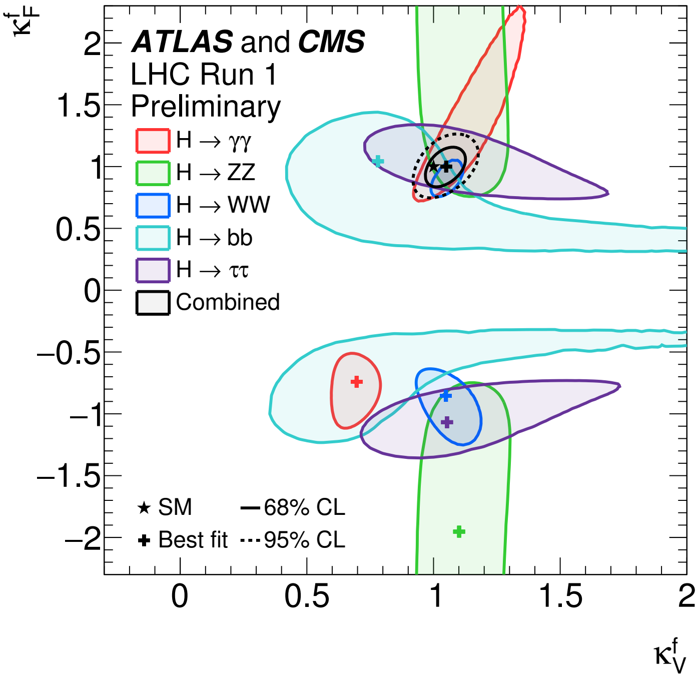
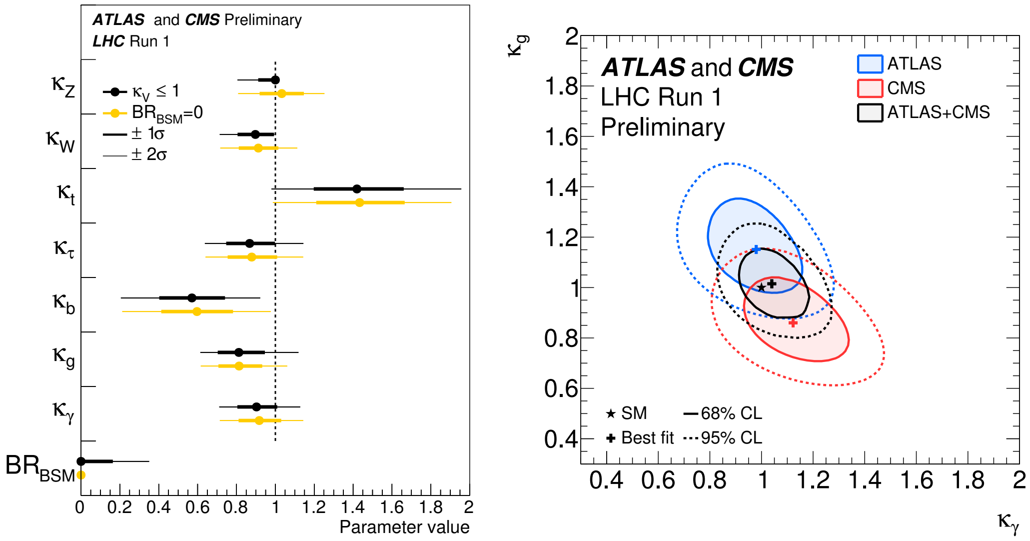


Figure 1.11: Likelihood scan in the plane (κ_V, κ_f), without assumptions on the sign of κ_f and κ_V. Result of the ATLAS and CMS combination together with the contribution coming from the individual channels. Solid line represents the contour of the 68% CL confidence region [55].



(a) (b)

Figure 1.12: (a) Fit results for the two parameterisations allowing BSM loop couplings. Black points represent the case with $\kappa_W \leq 1$ and $\kappa_Z \leq 1$ (denoted as $\kappa_V \leq 1$), yellow points the case without additional BSM contributions to the Higgs boson width, i.e. $BR_{BSM} = 0$. (b) Likelihood scan in the plane $(\kappa_g, \kappa_\gamma)$. No physics beyond the SM is assumed [55].

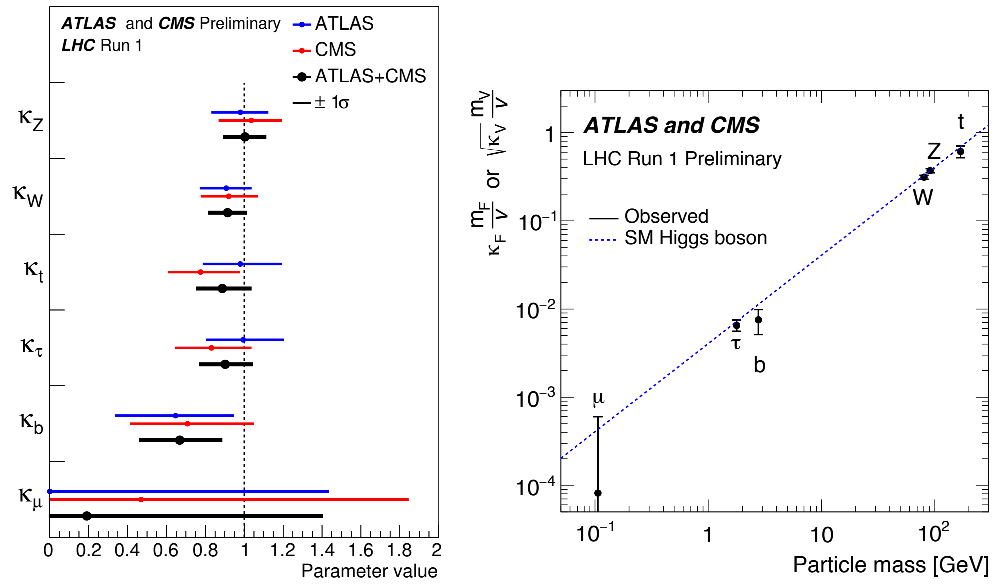
in a similar way. In this case the fit is performed using the parameters $\lambda_{lq} = \kappa_l/\kappa_q$, $\lambda_{Vq} = \kappa_V/\kappa_q$, $\kappa_{qq} = \kappa_q\kappa_q/\kappa_H$ Where $\kappa_H = \Gamma_H/\Gamma_H^{SM}$ and Γ_H is the total width of the Higgs boson. Result of this test is presented in Figure 1.14(b).

All the studies shown are consistent with the SM expectations.

1.2.6 Yukawa coupling with top quark

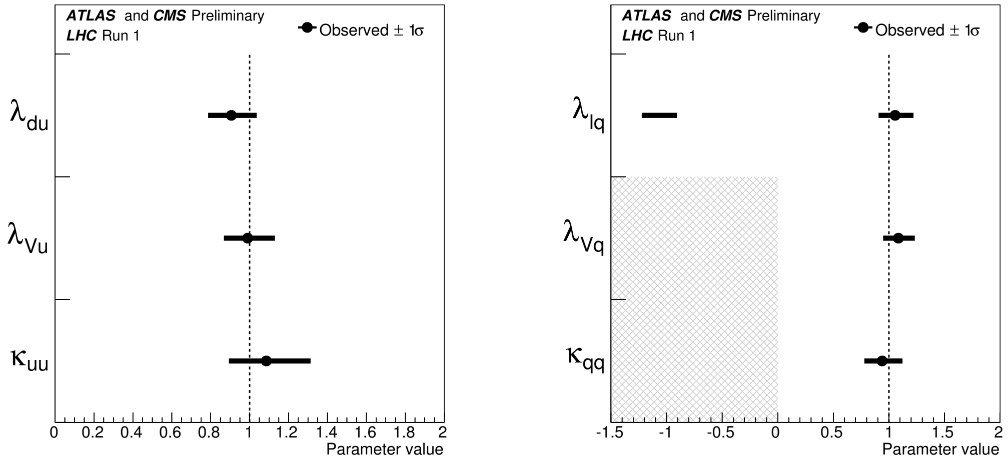
One of the most important measurements in the Higgs boson sector is the coupling with the top quark. From Equation 1.49 one can see that the coupling of the Higgs boson to fermions is proportional to their masses. As the top-quark is the heaviest known particle, with $m_{top} = 173.34 \pm 0.76$ GeV [57], the strength of its coupling is of order $C_{top} \sim 1$. A precise measurement of C_{top} can give hints on the scale of new physics [10].

A constrain on C_{top} may come from the measurement of the gluon-gluon fusion Higgs production process cross section. In this process, shown in Figure 1.15(a) at the tree level, the main contribution in the loop is given by the top-quark, with a small contribution also from b-quark and a destructive interference between the two. C_{top} can be constrained measuring the cross section of the process $gg \rightarrow ZH$. It is similar to gluon-gluon-fusion where the Z boson is radiated off the fermion loop. This process is sensitive also to the distructive interference between top-quark Yukawa coupling and Z-boson coupling to Higgs. The measurement of the



(a) (b)

Figure 1.13: (a) Best-fit values of coupling modifiers for the combination of ATLAS and CMS and separately for each experiment with the assumption of the absence of BSM particles in the loops, $BR_{BSM} = 0$, and $\kappa_j \geq 0$. (b)Representation of Higgs boson couplings to SM particles as function of the particle mass. Points are plotted according to two different scales: $\kappa_{F,i}\frac{m_{F,i}}{v}$ for the fermions; $\sqrt{\kappa_{V,i}}\frac{m_{V,i}}{v}$ for the vector bosons. The dashed line represents the SM expectation. The same values of the coupling modifiers shown in Figure (a) are used [55].



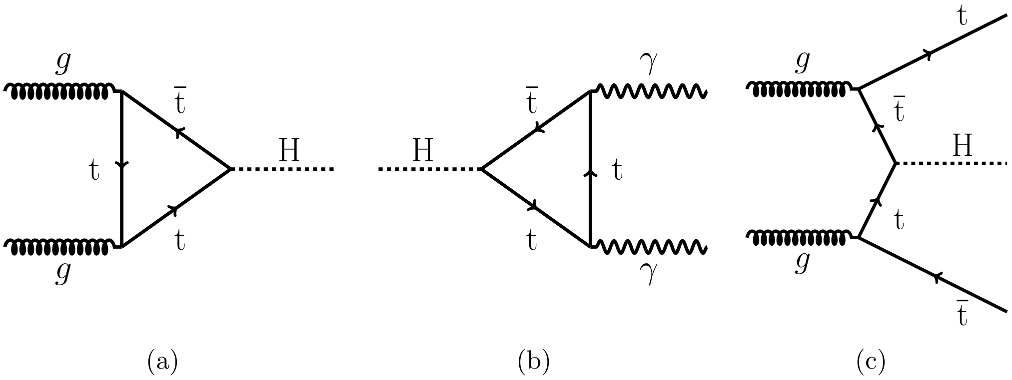
(a) (b)

Figure 1.14: 1σ intervals for combined results of (a) λ_{du}, λ_{Vu}, κ_{uu} parameters testing the up- and down-fermion coupling ratios. (b) λ_{lq}, λ_{Vq}, κ_{qq} parameters testing the lepton and quark coupling ratios [55].

branching ratio of the Higgs decay into a pair of photons, shown in Figure 1.15(b), is sensitive to the interference between C_{top} and W-boson coupling to Higgs (W loop contribution).

The top quark is the dominant contributor in the fermion loops. Nevertheless in these loops particles BSM may enter. To disentangle the top quark contribution it is necessary to search for final states that depend at the tree-level on the top-Higgs Yukawa coupling. A measurement of the rate of the Higgs boson production in association with a pair of top quarks ($t\bar{t}H$) process, shown in Figure 1.15(c), can provide a direct test of C_{top}.

Processes where the Higgs boson is produced together with only one top-quark (like $gb \to WtH$ and $qb \to tHq'$) strongly depend upon the sign of top-Higgs Yukawa coupling as well as upon its magnitude. In particular these are sensitive to the relative sign of C_{top} with respect to the coupling between the Higgs boson and the W boson. In the SM there is a destructive interference between diagrams with Higgs bosons emitted from top quark and W boson lines. If the two couplings have different signs the interference becomes constructive leading to an increase of the cross section.



(a) (b) (c)

Figure 1.15: (a): Feynman diagram for the gluon-gluon fusion Higgs production process. (b): Feynman diagram for the Higgs decay into a pair of photons. (c): Tree-level Feynman diagram for the production of the Higgs boson in association with a top quark pair ($t\bar{t}H$).

It is important to mention that evidence of a direct coupling of the Higgs boson to up-type fermions is still lacking, while coupling to down-type fermions has been observed, as summarized in [51, 55, 58].

Experimental results in $t\bar{t}H$ search

The ATLAS and the CMS collaboration have both preformed searches of the production of the Higgs boson associated with two top quarks. ATLAS collaboration has studied, the $H \to b\bar{b}$ decay channel [11], Higgs boson decaying into a pair of photons [12] and into WW, $\tau\tau$, ZZ with multi-leptonic final state [13].

The ATLAS analysis studying the final state with $H \rightarrow \gamma\gamma$ decay observe, after a fit to data, the best-fit strength parameter $\mu = \sigma^{t\bar{t}H}/\sigma^{t\bar{t}H}_{SM} = 1.3^{+2.5}_{-1.7}(\text{stat.})\ ^{+0.8}_{-0.4}(\text{syst.})$. The observed upper limit on μ is 6.7 times the SM expectation at 95% CL. These result is summarized in Figure 1.16.

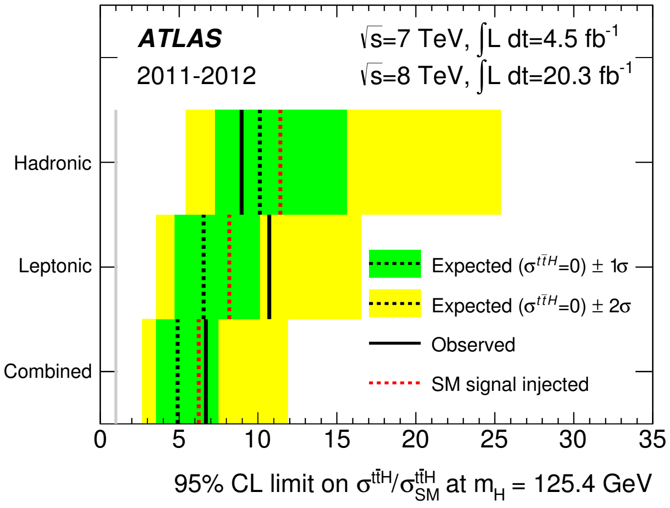


Figure 1.16: Results of the ATLAS search of $t\bar{t}H$, with $H \rightarrow \gamma\gamma$ [12]. Upper limit at 95% CL on the signal strength parameter. Results are shown separately for the different event categories considered in the analysis and for their combination.

The study of multi-leptonic final states performed by the ATLAS collaboration categorizes events according to the number of leptons (electron or muons) and τ reconstructed. The value of the strength parameter μ obtained from the best-fit to data is $\mu = 2.1^{+1.4}_{-1.2}$. The observed 95% CL upper limit is $\mu < 4.7$, while the expected value is $\mu < 2.4$. Results are summarized in Figure 1.17.

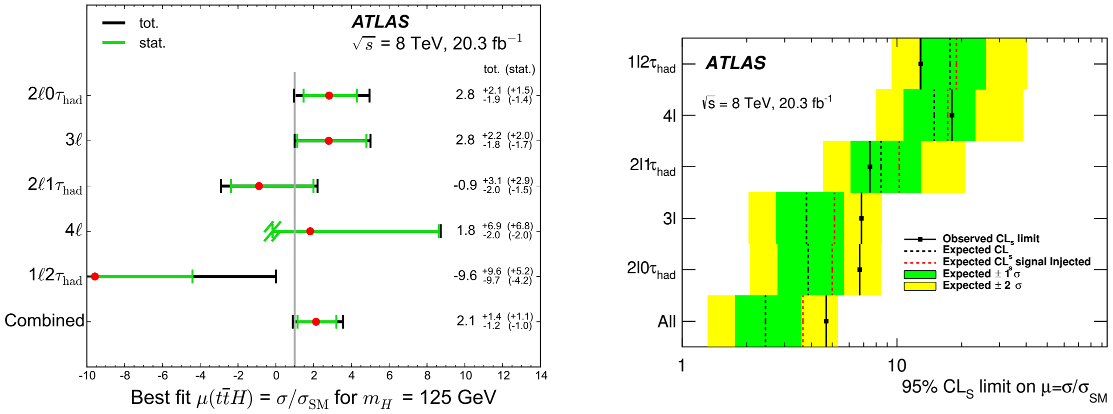
CMS collaboration has combined the search of several topologies into one measurement [14]. The topology with the Higgs decaying into a $b\bar{b}$ pair has been updated with the use of Matrix Element Method techniques. More detail about this updated version are given in Chapter 5. The combined analysis of CMS includes Higgs boson decaying to: $b\bar{b}$, $\gamma\gamma$, $\tau\tau$, WW, ZZ. Events selected for the analysis have at least one lepton [12] or two photons plus a number of jets depending on the decay channel studied. The main background is $t\bar{t}$ + jets production. The signal strength result obtained from a fit in selected data is $\mu = 2.8 \pm 1.0$ at 68% CL. An upper limit is set on μ at the value 4.5 at 95% CL, while the expected upper limit is 1.7. Figure 1.18 summarizes the signal strength results obtained by CMS.

The analyses focusing on the final state with $H \rightarrow b\bar{b}$ [11, 15] are discussed in more detail in Chapter 5.

ATLAS and CMS combined these analyses with several others [55] to measure, among other quantities, the coupling of the Higgs boson to the top quark. The

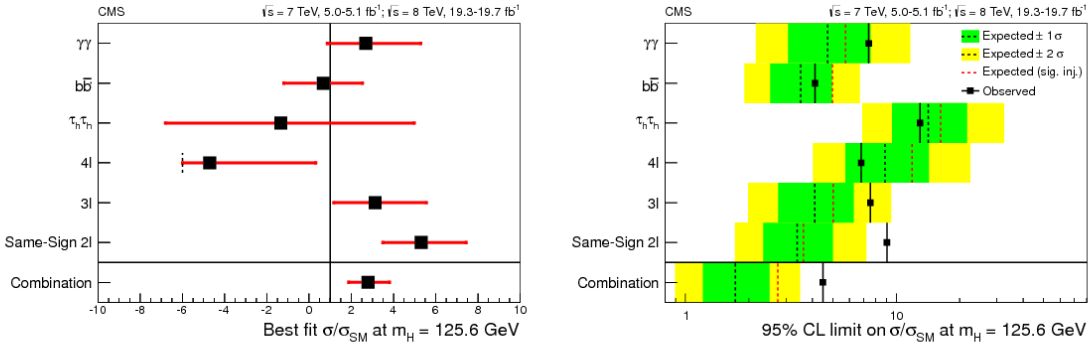
---

[12] For lepton here it is intended electron or muon.



(a) (b)

Figure 1.17: Results of the ATLAS search of $t\bar{t}H$ in multi-leptonic final states [13]. *Left*: Best fit values of the signal strength parameter $\mu = \sigma/\sigma_{SM}$. *Right*: Upper limits at 95% CL for the signal strength parameter. Results are shown separately for the different event categories considered in the analysis and for their combination at the bottom of the plot.



(a) (b)

Figure 1.18: Results of the CMS search for $t\bar{t}H$. *Left*: Best fit values of the signal strength parameter $\mu = \sigma/\sigma_{SM}$. *Right*: Upper limits at 95% CL on the signal strength parameter. Results are shown separately for the different event categories considered in the analysis and for their combination at the bottom of the plot. Note that the results obtained for the $b\bar{b}$ case have been updated in [15] and are shown in more detail in Chapter 5 .

result obtained is expressed as function of the coupling modifier κ_t, defined in Equation 1.54. The best-fit result is $\kappa_t = 0.89^{+0.15}_{-0.13}$, also shown in Figure 1.13(a).

The subject of this thesis work is the study and development of new techniques to perform measurment of final states that have never been studied before, with the aim of improving the current results.

Chapter 2

LHC collider and the ATLAS detector

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [59] is a particle accelerator located at CERN, the European Organization for Nuclear Research, and installed in the circular tunnel previously used for the LEP e^+e^- collider. The tunnel has a perimeter of 26.7 Km and is located underground at a depth varying from 45 m and 170 m on a plane inclined at 1.4%.

The LHC is designed to accelerate protons and heavy (Pb) ions and can produce collisions of proton-proton, lead-proton and lead-lead. LHC is the last step of the CERN's pre-existing accelerator chain, represented in Figure 2.1. At the beginning of the acceleration process, protons are extracted from a hydrogen bottle, after application of an electric field to tear the electron from the hydrogen atom. At this stage protons have an energy of 92 KeV, they are brought to an energy of 50 MeV with the linear accelerator (LINAC) and then injected into the Proton Syncroton Booster, which further increase the energy to 1.8 GeV. Proton beams go trough the Proton Syncroton, that shapes the proton beam in bunch trains with bunch spaced one from the other 25 ns (in 2012 it was 50 ns), and the Super Proton Syncroton to reach an energy of 450 GeV. Beams are subsequently injected in the LHC and accelerated up to the nominal energy. Nominal maximum energy for proton beams is 7 TeV, which corresponds to 14 TeV at the center of mass energy before collisions, with an instantaneous luminosity of 10^{34} cm^{-2}s^{-1}. For lead atoms, the design energy is 2.8 TeV per nucleon with an instantaneous luminosity of 10^{27} cm^{-2}s^{-1} The total luminosity delivered by the accelerator and the one recorded by the ATLAS experiment during 2012 data taking campaign is summarized in Figure 2.2.

To make the particles follow a circular trajectory, the LHC uses 1232 dipole magnets of 14.3 m length that can produce, for the designed working proton beams energy of 7 TeV, a magnetic field of 8.33 T. The operating temperature of the superconducting magnets is 1.9 K. Being a particle-particle collider, the LHC has two

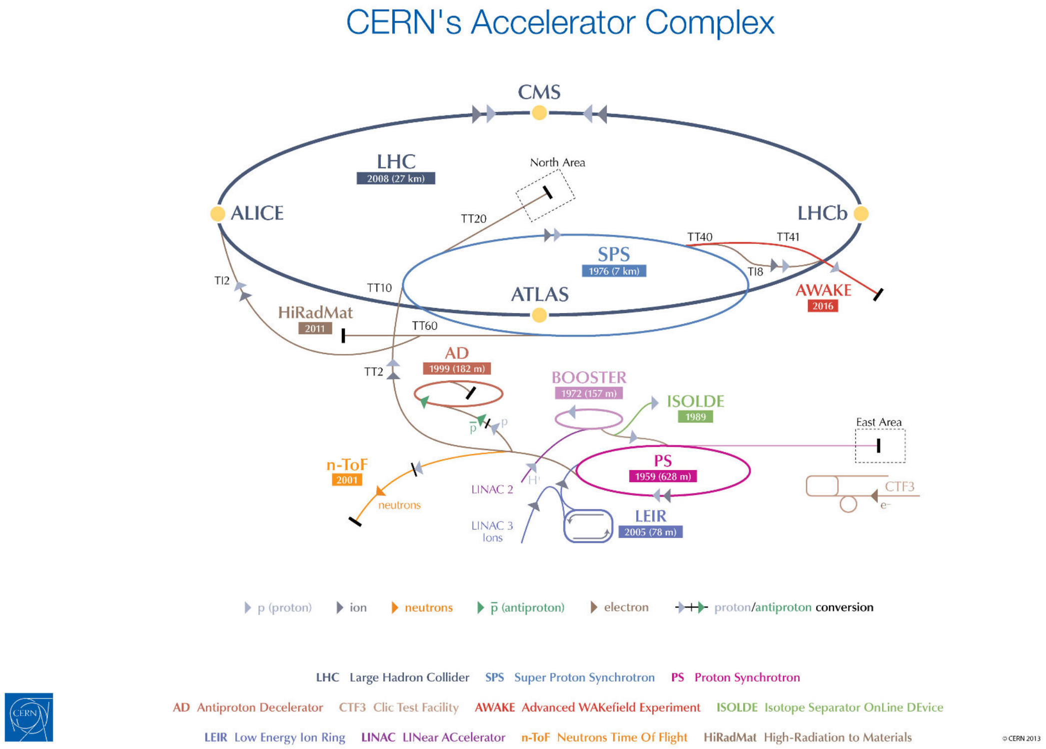


Figure 2.1: Schema of the CERN's accelerator chain and location of the associated four main experiments.

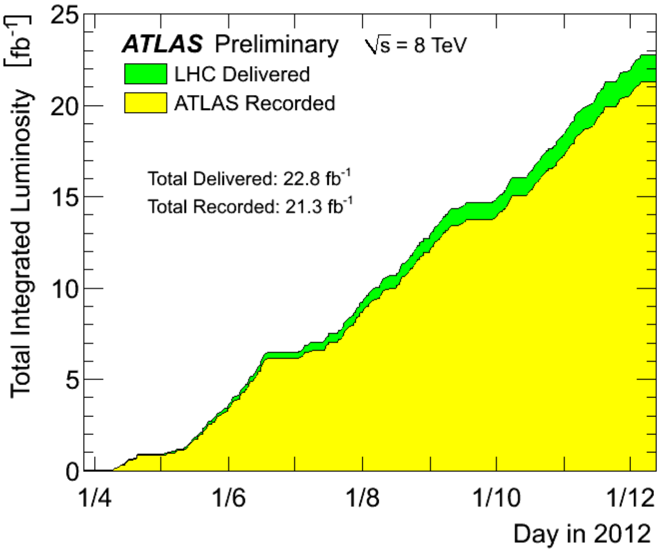


Figure 2.2: Integrated luminosity delivered by the LHC (green) and recorded by the ATLAS experiment (yellow) for pp collisions at $\sqrt{s} = 8$ TeV during 2012 data taking campaign [60].

different beam lines with magnetic field oriented in opposite direction to bend the beams. Due to the space constraints of the pre-existing tunnel and cost consideration of digging a new tunnel, a “two-in-one” design has been chosen for the dipoles. This means that the two sets of coils and beam channels share the same mechanical structure and cryostat. A schema of the section of a dipole is given in Figure 2.3. The main LHC working parameters in 2012 and the design one are shown in Table 2.1.

	2012	Design
Proton energy[TeV]	4	7
Dipole magnetic field[T]	4.76	8.33
Dipole operating temperature[K]	1.9	1.9
Instantaneous luminosity[cm^{-2}s^{-1}]	7.7×10^{33}	1.0×10^{34}
Number of bunches	1380	2808
Bunch spacing[ns]	50	25
Protons per bunch	1.6×10^{11}	1.15×10^{11}

Table 2.1: Value of some of the main LHC parameters during 2012 data taking campaign and the respective design values.

LHC hosts four main detectors[1], one on each of the four interaction points, as illustrated in Figure 2.1. One of these, The ALICE [61] experiment, is designed to detect ion-ion collisions. It exploits the unique physics potential of nucleus-nucleus interactions at LHC energies. Its aim is to study the physics of strongly interacting matter at extreme energy densities. The LHCb [62] experiment is designed to perform precision measurements of CP violation and rare decays of b- and c-hadrons. Two general-purpose experiments, ATLAS [63] and CMS [64], are located at the opposite sides of the accelerator ring and are designed to search for physics beyond the SM at the LHC energy and luminosity design value and to perform precision measurements of SM processes.

2.2 The ATLAS detector

The ATLAS (A Toroidal LHc ApparatuS) [63] is a general purpose LHC particle experiment. It is designed to perform a wide range of physics studies. One of the main objectives of the experiment are precision measurements of SM predictions like production and decays of the top quark, Higgs boson physics, electroweak physics and heavy-flavor physics. Searches for physics beyond the SM are also of fundamental importance, like search for dark matter and SUSY. Although the majority of the physics potential of ATLAS is in the study of *pp* collisions, a wide range of studies are also performed on proton-Lead and Lead-Lead collisions.

ATLAS is the biggest high energy physics detector ever build with 44 m of length and 25 m of height. It has a cylindrical shape and covers almost completely 4π

---

[1] ATLAS, CMS, ALICE and LHCb.

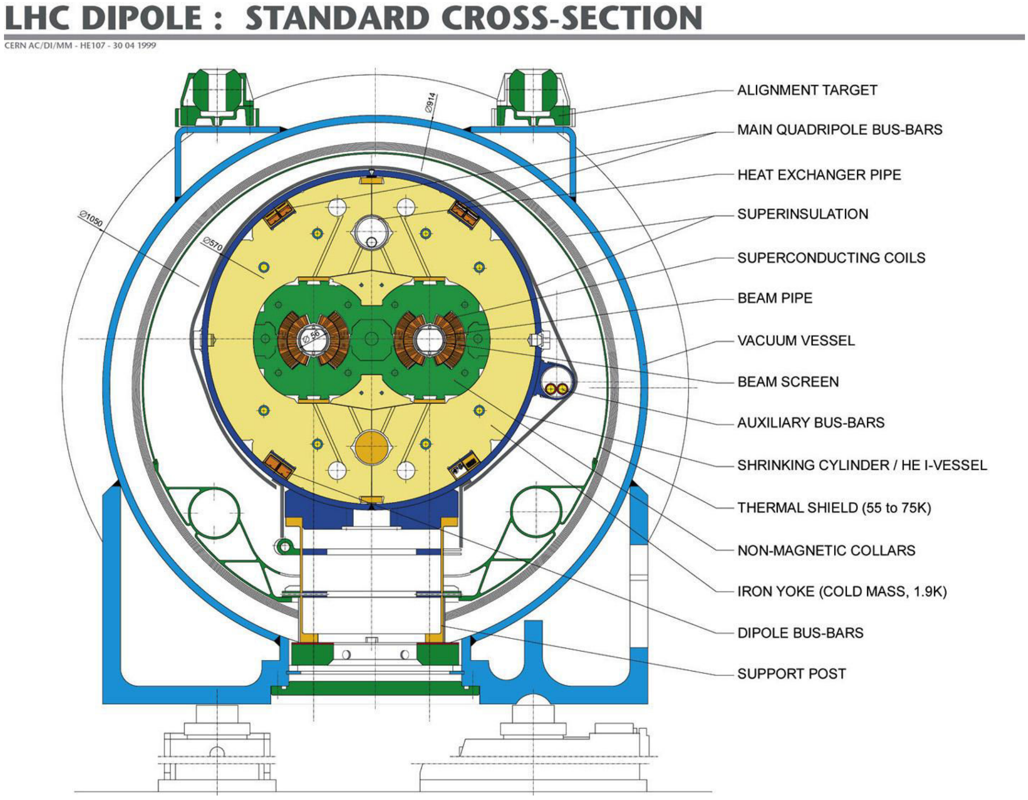


Figure 2.3: Schema of the cross-section of an LHC dipole magnet. To be noted the peculiar “two-in-one” design, with the two beam lines and magnetic coils sharing the same cryostat [59].

around the interaction point (IP) [2]. As can be seen in Figure 2.4 the ATLAS detector has an "onion-like" structure, with different sub-detectors organized in concentric layers, each one with the aim of measuring the properties of the different particles originating from the collisions and it is nominally forward-backward symmetric with respect to the IP. The inner most layer is the Inner Detector (ID), designed to detect charged particles. It is immersed in a magnetic field of 2 T delivered by a superconducting solenoid. This bends the trajectory of the charged particles, allowing the ID to measure their momenta. Outside the solenoid there are the calorimeters, to measure the energy of the particles. The outer layer is the muon spectrometer, designed to identify muons and participate to the measurement of their momentum. At 17 m in z direction from the IP and at 10 cm from the beam axis lies the LUCID (LUminosity measurement using Cerenkov Integrating Detector) detector. It detects inelastic pp scattering in the forward direction, and is the main online relative-luminosity monitor for ATLAS.

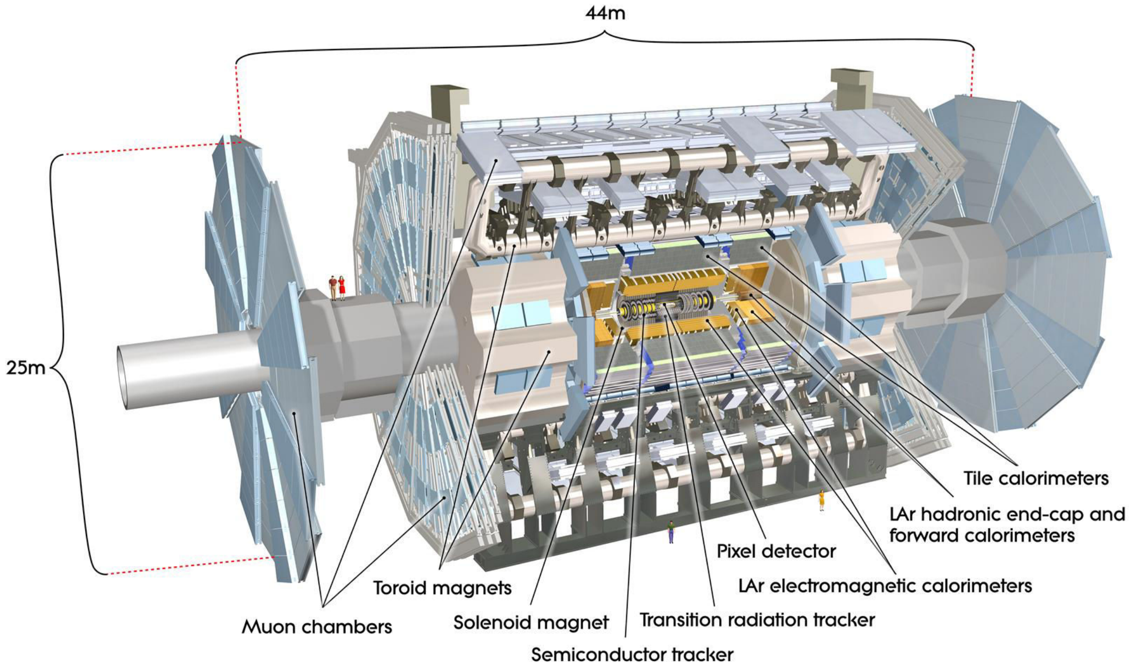


Figure 2.4: Cut-away view of the ATLAS detector.

2.2.1 ATLAS coordinate system

ATLAS uses a right-handed Cartesian coordinate system. The center of the coordinate system is the nominal IP at the center of the detector. The positive x-axis is pointing to the center of the LHC, the positive y-axis is pointing up-wards and the

---

[2] The point where the particle beams are designed to cross and interact.

z-axis is given by the direction of the beam pipe. Other important quantities are often used to describe the physics objects observed by the detector. The azimuthal angle with respect to the x-axis is called ϕ. θ is the polar angle with respect to the z-axis. A fundamental quantity is the pseudorapidity η defined as:

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

Objects lying in the (x,y) plane have $\eta = 0$, while $\eta = \infty$ points in the direction of the z-axis. The rapidity y_{rap} is defined as:

$$y_{rap} = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$$

Where E is the energy of the object and p_z its momentum component along the z-axis. The distance ΔR between two objects in the (η, ϕ) space is defined as

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

2.2.2 Magnetic system

ATLAS is equipped with four magnetic systems [65], one superconducting solenoidal magnet located at the center of the detector, and three toroidal magnets located outside the calorimeters.

Solenoid

The super-conducting solenoid magnet [66] is aligned with the beam axis and provides as mentioned above a magnetic field of 2 T along the z-axis of the detector. It contains completely the ID and has an inner radius of 2.46 m and a thickness of 10 cm. To preserve the performance of the calorimeters, the solenoid has been designed to have a material thickness as low as possible. The solenoid is contributing with a total of $\sim$ 0.66 radiation length (X_0) at normal incidence.

Toroids

A system of three large air-core toroids [67, 68] generates the magnetic field for the muon spectrometer. Each of the three toroids consists of eight coils assembled radially and symmetrically around the beam axis. The overall length of the barrel toroid system is 25.3 m, with inner and outer radius of 4.7 m and 10 m, respectively. It provides a magnetic field of $\sim$ 0.5 T. The two end-cap toroids line up with the central solenoid and are rotated by 22.5° with respect to the barrel toroid coil system. These generate a magnetic field of $\sim$ 1 T, required for optimizing the bending power in the end-cap regions of the muon spectrometer system. The end-cap toroids have inner and outer diameters of 1.65 m and 10.7 m respectively, with 5 m length.

2.2.3 Tracking system: Inner Detector

The tracking system (Tracker) is the closest detector to the interaction point, also called Inner Detector (ID), [69, 70]. It covers the region $|\eta| < 2.5$. It is fully contained within the solenoid magnet. The tracker is designed to measure charged particles trajectories and their momenta with high precision. It is used also to reconstruct the interaction vertices, allowing to distinguish primary vertices from secondary vertices originating from the decay of long lived particles. It is made of several sub-detectors allowing to have few high precision measurements close to the interaction point and a large number of lower precision measurements at higher radius. Each of the sub-detectors is divided into barrel and end-cap regions. In the barrel the detecting elements are arranged in cylindrical concentric layers around the beam axis. In the end-caps these are mounted on disks orthogonal to the beam pipe. Longitudinal cut view of the ID is presented in Figure 2.5, including the dimensions and relative positions of all the sub-detectors. The three dimensional view of the ID is shown in Figure 2.6(a).

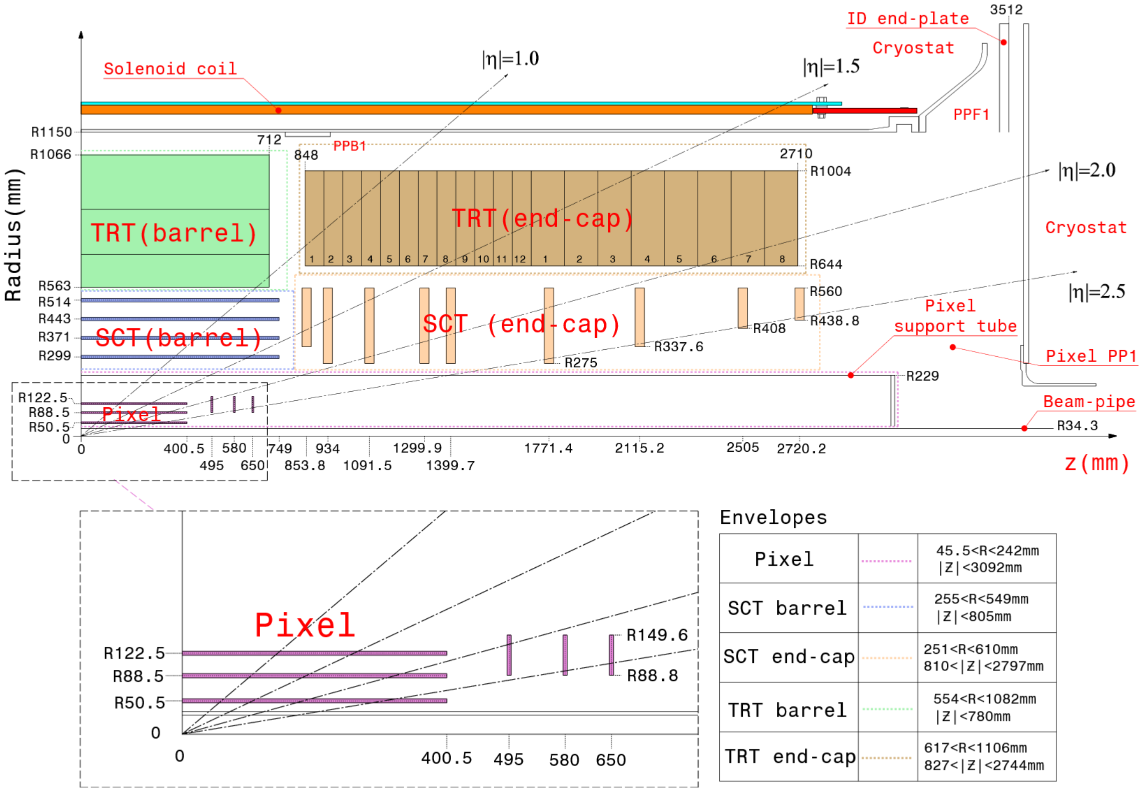
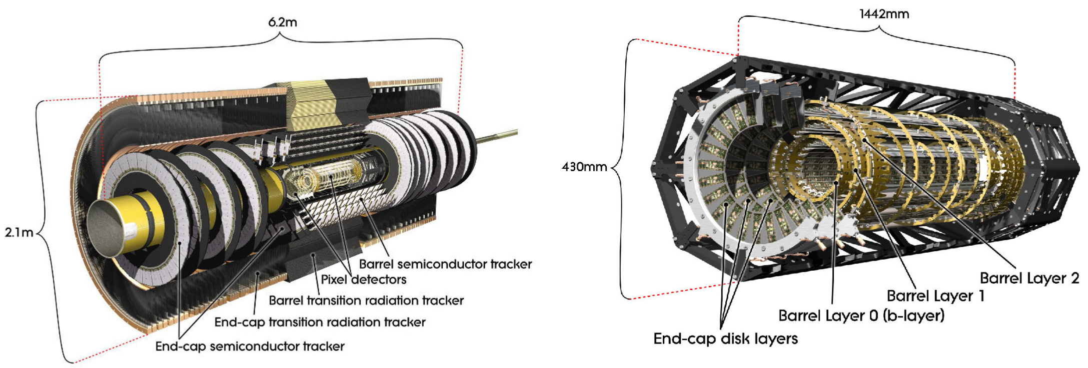


Figure 2.5: Longitudinal cut view of the Inner Detector [71].

Pixel detector

The pixel detector [72] is the innermost layer of the ATLAS detector and the closest to the interaction point. The barrel region is made of three concentric layers, the



(a) (b)

Figure 2.6: Schematic 3D-view of the ATLAS Inner Detector (a) and of the Pixel detector (b).

two endcap regions are identical and each is composed of three identical disk layers. A layout of the detector is shown in Figure 2.6(b). The total number of pixels is approximately 67 million in the barrel and 13 million in the end-caps, covering a total active area of about 1.7 m^2, with each pixel having a typical size of 50×400 $\mu\mathrm{m}^2$. The pixel detector is designed to provide at least three measure points on a charged track emerging from the collision region. It is capable of providing measurements with high granularity near the IP, where the density of charged tracks is maximal. It covers the range $|\eta| < 2.5$ and is capable of three-dimensional-vertexing with good resolution in the longitudinal z-axis, allowing primary vertex reconstruction of charged tracks with $\sigma(z) < 1$ mm and transverse impact parameter resolution of better than about 15 μm. The intrinsic accuracies are 10 μm in R-ϕ plane and 115 μm in the z-axis in the barrel, and 10 μm in R-ϕ plane and 115 μm in the z-axis in the disks. The pixel detector has also very good radiation-hardness, designed to resist to total dose of 500 kGy, and small material thickness to not degrade the performances of the other sub-detectors, ranging from 0.1 X_0 at $\eta = 0$ to a maximum of 0.6 X_0 for high η. In 2014 a fourth layer, called Insertable B-Layer (IBL) [16], has been added to the barrel region at a radius of 3.2 cm. For its installation the beam pipe has been replaced with one of 3 cm of radius. The one used in 2012 had 5 cm of radius.

SemiConductor Tracker

The SemiConductor Tracker (SCT) [73, 74] is the second layer of the ID and is built with silicon microstrip layers. The barrel region is made of four double-layered concentric cylinders and the end-caps are made of nine double-layered disks each giving at least four space-point measurements for each track. Barrel region of the detector uses small-angle (40 mrad) stereo strips modules to measure both coordinates. Strips are parallel to the beam axis (z), resulting in a precision of 17 μm in the R-ϕ coordinate and 580 μm in the z-coordinate. In the end-cap region, the detectors have a set of strips running radially and a set of stereo strips at an angle of 40 mrad.

Transition radiation tracker

The Transition Radiation Tracker [75] is the outermost component of the ID. It is a drift-tube tracker and transition radiation detector. It contributes significantly to the momentum measurement. It covers the region up to $|\eta| < 2.0$. It is made of 2 mm radius thin-walled proportional drift tubes, also called straw tube or straws, arranged parallel to the beam axis in the barrel region and radially in the end-cap regions. The straw tubes are filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. Straw tubes were chosen as detecting elements because they offer a high degree of modularity of the detector and can easily be integrated into a medium producing transition radiation without compromising the continuous tracking performances. They have an intrinsic accuracy of 130 μm. The combination of precision silicon-based trackers at small radii with the TRT, that can provide only R-ϕ information, gives very robust pattern recognition and high precision hit measurements in both R-ϕ and z coordinates. The lower precision per point compared to silicon detectors is compensated by the large number of measurements, typically $\sim$ 36 per crossing track, and longer measured track length.

2.2.4 Calorimeters

Calorimeters [76] aim at measuring the energy of the incident particles. In ATLAS these are located outside the solenoid magnet and cover the region $|\eta| < 4.9$. Three types of calorimeters are installed in ATLAS: Electromagnetic, Hadronic and Forward calorimeters. A schematic view of the ATLAS calorimeter system is given in Figure 2.7. The nominal resolutions for each sub-detector are:

$$\begin{aligned}
\frac{\sigma_E}{E} &= \frac{10\%}{\sqrt{E}} \oplus 0.7\% \quad ; \text{ Electromagnetic} \\
\frac{\sigma_E}{E} &= \frac{50\%}{\sqrt{E}} \oplus 3\% \quad ; \text{ Hadronic} \\
\frac{\sigma_E}{E} &= \frac{100\%}{\sqrt{E}} \oplus 10\% \quad ; \text{ Forward}
\end{aligned}$$

Electromagnetic calorimeter

The Electromagnetic calorimeter (EM) is designed for the measurement of electrons and photons. It is a sampling calorimeter with lead as passive absorber and liquid argon (LAr) as active material. It is designed with accordion-shaped kapton electrodes and lead absorber plates over its full coverage. The accordion geometry provides complete ϕ symmetry without azimuthal cracks. It covers the region $|\eta| < 3.2$, the

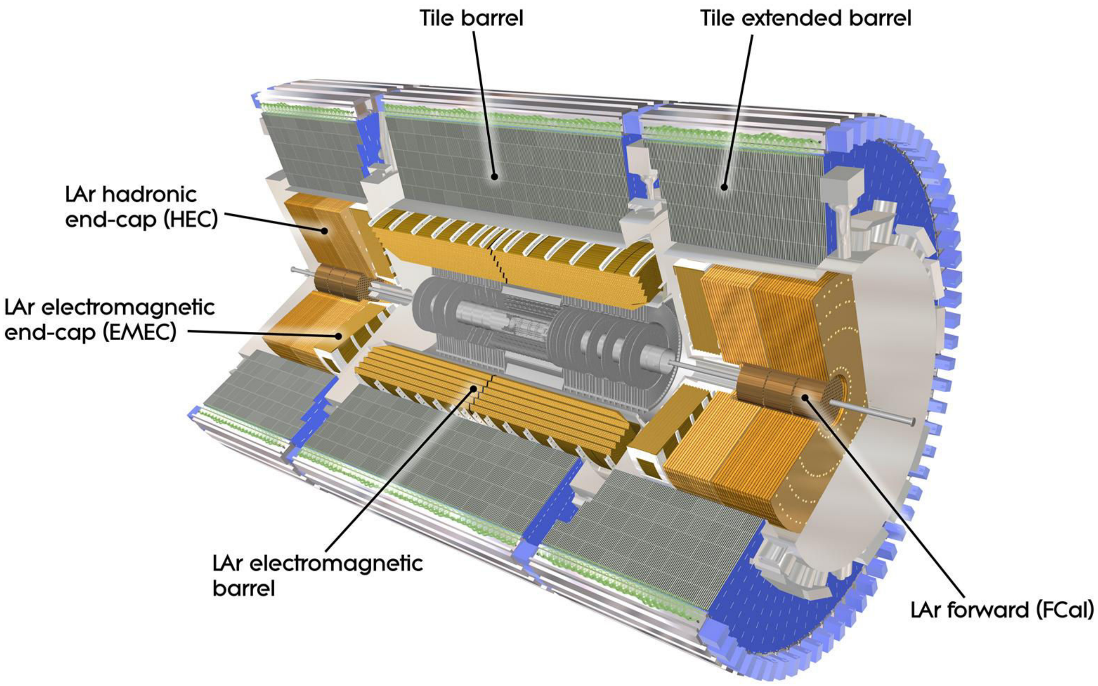


Figure 2.7: Schematic view of the ATLAS calorimeters.

barrel part covers $|\eta| < 1.475$, while the two end-caps cover $1.375 < |\eta| < 3.2$. The end-caps are further divided in two concentric wheels, covering respectively $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$. The central solenoid and the EM calorimeter share a common vacuum vessel, to reduce the material thickness and to avoid degradation of the calorimeter performances. It has finer granularity in the region overlapping with the ID to give better performance in photon and electron reconstruction. EM calorimeter has a material thickness of $> 22X_0$ in the barrel region and $> 24X_0$ in the end-caps. In the barrel the EM is segmented in three concentric layers of sampling, as can be seen in Figure 2.8. The innermost layer has fine η granularity to identify photons and electrons, that tends to have a narrower shower shape with respect to particle jets. In the region $|\eta| < 1.8$ a presampler detector is used to correct for the energy lost by electrons and photons before reaching the calorimeter. The presampler consists of an active LAr layer of thickness 1.1 cm in the barrel region and 0.5 cm in the end-cap.

Hadronic calorimeter

The hadronic calorimeter aims at measuring the energy of particle jets and covers the range $|\eta| < 3.2$. Two different technologies are used in the barrel and in the end-caps. The barrel, also called Tile Calorimeter, is a sampling calorimeter with steel tiles as absorber and plastic scintillating plates as active material. It covers the region $|\eta| < 1$, and the extended barrel covers $1 < |\eta| < 1.7$. In the Hadronic End-

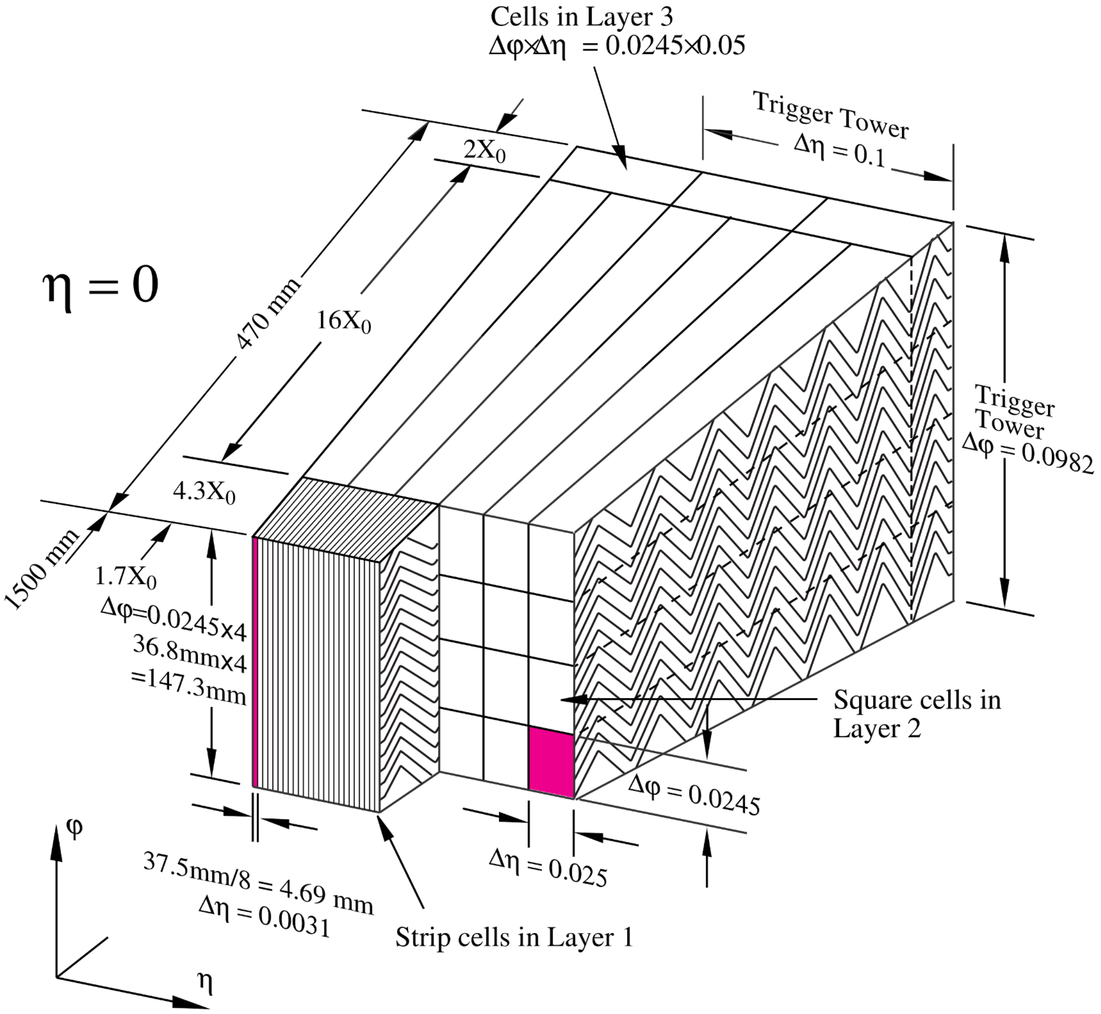


Figure 2.8: Schema of a LAr calorimeter barrel module. The different layers in which the detector is divided are clearly visible. Granularity in η and ϕ of cells and trigger towers are shown [76].

cap Calorimeter (HEC) a technology similar to the EM calorimeter is used, with copper as absorber and liquid argon as active material. These covers the region $1.5 < |\eta| < 3.2$.

Forward calorimeter

Forward Calorimeters (FCal) are located on each side of the ATLAS detector. These are designed to measure particle jets at high η values which allow to improve the missing transverse energy measurement. The FCal covers pseudorapidity range of $3.1 < |\eta| < 4.9$, it has a material thickness of approximately 10 interaction lengths and consists of three modules in each end-cap: the first, made of copper, is optimized for electromagnetic measurements, while the other two, made of tungsten, measure predominantly the energy of hadronic interactions.

2.2.5 Muon spectrometer

The Muon spectrometer [77] is the outermost layer of the ATLAS detector. It is instrumented with separate trigger and high-precision tracking chambers and is based on the magnetic deflection of muon tracks in the large superconducting air-core toroid magnets (see Section 2.2.2). The octagonal structure of the muon chambers reflects the symmetry of the magnetic field. The system consists of four sub-detectors build with different technologies: the Monitored Drift Tubes (MTD), the Cathode Strip Chambers (CSC), the Resistive Plate Chambers (RPC) and the Thin Gap Chambers (TGP). Figure 2.9 shows a schema of the muon spectrometer.

The MDT chambers are designed to perform the precision muon momentum measurement and covers the pseudorapidity range $|\eta| < 2.7$. These chambers consist of three to eight layers of drift tubes, operated at an absolute pressure of 3 bars, which lead to an average resolution of 80 μm per tube or about 35 μm per chamber. The drift tubes have diameter of 29.970 mm and contain a gas mixture of 93% Ar and 7% CO_2. After the gas ionization, the electrons are collected at the center on a 50 μm tungsten-rhenium wire.

In the region at high pseudorapidity ($2 < |\eta| < 2.7$) the CSC are used in the inner-most tracking layer due to their higher rate capability and time resolution. These are multi-wire proportional chambers with parallel wires, where central wire points in the radial direction, and cathode planes segmented into strips in orthogonal direction. This allows both coordinates to be measured from the induced-charge distribution. The resolution of a chamber is 40 μm in the bending plane and about 5 mm in the transverse plane.

The RPC and the TGC are used in the trigger system of the muon spectrometer and cover up to $|\eta| < 2.4$. The RPC covers the barrel region $|\eta| < 1.05$, it is a gaseous parallel electrode-plate detector where two resistive plates are kept parallel to each other at a distance of 2 mm. The electric field between the plates of about 4.9 kV/mm allows avalanches to form along the ionizing tracks towards the anode. TGC are multi-wire proportional chambers and cover the end-cap region $1.05 < |\eta| < 2.4$. These sub-detectors are also used to provide bunch-crossing identification and measure the muon coordinate in the direction orthogonal to that determined by the precision-tracking chambers.

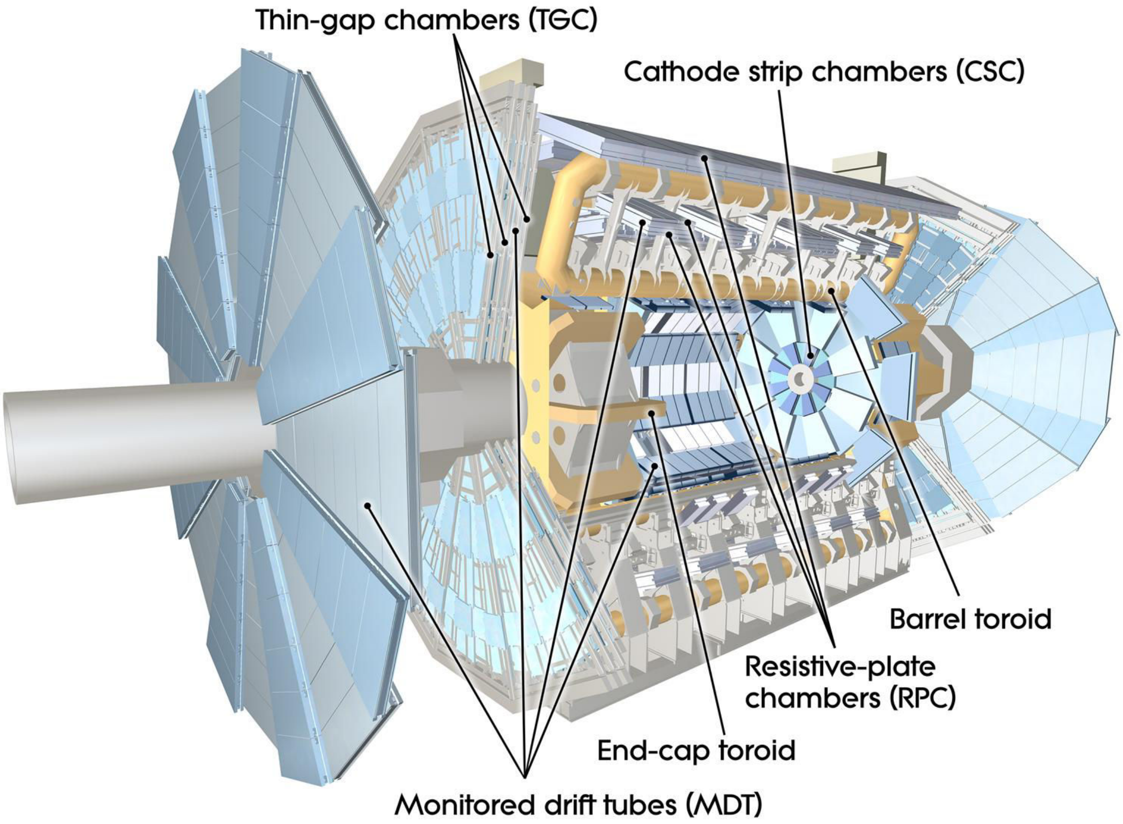


Figure 2.9: Schematic view of the ATLAS muon spectrometer.

2.2.6 Trigger systems

The LHC can deliver to the ATLAS experiment up to 40 millions of *pp* collisions per second, but the data acquisition system can put in permanent storage only few hundreds of events per second. To reduce this huge amount of collision events to an acceptable level, an on-line Trigger system, to select events with interesting physics properties and reject the huge number background collisions, is required. The ATLAS trigger system is described in detail in Chapter 4.

2.2.7 Computing Model

The big amount of data collected by the ATLAS experiment, of the order of 10000 Gb per year, and the huge number of Monte Carlo simulation events needed pose an important challenge in terms of storage needs (events have size of ~ 1.6 MB), and computation resources. The ATLAS computing model [78] should be able to allow the same level of data access and make available the same amount of computing resources to all members of the ATLAS Collaboration. The model makes substantial use of Grid Computing concepts, consisting of tiers of computing clusters suited to distribute the computing and storage loads among the different participating institutes.

Tier structure

The CERN computing center is called Tier-0 and is the first layer of the ATLAS analysis system. It is responsible for the archiving and distribution of the primary RAW data received from the detector. Tier-0 stores a complete copy of the RAW Data and also performs first-pass reconstruction producing derived data formats and copying these to the Tier-1 facilities. Tier-1 are computing centers located worldwide and host and provide long-term access and archiving of a subset of the RAW data, perform their reprocessing and provide ATLAS-wide access to the derived data formats. Tier-2 facilities spread worldwide into laboratories and institutes are meant to provide analysis capacity for physics working groups and subgroups. They have the responsibility for the official Monte Carlo production (simulated data are stored in the Tier-1s) and physics analysis. The development and refinement of calibration and reconstruction algorithms are also performed at the Tier-2 centers.

Event store

Data are stored as successively derived event representations, beginning with raw or simulated data and progressing through reconstruction into more refined event representations suitable for analysis. RAW data are events output of the detector, without any other treatment. A first-pass reconstruction produces ESD (Event Summary Data) and AOD (Analysis Object Data) files. The ESD data-format contains the reconstructed quantities measured by the detector (energy in the calorimeter cells, clusters information, tracks, vertices) as well as reconstructed physics objects (electrons, photon, jets, taus, muons). The event size of the ESD format is ~ 0.5 MB. AOD is a small-sized data format (0.15 MB per event), it is well suited for data distribution and only contains the physics objects. Derived Physics Data (DPD) format is a reduced data set with stricter event selection, reducing in size the information per object and dropping unwanted data objects. The physics analysis are performed on AOD or DPD data sets.

Chapter 3

Physics objects reconstruction

In this chapter the *physics objects* used in the full hadronic $t\bar{t}H$ ($H \to b\bar{b}$) analysis are described. Physics objects are the reconstructed objects that are identified as electrons, muons, particle jets, b-tagged jets, missing transverse energy and photons. The main objects used in the analysis are particle jets, described in Section 3.1, and b-tagged jets, described in Section 3.2. Electrons and muons are not directly used in the analysis, but are vetoed to avoid overlap with other $t\bar{t}H$ ($H \to b\bar{b}$) analysis channels when combining the results, their reconstruction is subject of Sections 3.3 and 3.4 respectively. Missing transverse energy (E_T^{miss}) and photons are not used in the analysis, but are briefly described for completeness in Sections 3.5 and 3.6.

3.1 Particle jets

In high energy physics experiments many of the processes studied have final states with quarks or gluons. These strong interacting particles loose their energy with the generation out of the vacuum of other strong interacting particles, in a process called *fragmentation*. In this way around the first energetic parton a cloud of particles with smaller energy is formed, and aggregate to form stable hadrons, like pions, kaons and protons. This process is called *hadronization*. These are reveled in the detector, mainly in the calorimeter system. The hadrons created in this process are almost collinear to the direction of the most energetic parton and are collimated in a narrow cone of particles called *jet*. Figure 3.1 shows a schematic view of the jet production process.

3.1.1 Jet reconstruction algorithms

A particle jet is not a simple object, like a reconstructed electron or a photon, but needs a more sophisticated algorithm to be reconstructed. In ATLAS jets are reconstructed starting from topological calorimeter clusters, and are subsequentially grouped together in jets via the jet algorithm “anti-k_t”.

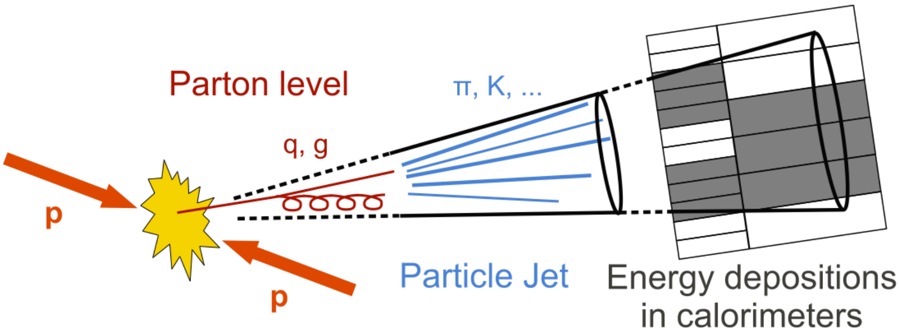


Figure 3.1: Schema of jet formation process.

Topological clustering

Reconstruction process starts from topological calorimeter clusters, also called *topo-clusters* [79], designed to follow the shower development. Topo-clusters have no predefined shape, as opposite to sliding-window clusters that have by construction rectangular or square shape.

Topological clustering is designed to group into clusters neighboring cells that have significant energies compared to the expected noise. The algorithm starts from a seed cell characterized by a signal-to-noise (S/N) ratio above four. Noise is evaluated as the absolute value of the energy deposited in the calorimeter cell divided by the RMS of the energy distribution measured in random triggered events. Cells adjacent to the seed or the forming cluster are added if satisfying the condition $S/N \geq 2$. In the final step all the cells surrounding the formed cluster are added.

To increase the separation of showers originating from close-by particles a splitting step is performed. The cells in the cluster are scanned searching for a local maxima with a threshold of 500 MeV. These maxima are used as seeds for new iteration of the topo-clustering algorithm, splitting the original cluster into more topo-clusters.

Anti-k_t algorithm

In the literature several jet algorithms are proposed, like k_t algorithm [80] and Cambridge/Aachen [81, 82]. Currently both ATLAS and CMS experiments use the so called “Anti-k_t” algorithm [83]. The algorithm takes as input a collection of entities, like calorimeter topo-clusters or reconstructed tracks or particles in a MC simulation, and group them according to particular criteria. To characterize these criteria it is necessary to define two quantities: d_{ij}, distance between two entities, and d_{iB}, distance between an entity and the beam:

$$d_{ij} = \min\left(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2}\right)\frac{\Delta_{ij}^2}{R^2} \tag{3.1}$$

$$d_{iB} = \frac{1}{k_{ti}^2} \tag{3.2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i + \phi_j)^2$ and k_{ti}, y_i and ϕ_i are respectively the transverse momentum, the rapidity and the azimuthal angle of the entity i. R is a radius parameter.

The algorithm proceeds iteratively in two steps. First, it looks for the smallest distance. If it is a d_{ij} it merges i and j adding their four-momenta, the new entity is added to the collection and i and j are removed. Else if it is d_{iB} it retains i as a jet and remove it from the collection. Second, distances are recalculated. These steps are repeated until there are no entities left. The execution of the algorithm requires a large number of operations to compute the distances and find the smallest, of the order of N^3 where N is the number of entities in the event. The FastJet C++ library [84, 85] has been implemented to reduce the time consumption to the order $N\log(N)$. It performs a factorization of the minimisation problem into momentum and geometrical parts reducing the size of the set of distances where to search for the minima.

The anti-k_T algorithm has the advantage of being infrared and collinear safe, meaning that the jet reconstruction process is stable against additional soft radiation and collinear splitting of the initial parton. This is not the case, for example for more simple cone algorithms [86]. Moreover, hard jets have a circular shape of radius R and only softer jets have more irregular shapes. This is not the case for the k_t and Cambridge/Aachen algorithms, that give irregular shapes to the jets.

3.1.2 Energy calibration

The jet reconstruction process uses as input topological clusters, as mentioned in the previous sub-section. The topo-clusters used can be reconstructed at two different energy scales [87]: ElectroMagnetic energy scale (EM), to measure with better precision the particles produced in the electromagnetic shower, and Local Cell signal Weighting (LCW), to correctly account for the response of the calorimeter to hadrons.

In LCW calibration, topo-clusters are classified as either electromagnetic or hadronic, primarily based on the measured energy density and the longitudinal shower depth. Based on this classification energy corrections are derived from single pion Monte Carlo simulations. Dedicated corrections take into account effects of non-compensation, signal losses due to noise threshold effects, and energy lost in non-instrumented regions. These are called “local” corrections due to the fact that they are applied directly to calorimeter clusters and are defined without reference to a jet definition. The anti-k_t jet algorithm runs using these calibrated topo-clusters as input.

In the $t\bar{t}H$ ($H \rightarrow b\bar{b}$) fully hadronic analysis LCW calorimeter jets are used.

In addition to this first calibration applied to the calorimeter clusters, an additional calibration is applied to recover Jet Energy Scale (JES) in the calorimeters to the true energy of the corresponding jet of stable particles entering the ATLAS detector. The calibration procedure consists of four steps. The first step is the pile-up[1] correction. An energy offset is evaluated to account for in-time pile-up, meaning additional collision recorded in the same event, and for out-of-time pile-up, meaning effects of past collisions influencing the energy deposit. The correction is derived from MC simulations as function of reconstructed primary vertices and expected average number of interactions in bins of η and transverse momentum. The second step is the correction of the origin of the jet, to make it point to the primary vertex instead of the nominal center of the ATLAS detector. The third step is the application of the proper calibration [88], consisting in corrections to pseudorapidity and energy of the jet. These corrections are derived in MC simulations. As last step a residual correction derived in situ is applied to jets reconstructed in data.

3.1.3 Jet selection in $t\bar{t}H$ analysis

In ATLAS jets are reconstructed, as mentioned in Section 3.1.1, with the anti-k_t algorithm with a distance parameter $R = 0.4$. In the $t\bar{t}H$ ($H \rightarrow b\bar{b}$) fully hadronic analysis, jets are required, after energy calibration, to have $p_T > 25$ GeV and $|\eta| < 2.5$. In addition, the five leading jets are required to have $p_T > 55$ GeV. These requirements allow to be in a region where the trigger behavior outside the trigger efficiency plateau is well understood, see Section 4.2.

During the jet reconstruction process no difference is made between electron candidates and jet candidates. To avoid double counting, any jet close within $\Delta R < 0.2$ to an accepted electron is discarded. Electrons that are at a distance $\Delta R < 0.4$ from a jet are also discarded.

Further selection is needed to reject jet candidates that are not originating from hard-scattering. Main sources of this background are events where one proton of the beam collides with the residual gas within the beam pipe, cosmic-ray muons overlapping in-time with collision events and calorimeter noise. Events are rejected if at least one jet is not associated to in-time real energy deposition in the calorimeter.

A dedicated observable, called Jet Vertex Fraction (JVF), is defined in order to increase the rejection of jets that are not originating from hard-scatter interaction. This is defined in Equation 3.3 as the ratio of the scalar sum of the p_T of the tracks matched to a jet (i.e. with a $\Delta R < 0.4$) and originating from a reconstructed vertex over the scalar sum of the p_T of all tracks matched to a jet.

$$\mathrm{JVF}(\mathrm{jet}_i, \mathrm{vtx}_j) = \frac{\sum_k p_T(\mathrm{trk}_k^{\mathrm{jet}_i}, \mathrm{vtx}_j)}{\sum_n \sum_l p_T(\mathrm{trk}_l^{\mathrm{jet}_i}, \mathrm{vtx}_n)} \tag{3.3}$$

where $p_T(\mathrm{trk}_k^{\mathrm{jet}_i}, \mathrm{vtx}_j)$ is the p_T of the track k originating from the vertex j and matched to the jet i. In the analysis jets with $p_T < 50$ and $|\eta| < 2.4$ are required to have JVF with respect to the primary vertex greater than 0.5.

---

[1] Presence of other pp interactions within the same bunch crossing.

3.2 Identification of b-jets

The identification of jets originating from the hadronization of b-quarks is called b-tagging and it is based on the peculiar properties of b-quark decays. The relatively long lifetime of b-hadrons, of the order of 1.5 ps, allows them to travel several millimeters before decaying. This has an effect on the structure of the jet, due the formation of Secondary Vertices (SV) separated from the primary vertex where the hard process occurred. For these reasons tracks within a b-jet tend to have larger impact parameter (d_0) with respect to the primary vertex than the ones originating from the primary vertex. A schema of the structure of a b-jet is presented in Figure 3.2.

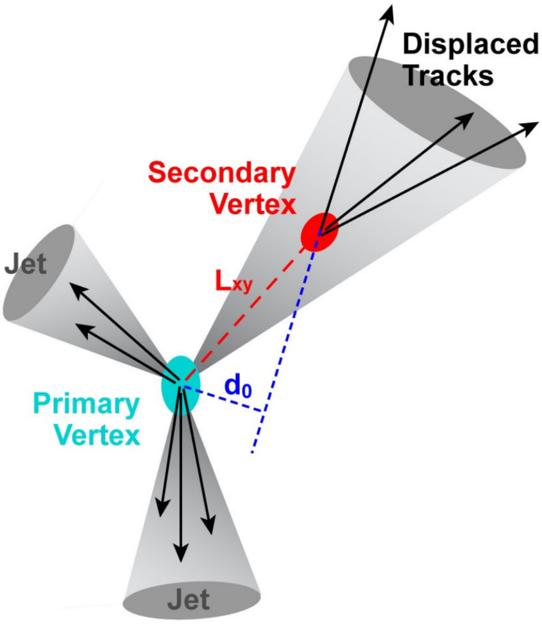


Figure 3.2: Schema of an events with two light jets, originating from the primary vertex, and a b-jet, originating from a secondary vertex. The impact parameter (d_0) of one of the tracks is also shown [89].

The ATLAS experiment uses a b-tagging algorithm called MV1 [90]. It is trained with b-jets as signal and light-flavour[2] jets as background, and computes a tag weight for each jet. MV1 combines into a neural network the information from three tagging algorithms:

- **IP3D**: The discriminating variable used is the two-dimensional distribution of transverse and longitudinal impact parameter significance of tracks within jets, d_0/σ_{d_0} and z_0/σ_{z_0} respectively. These are combined with a likelihood ratio technique in which inputs are compared to pre-defined distributions, obtained from Monte Carlo simulation, for both the b- and light jet hypotheses.

---

[2] All jets that are not originating from a b-quark are referred to as light-flavour jets if not specified otherwise.

- **SV1**: This algorithm exploits the properties of SVs, in particular it uses: the invariant mass of all tracks associated to the SV, the ratio of the sum of the energies of the tracks in the SV to the sum of the energies of all tracks in the jet, and the number of two-track SVs. These variables are combined using a likelihood ratio technique.

- **JetFitter**: It exploits the topology of subsequent b- and c-hadron decays inside the jet, using a Kalman filter to search for a common line connecting the primary vertex to b and c decay vertices. The discrimination between b-, c- and light jets is based on a likelihood using similar variables as in the SV1 tagging algorithm, and additional variables such as the flight length significance of the vertices.

3.2.1 Calibration of b-tagging algorithm

In order for the b-tagging algorithm to be used in physics analyses it is necessary to measure the difference of its performance in data and Monte Carlo simulation. The calibration procedure is performed measuring the efficiency of the algorithm to tag a jets originating from b-quarks, c-quarks and light flavour quarks. Calibration results are expressed in terms of scale factors to be applied to simulations to match the tagging rate observed in data.

The efficiency to tag jets originating from b-quarks is evaluated with the combinatorial likelihood method [91]. The calibration is performed in a data sample enriched in $t\bar{t}$ events. These are selected requesting the presence of two oppositely charged leptons in the final state, stemming from the leptonic decay of W boson from top quark decay. An unbinned likelihood is built in each of the four channels considered ($e\mu$ and $e^+e^- + \mu^+\mu^-$, 2 and 3 jets) and fitted to data. This approach takes into account different jet composition of each channel known from simulations, increasing the precision of the efficiency measurement.

The efficiency of the MV1 b-tagging algorithm to tag a jet originating from a c-quark is evaluated with the D^* method [90]. A data sample enriched in $D^{*\pm}$ is selected looking for the decay $D^* \to D_0 \to (K^-\pi^+)\,\pi^+$. The efficiency is measured via the comparison of the yield in data and MC in this sample before and after the b-tagging selection.

The efficiency to tag a jet originating from light-flavour partons (u, d, s quarks or gluons) is evaluated with the negative tag method [90]. Light-flavour jets are mistakenly tagged as b-jets mainly because of the finite resolution of the Inner Detector and the presence of tracks stemming from displaced vertices due to long-lived particles or material interactions. The negative tag rate is computed defining a negative version of the tagging algorithm which internally reverses the selections of the discriminant parameters.

In the $t\bar{t}H$ ($H \to b\bar{b}$) fully hadronic analysis the used MV1 b-tagging working point that is the one leading to a 60% efficiency in b-tagging selection, with a rejection factor of c-jets of ~ 10 and a rejection factor of light-jets of ~ 600. Systematic uncertainties on b-tagging calibrations are described in Section 6.9.2.

3.3 Electrons

Electron candidates [92] are reconstructed using the Electromagnetic calorimeter (EM) and the Inner Detector (ID), with the request that the energy deposition in the EM matches a reconstructed track. Only electrons with $|\eta| < 2.47$, limit of the ID, are used in the analysis. Electrons in the transition region between barrel and end-cap calorimeter ($1.37 < |\eta| < 1.52$) are also excluded.

Energy clusters in the EM are built using a sliding-window algorithm [79] that scans the calorimeter with a window of (3×5) towers[3] in the (η, ϕ) plane searching for local maxima. This search is triggered by the presence of a tower with energy above the threshold of 2.5 GeV. Clusters are required to be matched to a track originating from a vertex found in the beam interaction region. Additional selection are applied to increase rejection against fake electrons coming from hadron misidentification, photon conversion and semi-leptonic decay of heavy-hadrons. These selections are affecting both the calorimeter component of the electron candidate, with cuts on the energy leakage in the hadronic calorimeter and the ratio of the energy deposition in the different layer of the EM, and the track quality, with the request of minimum number of hits in the pixel detector, silicon detector and TRT. Candidates matched to photon conversion are also rejected.

To further increase the purity of the selection of real electrons additional criteria are applied. The candidate is required to be *isolated*, applying cuts on the energy deposit in a cone of $\Delta R = 0.2$ around the candidate. A similar cut is applied on the scalar sum of the tracks' p_T in a cone of $\Delta R = 0.3$. The track associated to the candidate is requested to have a longitudinal impact parameter with respect to the primary vertex smaller than 2 mm.

3.4 Muons

Reconstruction of muons [93] is performed using information from the muon spectrometer (MS) and from the ID. In ATLAS three types of algorithms are used to define the muon candidate:

- **Stand-Alone muons:** Trajectory of the muons is reconstructed only in the MS. The track is extrapolated from the MS to the interaction point to determine its properties, taking into account the estimation of the energy lost in the calorimeters.
- **Combined muons:** Tracks are reconstructed in the ID and in the MS independently. The candidate muon is obtained from the successful combination of the tracks in the two sub-detectors.
- **Segment-tagged muon:** Muon candidate is built from a track in the ID that is extrapolated to the MS and it is associated to at least one local track segment.

---

[3] Each calorimeter tower has a size of 0.025×0.025 in (η, ϕ).

Combined muons is the algorithm used in the analysis. A track-based Isolation requirements is applied to improve background rejection: the scalar sum of p_T of the tracks in a cone of variable size $\Delta R = 10\,\mathrm{GeV}/p_T^{\mu}$ around the muon must be less than 5% of the muon transverse momentum p_T^{μ}. Muons are also rejected if a jet is present within a ΔR distance of 0.4. The muon candidate is required to have a longitudinal impact parameter with respect to the primary vertex smaller than 2 mm.

3.5 Missing transverse momentum

Missing transverse momentum [94] is defined as the imbalance of the momentum in the plane perpendicular to the beam axis. This quantity is interpreted as one or more particles that escape the detector without interacting with it, like neutrinos. The missing transverse momentum is referred to with the symbol $\boldsymbol{E}_T^{\mathrm{miss}}$. Its magnitude, called *Missing transverse energy*, is referred to with the symbol E_T^{miss}.

In ATLAS the reconstruction of $\boldsymbol{E}_T^{\mathrm{miss}}$ is performed using information from the calorimeters and the muon spectrometer. Its two components are calculated as:

$$E_{x\ (y)}^{\mathrm{miss}} = E_{x\ (y)}^{\mathrm{miss,}\ calo} + E_{x\ (y)}^{\mathrm{miss,\ muon}} \tag{3.4}$$

The missing transverse energy and the ϕ coordinate of the missing momentum are evaluated as:

$$E_T^{\mathrm{miss}} = \sqrt{(E_x^{\mathrm{miss}})^2 + (E_y^{\mathrm{miss}})^2} \tag{3.5}$$

$$\phi^{\mathrm{miss}} = \arctan(E_y^{\mathrm{miss}}/E_x^{\mathrm{miss}}) \tag{3.6}$$

In the $t\bar{t}H$ $(H \to b\bar{b})$ fully hadronic analysis E_T^{miss} is not used.

3.6 Photons

Reconstruction of photons [95, 96] in ATLAS is very similar to the one of electrons, described in Section 3.3, using the same algorithms to reconstruct the energy clusters in the calorimeter. Photons are divided in two categories: *converted*, matched to at least one track originated from a vertex in the tracker volume, and *unconverted*, without any matching track. ID tracking system is able to reconstruct conversion vertices up to a radius of 80 cm in the transverse plane.

A series of rectangular cuts is applied to quantities related to the energy deposition in the different layer of the EM to distinguish photons from jets with a large electromagnetic component. Important role in background rejection is played by the first layer of the EM, segmented in the η direction (see Figure 2.8) in thin strips of width between 0.003 and 0.006, depending on η. These strips cover the calorimeter in the $|\eta| < 2.4$ region, excluding $1.4 < |\eta| < 1.5$. The selection is optimized independently for unconverted and converted photons to provide an identification efficiency of about 85% for photon candidates with transverse energy $E_T > 40$ GeV, and a background rejection factor of about 5000.

In the $t\bar{t}H$ ($H \rightarrow b\bar{b}$) fully hadronic analysis photons are not used.

Chapter 4

Trigger system

As described in Section 2.1, the LHC is able to deliver to the ATLAS experiment an instantaneous luminosity of 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, corresponding to $\sim$40 million bunch collisions per second. This huge amount of data is impossible to record with the available technology, which allows a recording rate of few hundreds events per second. Due to this bottle neck it is necessary to perform fast decisions in order to select events of physics interest to be recorded. This decision is taken by the trigger system.

4.1 Description of the trigger system

The ATLAS trigger system is made of three consecutive layers, a schema is shown in Figure 4.1: The Level 1 trigger (LVL1), the Level 2 trigger (LVL2) and the Event Filter (EF). The level 2 and event filter together form the High-Level Trigger (HLT). Each trigger level refines the decisions made at the previous level by applying additional selection criteria. The trigger selections are applied on physics objects[1] built with algorithms that have less resolution than the off-line ones[2]. This is due to the time constraint that does not allow to use computationally complex algorithms or the part of information from sub-detectors which is long to read or reconstruct. For an event to be recorded it has to pass selection criteria in each of the three trigger levels, these are connected in the so called *trigger chain*.

Since the work presented in this thesis makes use of jet-triggers, great attention is given to the description of such class of triggers, for a more complete information on triggers targeting other physics objects than jets the reader is directed to the Section 8 of [63].

4.1.1 Level 1

The LVL1 trigger is designed to select events that contain muons, electrons, photons, jets, τ-leptons decaying into hadrons, large missing transverse energy (E_T^{miss}) and

---

[1] Reconstructed particle, like for example: electrons, muons, photons or more complex objects like particle jets and missing transverse energy.

[2] For the definition of offline physics objects see Chapter 3.

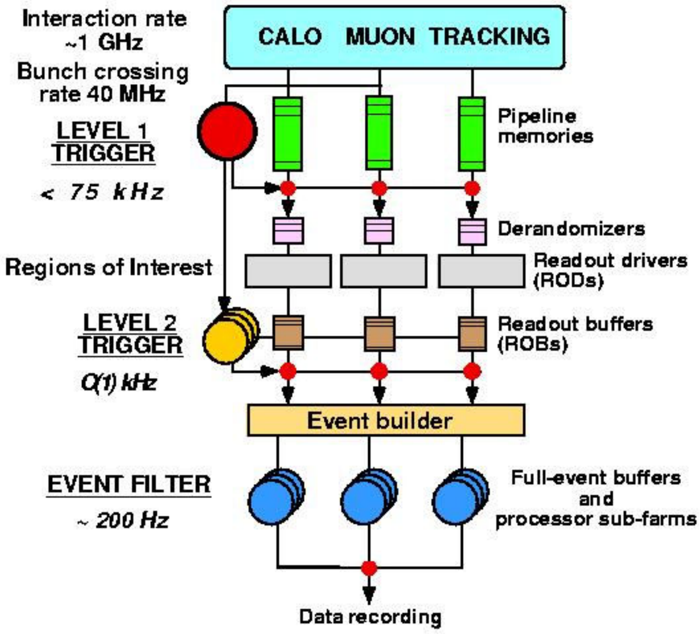


Figure 4.1: Schema representing the three level trigger of the ATLAS experiment in Run 1. The flow of the selection process is top to bottom of the diagram.

large total transverse energy[3]. It is a hardware based trigger and built on custom electronics. The maximum LVL1 accept rate which the ATLAS detector readout systems can handle is 75 kHz, and the LVL1 decision must reach the front-end electronics within 2.5 μs after the bunch-crossing with which it is associated. This time requirement has a direct impact on the amount of information that can be used for the decision. Only few sub-detectors with a reduced granularity with respect to offline can be used: the Resistive Plate Chambers and Thin-Gap Chambers for high-p_T muons and electromagnetic, hadronic and forward calorimeters for electromagnetic clusters, jets, τ-leptons, E_T^{miss}, and large total transverse energy. The ID is not used since the reading of all the channels and the track reconstruction do not fulfill the time requirement.

The LVL1 trigger takes its decision based only on multiplicity of trigger objects, no geometrical information is used. At this level a trigger object is simply a threshold that is passed in one or more than one sub-detector. During the processing of the trigger decision the information for all detector channels is retained in pipeline memories. The LVL1 trigger identifies the regions of the detector, in (η, ϕ) plane, that have interesting features. These are called Regions-of-Interest (RoI's) and are used to seed the HLT.

---

[3] Scalar sum of the transverse energy of the objects reconstructed in the event.

Calorimeter trigger

The calorimeter trigger is designed to handle $\sim$ 7000 analogue Trigger Towers (TT) of granularity 0.1×0.1 in $(\Delta\eta \times \Delta\phi)$, coarser with respect to its full granularity of: 0.025×0.0245 in $(\Delta\eta \times \Delta\phi)$. More details on the calorimeter are given in Section 2.2.4. The η range taken into account for the trigger decision depends on the physics object considered: electron, photon and τ triggers use $|\eta| < 2.5$, which is the fiducial limit for the ID and the electromagnetic calorimeter. Jet trigger uses the caloremetr up to $|\eta| < 3.2$. Total transverse energy and E_T^{miss} triggers extend to $|\eta| < 4.9$ using also the forward calorimeters. The jet trigger RoI is a 2×2 matrix of Jet Elements (JE), a JE corresponds to a 2×2 combination of TT, that is required to be a local maximum of energy deposition in the calorimeter with respect to its neighbors. The isolation requirement is applied to avoid multiple-counting of jet candidates. The jet algorithm identifies E_T sums within overlapping square windows of side 0.4, 0.6, or 0.8 in (η, ϕ) plane, consisting of 2×2 , 3×3, 4×4 JE respectively, as shown in Figure 4.2.

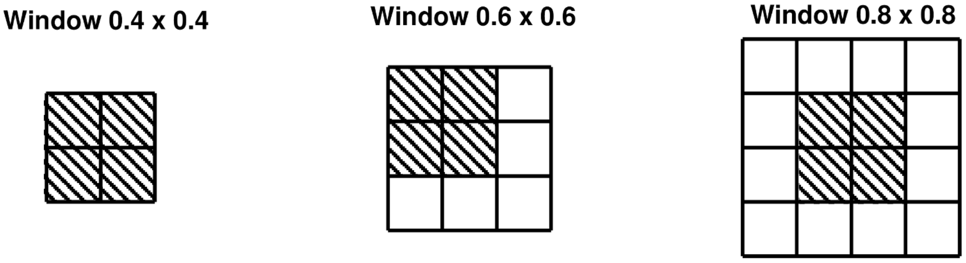


Figure 4.2: Schema of the different square window sizes for the jet algorithm at LVL1 trigger. Shaded area represents the RoI. The central plot shows one of the four possible configurations of the window 0.6×0.6 with respect to the RoI. In the case of window 0.8×0.8, the RoI is required to be at the center of the window to avoid the presence of two jets per window [63].

4.1.2 Level 2

The Level 2 trigger is software based, and runs on dedicated computer farms. The RoI constructed by the LVL1 trigger are used as input to the LVL2 trigger. This second trigger level uses the full granularity and full precision of the detector inside the RoIs, including information from the ID. This allows to save time with respect to reconstructing the full event, since the data constituting an RoI is $\sim 2\%$ of the data of the full event. The LVL2 selections are designed to reduce the trigger rate to approximately 3.5 kHz. Time constraints are less strict than LVL1 with an event processing time of about 40 ms, averaged over all events. LVL2 bases the event selection on specialized algorithms optimized for speed, with lower resolution than offline algorithms.

4.1.3 Event filter

The final stage of the event selection is carried out by the Event Filter (EF), which reduces the event rate to roughly 200 Hz. EF uses more complex algorithms, with respect to LVL2, very close to those used in the offline reconstruction software, within an average event processing time of the order of four seconds. This trigger level uses the information from the full detector at full granularity.

4.1.4 Trigger menus and prescaled triggers

In each run of the LHC the triggers that are used to collect data are organized in lists called "trigger menus". Each menu specifies the thresholds and the selection criteria at each of the three trigger levels for each trigger chain. It is compiled combining different physics signatures to match the needs of physics analysis within ATLAS.

An estimation of the background rejection capabilities at each selection stage and for each signature is one of the fundamental ingredients of the preparation of the menus. Not only the triggers that are used by physics analyses are selected to enter in the menu, but also supporting triggers that are needed for background studies or for the calibrations of other triggers or sub-detectors. These type of triggers in general have looser selection criteria or lower thresholds or request lower multiplicity of objects. This implies that a bigger amount of events are going to fulfill the trigger requirement saturating the trigger rates. To avoid this effect the supporting triggers are *prescaled*, meaning that these triggers are active only for a fraction of the collisions.

4.1.5 Jet and multi-jet trigger algorithms

The online algorithms used by the jet trigger slice have evolved during the Run 1 operation of the ATLAS detector. For $\sqrt{s} = 7$ TeV run algorithms at LVL2, trigger were seeded by LVL1 via RoIs, a simple cone-like jet algorithm with a radius of $R = 0.4$ was run at LVL2 exclusively within the RoIs. A different strategy, called L1.5, has been adopted and implemented in LVL2 trigger during the $\sqrt{s} = 8$ TeV run. The idea behind L1.5 triggers is to use data produced from the LVL1 trigger to access the full detector. Accessing LVL1 data allows to save time with respect to reading the calorimeter information at full granularity, moreover it allows to run more complex jet algorithms than the standard LVL1 or LVL2, as for example the anti-k_t algorithm via the FastJet package [83, 84, 85].

4.2 Trigger efficiency and scale factors determination

A detailed description of how trigger efficiencies and scale Factors (SF) are determined for the full hadronic $t\bar{t}H$ ($H \to b\bar{b}$) analysis is the subject of this section. A parametrization is derived and validated for trigger efficiencies in data and in

MC. Different closure tests to validate the method are shown in the following. The determination of a data-MC SF is presented. The section ends with closure tests performed on MC samples, used to assess the systematic uncertainty for this method.

4.2.1 Introduction

The non-perfect correlation between online and offline algorithms results in a loss of efficiency when the same requirements are imposed to objects reconstructed online and offline. Given that online objects have worse resolution than offline ones, behavior in regions close to the trigger activation threshold need to be understood to widen the use of these triggers in physics analysis. Studies are performed on the trigger efficiency, defined as the ratio of the events that satisfy both offline and online requirements over those passing online ones only. The trigger efficiency curves when described with respect to an offline quantity, are also called trigger "turn-on" curves.

As an example, Figure 4.3 shows the determined turn-on curves for some LVL1 and LVL2 jet triggers as function of offline jet E_T. It is important to notice that the turn-on curves for LVL1 are less steep than the LVL2 due to larger jet energy resolution. The *plateau* of the LVL1 trigger, defined as the region where the trigger has a flat efficiency (here $> 99\%$), that requires one object with $E_T > 15$ GeV (`L1_j15`) is reached only for offline jets with $E_T > 55$ GeV. Residual differences between data and MC are present, mainly away from the plateau region. A *scale factor* (SF), that takes the difference into account, is applied to compensate. Alternatively, offline selections are used to remove the region where there is disagreement.

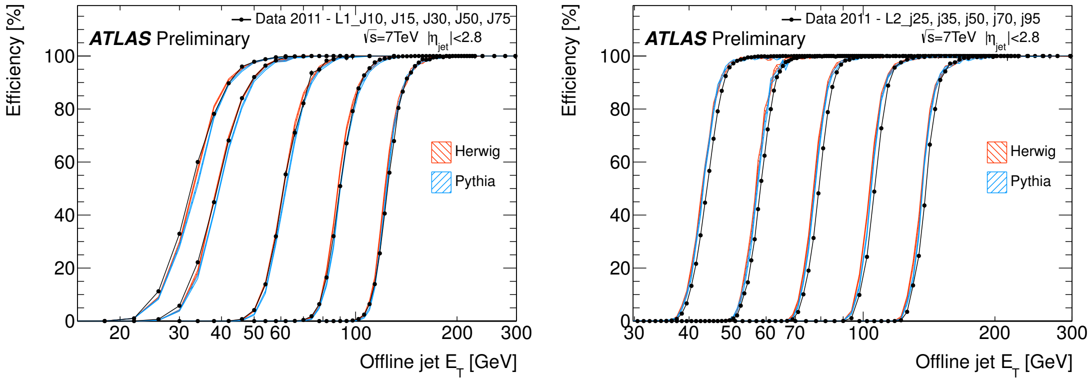


Figure 4.3: Efficiencies of LVL1 (*left*) and LVL2 (*right*) triggers chains in data and Pythia and Herwig MC events. For data, the efficiency is computed with respect to events taken by an independent trigger, 100% efficient in the relevant region. The turn-on curves for LVL1 are less steep than the LVL2 due to poor jet energy resolution. It can be noted that efficiency $\sim 100\%$ is reached for energies larger than the trigger thresholds [97] .

4.2.2 Jet trigger efficiency

The trigger used by fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$) analysis is a multi-jet trigger requiring four objects at LVL1 with a requirement on transverse energy of the LVL1 8×8 TT to be higher than 15 GeV, followed by 5 objects at LVL2 with a transverse energy higher than 15 GeV and finally 5 objects at EF level having transverse energy higher than 55 GeV.

Such a trigger is used in other analyses [98] in combination with harsh requirements on the offline reconstructed jet quantities in order to be away from possible sources of inefficiencies, which if badly modeled in MC simulations, could lead to incorrect evaluation of the acceptance. These requirements generally involve asking for:

- *N high-p_T* offline jets, where N is the multiplicity of the multi-jet trigger used, and for *high-p_T* it is intended that the p_T of the jets is large enough to be away from the trigger p_T-dependent turn-on region. Differences in the turn-on region are due to mis-modeling of the trigger jet resolution.

- *Isolation criteria* for offline jets. The isolation requirement allows to avoid dealing with the case where inefficiencies come from the modeling of the overlapping between TT and offline jets.

On the other hand, demanding offline jet at high p_T and vetoing event where non-isolated jets[4] are present could have the effect of cutting away up to half of the collected statistics of triggered signal events in analysis such as the fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$).

Therefore a procedure is needed to assess trigger efficiency between data and MC in the most accurate way in order to correct the possible mis-modeling in MC by using a SF, and assigning a systematic uncertainty to it. To do so an assumption that the trigger behavior of any sample of events can be inferred by the properties of all offline reconstructed objects is made, which individually contribute to the global efficiency. By taking into consideration the properties of all relevant reconstructed objects in the event, this approach derives a SF that can be applied universally to all MC samples.

Data and MC samples

For this study di-jet MC samples are used to compare directly with data and fully hadronic $t\bar{t}H$ MC events to investigate the sample dependency on the trigger efficiencies. Di-jet MC is generated using PYTHIA8 [99] in different slices of leading jet p_T as shown in Table 4.1. For $t\bar{t}H$ the same sample has been used as in the fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$) analysis, described in Section 6.4.1.

For data two different trigger chains are used:

---

[4] Two jets separated by a distance $\Delta R < 0.6$.

p_T range [GeV]	$\sigma \cdot \epsilon_f$ [nb]
0-20	$7.28 \cdot 10^7$
20-80	$3.60 \cdot 10^6$
80-200	$2.64 \cdot 10^4$
200-500	$5.44 \cdot 10^2$
500-1000	6.44
1000-1500	$3.97 \cdot 10^{-2}$
1500-2000	$4.16 \cdot 10^{-4}$
2000+	$4.06 \cdot 10^{-5}$

Table 4.1: List of Pythia8 di-jet MC samples and their corresponding cross section times filter efficiencies in each truth leading jet p_T slice. Filter efficiency is the fraction of events that pass from the general sample into the final simulated sample after the application of a generator-level filter, this avoid storing events that are not going to be used in the analysis.

- `EF_rd0_filled_NoAlg`: dedicated trigger item for efficiency determination. Trigger chain starts with a random trigger at LVL1. Subsequent levels do not apply further selection criteria. Jet's RoI are reconstructed at each trigger level.

- `EF_j110_a4tchad`: Jet trigger requiring at least one object at LVL1, with $E_T > 50$ GeV for 8×8 TT, at least one object at LVL2 with $E_T > 105$ GeV and at least one object at EF with $E_T > 110$ GeV. In events triggered by this high E_T single jet-trigger, in order to remove the trigger bias, all offline jets inside the hemisphere, defined as $\Delta\phi < \pi/2$, surrounding the object with highest E_T at EF level in the event are removed. Events with more than one jet with $E_T > 110$ GeV at EF level are rejected. These selections ensure not to pick up the jet that triggered the event. Notice that these jets are fully efficient for the trigger under study, so not having jets with high p_T to estimate the per jet trigger probability does not matter at the end.

For offline jets, the same selection criteria in terms of quality requirement is demanded as in the final analysis, as described in Chapter 3.

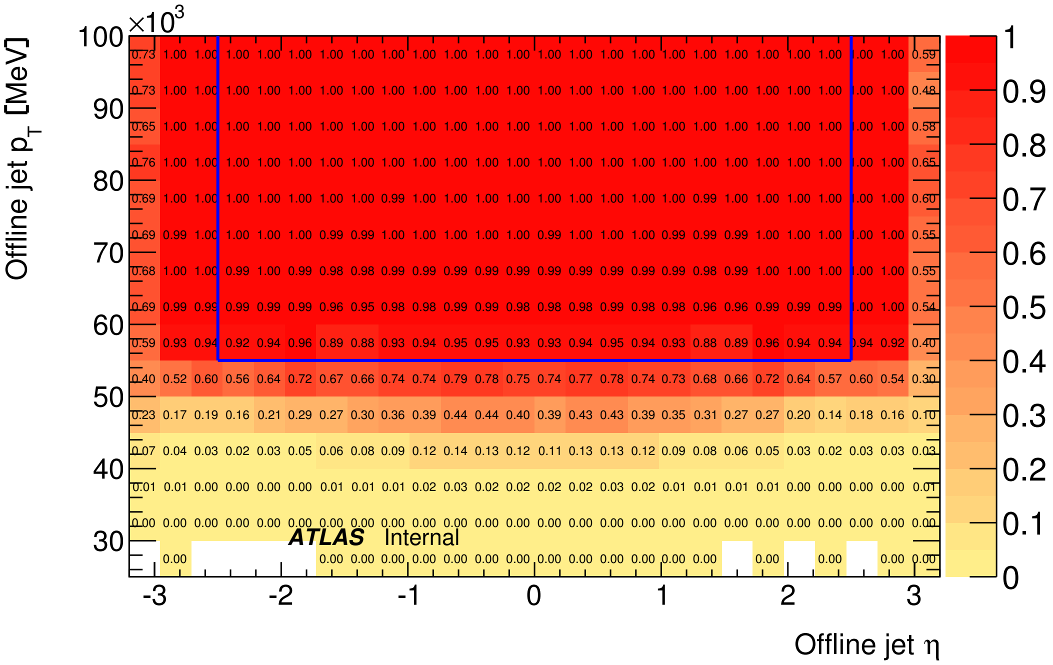
4.2.3 Single jet trigger efficiencies

The single jet trigger efficiency for the jet trigger chain considered is defined as the probability for an offline jet to be associated to a `L1_J15` → `L2_j15` → `EF_j55` chain. The association is defined in terms of finding a LVL1 8×8 TT of 15 GeV, a LVL2 object with $E_T > 15$ GeV and an EF object with $E_T > 55$ GeV in a ΔR cone of 0.4 radius. Events that have two offline jets reconstructed within a $\Delta R < 0.6$ are vetoed to focus only on estimating the effect of single jet difference and not cumulative behavior which can arise from close-by effects.

With this definition the object matching efficiency and the trigger efficiency are taken into account at the same time. The single offline jet trigger efficiency is then defined as:

$$\epsilon_{\text{trig}}(p_T, \eta) = \frac{N_{trigger}(p_T, \eta)}{N(p_T, \eta)} \tag{4.1}$$

Where $N(p_T, \eta)$ is the number of isolated offline jets in each (p_T, η) bin, and $N_{trigger}$ is the sub-sample of offline jets that are matched to a complete trigger chain: the offline jet is matched independently to a `L1_J15`, `L2_j15` and `EF_j55` trigger object. Figure 4.4 shows the efficiency for Monte Carlo simulated fully hadronic $t\bar{t}H$ events with this definition.



0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.59
0.73	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.48
0.65	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.58
0.76	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.65
0.69	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.60
0.69	0.99	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.55
0.68	1.00	1.00	0.99	1.00	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	1.00	1.00	1.00	1.00	1.00	0.55
0.69	0.99	0.99	0.99	0.99	0.99	0.96	0.95	0.98	0.98	0.99	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.96	0.96	0.99	0.99	0.99	1.00	1.00	0.54
0.59	0.93	0.94	0.92	0.94	0.96	0.89	0.88	0.93	0.94	0.95	0.95	0.93	0.93	0.94	0.95	0.94	0.93	0.88	0.89	0.96	0.94	0.94	0.94	0.92	0.40
0.40	0.52	0.60	0.56	0.64	0.72	0.67	0.66	0.74	0.74	0.79	0.78	0.75	0.74	0.77	0.78	0.74	0.73	0.68	0.66	0.72	0.64	0.57	0.60	0.54	0.30
0.23	0.17	0.19	0.16	0.21	0.29	0.27	0.30	0.36	0.39	0.44	0.44	0.40	0.39	0.43	0.43	0.39	0.35	0.31	0.27	0.27	0.20	0.14	0.18	0.16	0.10
0.07	0.04	0.03	0.02	0.03	0.05	0.06	0.08	0.09	0.12	0.14	0.13	0.12	0.11	0.13	0.13	0.12	0.09	0.08	0.06	0.05	0.03	0.02	0.03	0.03	0.03
0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.02	0.02	0.02	0.02	0.03	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.00					0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		0.00		0.00		0.00	0.00

Figure 4.4: Single jet trigger efficiency for the chain `L1_J15` → `L2_j15` → `EF_j55` in MC simulated $t\bar{t}H$ events as a function of the offline jet p_T and η. Statistical uncertainty is of the order of few permille. Blue line represent the selection of jets with $p_T > 55$ GeV and $|\eta| < 2.5$. This is the selection applied to the five leading jets in the fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$.

4.2.4 Trigger efficiency formalism

After estimating the trigger parametrization in terms of single jet properties, the formalism necessary to predict the number of events passing the multi-jet triggers needs to be defined. The two-dimensional single jet trigger efficiency, as defined in Equation 4.1, is interpreted as the probability for a single jet to fire a trigger chain. In the case under study this corresponds to a sequence of accepted LVL1, LVL2 and EF objects in the same RoI. The overall number of events passing the multi-jet trigger is estimated weighting each event with the probability of the event to pass the multi-jet trigger. This latter is obtained by combining the probability $\epsilon_{\text{trig}}(p_T, \eta)$, as defined in Equation 4.1, of each jet in the event. In the case under study this corresponds to the probability for an event to fire at least 5 trigger chain:

$$P_{\geq 5} = 1 - (P_{=0} + P_{=1} + P_{=2} + P_{=3} + P_{=4}), \tag{4.2}$$

where $P_{=0}$, $P_{=1}$, $P_{=2}$, $P_{=3}$ and $P_{=4}$ are the probability that the events has exactly 0, 1, 2, 3, or 4 trigger chains fired defined as:

$$P_{=0} = \prod_{i=1} (1 - \epsilon_i), \tag{4.3}$$

$$P_{=1} = \sum_{j=1} \left(\epsilon_j \prod_{i \neq j} (1 - \epsilon_i) \right), \tag{4.4}$$

$$P_{=2} = \sum_{j=1} \sum_{l=j+1} \left(\epsilon_j \epsilon_l \prod_{i \neq j,l} (1 - \epsilon_i) \right), \tag{4.5}$$

$$P_{=3} = \sum_{j=1} \sum_{l=j+1} \sum_{m=l+1} \left(\epsilon_j \epsilon_l \epsilon_m \prod_{i \neq j,l,m} (1 - \epsilon_i) \right), \tag{4.6}$$

and

$$P_{=4} = \sum_{j=1} \sum_{l=j+1} \sum_{m=l+1} \sum_{n=m+1} \left(\epsilon_j \epsilon_l \epsilon_m \epsilon_n \prod_{i \neq j,l,m,n} (1 - \epsilon_i) \right). \tag{4.7}$$

Where ϵ stands for $\epsilon_{\text{trig}}(p_T, \eta)$ and indeces i, j, k, l, m, n run over the number of jets in the event.

4.2.5 Validation of isolated jet turn on curves

A first test consists of checking normalization predicted in several jet and b-jet multiplicity bins. The selections used are the same as in the fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$) analysis, described in Section 6.6. Since this is an internal consistency check of the method, the same, in this case the fully hadronic $t\bar{t}H$, sample is used, splitted in two halves, one used to estimate the trigger efficiency and the second half to validate the trigger prediction. The results are shown in Table 4.2, the uncertainties shown are statistical only and consist of the sum in quadrature of the uncertainty coming from the sample used to derive the prediction and the uncertainty derived from the trigger efficiency parametrization. This latter has been derived by averaging the effect of shifting the efficiency 1 σ up and 1 σ down coherently in each p_T and η bins. Different distributions of topological variables are shown in Figure 4.5 with a good agreement between prediction and `EF_5j55_a4tchad_L2FS` trigger requirement.

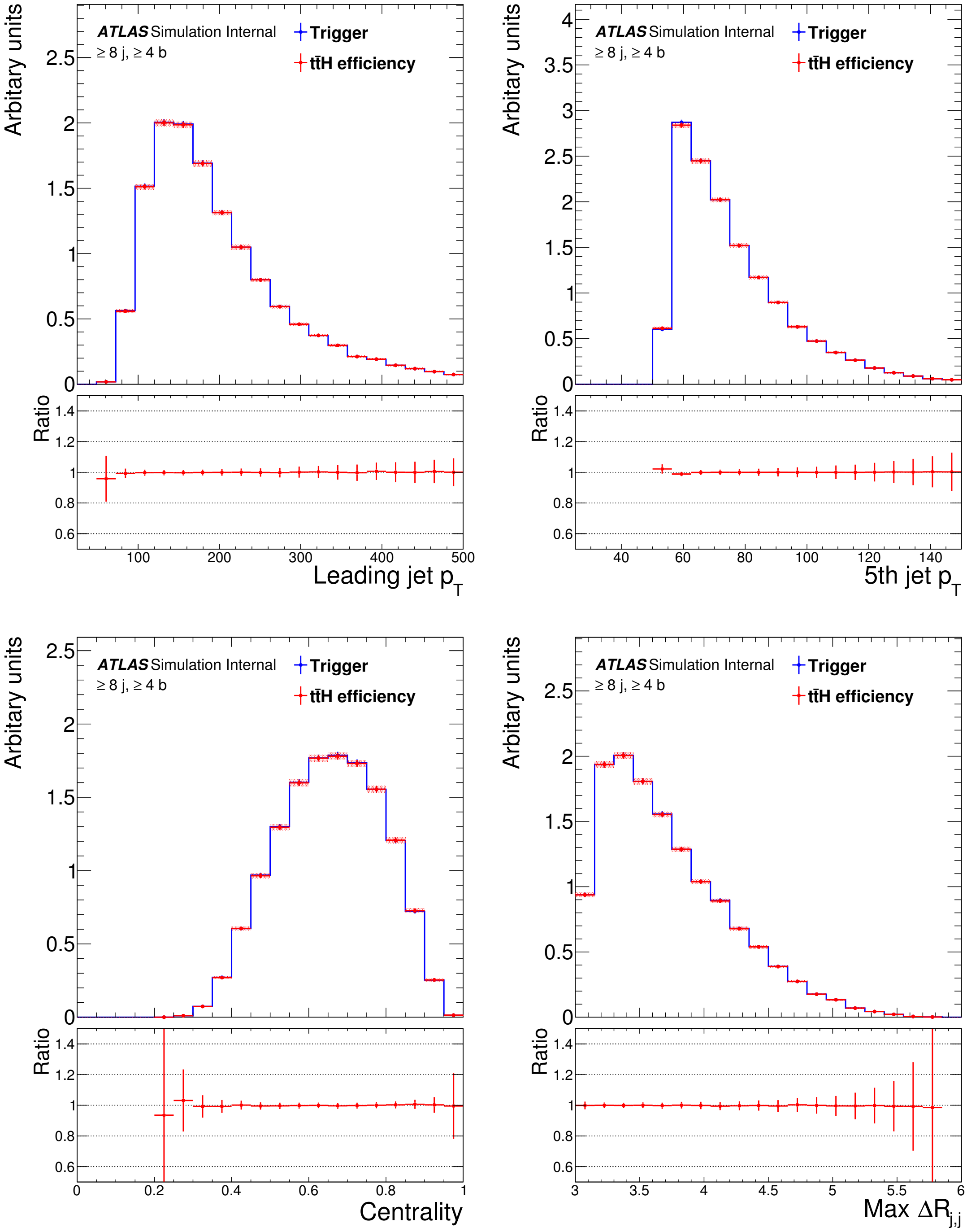


Figure 4.5: Leading jet p_T, 5th leading jet p_T, Centrality (i.e. Scalar sum of the p_T of all jets divided by the sum of E of all jets) and maximum of ΔR between any jet distributions in events passing the offline selection after `EF_5j55_a4tchad_L2FS` trigger requirement (blue histogram), and prediction by applying event weights (red histogram). Bottom panel shows the ratio between the prediction and the requirement of the trigger. Plots correspond to jet multiplicity ≥ 8 and b-tagged jet multiplicity ≥ 4.

Jet multiplicity	b-tagged jet multiplicity	`EF_5j55`	Prediction
6	$== 3$	8.97 ± 0.06	8.97 ± 0.06
	≥ 4	2.05 ± 0.02	2.04 ± 0.02
7	$== 3$	16.97 ± 0.08	16.92 ± 0.09
	≥ 4	4.95 ± 0.04	4.93 ± 0.04
≥ 8	$== 3$	36.77 ± 0.13	36.70 ± 0.13
	≥ 4	13.14 ± 0.06	13.12 ± 0.06

Table 4.2: Prediction using fully hadronic $t\bar{t}H$ derived per jet trigger efficiencies compared to the number of events passing the trigger `EF_5j55_a4tchad_L2FS`. Uncertainties quoted here are statistical only. The number of events are normalized to the integrated luminosity collected by the `EF_5j55_a4tchad_L2FS` trigger.

4.2.6 Data MC comparison

To perform a trustable comparison of the trigger behavior in data and MC it is necessary to verify that the two samples are similar with respect to fundamental kinematic variables. This is to exclude that eventual efficiency differences are originating from kinematic characteristics rather than trigger simulation in MC. Figure 4.6 shows distributions of the leading jet p_T, the second leading jet p_T and the scalar sum of the p_T of all jets in the event (H_T) in data passing `EF_rd0_filled_NoAlg` and Pythia di-jet MC for events with at least one jet with p_T >25 GeV and $|\eta| < 3.2$. The agreement is not expected to be perfect, since Pythia di-jet MC is not designed to describe this type of events. Nevertheless distributions are compatible, allowing the following steps of the study.

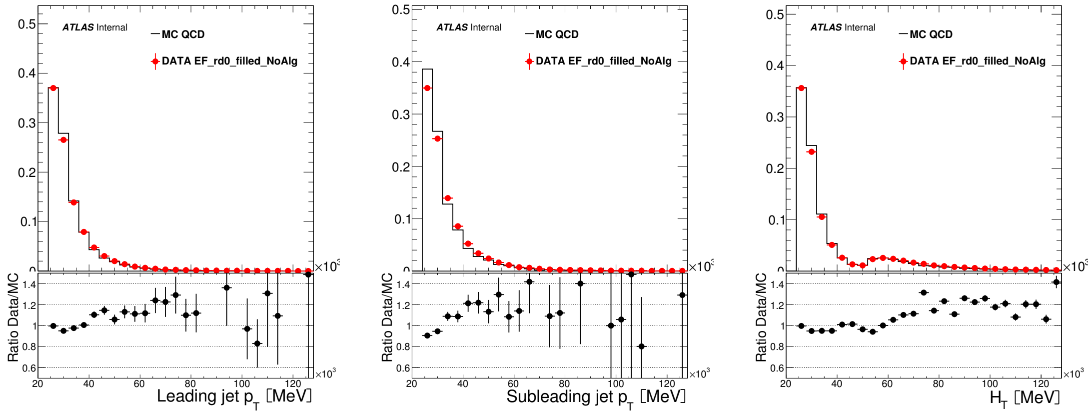


Figure 4.6: Distributions of leading jet p_T, second leading jet p_T and H_T in data events passing `EF_rd0_filled_NoAlg` and in Pythia di-jet MC. The agreement is not expected to be perfect, since Pythia di-jet MC is not designed to describe this type of events. Nevertheless distributions are compatible.

Single jet efficiencies are shown for the Pythia di-jet sample, in Figure 4.7 and in randomly triggered data, in Figure 4.8. The effect of using efficiencies derived from fully hadronic $t\bar{t}H$ MC events, di-jet MC and data passing `EF_rd0_filled_NoAlg` to perform a prediction of the trigger selection has been studied in the fully hadronic $t\bar{t}H$ sample. Results can be seen in Table 4.3 where there is a good agreement between the predictions made with the efficiencies evaluated in MC events, while prediction with efficiency evaluated in data is lower than the other two. The latter prediction presents the largest statistical uncertainties due to the small statistics collected by `EF_rd0_filled_NoAlg` trigger. Figure 4.9 shows the prediction of the shapes of 5^{th} jet p_T and of H_T. For events where the 5^{th} jet has $p_T < 65$ GeV a disagreement is visible between the predictions made with efficiency from data and efficiency from fully hadronic $t\bar{t}H$ MC sample. A slight difference is also seen, in the same region, between predictions made with efficiency from Pythia di-jet MC and efficiency from fully hadronic $t\bar{t}H$ MC sample. The same differences are visible also in the low part of the H_T spectrum.

For all the variables the prediction made using the efficiency evaluated in data is systematically lower than other predictions. This is an effect of the large statistical uncertainty of the data efficiency map. Since in the plateau region efficiency is ~1, the statistical error has the effect of reducing it to values < 1, this is clearly visible in Figure 4.8 where several bins in the plateau region have value < 1.

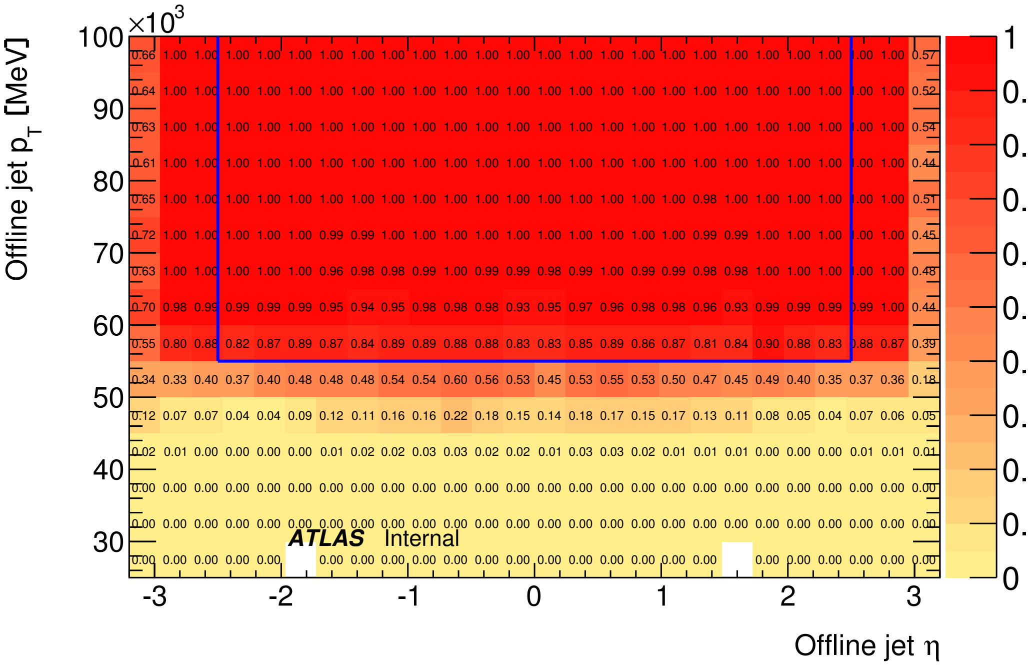


Figure 4.7: Single jet trigger efficiency for the chain `L1_J15` → `L2_j15` → `EF_j55` in Pythia di-jet MC events as a function of the offline jet p_T and η. Blue line represent the selection of jets with $p_T > 55$ GeV and $|\eta| < 2.5$. This is the selection applied to the five leading jets in the fully hadronic $t\bar{t}H$ $(H \rightarrow b\bar{b})$.

To reduce this statistical uncertainty, events passing `EF_j110_a4tchad` are used

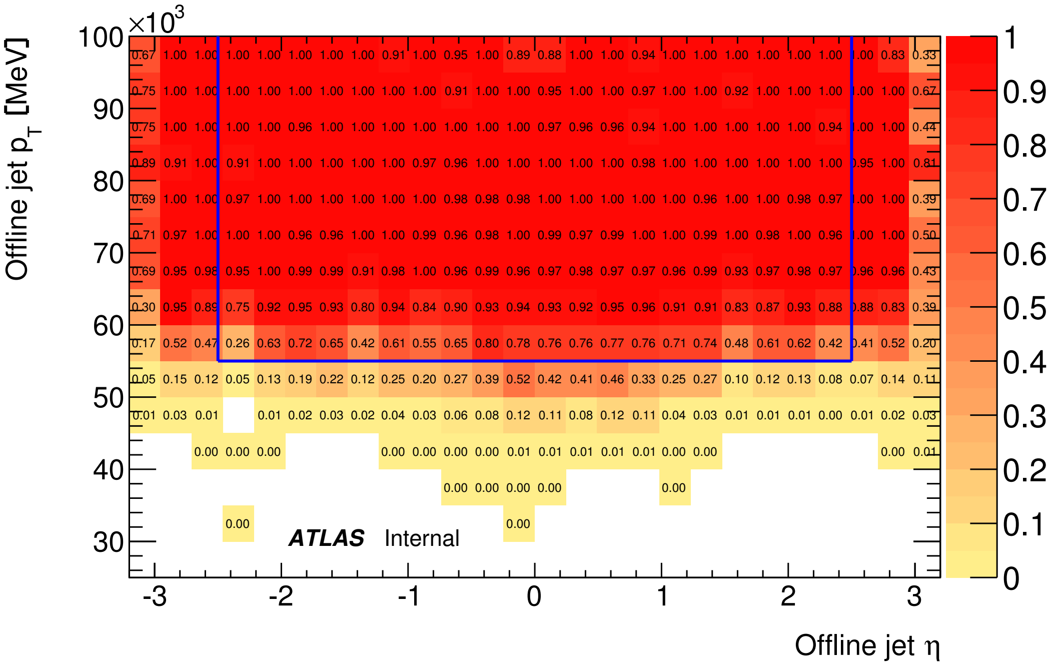


Figure 4.8: Single jet trigger efficiency for the chain `L1_J15` → `L2_j15` → `EF_j55` in data events passing `EF_rd0_filled_NoAlg` as a function of the offline jet p_T and η. Blue line represent the selection of jets with $p_T > 55$ GeV and $|\eta| < 2.5$. This is the selection applied to the five leading jets in the fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$).

Jet multiplicity	b-tagged jet multiplicity	Prediction$_{t\bar{t}H}$	Prediction$_{\text{PYTHIA}}$	Prediction$_{data\ rndm}$
6	$== 3$	8.97 ± 0.06	8.76 ± 0.08	7.48 ± 0.28
	≥ 4	2.04 ± 0.02	2.00 ± 0.02	1.71 ± 0.07
7	$== 3$	16.92 ± 0.09	16.58 ± 0.12	14.58 ± 0.45
	≥ 4	4.93 ± 0.04	4.84 ± 0.04	4.27 ± 0.13
≥ 8	$== 3$	36.70 ± 0.13	36.20 ± 0.17	33.29 ± 0.67
	≥ 4	13.12 ± 0.06	12.95 ± 0.07	11.98 ± 0.23

Table 4.3: Prediction of fully hadronic $t\bar{t}H$ events passing `EF_5j55_a4tchad_L2FS` trigger requirement, estimated using single jet trigger efficiencies derived respectively from $t\bar{t}H$ MC sample itself (Prediction$_{t\bar{t}H}$) on PYTHIA MC di-jet events (Prediction$_{\text{PYTHIA}}$) and on randomly triggered data (Prediction$_{data\ rndm}$). Uncertainties quoted here are statistical only. The number of events are normalized to the integrated luminosity collected by the `EF_5j55_a4tchad_L2FS` trigger.

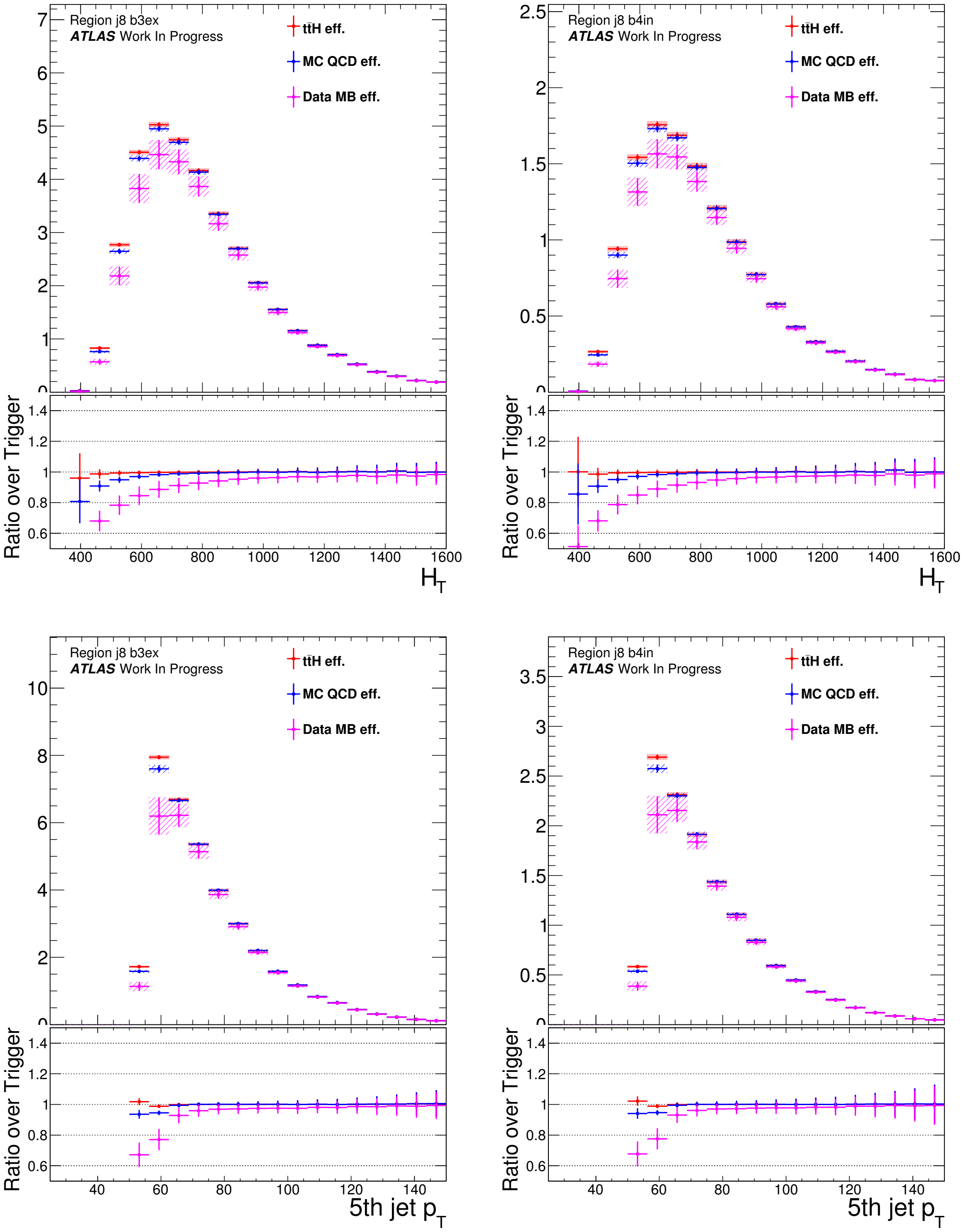


Figure 4.9: Distributions of predicted H_T and 5^{th} leading jet p_T in fully hadronic $t\bar{t}H$ MC events passing the offline selection. Different predictions are derived applying event weights using efficiencies estimated from $t\bar{t}H$ sample (red points), Pythia MC di-jet (blue points) and random triggered data (purple points). Bottom panel shows the ratio between the prediction and the requirement of the the `EF_5j55_a4tchad_L2FS` trigger. Plots correspond to $\geq$ 8 jet multiplicity bin and exactly 3, *left*, and $\geq$ 4, *right*, b-tagged jet multiplicity bins. A systematic undershooting is observed in the prediction made with efficiency from data in all the variables range. This is an effect of the large statistical uncertainty of the data efficiency map. Since in the plateau region efficiency is ∼1, the statistical error has the effect of reducing it to values < 1.

to derive the data efficiency map. To do so it is necessary to make sure that the trigger bias is correctly taken into account, since in this dataset, jets are actually used to decide whether an event is recorded or not. For this reason single jet efficiencies are evaluated using only jets that are not used by the trigger to take decisions. This is done, as explained in Section 4.2.2, by looking at jets in the opposite hemisphere from the only object having E_T larger than 110 GeV at EF level. The assumption that the single jet trigger efficiency is correctly estimated in this way is tested on Pythia di-jet MC events where it is possible to compare the trigger efficiency evaluated with and without `EF_j110_a4tchad` trigger requirement. The ratio of these two efficiencies is shown in Figures 4.10. This ratio is compatible with unity within the error.

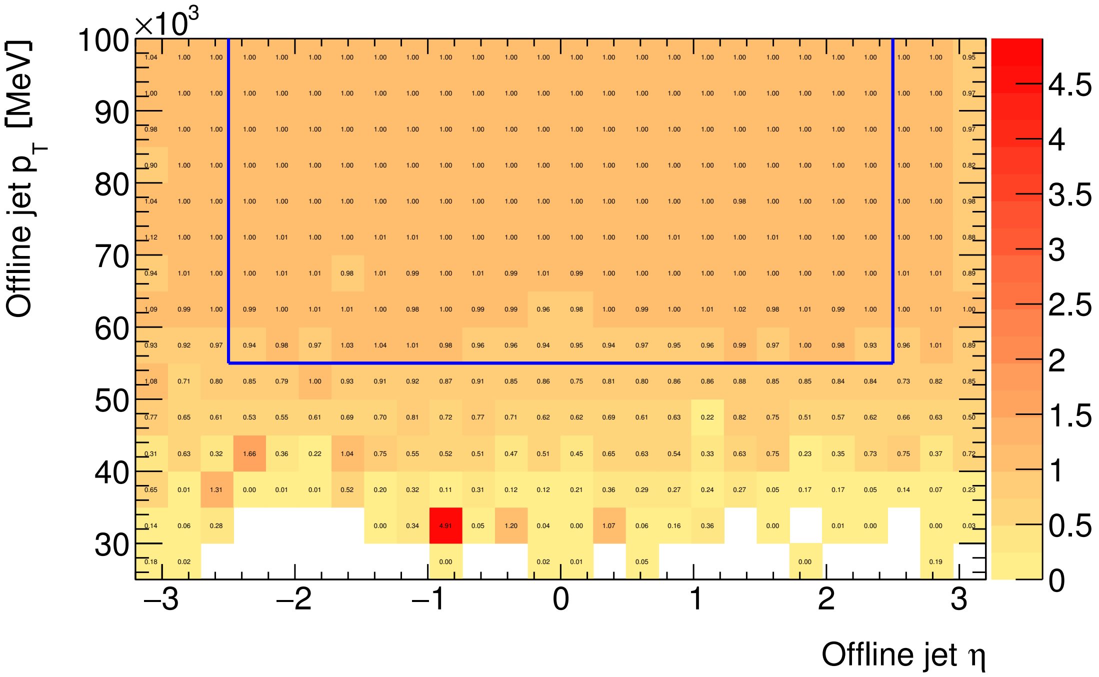


Figure 4.10: Ratio of single jet trigger efficiencies derived in Pythia di-jet MC events selected with no trigger and by the selection of `EF_j110_a4tchad` trigger, for the chain `L1_J15` → `L2_j15` → `EF_j55` as a function of the offline jet p_T and η. Blue line represent the selection of jets with $p_T > 55$ GeV and $|\eta| < 2.5$. This is the selection applied to the five leading jets in the fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$.

Since no bias is observed in the method, it is possible to derive an `EF_j110_a4tchad` based efficiency map in data, as shown in Figure 4.11. The comparison of the propagation of the single jet trigger efficiencies of the normalization of predicted fully hadronic events is shown in Table 4.4 and for the effect on the shape in Figure 4.12. The undershooting of the data prediction has disappeared and also the disagreement in the low p_T region has reduced. The residual shape difference is due to the effect of the different description of the trigger behavior in data and MC.

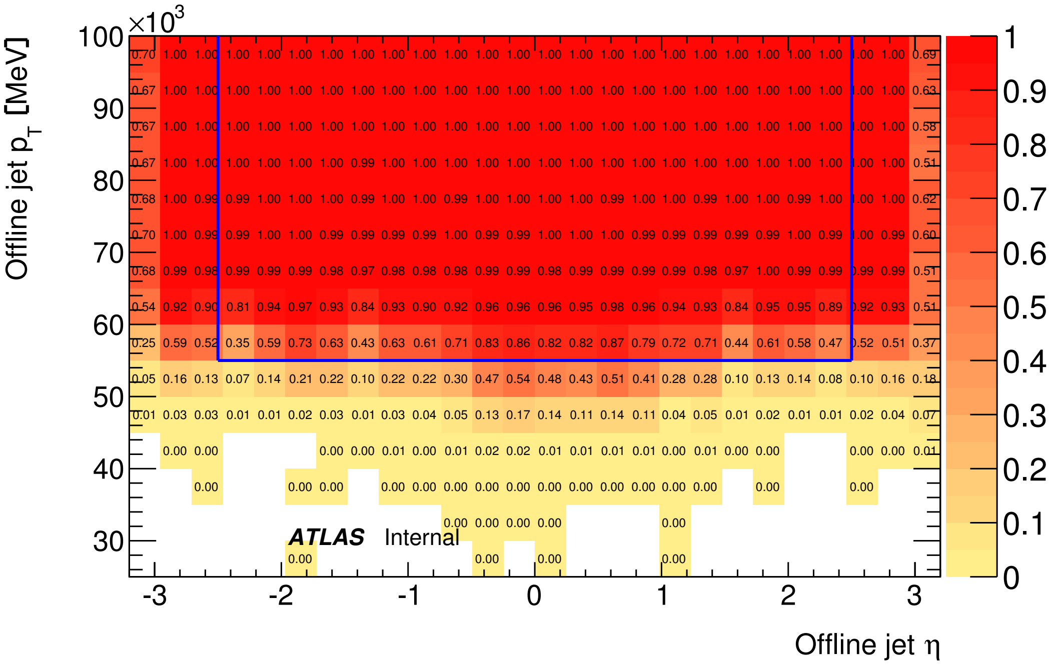


Figure 4.11: Single jet trigger efficiency derived in data events selected by the `EF_j110_a4tchad` trigger for the chain `L1_J15` → `L2_j15` → `EF_j55` as a function of the offline jet p_T and η. Only jets that did not contributed to the `EF_j110_a4tchad` trigger decision are considered. Blue line represent the selection of jets with $p_T > 55$ GeV and $|\eta| < 2.5$. This is the selection applied to the five leading jets in the fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$.

# jets	# b-jets	Pred.$_{\textsc{Pythia}}$	Pred.$_{\textsc{Pythia}\ ST}$	Pred.$_{data\ rndm}$	Pred.$_{data\ ST}$
6	== 3	8.76 ± 0.08	8.84 ± 0.07	7.48 ± 0.28	8.26 ± 0.07
	≥ 4	2.00 ± 0.02	2.02 ± 0.02	1.71 ± 0.07	1.89 ± 0.02
7	== 3	16.58 ± 0.12	16.71 ± 0.10	14.58 ± 0.45	15.77 ± 0.11
	≥ 4	4.84 ± 0.04	4.87 ± 0.04	4.27 ± 0.13	4.61 ± 0.04
≥ 8	== 3	36.20 ± 0.17	36.41 ± 0.15	33.29 ± 0.67	35.02 ± 0.16
	≥ 4	12.95 ± 0.07	13.02 ± 0.07	11.98 ± 0.23	12.56 ± 0.07

Table 4.4: Prediction of fully hadronic $t\bar{t}H$ events passing `EF_5j55_a4tchad_L2FS` trigger, estimated using single jet trigger efficiencies derived respectively on Pythia MC di-jet events without trigger requirement (Pred.$_{\textsc{Pythia}}$) and after asking single jet `EF_j110_a4tchad` trigger (Prediction$_{\textsc{Pythia}\ ST}$), on randomly triggered data (Prediction$_{data\ rndm}$) and on data passing `EF_j110_a4tchad` trigger (Pred.$_{data\ ST}$). Uncertainties quoted here are statistical only. The number of $t\bar{t}H$ events are normalized to the integrated luminosity collected by the `EF_5j55_a4tchad_L2FS` trigger.

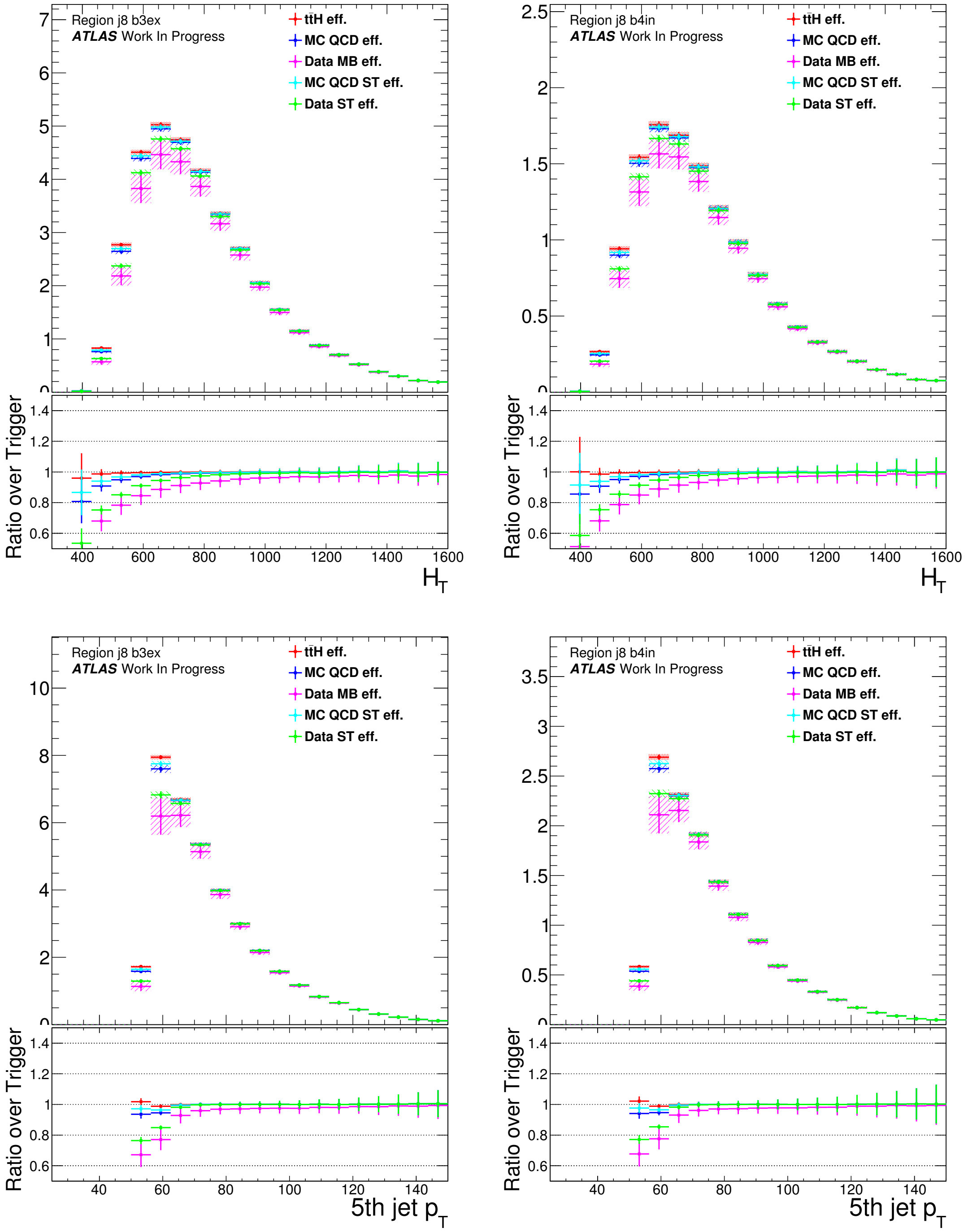


Figure 4.12: Distributions of predicted H_T and 5^{th} leading jet p_T in fully hadronic $t\bar{t}H$ MC events passing the offline selection. Different predictions are derived applying event weights using efficiencies estimated from $t\bar{t}H$ sample (red points), Pythia MC di-jet (blue points) and random triggered data (purple points), using `EF_j110_a4tchad` for Pythia MC di-jet (cyan points) and data (green points). Bottom panel shows the ratio between the prediction and the requirement of `EF_5j55_a4tchad_L2FS` trigger. Plots correspond to ≥ 8 jet multiplicity bin and exactly 3, *left*, and ≥ 4, *right*, b-tagged jet multiplicity bins. The systematic undershooting of the prediction made with efficiency from data seen in Figure 4.9 has been corrected.

4.2.7 Data Monte Carlo Scale Factor

The ratio between efficiencies evaluated in data and Pythia di-jet MC events is defined as the SF and it is shown in Figure 4.13. For $t\bar{t}H$ events the scaled efficiency becomes:

$$\epsilon' = \epsilon_{\text{trig}}^{t\bar{t}H} \cdot \text{SF}_{\text{trigger}} = \epsilon_{\text{trig}}^{t\bar{t}H} \cdot \frac{\epsilon_{\text{Data}}}{\epsilon_{\text{trig}}^{\text{Pythia}}}. \tag{4.8}$$

For the evaluation of the SF the efficiencies are derived using the `EF_j110_a4tchad` trigger to avoid suffering from statistical fluctuations.

The residual difference between $\epsilon_{\text{trig}}^{t\bar{t}H}$ and $\epsilon_{\text{trig}}^{\text{Pythia}}$ is taken as systematic uncertainty on the trigger SF. The effect on the normalization when applying SF_{trigger} on $t\bar{t}H$ events is shown in Table 4.5, whereas the effect on the shapes can be seen in Figure 4.14. The shape differences observed in the low p_T region are due to the effect of the SF application, giving a behavior similar to the one observed when using the efficiency evaluated in data. Regions with no shape difference are not affected by the SF.

# jets	# b-jets	Prediction$_{t\bar{t}H}$	Prediction$_{t\bar{t}H}$ × SF
6	== 3	8.97 ± 0.06	8.38 ± 0.13
	≥ 4	2.05 ± 0.02	1.92 ± 0.03
7	== 3	16.97 ± 0.08	15.99 ± 0.23
	≥ 4	4.95 ± 0.04	4.68 ± 0.07
≥ 8	== 3	36.77 ± 0.13	35.39 ± 0.39
	≥ 4	13.14 ± 0.06	12.69 ± 0.14

Table 4.5: Predicted fully hadronic $t\bar{t}H$ events passing `EF_5j55_a4tchad_L2FS`, estimated using single jet trigger efficiencies derived from the same samples with and without SF applied. The number of $t\bar{t}H$ events are normalized to the integrated luminosity collected by `EF_5j55_a4tchad_L2FS` trigger.

4.2.8 Conclusions

To conclude, the study presented in this section is aimed at defining single-jet (ϵ_{trig}) trigger efficiencies, evaluated in data and MC as a function of the jet p_T and η. As shown in Section 4.2.5 ϵ_{trig}, once interpreted as the probability for the jet to fire, a single trigger chain is able to precisely predict the multi-jet trigger behavior. To recover the limited statistics present in data recorded with unbiased triggers, a dedicated data sample is used to estimate ϵ_{trig} defined by a single jet trigger and where only jets not entering in the trigger decision are considered. In such a sample, two-dimensional ϵ_{trig} are derived and then used to define a data-MC trigger SF (SF_{trig}). When used in a MC event, SF_{trig} corrects the difference in trigger response between MC and data. At the end, the residual difference between ϵ_{trig} in $t\bar{t}H$ and Pythia di-jet MC events, where the latter is used to derive SF_{trig}, is taken as the systematic uncertainty for this method.

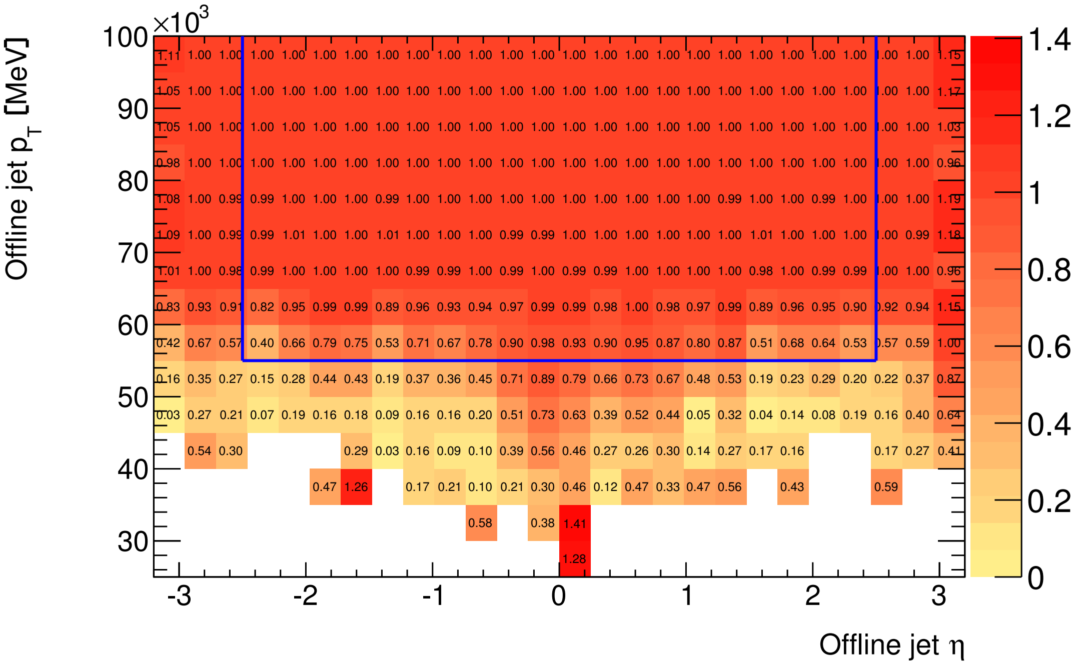


Figure 4.13: Ratio of single-jet trigger efficiencies derived in data and MC events selected by the `EF_j110_a4tchad` trigger, for the chain `L1_J15` → `L2_j15` → `EF_j55` as a function of the offline jet p_T and η. Blue line represent the selection of jets with $p_T > 55$ GeV and $|\eta| < 2.5$. This is the selection applied to the five leading jets in the fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$.

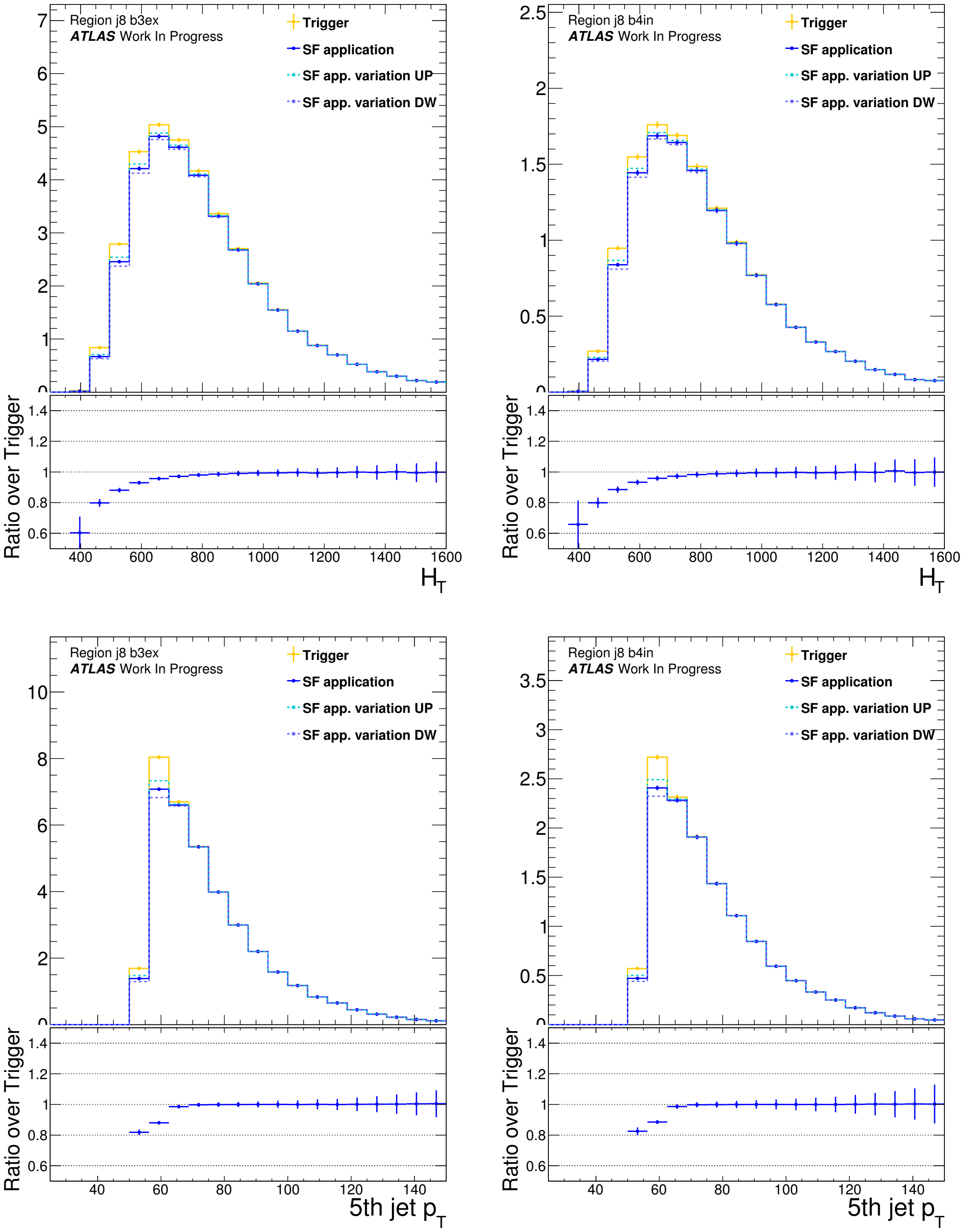


Figure 4.14: Distributions of predicted H_T and 5^{th} leading jet p_T in fully hadronic $t\bar{t}H$ MC events passing the offline selection. Comparison between predictions derived applying event weights using efficiencies estimated on the same $t\bar{t}H$ sample after the application of the Scale Factor (blue points) and after `EF_5j55_a4tchad_L2FS` trigger requirement (yellow points). Systematic uncertainty is also shown. Bottom panel shows the ratio between the prediction and the requirement of `EF_5j55_a4tchad_L2FS` trigger. Plots correspond to ≥ 8 jet multiplicity bin and exactly 3, *left*, and ≥ 4, *right*, b-tagged jet multiplicity bins.

4.3 Flavor tagging at trigger level

The multi-jet production is by far the process that contributes most to jet production at LHC. This overwhelming background requires stringent selections at the trigger level to avoid saturation of the trigger rates by uninteresting events, and leads to the requirement of high E_T threshold and high multiplicity. The downside of this harsh selection is a loss in efficiency of interesting signal events with multi-jet signature, like $t\bar{t}H$ in fully hadronic final state. To compensate for this effect, a flavor tagging algorithm has been adopted and applied at the LVL2 and the EF with the aim of discriminating particle jets originating from b-quarks from other jets. The online b-tagging selection allows to decrease the E_T thresholds and the multiplicity without increasing trigger rates.

The online b-tagging selection is performed only inside the RoI defined by the LVL1 trigger, even after the introduction of L1.5 trigger. This is to reduce the time needed to read the ID data and run the tracking algorithms. The b-tagging algorithm running online is a simplified version of the software running offline. The physics principle of b-tagging has been explained in Section 3.2. In particular a combination of two likelihood-based algorithms, exploiting the impact parameter significance distribution (IP3D) and the secondary vertex properties (SV1) were used during the 2012 data taking campaign. Figure 4.15 shows the comparison of this tagger in data and MC.

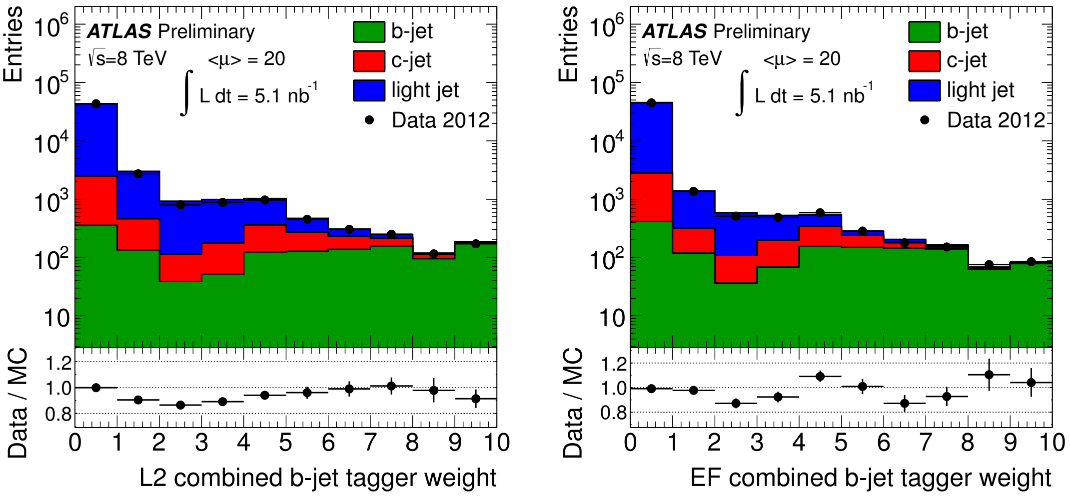


Figure 4.15: Jet weight distribution for the tagger based on the combination of the impact parameter significance (IP3D) and the secondary vertex likelihood-based (SV1) taggers for: *left*, LVL2 objects with $p_T > 50$ GeV and $|\eta| < 2.5$, and *right*, EF objects with $p_T > 55$ GeV and $|\eta| < 2.5$. Only statistical errors are shown [100]. Light and c-jets accumulate at low value of the tagger weight, while high values are mainly populated by b-jets as expected. Overall good agreement between data and MC is observed.

In the fully hadronic $t\bar{t}H$ ($H \rightarrow b\bar{b}$) analysis, these b-jet triggers have not been used due to a problem in the implementation of the b-tagging algorithm during Run 1 that lowered tagging efficiency. This problem has been solved and b-jet triggers are now in place and are used in data taking in Run 2 of the LHC.

Chapter 5

The $t\bar{t}H$ ($H \to b\bar{b}$) search: current results

The search of the Higgs boson production in association with a top anti-top quark pair, is a complex analysis. This search has been performed by the ATLAS and CMS collaborations [11, 15]. The fully hadronic $t\bar{t}H$ ($H \to b\bar{b}$) analysis shares most of the methodology with these, hence to help understand its peculiarities this chapter is dedicated to their review.

5.1 General introduction

Even though, for a SM Higgs boson with a mass of 125 GeV, the branching ratio $H \to b\bar{b}$ is approximately 56%, making it the main decay channel, up to this date this process has not yet been observed. The search $H \to b\bar{b}$ via gluon fusion presents an overwhelming $b\bar{b}$ background which surpasses the signal by several orders of magnitude. The associated production of the Higgs boson with a vector boson [101, 102, 103] allows to increase the signal-to-background ratio making the search for this decay channel a feasible one.

The Higgs boson production in association with a top anti-top quark pair is also a viable option to exploit this decay channel. The $t\bar{t}H$ process for SM Higgs boson mass of $M_H = 125$ GeV and $\sqrt{s} = 8$ TeV has a cross section of:

$$\sigma_{t\bar{t}H} = 0.1293 \text{ pb } {}^{+3.9\%}_{-9.3\%} \text{ (QCDScale)} \quad \pm 8.1\% (\text{PDF} + \alpha_s)$$

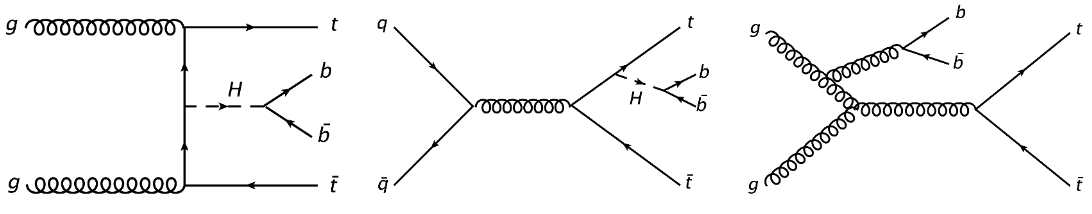
and it is known at the next-to-leading order of QCD [104, 30]. For this calculation an on-shell top-quark mass of $M_t = 172.5$ GeV is used. The Yukawa coupling of the top-quark is set to $C_t = M_t \left(\sqrt{2} G_F\right)^{1/2}$. Factorization ($\mu_F$) and renormalization ($\mu_R$) scales are set to $\mu_R = \mu_F = \mu_0 = M_t + M_H/2$. To test the $\sigma_{t\bar{t}H}$ dependence on the choice of the PDF, different PDFs have been used to perform the calculation: MSTW2008 [105, 106], CTEW6.6 [107] and NNPDF2.0 [108]. The stated uncertainties are estimated respectively with a variation of a factor of two in the scale μ_0 and propagating the 68% CL uncertainties on the PDFs and strong coupling constant α_s.

An example of the tree level partonic processes contributing to $t\bar{t}H$ production are given in Figure 5.1(a) and 5.1(b). One of the main physics backgrounds comes from top-antitop quark pairs production associated with additional jets ($t\bar{t}$ + jets). The cross section of this process has been evaluated to the next-to-next-to-leading order in QCD and includes soft-gluon resummation at next-to-next-to-leading logarithmic contribution [109], for $\sqrt{s} = 8$ TeV pp collisions:

$$\sigma(t\bar{t} + \text{jets}) = 245.8^{+6.2}_{-8.4}\ (\text{scales})\ ^{+6.2}_{-4.4}\ (\text{PDF})\ \text{pb}$$

In particular, the sub-process $t\bar{t} + b\bar{b}$, Figure 5.1(c), represents the main irreducible background for any $t\bar{t}H$ ($H \to b\bar{b}$) search, as it shares most of the experimental signature of the signal. Its cross section is complex to evaluate theoretically and in current $t\bar{t}H$ analysis this process is simulated by MC generators via parton-shower. Precise measurements of $t\bar{t} + b\bar{b}$ cross section is fundamental to help constraining the theoretical uncertainties affecting all $t\bar{t}H$ ($H \to b\bar{b}$) analyses. The CMS collaboration has performed a measurement [110] of the ratio of $\sigma(t\bar{t}+b\bar{b})$ over the inclusive $t\bar{t}$ + jets cross section $\sigma(t\bar{t} + jets)$ and obtained the following result:

$$\frac{\sigma(t\bar{t} + b\bar{b})}{\sigma(t\bar{t} + jets)} = 0.022 \pm 0.003(\text{stat}) \pm 0.005(\text{syst})$$



(a) (b) (c)

Figure 5.1: Representative tree-level Feynman diagrams for: (a) and (b), production of the Higgs boson in association with a top-quark pair ($t\bar{t}H$) and with the Higgs boson decaying into a $b\bar{b}$ pair; (c), for the main irreducible background of the analysis, the production of two top-quarks and two b-quarks ($t\bar{t} + b\bar{b}$).

5.2 Object reconstruction

Objects used in the $t\bar{t}H$ analyses are, regardless the decay channel: jets, electrons, muons and missing energy.

5.2.1 CMS

CMS uses particle-flow algorithm to perform the global event reconstruction [111, 112]. Only charged particles originating from the primary vertex are considered when reconstructing physics objects. The primary vertex is the one that has highest value of $\sum p_{Ti}^2$, where i spans all the particles associated to the vertex. The missing transverse momentum $\overrightarrow{p}_T^{\text{miss}}$ is defined as the negative of the vectorial sum of the transverse momentum of all the neutral and charged particles. Its magnitude is referred to as E_T^{miss}.

Muons are reconstructed combining information from silicon detector and muon system [113]. Electrons are required to have matching between track in the silicon tracker with an energy deposition in the electromagnetic calorimeter [114]. Isolation criteria are applied to both electrons and muons to reduce pile-up contribution.

Jets are reconstructed using anti-k_t algorithm [83, 84, 85] with a radius of 0.5. Jets used in the analysis are required to have $|\eta| < 2.5$. Jets are identified to be originating from bottom quarks via the CSV algorithm [115] that takes into account information from properties of secondary vertices and impact parameters of the tracks with respect to the primary vertex. In the analysis a jet is considered to be b-tagged if it passes the CSV medium working point, that provides efficiency of 70% to tag jets originating from bottom quarks, 20% to tag jets from charm quarks and $\sim 2\%$ of jets originating from light-quarks and gluons.

5.2.2 ATLAS

For an extensive discussion on ATLAS object reconstruction the reader is directed to Chapter 3. The working point for the jet b-tagging algorithm used is 70% efficient to tag a jet originating from a bottom-quark, 20% one from a charm-quark and 1% a jet from a light-quark.

5.3 Event selection

ATLAS selects events using a combination of single lepton triggers with different p_T thresholds to maximize the overall efficiency. In case of low p_T threshold there is an additional requirement on the *isolation* of the lepton candidate. This leads to inefficiencies at high p_T that are recovered by using triggers with high p_T thresholds. For electrons two triggers are used, both requiring at least one candidate with isolation and $p_T > 25$ GeV or $p_T > 60$ GeV without isolation. Similarly muon-trigger requires at least one candidate with isolation and $p_T > 25$ GeV or with $p_T > 36$ GeV without isolation.

CMS uses single lepton and di-lepton triggers. All the triggers require isolation of the lepton candidate. Accepted events should have at least one muon with $p_T >$ 24 GeV or at least one electron with $p_T > 27$ GeV. Di-electron trigger requires at least two electrons with $p_T > 17, 8$ GeV.

Offline selections are driven by efficiency of online selections. To avoid trigger efficiency turn-on regions where differences between simulations and data can arise, offline selections aim at having objects in the plateau region. ATLAS requires leading and sub-leading leptons to have $p_T > 25$ GeV and $p_T > 15$ GeV respectively, both with $|\eta| < 2.5$. CMS requires a tighter selections: leading lepton is required to have

$p_T > 30$ GeV and subleading $p_T > 20$ GeV. Muon acceptance is $|\eta| < 2.1$, while for electrons it is $|\eta| < 2.5$.

5.4 Classification of $t\bar{t}$ + jets events

As mentioned in previous section $t\bar{t}$ + jets is the dominant background. To study and describe its behaviour in more accurate way, it is divided in categories according to the flavor of the jets not originating from the $t\bar{t}$ pair. This separation is helpful also in the assessment of systematic uncertainties, since the uncertainty on the cross sections of the processes $t\bar{t}$ + light-jets and $t\bar{t}$ + heavy+flavor are know with different precision.

ATLAS defines three categories using a matching procedure to link jets and partons. Jets considered in the procedure are slightly softer than the one used in the analysis. These have radius $R = 0.4$, $p_T > 15$ GeV and $|\eta| < 2.5$. In the analysis the jet minimum p_T is 25 GeV. If at least one of such jets is found at a distance $\Delta R < 0.4$ from a bottom quark not originating from the $t\bar{t}$ pair the event is labeled as $t\bar{t} + b\bar{b}$. If an event, not classified as $t\bar{t} + b\bar{b}$, has at least one particle jet at $\Delta R < 0.4$ from a charm quark it is labeled $t\bar{t} + c\bar{c}$. Events that are not entering these two categories are labeled $t\bar{t}$ + light-jets.

CMS defines four categories using a matching criterion in which, objects are considered matched if $\Delta R < 0.5$. Events with at least two jets matched to two b quarks not originating from the $t\bar{t}$ pair decay are labeled as $t\bar{t} + b\bar{b}$. Events with only one jet matched to a b quark are labeled as $t\bar{t}$ + b. Events not entering in the previous categories and with at least one jet matched to a charm quark are labeled $t\bar{t} + c\bar{c}$. Events not entering any of the previous categories are labeled $t\bar{t}$ + lf (light-flavor).

Events in the categories $t\bar{t} + b\bar{b}$, $t\bar{t}$ + b and $t\bar{t} + c\bar{c}$, are referred generically as $t\bar{t}$ + heavy-flavor.

5.5 Signal modeling

CMS uses Pythia 6.426, LO generator, to describe the $t\bar{t}H$ process. Normalization is taken from NLO calculation of the cross section, assuming SM Higgs boson with mass $m_H = 125$ GeV.

ATLAS models the signal using NLO matrix elements from HELAC-Oneloop and interfaced using Powheg-Box to the Monte Carlo programs, Pythia 8.1 which is used for the showering. The signal sample is generated inclusively for the Higgs boson decays. Normalization is obtained from (N)NLO calculations.

5.6 $t\bar{t}$ + jets background modeling

CMS uses Madgraph 5.1.3 with tree-level matrix elements to describe the $t\bar{t}$ + jets background. Normalization is obtained from NNLO cross section calculation with soft-gluon resummation and next-to-next-to-leading logarithmic accuracy. ATLAS uses Powheg-Box 3.0. NLO generator, interfaced to Pythia 6.425 for the showering. Normalization is obtained from top++2.0 NNLO cross section calculation with soft-gluon resummation and next-to-next-to-leading logarithmic accuracy.

To improve the agreement between data and simulation, $t\bar{t}$ + light-jets and $t\bar{t}+c\bar{c}$ events from Powheg + Pythia are corrected to reproduce the distributions of the top quark p_T and the $t\bar{t}$ system p_T measured in data [116]. More detailed description and results of the procedure are presented in Section 6.4.3. Given the fact that $t\bar{t}+b\bar{b}$ events are the most important irreducible background a series of detailed studies has been performed to increase the quality of its description. Since these are common to the $t\bar{t}H$ ($H \to b\bar{b}$) fully hadronic analysis, these are describe in Section 6.4.3.

5.7 Analysis strategy

The two collaborations divide the $t\bar{t}H$ analyses in two channels according to the number of leptons, *single-lepton* channel (SL) for events having exactly 1 lepton, *dilepton* channel (DL) for events with exactly 2 leptons of opposite electric charge. Here the word lepton is referred only to muons and electrons, since the identification in the detector for these is more efficient than for taus. The analyses do not have dedicated selections for the identification of events with the presence of taus. Matrix element method (MEM), described in Section 5.9, is used to increase discriminating power between signal and $t\bar{t} + b\bar{b}$ irreducible background. No attempt of reconstructing the $t\bar{t}H$ system is made.

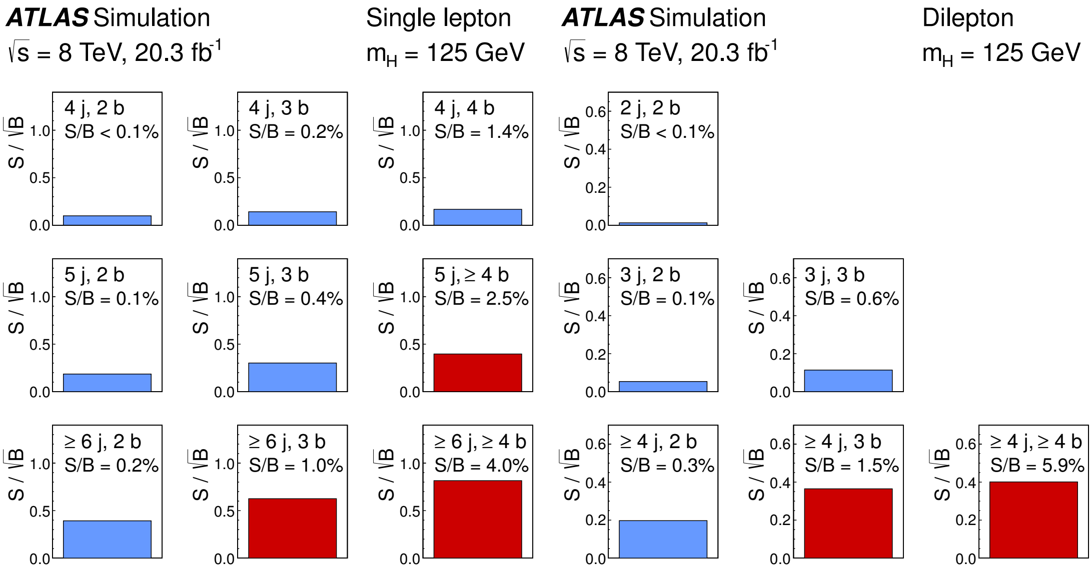
5.7.1 ATLAS

Each of the two decay channels, SL and DL, is divided in regions according to jet and b-tagged jet multiplicity. A total of nine regions are defined for the SL channel ranging from exactly four jets to at least six jets and from exactly two to at least four b-tagged jets. DL channel is divided in six regions, where the minimum number of jets or b-tagged jets considered is two.

Some of the regions are background dominated and are fundamental to constraint systematic uncertainties when performing the statistical analysis. For example, the regions with exactly 2 b-tagged jets provide strong constraints on $t\bar{t}$ + light-jets modeling systematics. The category with exactly 4 jets and exactly 3 b-tagged jets in the SL channel gives constraints on c-tagging due to $W^+ \to cs$ decays from $t\bar{t}$ + light-jets. The S/B and $S/\sqrt{B}$ for all these regions are shown in Figure 5.2. Event yields are presented in Figure 5.3.

5.7.2 CMS

CMS categorizes its events using the number of jets, leptons and the value of a discriminator $\mathcal{F}$ (see Equation 5.2). Jets are taken into account if they have $p_T > 30$ GeV and $|\eta| < 2.5$. Events are required to have at least four jets in the di-lepton



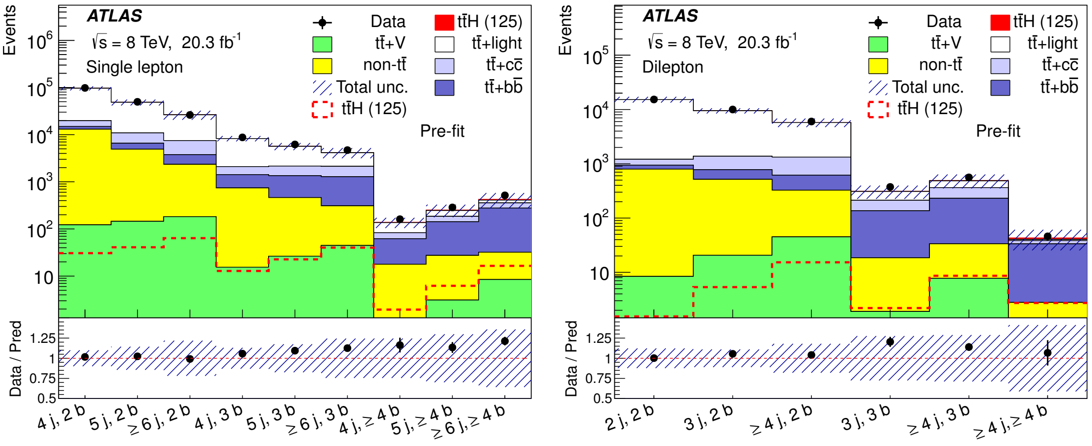
(a) (b)

Figure 5.2: $S/\sqrt{B}$ ratio for each of the regions considered in the ATLAS analysis. SM cross section and branching fractions are calculated assuming a Higgs boson mass of 125 GeV. (a) and (b) plots refer to the SL and DL channels respectively. Rows show values for specific jet multiplicity and columns show b-jet multiplicity. Signal-rich regions are colored in red, while the rest are shown in blue. The S/B ratio for each region is also shown [11].

channel and at least five jets in the single-lepton channel. The latter one is further divided in sub-categories according to the properties of the un-tagged jets in the event. One category is defined for events where exactly six jets are present and the mass of the two un-tagged is compatible with the mass of the W boson, i.e. in the range [60,100] GeV. The mass range tightens with the increase of jet multiplicity to account for combinatoric background. If the jets are not fulfilling the mass requirement, events are classified in a second category. Third category is defined for events with exactly five jets.

Categorization follows the physics interpretation of the events. First case is associated with the full reconstruction of $W \rightarrow q\overline{q}'$ decay. In second category the W is considered not be reconstructed, due to the fail of the mass window requirement. In the third case also the W is assumed not fully reconstructed. The dilepton channel is not divided in further categories. It is interpreted as each of the four jets are coming from one of the four bottom quarks present in the final state.

To increase the background rejection a variable is defined to select events that are more likely to be originating from heavy-flavour final states. A variable $\mathcal{F}$ is build using information from the flavour discriminator CSV [115]. In the single-lepton channel only six jets with the highest CSV value are taken into account, in the dilepton channel only the leading four. The values of CSV for the considered jets are



(a) (b)

Figure 5.3: Comparison of data to background prediction in ATLAS. In the plots each bin corresponds to a region of the analysis. (a) and (b) plots refer to the SL and DL channels respectively. $t\bar{t}H$ Signal is normalized to the SM prediction. The dashed area corresponds to the total uncertainty on the yields [11].

denoted as $\overrightarrow{\xi}$. The likelihood to observe $\overrightarrow{\xi}$ is evaluated under two hypotheses: $t\bar{t}$ + two heavy-flavour jets ($t\bar{t}$ + hf) or $t\bar{t}$ + two light-flavour jets ($t\bar{t}$ + lf). Correlation among jets in the same events are neglected. Example of how the likelihood is evaluated is given for the single lepton case in events with six jets:

$$f(\overrightarrow{\xi}|t\bar{t}+\text{hf}) = \sum_{i_1}\sum_{i_2\neq i_1}\cdots\sum_{i_6\neq i_1,\ldots,i_5}\left\{\prod_{k\in\{i_1,i_2,i_3,i_4\}} f_{\text{hf}}(\xi_k)\prod_{m\in\{i_5,i_6\}} f_{\text{lf}}(\xi_m)\right\} \tag{5.1}$$

Where ξ_i is the CSV discriminator for the ith jet and f_{hf} and f_{lf} are respectively the probability density function of observing ξ_i when the ith jet is originating from a heavy or light quark respectively. To simplify the evaluation of the likelihood, f_{lf} takes into account quarks u, d, s and gluons, but not quark c. Equation 5.1 is extended to cases with five jets, for single lepton case, or four jets, for dilepton case, associating four jets to the heavy-flavour partons and the other jets to the light-flavour partons. The likelihood $f(\overrightarrow{\xi}|t\bar{t}+\text{lf})$ is the same as Equation 5.1, but exchanging f_{lf} with f_{hf}. The variable $\mathcal{F}$ used in the event selection is given by the ratio:

$$\mathcal{F} = \frac{f(\overrightarrow{\xi}|t\bar{t}+\text{hf})}{f(\overrightarrow{\xi}|t\bar{t}+\text{hf}) + f(\overrightarrow{\xi}|t\bar{t}+\text{lf})} \tag{5.2}$$

Events are rejected if $\mathcal{F} < \mathcal{F}_L$, where $\mathcal{F}_L$ varies between 0.85 and 0.97 depending on the channel and jet multiplicity. A selection is made to sort the remaining events requiring $\mathcal{F} > \mathcal{F}_H$, where $\mathcal{F}_L < \mathcal{F}_H < 1.0$. The ones passing the cut are classified as high-purity (H) and the other as low-purity. $\mathcal{F}_H$ depends on the lepton and jet multiplicity. Regions with low-purity are used to constrain systematic uncertainties while high-purity regions are more signal rich.

The event yields after the selections described in this section are presented in Figure 5.4.

5.8 Discriminating variables

ATLAS uses output of a Neural Network (NN) as final discriminant in regions with higher sensitivity. NNs are trained independently in each region to discriminate $t\bar{t}H$ signal against backgrounds. These are used in event categories (5j,$\geq$ 4b), ($\geq$ 6j,3b), ($\geq$ 6j,$\geq$ 4b) of SL channel and in (3j, 3b), ($\geq$ 4j, 3b) ($\geq$ 4j,$\geq$ 4b) of DL channel. Region (5j, 3b) of SL channel uses a NN to discriminate $t\bar{t}$ +light against $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$. Several type of variables are entering the NN evaluation: object kinematics, global event variables, event shape variables, object pair properties and matrix element method variables. The latter are used only in regions ($\geq$ 6j,3b), ($\geq$ 6j,$\geq$ 4b) of SL channel and will be described briefly in Section 5.9. In the other regions the variable H_T is used in the DL channel and H_T^{had} in the SL channel, where H_T is the scalar sum of p_T of the jets and of the leptons in the event and H_T^{had} is the scalar sum of the p_T of the jets exclusively.

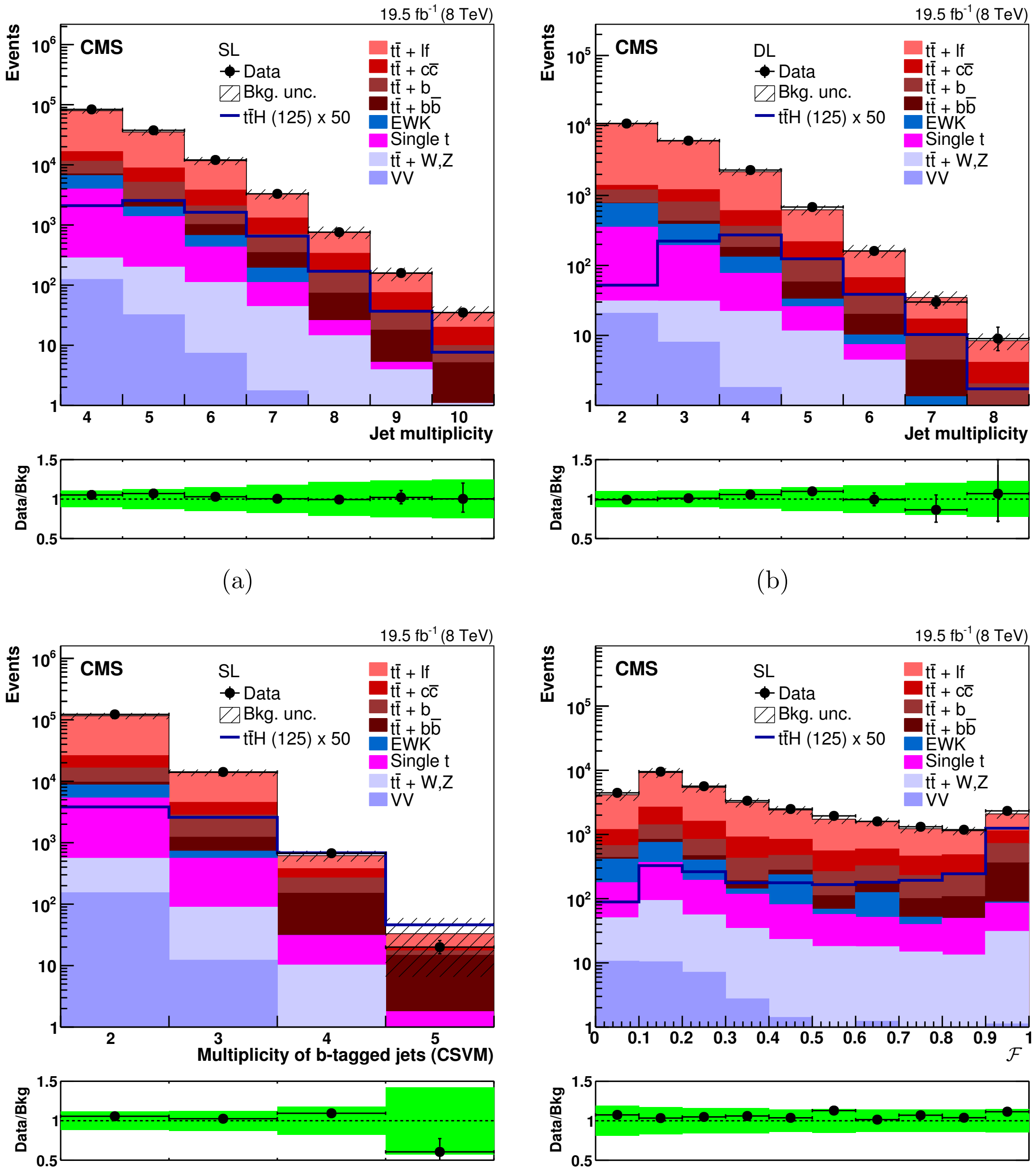
CMS uses two discriminating variables called $P_{s/b}$ and $P_{h/l}$. Variable $P_{s/b}$ is defined in Equation 5.4, it is carrying the information about the kinematics and the dynamics of the event via the Matrix Element Method, described in Section 5.9. It is designed to discriminate $t\bar{t}H$ signal events against $t\bar{t}+b\bar{b}$ background. Variable $P_{h/l}$ is carrying information about flavour tagging and it is designed to discriminate heavy-flavour component of $t\bar{t}$ background against light-flavour component.

$$P_{h/l} = \frac{f(\overrightarrow{\xi}|t\bar{t}+\mathrm{hf})}{f(\overrightarrow{\xi}|t\bar{t}+\mathrm{hf}) + k_{h/l} f(\overrightarrow{\xi}|t\bar{t}+\mathrm{lf})} \tag{5.3}$$

Only jets considered in the evaluation of $P_{s/b}$ are entering in this calculation, making $P_{h/l}$ different from $\mathcal{F}$. $k_{h/l}$ is a positive constant that has different values in the different regions, it is defined to maximize the sensitivity due to the use of binned distributions and its value has no influence in the separating power of the variable. The two discriminants are used together in the statistical analysis via a two dimensional likelihood fit to extract the result.

5.9 Matrix element method

As discussed briefly at the beginning of this section the background of the $t\bar{t}H$ analysis is dominated by the $t\bar{t}$ + jets production. In particular the $t\bar{t}+b\bar{b}$ process is the main irreducible background having the same signature as the signal. To increase



(a) (b)

(c) (d)

Figure 5.4: Comparison of data to background prediction in CMS. Plots on the top row show jet multiplicity distribution in events required to have at least two jets passing the CSV medium working point in the SL (*left*) and DL (*right*) channels. Plots in the bottom row refer to SL category. (c) shows the multiplicity of jets passing the CSV medium working point in events with at least four jets. (d) shows the distribution of the variable $\mathcal{F}$, defined in Equation 5.2, for events with at least six jets and at least one jet passing the CSV medium working point. The dashed area in the top panel and the solid green one in the bottom panel correspond to the statistical plus systematic uncertainty [15].

the discriminating power against this background both ATLAS and CMS collaborations use the Matrix Element Method (MEM) [117]. This method is described briefly in this section.

The method links theoretical calculation for the signal and background processes to the observed quantities and allows for each event to evaluate the probability to be compatible with one of the two hypothesis. The method to build a probability density function takes into account the matrix elements of the process and uses transfer functions to connect partons to observed quantities. Then this quantity is integrated over the phase space of the initial- and final-state particles. Probability density functions are determinated for each of the possible assignment of the observed objects to the final-state partons and summed to build a likelihood function for the process. The only background hypothesis considered is $t\bar{t} + b\bar{b}$, due to the fact that it is the main background. Nevertheless separation power against the other background $t\bar{t}$ + jets is present.

The discriminant variable used by the two analysis is the likelihood ratio, called $D1$ in ATLAS and $P_{s/b}$ in CMS:

$$D1 \equiv P_{s/b} = \frac{\mathcal{L}_{t\bar{t}H}}{\mathcal{L}_{t\bar{t}H} + \alpha \mathcal{L}_{t\bar{t}+b\bar{b}}} \tag{5.4}$$

Where $\mathcal{L}_{t\bar{t}H}$ and $\mathcal{L}_{t\bar{t}+b\bar{b}}$ are the likelihood functions for the signal and the background hypothesis, α is a relative normalization factor defined to optimize the performance due to the finite bin size of the variable distribution. ATLAS uses as additional variable also the logarithm of summed signal likelihoods ($\mathcal{L}_{t\bar{t}H}$), called $SSLL$.

5.10 Systematic uncertainties

Since the $t\bar{t}H$ ($H \to b\bar{b}$) fully hadronic analysis shares the majority of the systematic uncertainties with the other $t\bar{t}H$ ($H \to b\bar{b}$) decay channels described in this subsection and to avoid repetition, only the most relevant ones and those specific to non-fully hadronic analysis will be reported here.

Systematic uncertainties are affecting both normalization and shape of the different backgrounds. To reduce the impact on the analysis sensitivity, uncertainties are embedded in the statistical analysis as nuisance parameters (NP). Each NP is correlated across all the signal regions to allow the background dominated regions to constrain the uncertainties in the signal rich regions.

5.10.1 Uncertainties on physics objects

Main uncertainties affecting jets are the Jet Energy Scale (JES) and the b-tagging calibration. ATLAS divides JES in 22 uncorrelated components with different p_T and η dependencies [88]. For CMS JES has an effect ranging from 1% to 8% depending on p_T and $|\eta|$ [118]. Calibration of b-tagging discriminator is described by ATLAS using 6 uncorrelated components for b-tagging, 6 for c-tagging and 12 for light-tagging [91]. For CMS uncertainty on b-tagging calibrations is ranging from 2% to 17%.

Uncertainties on the leptons in ATLAS are originating from trigger, reconstruction, identification, isolation and lepton momentum scale and resolution. These are treated via 5 NP for electrons and six for muons. CMS estimates them to have an effect of 2% for both electrons and muons.

5.10.2 Background modeling and theoretical uncertainties

Uncertainty on luminosity is 2.6% for CMS [119] and 2.8% for ATLAS [120] and applies to all Monte Carlo predictions.

Modeling of $t\bar{t}$ + jets processes plays a fundamental role in the $t\bar{t}H$ analyses. Variation of renormalization scale and PDF have been computed and taken into account as NPs. Parton shower and hadronization model uncertainties are taken into account interfacing the Powheg generator with Pythia and Herwig.

Both collaborations use $t\bar{t}$ p_T re-weighting, but the way the weight is evaluated is different, also the systematic uncertainty estimation is different between the two. ATLAS propagates the nine largest uncertainties from the measurement used to extract the re-weight [116]. This is because they were giving 95% contribution of the total uncertainty of the measurement. CMS considers as uncertainty the non application and the doubling of the re-weighting effect.

ATLAS performed a series of dedicated studies to better understand the behaviour of $t\bar{t}+b\bar{b}$ process. Since these studies will also be used in the $t\bar{t}H$ ($H \rightarrow b\bar{b}$) fully hadronic analysis, they are described here briefly and will be reported in detail in Section 6.9.5. Variation of resummation, factorization and renormalization scales are propagated to the re-weight that is applied to $t\bar{t} + b\bar{b}$ events. Effects of PDF choice, multi-parton interaction and final state radiation are also taken into account. Comparison of Powheg + Pythia and SherpaOL predictions gives and uncertainty on the normalization of $t\bar{t} + b\bar{b}$ of 50%. Same value is considered for $t\bar{t} + c\bar{c}$ process because there are no NLO calculation available to perform a similar test. These two uncertainties are the leading and third leading ones for the ATLAS analysis. CMS considers 3 independent uncertainties of 50% on the $t\bar{t}+b\bar{b}$, $t\bar{t} + b$ and $t\bar{t}+c\bar{c}$ normalizations [121].

Figure 5.5 shows the ranking of the leading systematic uncertainties for the ATLAS analysis.

5.11 Results

The two collaboration use similar approach to combine the information from all the discriminant distributions to asses limits on the presence of the $t\bar{t}H$ signal. In the statistical analysis a binned likelihood function $\mathcal{L}(\mu, \theta)$ is constructed. It covers all the signal regions and depends on the signal strength parameter μ, defined as the ratio of the observed $t\bar{t}H$ cross section over the standard model prediction, and the set of nuisance parameters θ. A fit is applied to data to find the combination of θ and

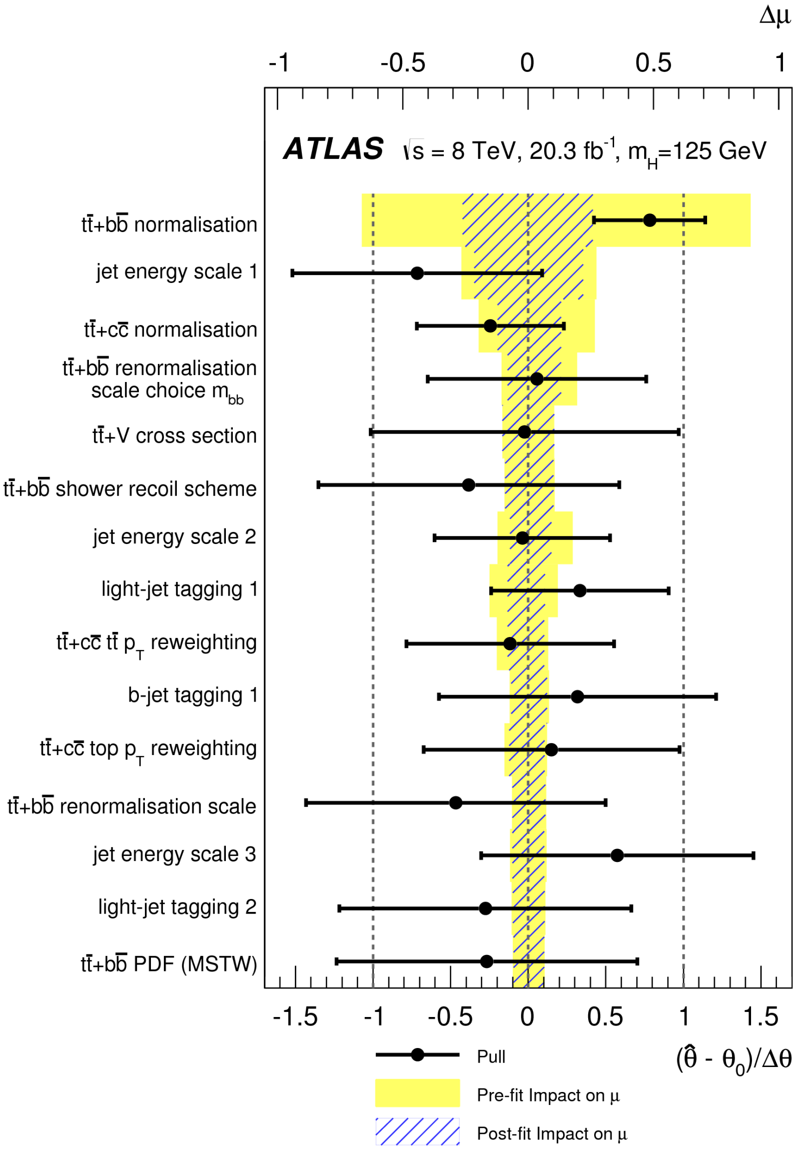
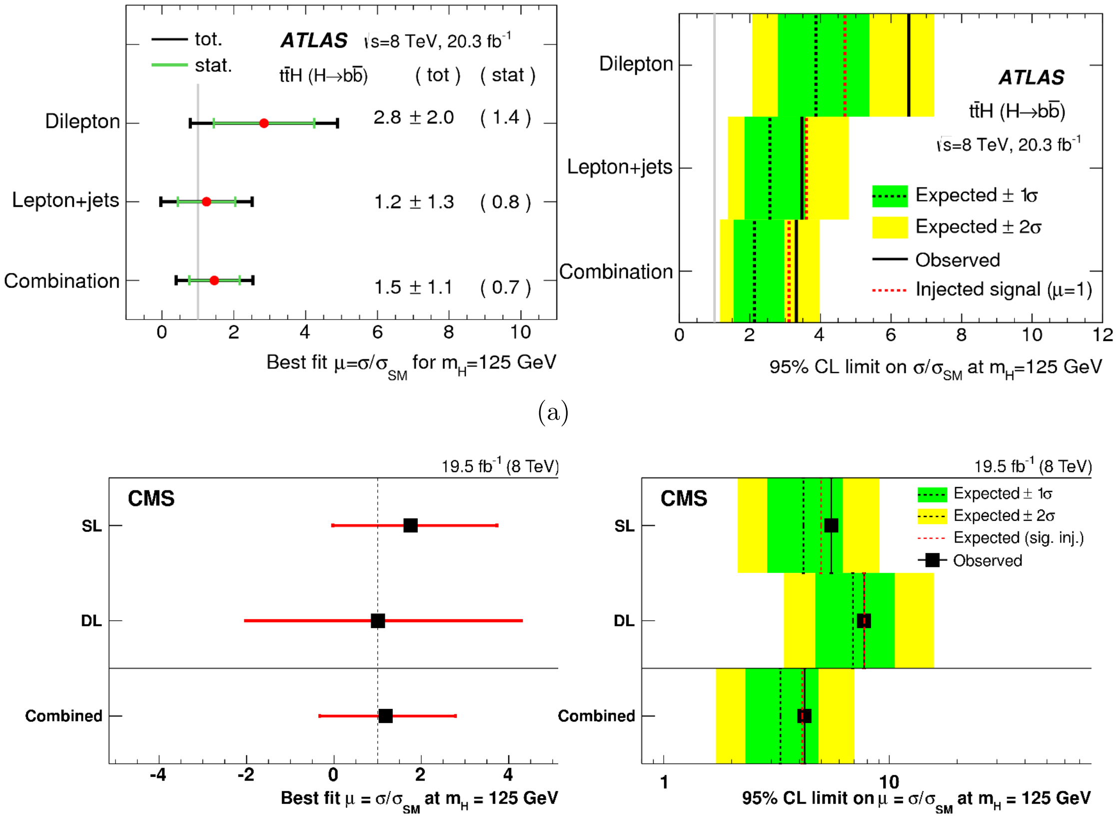


Figure 5.5: Best fit values of the nuisance parameters with the largest impact on the measured signal strength μ. Black points are drawn according to the bottom axis and represent the deviation of the fitted nuisance parameters $\hat{\theta}$ from the pre-fit value θ_0 in units of the pre-fit standard deviation $\Delta\theta$. Error bars represent the post-fit uncertainty σ_θ. i.e. if data are constraining the uncertainties, the bars are smaller than 1, otherwise these are close to 1. Yellow and dashed blue bands represent the pre- and post-fit impact on μ of the nuisance parameter. Bands are drawn according to the top axis [11].

μ that maximizes the likelihood. This procedure reduces the impact of systematic uncertainties exploiting the regions background dominated to constrain them and reduce the degradation of sensitivity. The statistical analysis technique is the same as the one used in the $t\bar{t}H$ fully-hadronic search, so more details can be found in section 6.10. ATLAS measures a signal strength μ of 1.5 ± 1.1 [11] using 20.3fb^{-1} of pp collision data at $\sqrt{s} = 8$ TeV. A $t\bar{t}H$ cross section 3.4 times the SM expectation is excluded at 95% confidence level (CL), while signal 2.2 times SM prediction is expected to be excluded. Plots summarizing the results are shown in Figure 5.6(a).

CMS measures a signal strength of $1.2^{+1.6}_{-1.5}$ [15] using using 19.5fb^{-1} of pp collision data at $\sqrt{s} = 8$ TeV. Observed exclusion limit at 95% CL is 4.2 times SM prediction, while expected is 3.3. Plots summarizing the results are shown in Figure 5.6(b).



(a)

(b)

Figure 5.6: *Left*: Plots show the signal strength $\mu = \sigma/\sigma_{SM}$ result and its uncertainties of the ATLAS [11] (a) and CMS [15] (b) analyses. *Right*: Plots show 95% CL upper limits on μ. Solid line represents the observed limit, red dashed line the expected exclusion for $\mu = 1.$, and black dashed line represents the expected exclusion for $\mu = 0$. For the latter also 68% and 95% confidence intervals are shown. Results are shown for single-lepton and dilepton channels separately and for their combination.

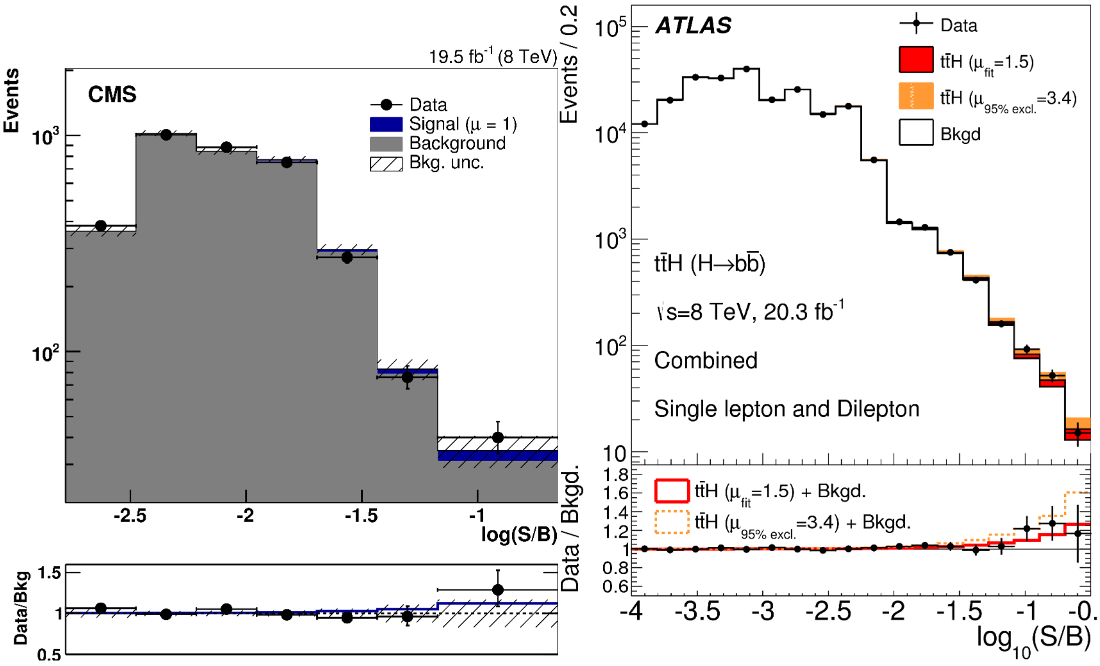
Results on upper limits for both experiments are summarized in Table 5.1. Best fit values for the signal strength μ are shown in Table 5.2. Figure 5.7 shows the event yield as function of the decimal logarithm of signal to background ratio: log(S/B). The plots are constructed taking into account all the bins entering in each of the region considered by the respective analyses. The aim of these plots is to show the overall good agreement between background modelization and observed data and visualize the regions where the signal is supposed to show up.

95% CL upper limit	Observed	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Median ($\mu = 1$)
ATLAS							
Single lepton	3.6	1.4	1.9	2.6	3.7	4.9	3.6
Dilepton	6.7	3.0	4.1	5.8	5.8	7.7	4.7
Combination	3.4	1.6	2.2	3.0	3.0	4.1	3.1
CMS							
Single lepton	5.5	2.1	2.9	4.2	6.2	9.1	5.0
Dilepton	7.7	3.4	4.7	6.9	10.6	15.8	7.8
Combination	4.2	1.7	2.3	3.3	4.9	7.0	4.1

Table 5.1: 95% CL upper limits on signal strength $\mu = \sigma/\sigma_{SM}$ are listed for individual channels of the ATLAS and CMS analyses and for the respective combinations. Second column shows the observed limits. Central part of the table lists the expected limits under the background-only hypothesis together with $\pm 1\sigma$ and $\pm 2\sigma$ intervals. Last column shows the expected 95% CL upper limit when SM prediction for $\sigma(t\bar{t}H)$ is assumed.

Best fit μ	ATLAS	CMS
Single lepton	$+1.2^{+1.3}_{-1.3}$	$+1.7^{+2.0}_{-1.8}$
Dilepton	$+2.8^{+2.0}_{-2.0}$	$+1.0^{+3.3}_{-3.0}$
Combination	$+1.5^{+1.1}_{-1.1}$	$+1.2^{+1.6}_{-1.5}$

Table 5.2: Best ATLAS and CMS fit values of signal strength μ with their uncertainties for individual channels of the analyses and for the respective combinations.



(a) (b)

Figure 5.7: Event yields as function of decimal logarithm of signal to background ratio: $\log(S/B)$. All the regions entering the respective analyses are considered. (a) CMS [15] shows the yield obtained from a fit to data with the constraint of signal strength $\mu = 1$. $t\bar{t}H$ signal is shown as blue histogram stacked over the background and as a blue line in the bottom panel. (b) ATLAS [11] shows background and signal yields obtained from the fit to data. $t\bar{t}H$ signal is shown as red area. The yield for the 95% CL excluded cross section is also shown.

Chapter 6

Fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$

This chapter describes the analysis work performed in the framework of this doctoral thesis for the search of the Higgs boson production in association with a top anti-top quark pair, in the fully hadronic final state. A brief discussion on the prospects for this physics channel in Run 2 concludes the chapter.

It is important to mention that the analysis described in this document is the first fully hadronic $t\bar{t}H$ $(H \to b\bar{b})$ search ever performed.

6.1 Introduction

The fully-hadronic $t\bar{t}H$ $(H \to b\bar{b})$ analysis presents several advantages with respect to the lepton based $t\bar{t}H$ $(H \to b\bar{b})$ analyses. It allows to gain in signal statistics since it can exploit the high branching fraction of top-quark into hadrons. For the $t\bar{t}$ pair this is summarized in Figure 6.1. An other advantage of the analysis is the absence of missing energy in the final state, allowing in principle to being able to reconstruct in the detector all the decay products of the $t\bar{t}H$ system.

In Section 6.3 the preselection of events is described and in Section 6.4 the description of the Monte Carlo samples used in the analysis and of the Tag Rate Function method used to reduce their statistical uncertainties are presented.

The major challenge of the analysis is the presence of the overwhelming multijet background which comes with a cross section six order of magnitude higher than the signal. This background is estimated with an ad-hoc data-driven method described in details in Section 6.5. Validation studies for this method are also reported.

Section 6.6 describes the classification of the events that are used in the analysis and analysis strategy. The multivariate technique used to discriminate signal and background is introduced in Section 6.7. The description of the systematic uncertainties considered in the analysis is given in Section 6.9. The result of the statistical analysis is presented in Section 6.10.

This search is based on 20.3 fb^{-1} of data collected by the ATLAS experiment in pp collisions at $\sqrt{s} = 8$ TeV between April and October 2012.

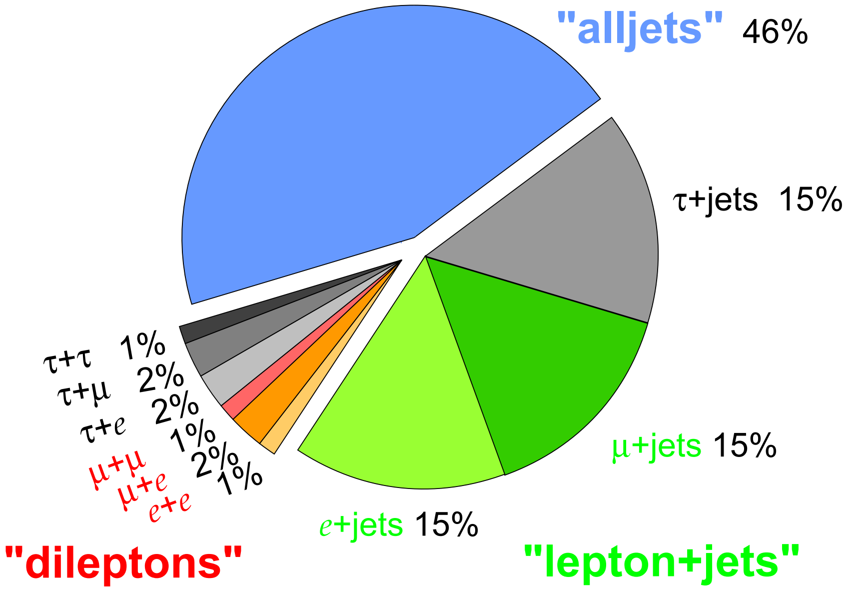


Figure 6.1: Pie chart representing the branching ratios of a top-antitop quark pair. The light blue represents the full-hadronic BR of 46%, the di-leptonic BR (without τ lepton) is shown in 3 gradations of red, with a total of 4%, and in in 2 gradations of green the lepton (e or μ) + jets BR of 30% [122].

6.2 Object reconstruction

The main objects used for this analysis are hadronic jets and b-tagged jets. Electrons and muons are used to assure that events entering the fully hadronic $t\bar{t}H$ $(H \rightarrow b\bar{b})$ data sample are not used by the semi- and dilepton $t\bar{t}H$ $(H \rightarrow b\bar{b})$ analyses (orthogonality). orthogonality with the analysis studying other $t\bar{t}H$ $(H \rightarrow b\bar{b})$ decay channels. A detailed description of how these objects are defined and what algorithms are used is given in Chapter 3.

6.3 Event preselection

The data sample used by the fully hadronic $t\bar{t}H$ $(H \rightarrow b\bar{b})$ analysis has been collected by a multi-jet trigger. This requires four jets at LVL1 with transverse energy of the LVL1 8×8 towers (each covering 0.1×0.1 in $\eta - \phi$) to be larger than 15 GeV, followed by requiring five jets reconstructed at LVL2 with a transverse energy larger than 15 GeV and finally five jets at EF having $p_T > 55$ GeV. The determination of the trigger efficiencies and scale factors are described in Chapter 4.

After the event is accepted by the trigger, it is required to have at least one reconstructed vertex with at least five associated tracks, consistent with the beam collision region in the $x - y$ 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 momenta of its associated tracks. Events are discarded if any jet with $p_T > 20$ GeV is independently identified as out-of-time activity from a previous pp collision or as calorimeter noise [123].

Off-line, the five leading jets are required to have $p_T > 55$ GeV and $|\eta| < 2.5$. Any extra jet must have $p_T > 25$ GeV and must be in the same rapidity range. Events with well identified (isolated) muons or electrons are discarded to ensure orthogonality with the other $t\bar{t}H$ ($H \rightarrow b\bar{b}$) analyses.

6.4 Background and signal modeling

After event preselection the main background consists of multi-jet production and by $t\bar{t}$+jets production. Small background contributions come from the production of a single top quark production and from the associated production of a vector boson and a $t\bar{t}$ pair. While the multi-jet background is estimated from data, all other backgrounds and the $t\bar{t}H$ signal are estimated from simulation and then normalized to the theoretical cross sections.

6.4.1 Signal modeling

The $t\bar{t}H$ signal process is modeled using matrix elements obtained from the HELAC-Oneloop package [124] that corresponds to the NLO QCD accuracy. Powheg-Box [125, 126, 127] serves as an interface to the shower Monte Carlo programs. The samples which have been created using this approach are referred to as PowHel samples [128]. The Higgs boson decays inclusively in all these samples. The CT10NLO PDF sets are used and the factorization (μ_{F}) and renormalization (μ_{R}) scales are set to $\mu_0 = \mu_{\mathrm{F}} = \mu_{\mathrm{R}} = m_t + m_H/2$. Pythia 8.1 [99] with the CTEQ6L1 set of parton distribution functions and AU2 underlying event tune [129] is used to add the event pile-up and the underlying event. The top-quark mass is set to 172.5 GeV. The cross section normalization and the Higgs boson decay branching fractions for these $t\bar{t}H$ samples are taken from the NLO theoretical calculations from Ref. [104]. Higgs boson mass is set to $m_H = 125$ GeV. The sample is inclusive in top-quark decays.

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

The simulation of events representing the $t\bar{t}$ + jets process are categorized into three non-overlapping samples: $t\bar{t} + b\bar{b}$, $t\bar{t} + c\bar{c}$, and $t\bar{t}$ + light. The labelling is based on a hadron-matching algorithm to particle jets. All particle jets with $p_T > 15$ GeV and $|\eta| < 2.5$ which are matched using a $\Delta R < 0.4$ to a b-hadron with $p_T > 5$ GeV not originating from a top quark decay are labeled as b. If the event has at least one jet matched to a b-hadron, the event is labeled as $t\bar{t} + b\bar{b}$. Events that have no b-hadron match but at least one c-hadron match are labeled $t\bar{t} + c\bar{c}$. All other events are labeled as $t\bar{t}$ + light. The main purpose of dividing the $t\bar{t}$ + jets samples into categories is to apply to each sub sample its systematic uncertainties.

6.4.3 $t\bar{t}$ + jets background modeling

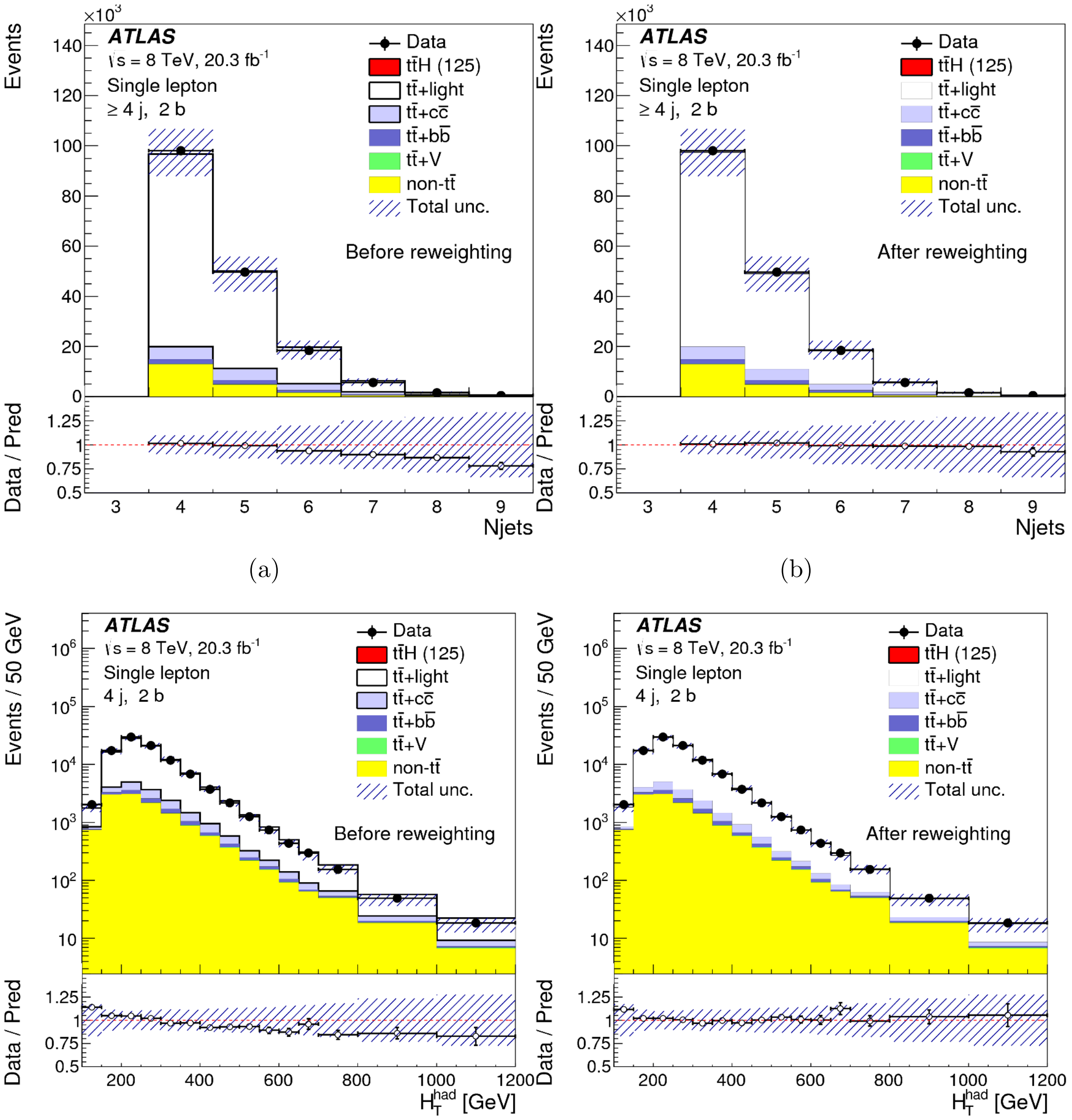
The $t\bar{t}$ sample is generated using the Powheg NLO generator [125, 126, 127] with the CT10 PDF set [130], assuming the value of the top quark mass of 172.5 GeV. It is interfaced to Pythia 6.425 [131] with the CTEQ6L1 set of parton distribution functions and Perugia2011C underlying event tune. The sample is normalized to the theoretical calculation performed at next-to-next-to leading order (NNLO) in QCD that includes resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms with top++2.0 [109, 132, 133, 134, 135, 136] yielding a cross section of 253^{+13}_{-15} pb for $\sqrt{s} = 8$ TeV.

As $t\bar{t}$ + jets is the dominant background, a series of detailed studies have been performed by ATLAS to ensure the best possible description for such events. A reweighting is applied to Powheg + Pythia simulations to correctly model data. This procedure is applied to $t\bar{t}$ + light-jets and $t\bar{t}$+$c\bar{c}$ and the reweighting is evaluated from the ratio of the measured differential cross section at $\sqrt{s} = 7$ TeV in data and simulation of top quark p_T and $t\bar{t}$ system p_T [116]. The impact of the reweighting procedure is shown on Figure 6.2 where one can see a good agreement between observation and prediction of the number of jets, correlated to the $t\bar{t}$ system p_T reweighting, and jet p_T related variables, correlated to top quark p_T reweighting.

A comparison with the latest available simulations has been performed for the $t\bar{t}$ + $b\bar{b}$ component. For this purpose a finer categorization of $t\bar{t}$ + heavy-flavour component than the one described in Section 6.4.2 has been used. If two particle jets are both matched to an extra b-quark or extra c-quark each, the event is referred to as $t\bar{t}$ + $b\bar{b}$ or $t\bar{t}$ + $c\bar{c}$; if a single particle jet is matched to a single bottom or charm quark the event is referred to as $t\bar{t}$ + b or $t\bar{t}$ + c; if a single particle jet is matched to a $b\bar{b}$ or $c\bar{c}$ pair, the event is referred to as $t\bar{t}$ + B or $t\bar{t}$ + C respectively. Prediction from Powheg + Pythia has been compared with Madgraph + Pythia, LO generator, showing an overall good agreement between the two. Comparison of Powheg + Pythia with Sherpa OpenLoops, NLO generator, shows differences in some categories. This sample is generated following the four-flavour scheme using the Sherpa 2.0 pre-release and the CT10 PDF set. The renormalization scale (μ_R) 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 factorization and resummation scales are both set to $(E_{\mathrm{T},t} + E_{\mathrm{T},\bar{t}})/2$. Detail of the comparison is shown in Figure 6.3. In particular SherpaOL predicts higher contribution when the production of a second $b\bar{b}$ pair is required. This reflects in the distributions of low region of p_T and mass of $b\bar{b}$ pair and in the p_T of top quark and $t\bar{t}$ system. Since SherpaOL is expected to model better $t\bar{t}+b\bar{b}$ process a reweighting with respect to SherpaOL variables is applied to Powheg + Pythia to account for these effects.

6.4.4 Electroweak backgrounds modeling

The samples of single top quark backgrounds corresponding to the s-channel and Wt production mechanisms are generated with Powheg [125, 126, 127] using the CT10 PDF set [130]. In the case of the Wt-channel, the nominal sample uses the



(a) (b)

(c) (d)

Figure 6.2: Effect of the reweighting of the $t\bar{t}$ system p_T and of the top quark p_T in Powheg + Pythia $t\bar{t}$ sample. Plots on the left are before the reweighting, the one on the right are after. Top row shows the jet multiplicity distribution, bottom row shows H_T^{had} distribution. Plots are made with the $t\bar{t}H$ $(H \to b\bar{b})$ semi-leptonic decay channel [11].

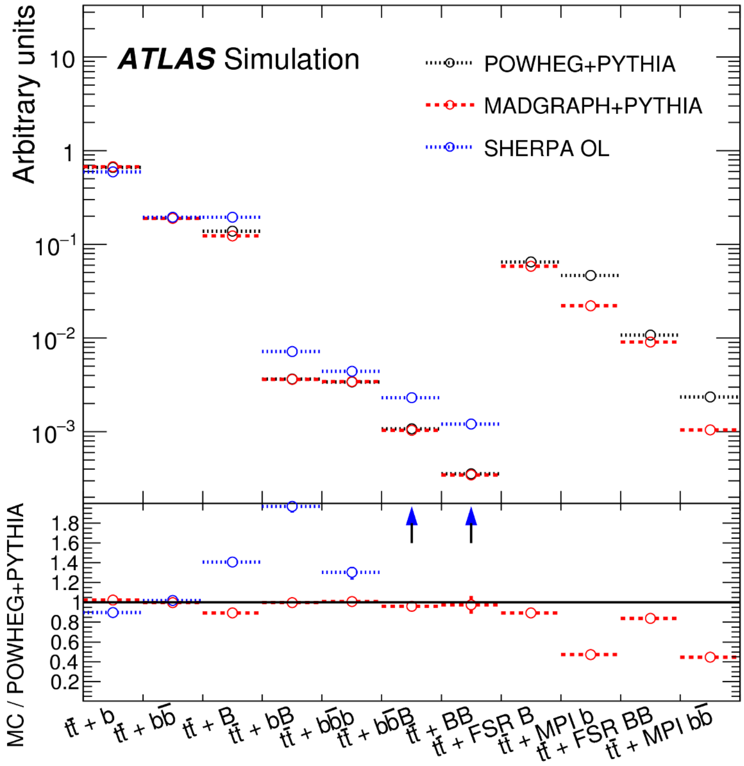


Figure 6.3: Relative contributions of different categories of $t\bar{t}$ + heavy-flavour events in Powheg + Pythia, Madgraph + Pythia and SherpaOL samples. Labels "tt+MPI" and "tt+FSR" refer to events where heavy flavour jets are produced via multi-parton interaction (MPI) or final state radiation (FSR), respectively. These contributions are not included in the SHERPA OL calculation. Arrows indicate that the points lie outside of the scale [11].

Sample	Generator	PDF	Shower	Normalization
$t\bar{t}H$	HELAC-Oneloop	CT10	Pythia 8.1	NLO
$t\bar{t}$ + jets	PowHeg	CT10	Pythia 6.425	NNLO+NNLL
Single top (s-channel, Wt)	PowHeg	CT10	Pythia 6.425	aNNLO
Single top (t-channel)	Acer	MRST	Pythia 6.425	aNNLO
$t\bar{t}V$	Madgraph	CTEQ6L1	Pythia 6.425	NLO

Table 6.1: A summary of basic event generator parameters used to simulate various processes

diagram removal approach to handle the interference starting at NLO. AcerMC [137] is used to simulate the t-channel single top events, with MRSTMCAL PDF set.

All samples are interfaced to PYTHIA 6.425 [131] with the CTEQ6L1 [138] set of parton distribution functions and Perugia2011C [139] underlying event tune. Overlaps between the $t\bar{t}$ and Wt final states are removed [140]. The single top quark samples are normalized to the approximate NNLO theoretical cross sections [141, 142] using the MSTW2008 NNLO PDF set.

The samples of $t\bar{t}V$ ($V = W, Z$) events are generated with the MADGRAPH v5 LO generator [143] and the CTEQ6L1 PDF set. PYTHIA 6.425 with AUET2B tune is used for showering. The $t\bar{t}V$ samples are normalised to the NLO cross section predictions [144, 145].

6.4.5 Common treatment of MC samples

The parton-shower systematics are calculated using POWHEG interfaced with HERWIG. All the event generators using HERWIG are also interfaced to JIMMY v4.31 [146], to simulate the underlying event. All simulated samples utilize PHOTOS 2.15 [147] to simulate photon radiation and TAUOLA 1.20 [148] to simulate τ decays. Finally, all simulated samples include multiple pp interactions and are processed through a simulation [149] of the detector geometry and response using GEANT4 [150], with the exception of the signal samples, for which a fast simulation of the calorimeter response is used. The t-channel single top sample was simulated using AF-II. All simulated samples are processed through the same reconstruction software as the data. Simulated events are corrected so that the object identification efficiencies, energy scales and energy resolutions match those determined in data control samples. Table 6.1 provides a summary of the basic parameters of the MC samples used in the analysis.

6.4.6 Tag Rate Function method in MC ($\mathrm{TRF_{MC}}$)

When requiring high jet and b-tag multiplicity in the analysis, the available MC sample is significantly reduced, leading to large fluctuations in the resulting distributions. This can negatively affect the sensitivity of the analysis, as the corresponding statistical uncertainties on the background templates need to be taken into account in the determination of exclusion limits, and lead to unreliable systematic uncertainties in the predicted distribution shapes. In addition, the observed limits may be biased, depending on how the MC distributions fluctuate with respect to the data in the signal region.

In order to mitigate this problem,the $\mathrm{TRF_{MC}}$ method is applied to all MC samples used in the analysis. This allows the use of all events in the pre-b-tagged sample to predict the normalization and shape after b-tagging. The principle of this method, and the various studies performed to validate it for the sub-leading background in the analysis ($t\bar{t}$), and the $t\bar{t}H$ signal are discussed in this section.

Description of the method

The $\mathrm{TRF_{MC}}$ method use the probability ϵ for a jet to be b-tagged. This is usually parametrized as a function of the jet p_T, η, and flavor, f, as well as of the b-tag operating points, OP. To simplify the notation in what follows the efficiency for a given jet k will be referred as ϵ_k, even though it is intended that it should be read as $\epsilon(p_{T,k}, \eta_k, f_k)$. Using $\mathrm{TRF_{MC}}$ method one can calculate the probability $P^{ev}(j, n_b)$ ($P^{ev}(j, \geq n_b)$) that in an event m of n_j jets exactly (at least) $n_b \leq n_j$ jets are b-tagged, starting from the individual probability for each jet k of the events to be b-tagged (ϵ_k). The different probabilities are defined as described in Section 4.2.4:

$$P^{ev}(j,0) = P_{=0} \ \ , \ \ P^{ev}(j,1) = P_{=1} \ \ , \ \ P^{ev}(j,2) = P_{=2} \ \ , \ \ P^{ev}(j,3) = P_{=3} \tag{6.1}$$

and

$$P^{ev}(j, \geq 4) = 1 - P^{ev}(j,3) - P^{ev}(j,2) - P^{ev}(j,1) - P^{ev}(j,0). \tag{6.2}$$

When writing these formulas the assumption that the probabilities ϵ_k of tagging different jets in the events are uncorrelated is made. The above probabilities P^{ev} can be used as event weights, allowing to use all events in the pre-b-tagged sample to predict the normalization and the shape after b-tagging.

To estimate the number of events, N_n^{Pred}, present in a subsample where each event has exactly n b-tagged jets it is necessary to sum the $N_{\geq 0}$ probabilities for each event to have this b-tagging requirement:

$$N_n^{Pred} = \sum_{i=0}^{i=N_{\geq 0}} P^{ev}(i,n) \tag{6.3}$$

where $N_{\geq 0}$ indicates the number of events having at least 0 b-tagged events, i.e. the total sample of events before b-tagging requirement.

By using the $\mathrm{TRF_{MC}}$ method it is possible to simulate the presence of n b-tagged jets in the event. The implementation pass by the evaluation of the n_c probabilities w_i, computed with the $\mathrm{TRF_{MC}}$, for each combination of n b-tagged jets on n_j jets in the event , i.e.:

$$n_c = \binom{n_j}{n} \tag{6.4}$$

A random number r uniformly distributed between 0 and W, where $W = \sum_{i=1}^{n_c} w_i$ is thrown. The k^{th} combination such as $\sum_{i=1}^{k-1} w_i < r \leq \sum_{i=1}^{k} w_i$ is chosen, and the jets considered as b-tagged in the k^{th} combination are considered as b-tagged in the evaluation of observables that make direct use of the b-tagging information, such as the leading b-tagged jet p_T, average ΔR among b-tagged jets, etc.

Comparison of predictions using $\mathrm{TRF}_{\mathrm{MC}}$ and direct b-tagging

The above assumption on uncorrelated TRF's, as well as their correct determination has to be verified by comparing the number of events and several distributions of different physical quantities obtained by applying direct b-tagging with those obtained without b-tagging but using P^{ev} as event weight.

This comparison is carried out on $t\bar{t}$ Powheg +Pythia MC events subdivided for their heavy flavor content in $t\bar{t}$ + light, $t\bar{t}$ + $c\bar{c}$ and $t\bar{t}$ + $b\bar{b}$, as described in Section 6.4.2. The result is shown below in the N_b with $b = 1, 2, 3$ and ≥ 4 categories, Tables 6.2, 6.3 and 6.4 compare the total number of events obtained by direct b-tagging and applying event weights as described above for events passing the analysis event selection for different jet multiplicities. Table 6.5 shows the same comparison for the $t\bar{t}H$ sample. The agreement is overall good in all the samples and all the regions. Small tension is seen in Table 6.5 for $t\bar{t}H$ events. To compensate for such residual differences all the MC samples are corrected with a normalization factor.

	$t\bar{t}$+light		$t\bar{t}c\bar{c}$		$t\bar{t}b\bar{b}$	
N_b	Direct b-tag	N_b	Direct b-tag	N_b	Direct b-tag	N_b
1	130811	129585 ± 251	17347	17236 ± 91	4892	4694 ± 43
2	87004	90082 ± 190	11515	11835 ± 68	4903	4950 ± 45
3	8336	8738 ± 25	1617	1750 ± 12	1855	1966 ± 22
≥ 4	253	250 ± 1	96	97 ± 1	241	252 ± 5

Table 6.2: Number of events in different n_b-tag categories obtained by direct b-tagging and by applying event weights based on $\mathrm{TRF}_{\mathrm{MC}}$ to events with exactly 6 reconstructed hadronic jets in the events. The comparison is shown for different MC samples, $t\bar{t}$+light, $t\bar{t}c\bar{c}$ and $t\bar{t}b\bar{b}$.

	$t\bar{t}$+light		$t\bar{t}c\bar{c}$		$t\bar{t}b\bar{b}$	
N_b	Direct b-tag	N_b	Direct b-tag	N_b	Direct b-tag	N_b
1	107853	106751 ± 227	20056	19608 ± 96	5494	5449 ± 45
2	73548	75874 ± 174	13751	14324 ± 75	6161	6166 ± 49
3	7442	7817 ± 24	2196	2345 ± 14	2702	2760 ± 26
≥ 4	236	247 ± 1	119	153 ± 1	463	462± 7

Table 6.3: Number of events in different n_b-tag categories obtained by direct b-tagging and by applying event weights based on $\mathrm{TRF}_{\mathrm{MC}}$ to events with exactly 7 reconstructed hadronic jets in the events. The comparison is shown for different MC samples, $t\bar{t}$+light, $t\bar{t}c\bar{c}$ and $t\bar{t}b\bar{b}$.

	$t\bar{t}$+light		$t\bar{t}c\bar{c}$		$t\bar{t}b\bar{b}$	
N_b	Direct b-tag	N_b	Direct b-tag	N_b	Direct b-tag	N_b
1	89437	88155 ± 205	24111	23756 ± 104	7045	6947 ± 49
2	61728	63904 ± 159	17544	18187 ± 84	8607	8561 ± 58
3	6552	6959 ± 23	3158	3338 ± 18	4328	4416 ± 34
≥ 4	214	247 ± 1	267	266 ± 2	1013	1008 ± 12

Table 6.4: Number of events in different n_b-tag categories obtained by direct b-tagging and by applying event weights based on TRF_{MC} to events with at least 8 reconstructed hadronic jets in the events. The comparison is shown for different MC samples, $t\bar{t}$+light, $t\bar{t}c\bar{c}$ and $t\bar{t}b\bar{b}$.

	$t\bar{t}H$					
	Exactly 6 jets		Exactly 7 jets		At least 8 jets	
N_b	Direct b-tag	N_b	Direct b-tag	N_b	Direct b-tag	N_b
1	28260	27429 ± 104	37870	37037 ± 119	65322	63927 ± 154
2	29766	29962 ± 108	41857	42439 ± 127	74666	75596 ± 168
3	14766	15289 ± 67	22957	23400 ± 83	43867	44354 ± 115
≥ 4	3609	3881 ± 25	6934	7043 ± 35	15265	15939 ± 57

Table 6.5: Number of events in different n_b-tag categories obtained by direct b-tagging and by applying event weights based on TRF_{MC} to events with exactly 6, exactly 7 and at least 8 reconstructed hadronic jets in the events. The comparison is shown for $t\bar{t}H$ sample.

Figure 6.4 represents the comparison of distributions for the first 4 leading jets in the $t\bar{t} + b\bar{b}$ sample for events with exactly 6 jets and 1 to 4 b-tagged jets, Figure 6.5 represents the comparison of distributions for the first 4 leading b-tagged jets in the $t\bar{t} + c\bar{c}$ sample for events with exactly 7 jets and 1 to 4 b-tagged jets, and Figure 6.6 represents the comparison of distributions for aplanarity[1], centrality[2], H_T and aplanarity computed using only b-tagged jets (called aplab), in the $t\bar{t}$ + light-jets sample for events with at least 8 jets and 1 to 4 b-tagged jets. Finally Figure 6.7 represents the comparison of distributions for H_T in $t\bar{t}H$ events. Overall good agreement is observed for all the MC samples.

---

[1] $1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor.

[2] Sum of the p_T divided by the E for all the jets.

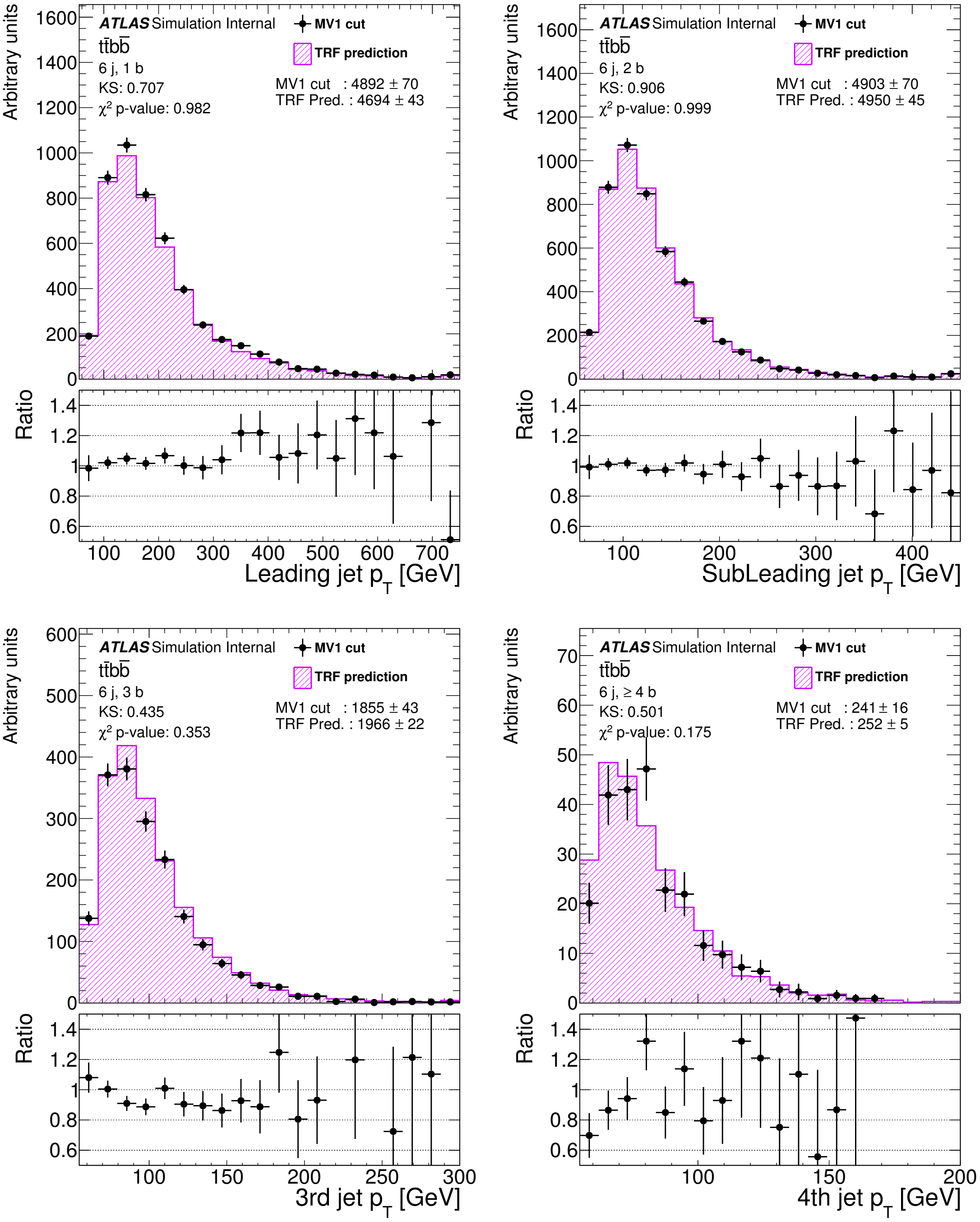


Figure 6.4: Leading, sub-leading, third and fourth leading jet p_T distributions in events obtained by direct b-tagging (black circles) and by applying event weights (purple histogram) as well as their ratios (bottom panels), for $t\bar{t}b\bar{b}$ events with exactly 6 reconstructed jets for different b-tagging multiplicities, respectively with $n_b = 1, 2, 3$ and $n_b \geq 4$.

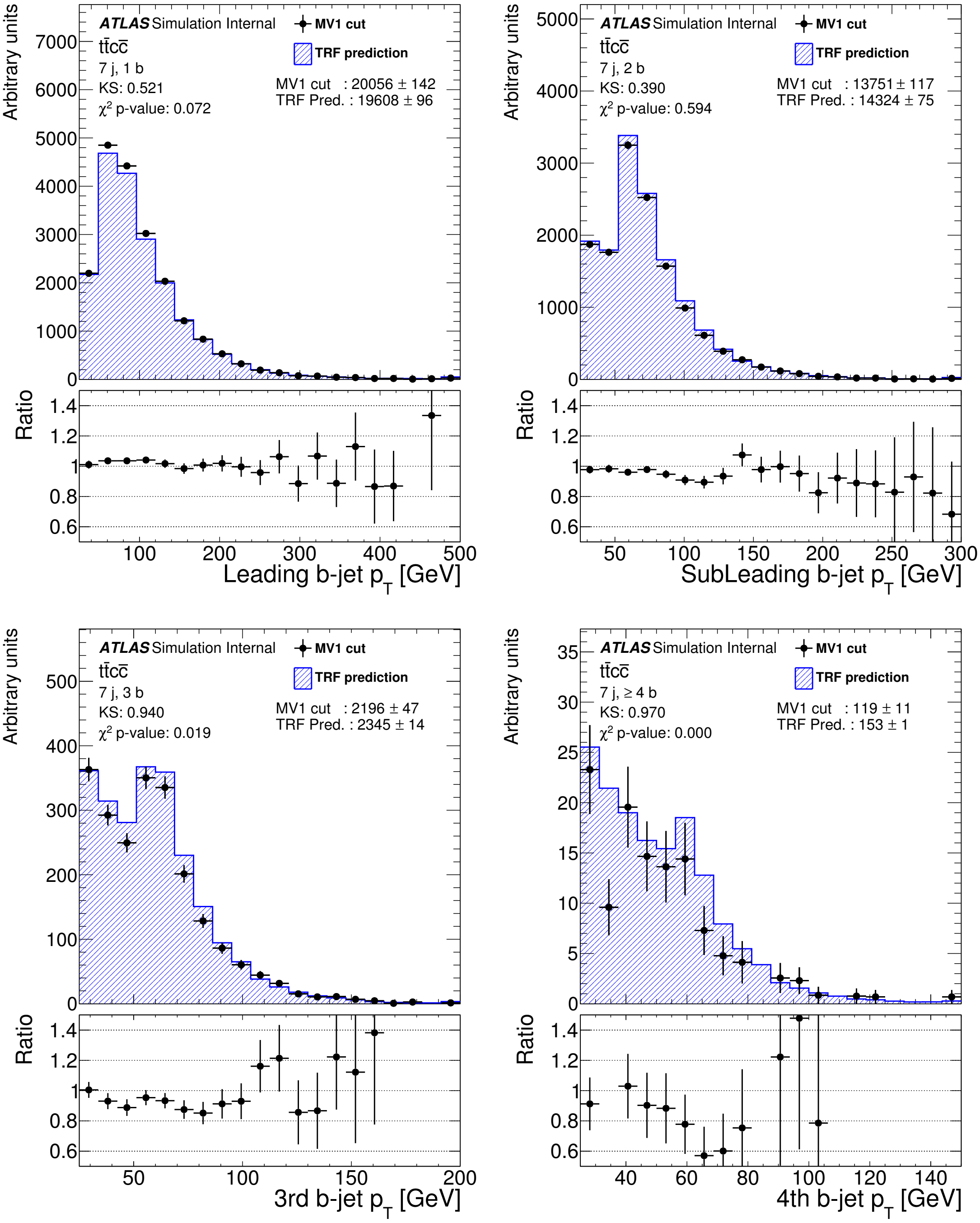


Figure 6.5: Leading, sub-leading, third and fourth leading b-tagged jet p_T distributions in events obtained by direct b-tagging (black circles) and by applying event weights (blue histogram) as well as their ratios (bottom panels), for $t\bar{t}b\bar{b}$ events with exactly 7 reconstructed jets for different b-tagging multiplicities, respectively with $n_b = 1, 2, 3$ and $n_b \geq 4$.

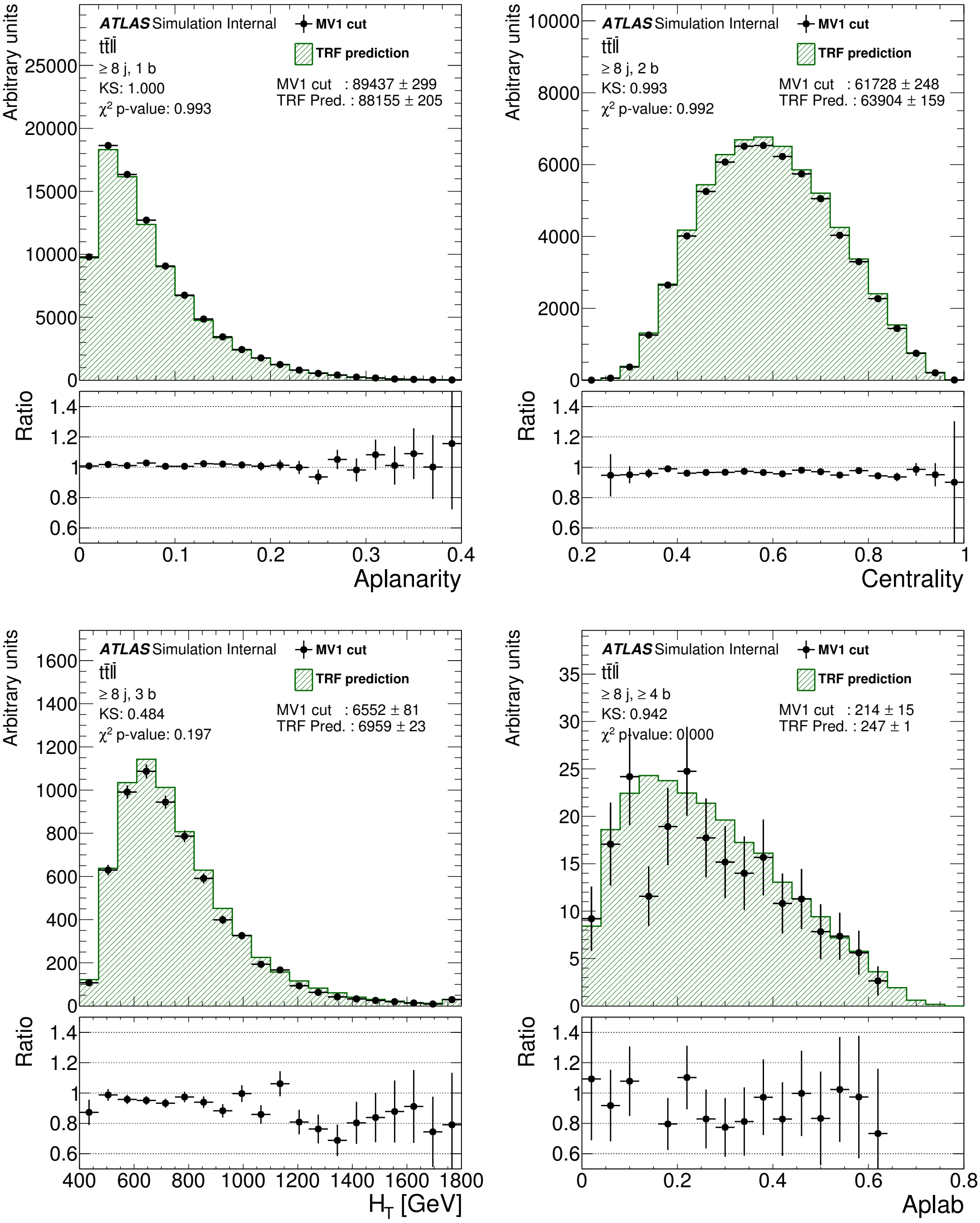


Figure 6.6: Aplanarity, centrality, H_T and aplanarity computed using only b-tagged jets, distributions in events obtained by direct b-tagging (black circles) and by applying event weights (green histogram) as well as their ratios (bottom panels), for $t\bar{t}$+light events with at least 8 reconstructed jets for different b-tagging multiplicities, respectively with $n_b = 1, 2, 3$ and $n_b \geq 4$.

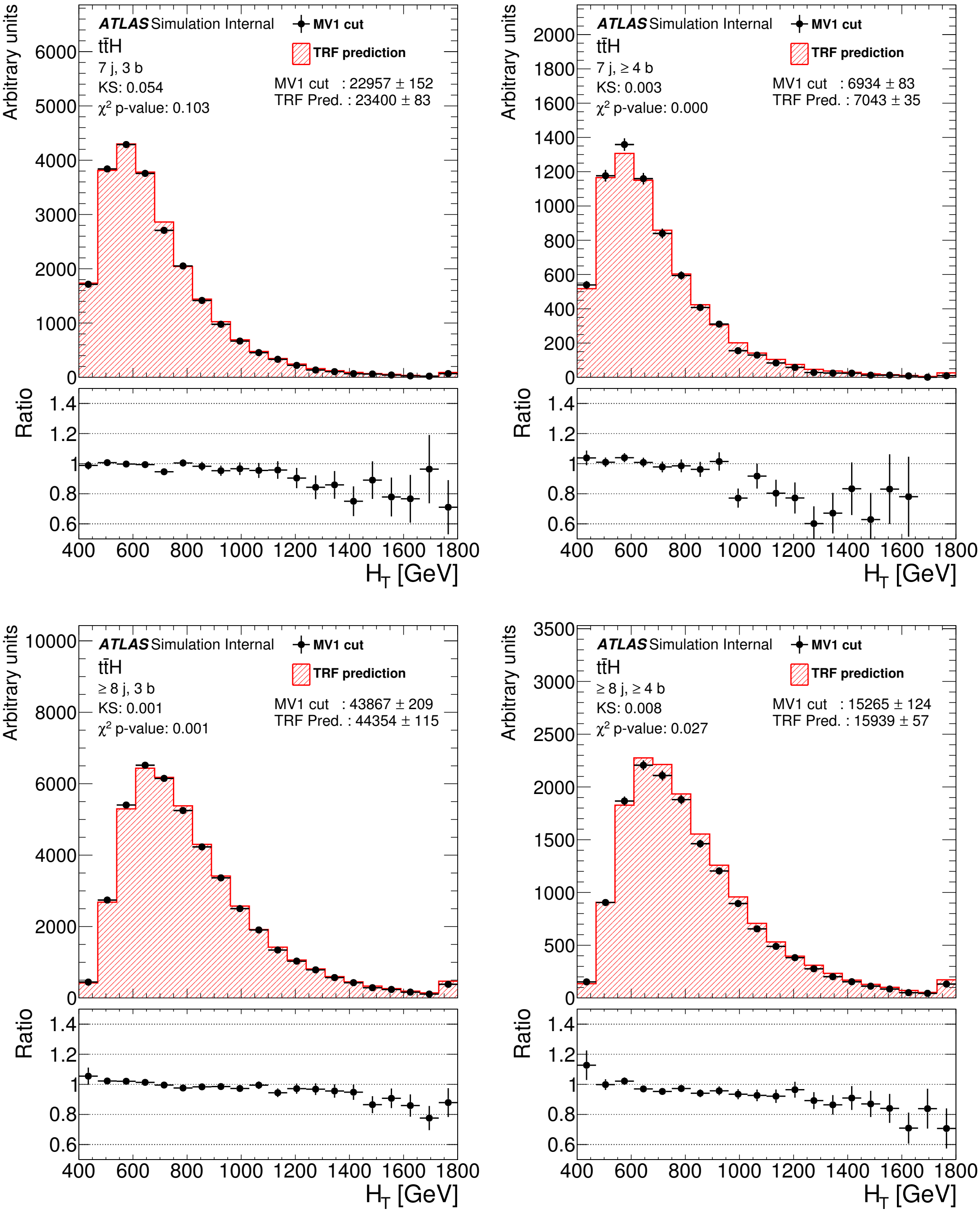


Figure 6.7: H_T distribution in events obtained by direct b-tagging (black circles) and by applying event weights (red histogram) as well as their ratios (bottom panels), for $t\bar{t}H$ events with exactly 7 and at least 8 reconstructed jets of which exactly 3 and $\geq$ 4 b-tagged jets.

6.5 Multijet background estimation: TRF$_{\mathrm{MJ}}$

The dominant background after the $t\bar{t}H$ $(H \to b\bar{b})$ fully-hadronic event selection is multijet production. To estimate it a data-driven method is developed, referred to as TRF$_{\mathrm{MJ}}$, which is based on the probability for a multijet event to be b-tagged, $\varepsilon_{\mathrm{MJ}}$, derived as a function of the jet p_T, $|\eta|$ and a variable correlated with the distance of any other b-tagged jet in the event. Using events with exactly 2 b-tagged jets, the multijet background is estimated for events with higher b-tag multiplicity, 3 and 4 b-tagged jets, the *signal-rich* region where the $t\bar{t}H$ search is performed. Figure 6.8 shows a scheme of the application of the TRF$_{\mathrm{MJ}}$ method.

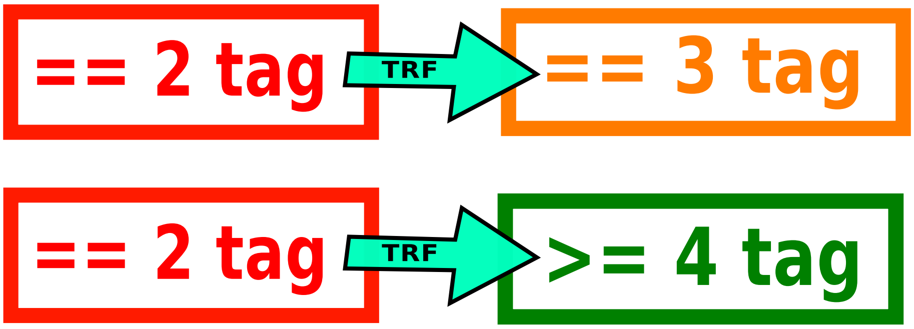


Figure 6.8: TRF$_{\mathrm{MJ}}$ method uses events with exactly 2 b-tagged jets to estimate the multijet background in events with higher b-tag multiplicity, 3 and and at least 4 b-tagged jets, regions where the $t\bar{t}H$ search is performed.

This method, similar to the TRF$_{\mathrm{MC}}$ method described in Section 6.4.6, presents three main differences:

- **No knowledge of the flavor of the jet:** since the probability $\varepsilon_{\mathrm{MJ}}$ is extracted in data, the flavour of the jet is undefined. This implies that the flavor composition of the sample where the $\varepsilon_{\mathrm{MJ}}$ is extracted (TRF$_{\mathrm{MJ}}$*estimation region*) has to be as similar as possible to the region where the TRF$_{\mathrm{MJ}}$ method is applied. Also, information about correlation between b-tagged jets, i.e. their relative angular distribution, carry valuable information about flavor composition and are used in the $\varepsilon_{\mathrm{MJ}}$ description.

- **Knowledge of the amount of multijet background in a given point of the phase space:** The amount of multijet background events in a given low b-tagged jet multiplicity region (TRF$_{\mathrm{MJ}}$*multijet extraction region*) where the method is applied has to be estimated a priori. This is done in the regions with exactly 2 b-tagged jets, where the multijet background is defined by subtracting the MC background from data.

- **Predict multijet in a orthogonal sample from where it is extracted:** The TRF$_{\mathrm{MJ}}$ extraction region needs to be independent from the signal regions in order to define the multijet background without making any assumption on the presence of signal.

6.5.1 TRF$_{\mathrm{MJ}}$: estimation region

A dedicated data sample, collected requiring a set of high-p_T single jet triggers and low multiplicity multi-jet triggers, listed in Table 6.6 is used to select a multijet dominated region, where the presence of $t\bar{t}$ production is negligible. The presence of at least 3 jets with $p_T > 25$ GeV and $|\eta| < 2.5$, out of which there were at least 2 b-tagged jets is also requested. It has been tested that only 10% of events of this data sample coincides with those used for the search of the $t\bar{t}H$ signal. After excluding the two jets with the highest b-tagging weight from the events, the probability is evaluated that one of the remaining jets in the event is b-tagged. The efficiency $\varepsilon_{\mathrm{MJ}}$ obtained with this method is presented in Figure 6.9.

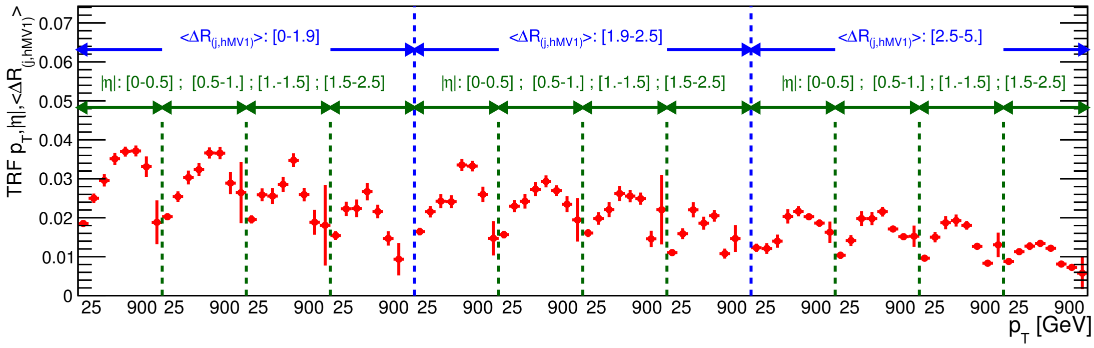


Figure 6.9: One dimensional view of the dependence of $\varepsilon_{\mathrm{MJ}}$ as a function of the jets' p_T and $|\eta|$ and $\left\langle \Delta R_{(j,\mathrm{hMV1})} \right\rangle$. Where $\left\langle \Delta R_{(j,\mathrm{hMV1})} \right\rangle$ is the average of the distances of the jet from the two jets with highest MV1 weight in the event.

Trigger name	LVL1 thr. [GeV]	LVL2 thr. [GeV]	EF thr. [GeV]	L [fb^{-1}]
`EF_j360_a4tchad`	75	165	360	20.3
`EF_j280_a4tchad`	75	165	280	1.16
`EF_j220_a4tchad`	75	165	220	0.26
`EF_j180_a4tchad`	75	165	180	0.079
`EF_j145_a4tchad`	75	140	145	0.036
`EF_3j55_a4tchad`	3 × 15	3 × 15	3 × 55	0.029

Table 6.6: List of triggers used for selecting the TRF$_{\mathrm{MJ}}$ extraction region, together with the integrated luminosity collected by each trigger. The E_T thresholds at each trigger level are also shown. `EF_3j55_a4tchad` is a multi-jet trigger requiring at least three objects at each trigger level.

6.5.2 Algorithm used for the application of the TRF$_{\mathrm{MJ}}$

The TRF$_{\mathrm{MJ}}$ method is carried out by constructing a pseudo data sample containing n jets ($n \geq 6$) out of which n_b are b-tagged ($n_b \geq 3$). The method is applied on a real data sample where in each event 2 jets are already b-tagged. At the analysis level, having subtracted the contribution of the non multijet background obtained with MC simulation, a weight $w(n, n_b)$ built from $\varepsilon_{\mathrm{MJ}}$ is applied to each event of n jets. Moreover, in each pseudo event $\varepsilon_{\mathrm{MJ}}$ is used to promote n_b-2 un-tagged jets to b-tagged jets. The dependence of $w(n, n_b)$ on $\varepsilon_{\mathrm{MJ}}$ is analogous to that described (for P^{ev}) in Section 6.4.6, with the difference that the derivation of $w(n, n_b)$ takes into account that the starting sample contains 2 b-tagged jets.

For any event in $N(\geq 2b)$, where $N(n_b)$ represents a set of events with n_b b-tagged jets, it is possible to estimate the probability to be an event with exactly 3 b-tagged jets. This particular configuration is referred to as $2b + 1b$ to indicate that what is predicted is the existence of another b-tagged jet in addition to the two whose existence have been verified (b_1 and b_2).

$$P_i(\geq 2 \rightarrow 2b + 1b) = \sum_{j \neq b_1, b_2}^{n_i} \varepsilon_j^i \cdot \prod_{k \neq j, b_1, b_2} (1 - \varepsilon_k^i), \tag{6.5}$$

where ε_j^i is the value of $\varepsilon_{\mathrm{MJ}}$ for a given jet j in the ith event having a number of jets equal to n_i. The number of events with exactly 3 b-jets is given by

$$N(2b + 1b) = \sum_{i}^{N(\geq 2b)} P_i(\geq 2 \rightarrow 2b + 1b) \tag{6.6}$$

Since in general ε_j^i depends on the relative position of the jet j with the already b-tagged jets in ith the event, the position of the third b-tagged jets has to be predicted for each of the $N(\geq 2b)$ event in order to predict $N(2b + mb)$ with $m \geq 2$. This is done by choosing one particular configuration of b-tagged jets, using the same technique described in Section 6.4.6, the chosen configuration consists of two b-tagged jets and one *promoted* jet. The promoted jet is only used to evaluate ε_j^i. Once this is done, $N(\geq 4b)$ is estimated by:

$$N(\geq 2b + 2b) = \sum_{i}^{N(\geq 2b)} P_i(\geq 2b + 1b \rightarrow \geq 2b + 2b) \tag{6.7}$$

where:

$$P_i(\geq 2b + 1b \rightarrow \geq 2b + 2b) = 1 - P_i(\geq 2b + 1b \rightarrow 2b + 1b), \tag{6.8}$$

and

$$P_i(\geq 2b + 1b \rightarrow \geq 2b + 2b) = \sum_{j \neq b_1, b_2, p_1}^{n_i} \varepsilon_j^i \cdot \prod_{k \neq j, b_1, b_2, p} (1 - \varepsilon_k^i) \tag{6.9}$$

where p is the index of the promoted jet. For any event with ≥ 2 b-tagged jets, the probability of obtaining an event with exactly 3 or at least 4 b-tagged jets is calculated using the TRF$_{\mathrm{MJ}}$ method.

Application of $\mathbf{TRF_{MJ}}$ to a sample with exactly 2 *b*-tagged jets

In the analysis regions, the multijet background is defined in data as the difference between data and the known MC background in a region with exactly 2 b-tagged jets and the events with ≥ 2 or more b-tagged jets are estimated from the region with exactly 2 b-tagged jets using the inverse of $\varepsilon_{\mathrm{MJ}}$. Therefore, the $N(\geq 2b)$ events with at least 2 b-tagged jets need to be estimated in order to use the formalism described in 6.5.2. This is done by using the inverse of the probability $P_{\geq 2b \rightarrow 2b}$ estimated on the $N(2b)$ events with exactly 2 b-tagged jets. So, from:

$$P_i(\geq 2b \rightarrow 2b) = \prod_{j \neq b_1, b_2}^{n_i} (1 - \varepsilon_j), \tag{6.10}$$

we can infer $N(\geq 2b)$ by weighting each of the $N(2b)$ event by the inverse of $P_i(\geq 2b \rightarrow 2b)$:

$$N(\geq 2b) = \sum_{l=1}^{N(2b)} \frac{1}{P_l(\geq 2b \rightarrow 2b)} = \sum_{l=1}^{N(2b)} \frac{1}{\prod\limits_{j \neq b_1, b_2}^{M_l} (1 - \varepsilon_j^l)}, \tag{6.11}$$

where l runs over the $N(2b)$ events with exactly two b-tagged jets. Then Equation 6.6 becomes

$$N(2b + 1b) = \sum_{l}^{N(2b)} \frac{P_l(\geq 2b \rightarrow 2b + 1b)}{P_l(\geq 2b \rightarrow 2b)} \tag{6.12}$$

Then the same steps need to be taken to go to higher multiplicity, extending this notation to Equation 6.7 one needs to estimate $N(\geq 2b + 1b)$ as:

$$N(\geq 2b + 1b) = \sum_{l}^{N(2b)} \frac{P_l(\geq 2b \rightarrow 2b + 1b)}{P_l(\geq 2b \rightarrow 2b) P_l(\geq 2b + 1b \rightarrow 2b + 1b)} = \sum_{l}^{N(2b)} w(n_l, 3) \tag{6.13}$$

and finally:

$$N(\geq 2b + 2b) = \sum_{l}^{N(2b)} \frac{P_l(\geq 2b \rightarrow 2b + 1b) P_l(\geq 2b + 1b \rightarrow \geq 2b + 2b)}{P_l(\geq 2b \rightarrow 2b) P_l(\geq 2b + 1b \rightarrow 2b + 1b)} = \sum_{l}^{N(2b)} w(n_l, \geq 4) \tag{6.14}$$

6.5.3 Validation of $\mathbf{TRF_{MJ}}$ predictions in data

To validate the $\mathrm{TRF_{MJ}}$ method a test is performed using the same data sample where $\mathrm{TRF_{MJ}}$ is estimated, described in Section 6.5.1. The sample is divided in two, $\mathrm{TRF_{MJ}}$ is then estimated in one sub-sample and applied to the other. Table 6.7 shows the comparison between the number of predicted events and the observed ones, for different jet and b-tagged jet multiplicity bins. Good agreement is observed, the maximum difference is 5% for events with at least four jets and exactly three b-tagged jet.

Figures 6.10 and 6.11 show the distributions of the leading and subleading jet p_T where, despite the low statistics, an overall good agreement is observed between direct b-tagging and the prediction by $\mathrm{TRF}_{\mathrm{MJ}}$. Figure 6.12 shows H_T and the object missing energy, S_T. For these distributions it is possible to see a mismodelling which is covered by the systematics of the method as it is explained in Section 6.9.3. Finally Figure 6.13 shows distributions of correlation between b-tagged jets.

N_{b-tags}		$N_{jets} == 3$	$N_{jets} \geq 4$
3	data	641	7585
	$\mathrm{TRF}_{\mathrm{MJ}}$ prediction	632 ± 4	7952 ± 25
≥ 4	data		425
	$\mathrm{TRF}_{\mathrm{MJ}}$ prediction		452 ± 2

Table 6.7: Number of events in different N_b-tag categories containing N_j jets obtained by direct b-tagging (data) and predicted from the sample containing exactly 2 b-tagged jets by applying the $\mathrm{TRF}_{\mathrm{MJ}}$ method.

6.5.4 Validation of $\mathrm{TRF}_{\mathrm{MJ}}$ predictions in di-jet MC

A test similar to the one performed in data, using the same definition of the $\mathrm{TRF}_{\mathrm{MJ}}$ estimation region (Section 6.5.3), is performed using PYTHIA8 [99] di-jet MC, to test the method in a more controlled sample. This MC sample has been described in Section 4.2.2.

MC events are required to pass the trigger `EF_j360_a4tchad`, the most representative among the ones used in the data $\mathrm{TRF}_{\mathrm{MJ}}$ estimation region.[3] As for data the offline preselection consists of requiring ≥ 3 jets with $p_T \geq 25$ GeV and ≥ 2 b-tagged jets and the $\mathrm{TRF}_{\mathrm{MJ}}$ is parametrized with respect to the offline jet's p_T and $|\eta|$ and $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$, i.e. the average of the distances of the jet from the two jets with highest MV1 weight in the event, see Figure 6.14.

Table 6.8 shows the normalization obtained with the $\mathrm{TRF}_{\mathrm{MJ}}$ method compared to direct tagging for all different p_T slices making up the PYTHIA8 di-jet MC sample. It can be noted that for p_T smaller than 1 TeV the agreement between $\mathrm{TRF}_{\mathrm{MJ}}$ prediction and direct b-tagging is good within the statistical uncertainty, while for high p_T the agreement worsen, linked to the fact that the $\mathrm{TRF}_{\mathrm{MJ}}$ map used for this exercise stops at 900 GeV (Figure 6.14), below the generated leading truth jet p_T for these samples. Also no events from samples with truth leading jet p_T below 200 GeV are present in the table since those do not pass the single jet trigger threshold used.

For the sample with truth leading jet p_T between 500 GeV and 1 TeV the distributions for the leading jet and leading b-tagged jet p_T are shown in Figure 6.15 for events with at least 4 jets.

---

[3] Full list of the triggers used in the $\mathrm{TRF}_{\mathrm{MJ}}$ extraction region are presented in Table 6.6.

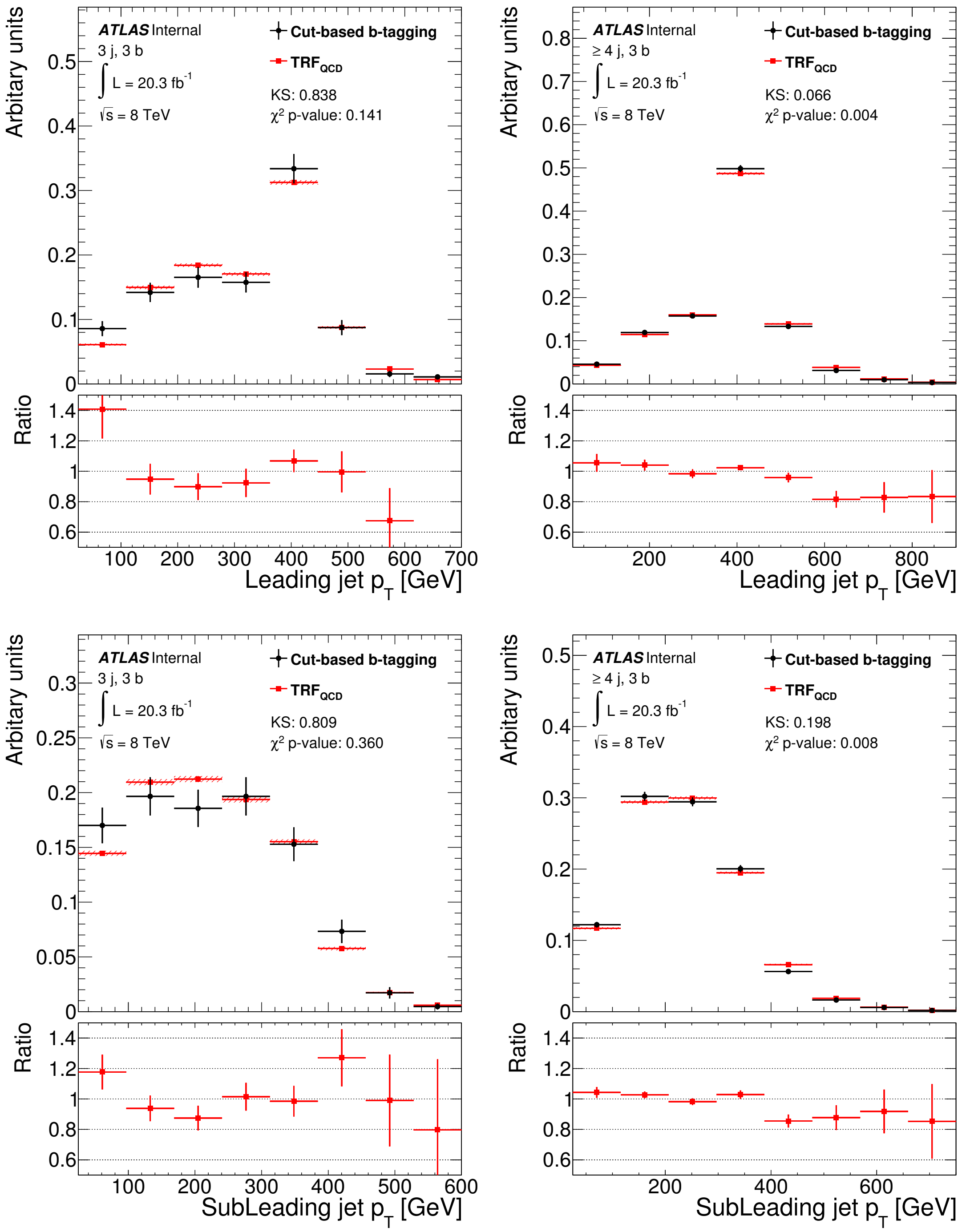


Figure 6.10: Comparison of shapes of the leading (*top*) and subleading (*bottom*) jet p_T distributions obtained with cut-based b-tagging (black circles) and with the TRF_{MJ} method (red squares) for events with exactly 3 jets and exactly 3 b-tagged jets (*left*) and at least 4 jets and exactly 3 b-tagged jets (*right*).

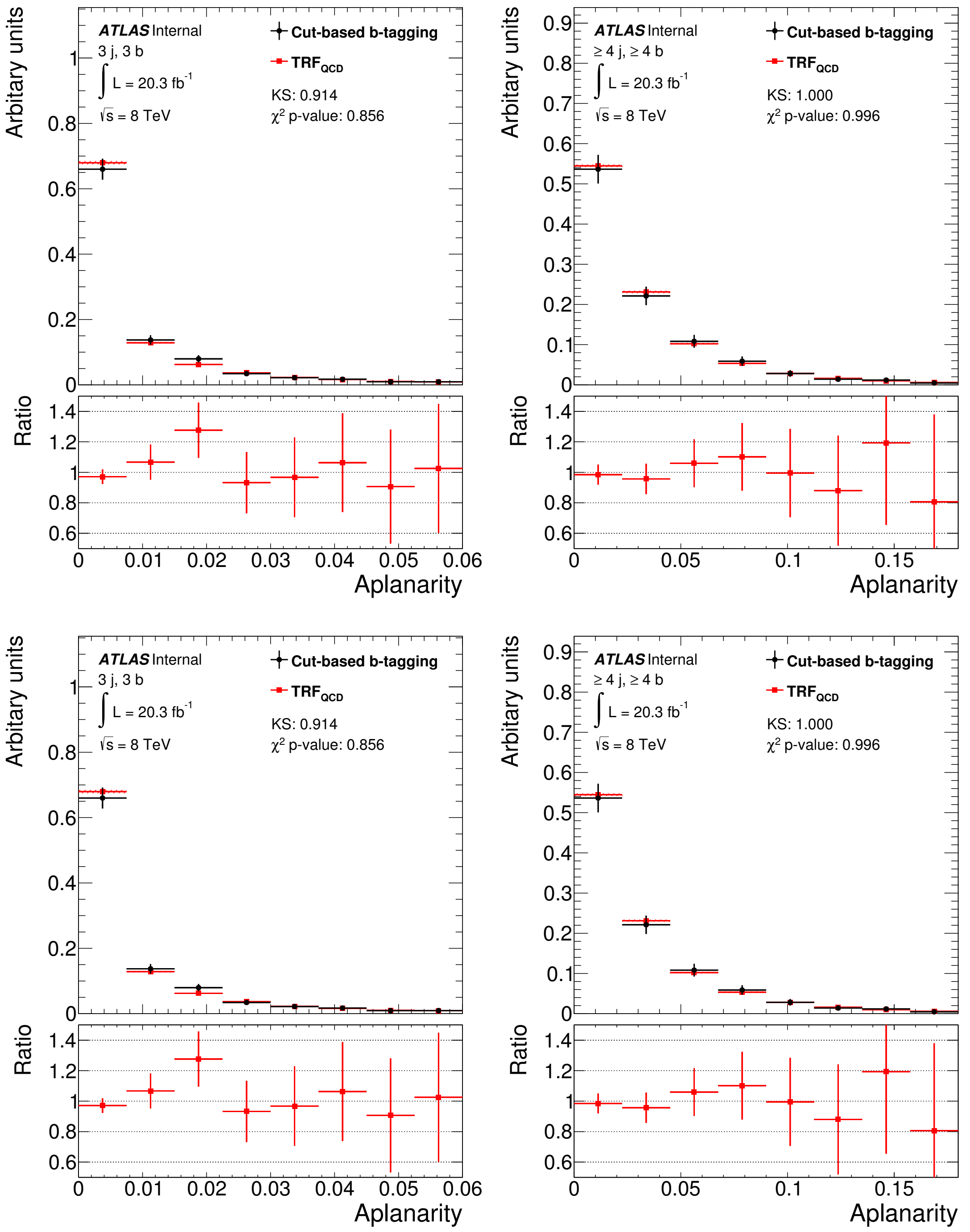


Figure 6.11: Comparison of shapes of aplanarity (*top*) and the same quantity built by only using b-tagged jets (*bottom*) in events obtained with cut-based b-tagging (black circles) and with the TRF$_{\mathrm{MJ}}$ method (red squares) for events with exactly 3 jets and exactly 3 b-tagged jets (*left*) and at least 4 jets and at least 4 b-tagged jets (*right*).

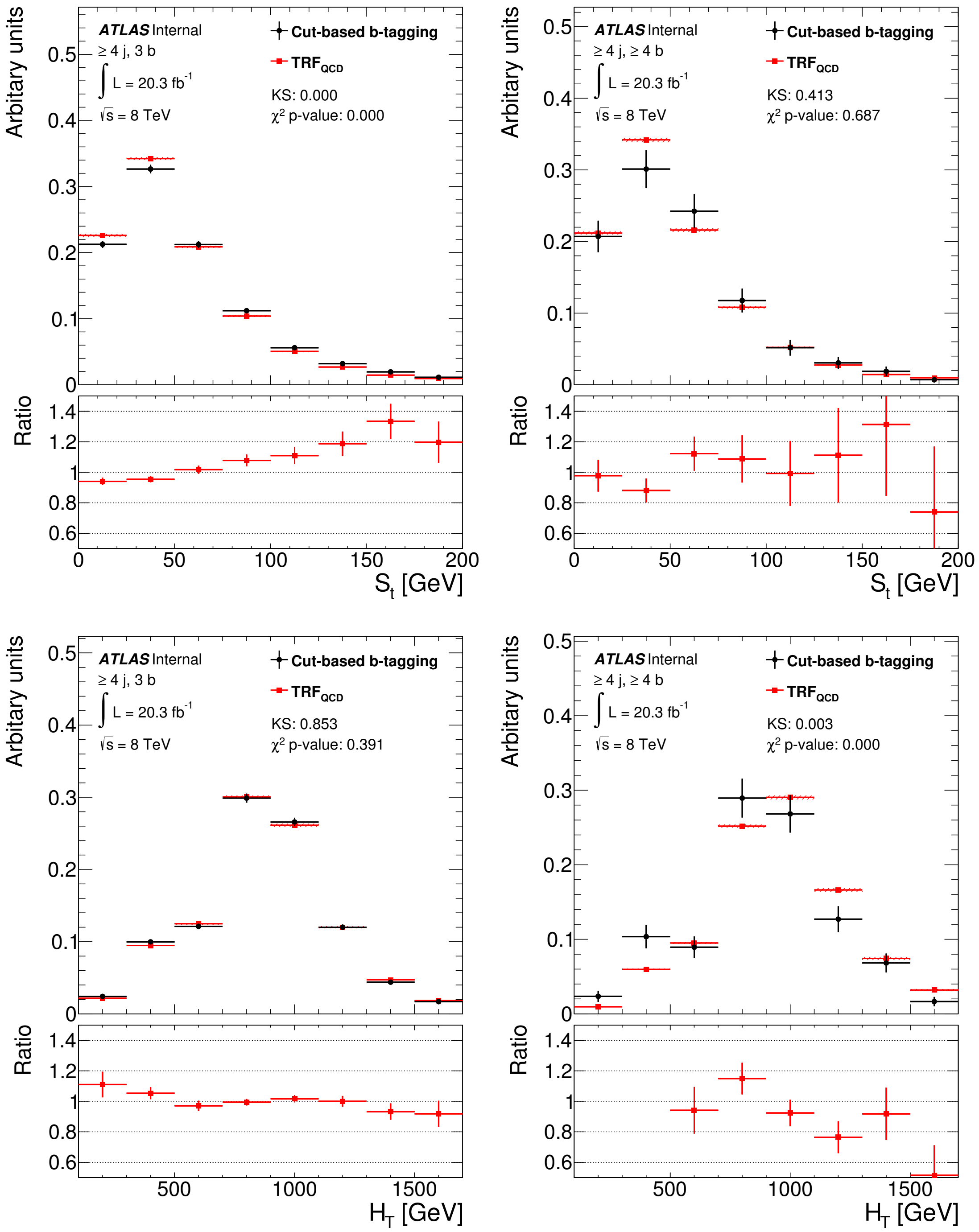


Figure 6.12: Comparison of shapes for the object missing energy, S_T, (*top*) and H_T (*bottom*) in events obtained with cut-based b-tagging (black circles) and with the TRF_{MJ} method (red squares) for events with at least 4 jets and exactly 3 b-tagged jets (*left*) and at least 4 jets and at least 4 b-tagged jets (*right*).

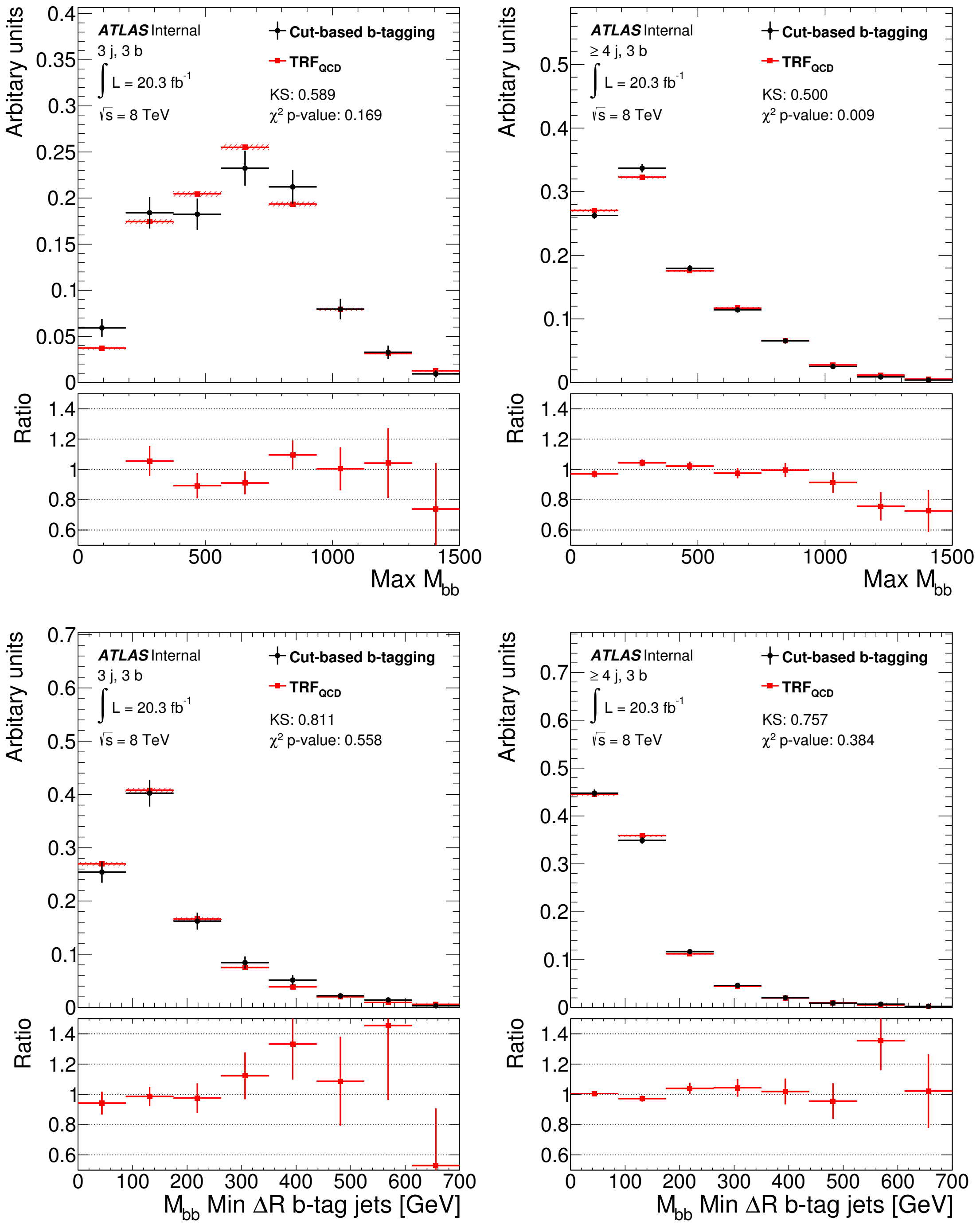


Figure 6.13: Comparison of shapes for the invariant mass of the two b-tagged jets having the max invariant mass (*top*) and the minimum ΔR distance (*bottom*) in events obtained with cut-based b-tagging (black circles) and with the TRF$_{\text{MJ}}$ method (red squares) for events with exactly 3 jets and exactly 3 b-tagged jets (*left*) and at least 4 jets and exactly 3 b-tagged jets (*right*).

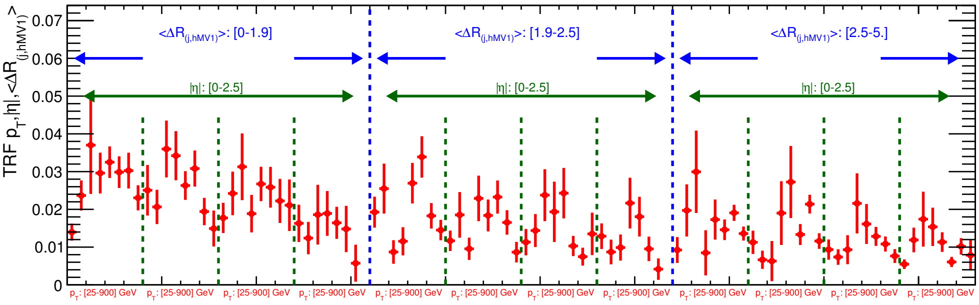


Figure 6.14: One dimensional view of the dependence of $\varepsilon_{\mathrm{MJ}}$ derived in Monte Carlo PYTHIA8 di-jets events as a function of jet p_T and $|\eta|$ and $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$.

	All samples	leading truth jet p_T [GeV] 200-500	500-1000	1000-1500 $(\times 10^{-3})$	1500-2000 $(\times 10^{-6})$	2000+ $(\times 10^{-7})$
		3 b-tagged jets				
$\mathrm{TRF}_{\mathrm{MJ}}$	15.5 ± 0.1	12.9 ± 0.1	2.59 ± 0.03	15.8 ± 0.2	250 ± 3	48 ± 0.4
Direct b-tagging	14.6 ± 0.5	12.3 ± 0.5	3.10 ± 0.03	13.4 ± 0.6	207 ± 9	39 ± 1.2
		≥ 4 b-tagged jets				
$\mathrm{TRF}_{\mathrm{MJ}}$	0.89 ± 0.01	0.71 ± 0.01	17.3 ± 0.2	1.18 ± 0.02	18.2 ± 0.3	3.3 ± 0.04
Direct b-tagging	0.9 ± 0.1	0.8 ± 0.1	14 ± 2	0.9 ± 0.1	10 ± 2	2.0 ± 0.3

Table 6.8: Number of events with at least 4 reconstructed jets and exactly 3 or at least 4 b-tagged jets, obtained by direct b-tagging in PYTHIA di-jet MC events compared with the prediction of the $\mathrm{TRF}_{\mathrm{MJ}}$ method. Numbers in the second column do not correspond to the sum of the other columns since the former are evaluated taking into account the different cross sections of each p_T slice sub-sample.

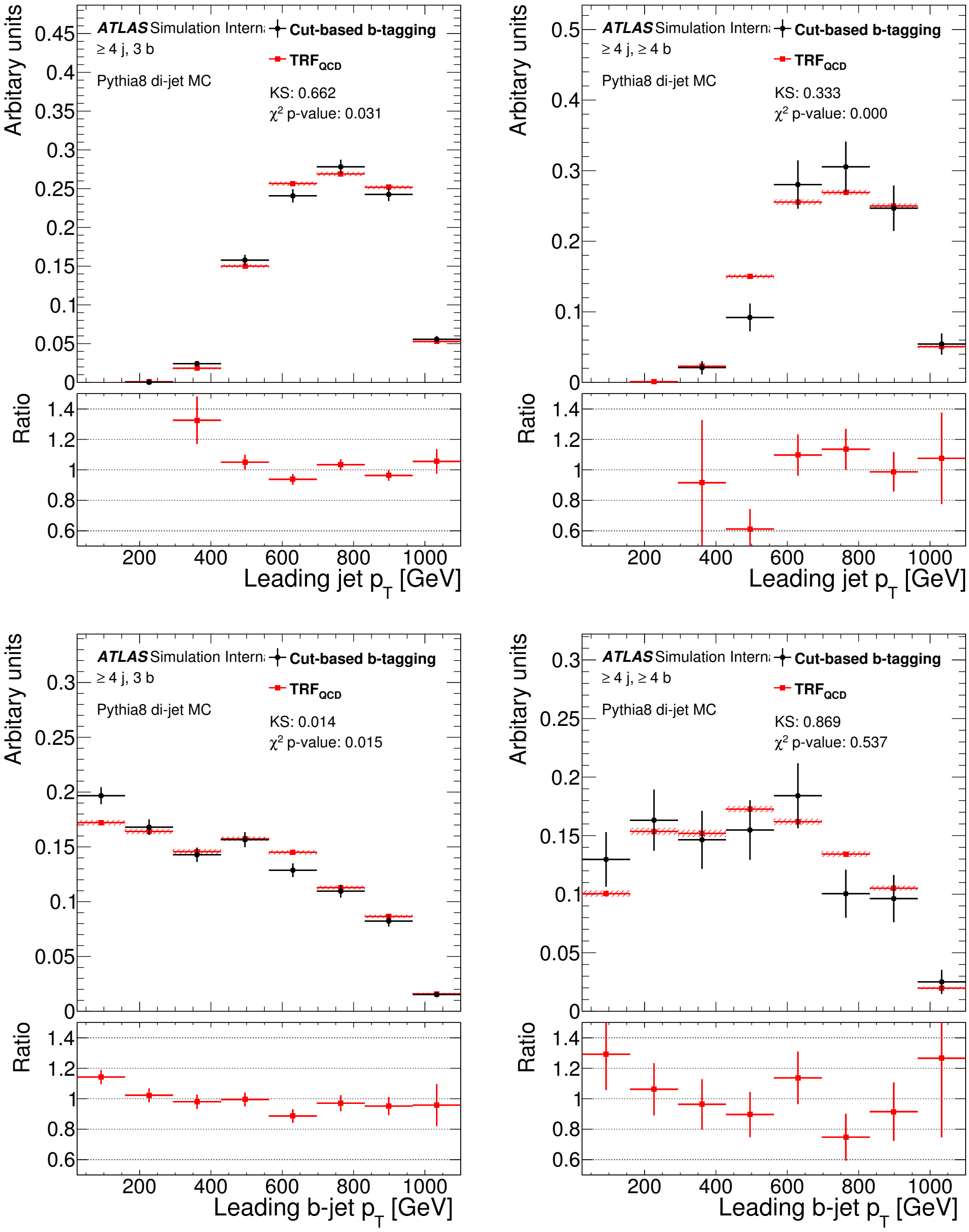


Figure 6.15: Comparison of shapes of leading (*top*) and subleading (*bottom*) jet p_T distributions in Pythia8 di-jet MC events obtained by direct b-tagging (black circles) and by applying the TRF$_{MJ}$ method (red squares), for events with at least 4 jets and exactly 3 b-tagged jets (*left*) or at least 4 b-tagged jets (*right*) in events generated with truth leading jet p_T between 500 GeV and 1 TeV.

6.6 Signal and control regions definition

The pre-selected events are categorized according to the number of jets with $p_T >$ 25 GeV and on the number of b-tagged jets. Only events with at least 6 jets are kept for the analysis mostly because the trigger used requires the presence of at least 5 jets at HLT. Nine regions are defined in the analysis: (6j, 2b), (7j, 2b), (≥8j, 2b). (6j, 3b), (7j, 3b), (≥8j, 3b). (6j, ≥4b), (7j, ≥4b), (≥8j, ≥4b). Where (mj, nb) means a region with m jets of which n are b-tagged.

The multi-jet background content in each region is estimated by applying $\mathrm{TRF}_{\mathrm{MJ}}$ to the multijet *extraction region*, defined for each jet multiplicity region as the event with exactly 2 b-tagged jets. Since in this region there is a non-negligible fraction of top quark production background, here the multijet background is defined as:

$$N_{\mathrm{MJ}}(2b) = N(2b) - N_{\mathrm{MC}}^{\mathrm{bkg}}(2b). \tag{6.15}$$

Where $N_{\mathrm{MJ}}(2b)$ is the number of multijet content, $N(2b)$ the number of data and $N_{\mathrm{MC}}^{\mathrm{bkg}}(2b)$ the number of MC simulated top quark background.

This implies that, from an operational point of view, when extrapolating multijet background to high b-tag multiplicity $\mathrm{TRF}_{\mathrm{MJ}}$ needs to run over both data and MC events, i.e. by defining $\mathrm{TRF}_{\mathrm{MJ}}^{2\rightarrow n}$ as the function that, by applying the $\mathrm{TRF}_{\mathrm{MJ}}$ method estimates the content in the n b-tagged jet region, then:

$$N_{\mathrm{MJ}}(nb) = \mathrm{TRF}_{\mathrm{MJ}}^{2\rightarrow n}(N(2b)) - \mathrm{TRF}_{\mathrm{MJ}}^{2\rightarrow n}(N_{\mathrm{MC}}^{\mathrm{bkg}}(2b)). \tag{6.16}$$

A schema representing how the analysis is organized is illustrated in Figure 6.16. In the high b-tag jets multiplicity regions, namely with exactly 3 and ≥ 4 b-tagged jets, a Boosted Decision Tree (BDT) is applied to separate signal and background, this is described in the next Section.

The multijet content in the region with exactly 2 b-tagged jet is defined following Equation 6.15 while it is estimated using Equation 6.16 in higher b-tagged jet multiplicity regions. Figure 6.17 shows the $S/\sqrt{B}$ ratio, where S and B denote the expected signal (assuming SM cross sections and branching ratios at $\sqrt{s} = 8$ TeV, and $M_H = 125$ GeV) and background, respectively, in each region. The relative proportions of the background yields obtained in each topology are shown in Figure 6.18. Following the same convention, the ratio of $t\bar{t}H$ events labeled by their Higgs decay is presented in Figure 6.19.

6.7 Boosted Decision Tree (BDT)

The final discriminant used in the analysis is a ROOT [151] based TMVA BDT [152] trained to separate the signal ($t\bar{t}H$) from all the backgrounds.

A BDT is a multivariate classifier having a binary tree structure. The boosting algorithm is a procedure that combines many "weak" classifiers to achieve a final powerful classifier, which is given by an average of the individual decision trees[153, 154].

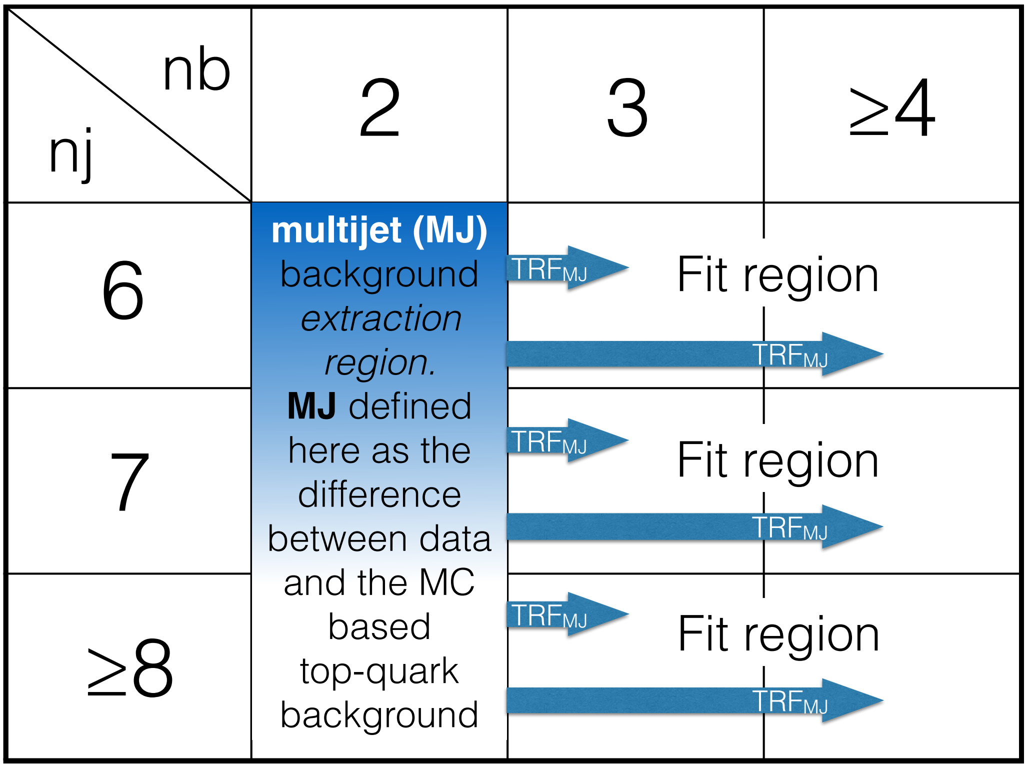


Figure 6.16: Schema of the fully hadronic $t\bar{t}H$ analysis. The multijet background is defined in the exactly 2 b-tagged region and then extrapolated to higher b-tagged jet multiplicity regions by means of the TRF$_{\mathrm{MJ}}$.

A separate BDT is defined and optimized in each of the regions entering the fit. In the training of the BDTs, the background is the sum of the data-driven multijet prediction, as described in Section 6.5, and the background MC samples. The expected normalization of each of the background component is taken into account during the training. In this way the BDT discrimination is more effective against the more abundant background (i.e. multijet production) and not against the component with the greatest number of simulated events (i.e. $t\bar{t}$ + jets). The signal is the $t\bar{t}H$ PowHel sample, described in Section 6.4.1. Half of the events of both signal and background sample are selected randomly and used to train the BDT. The other half is used to test the BDT behaviour to avoid the presence of overtraining.

To choose the most discriminating observables to be used as input for the BDTs, an iterative procedure has been defined. Initially a set of 90 variables is considered; all are used as input in one BDT. The variables which give the least separation[4] and have the highest correlation to each other are removed, reducing the number of candidate variables from $\sim$ 90 to 35. These preselected variables are iteratively added to the inputs of the developing BDT one by one. They are retained if they are

---

[4] Looking at the distributions.

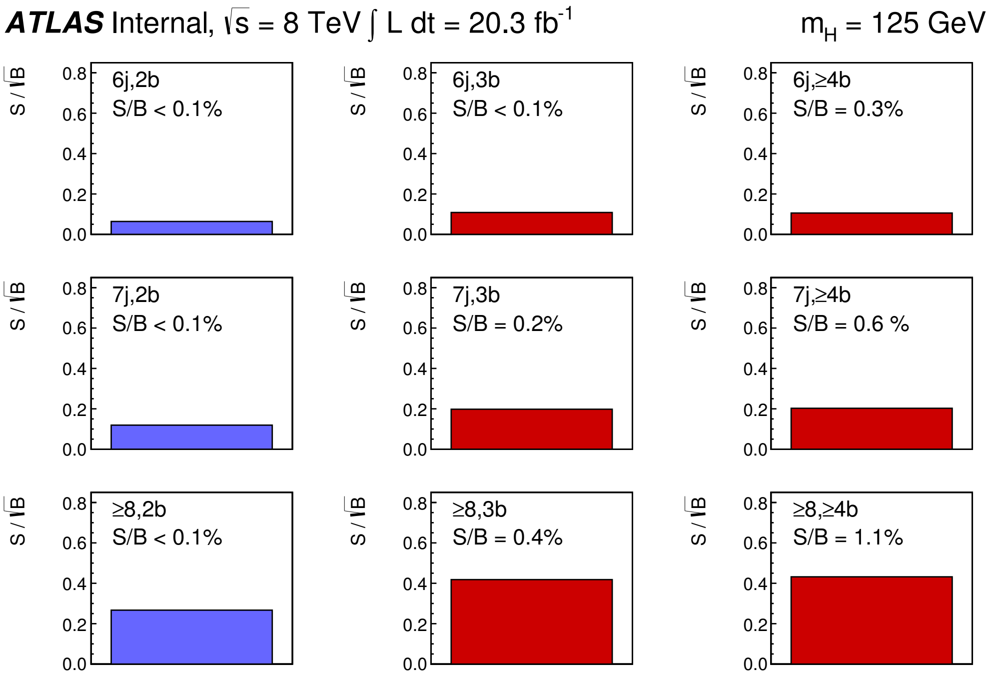


Figure 6.17: $S/\sqrt{B}$ ratio for each of the topologies under consideration after preselection (assuming SM cross sections and branching ratios at $\sqrt{S} = 8$ TeV, and $m_H = 125$ GeV). Each row shows the plots for a specific jet multiplicity (6, 7, ≥ 8), and the columns show the b-tagged jet multiplicity (2, 3, ≥ 4). Signal regions are shown in red.

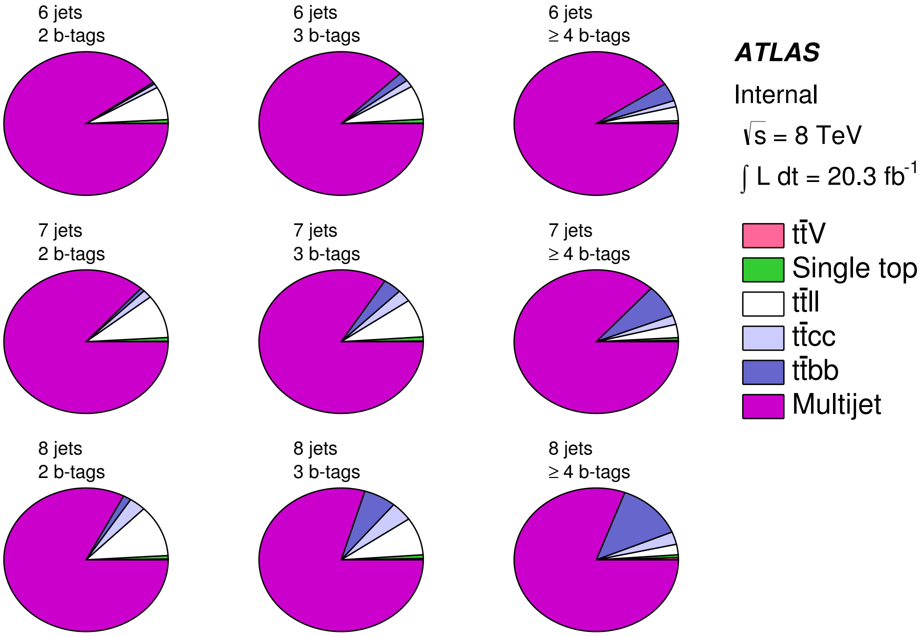


Figure 6.18: Pie-chart of the predicted background events after preselection, simulated by MC and predicted with the $\mathrm{TRF}_{\mathrm{MJ}}$ method.

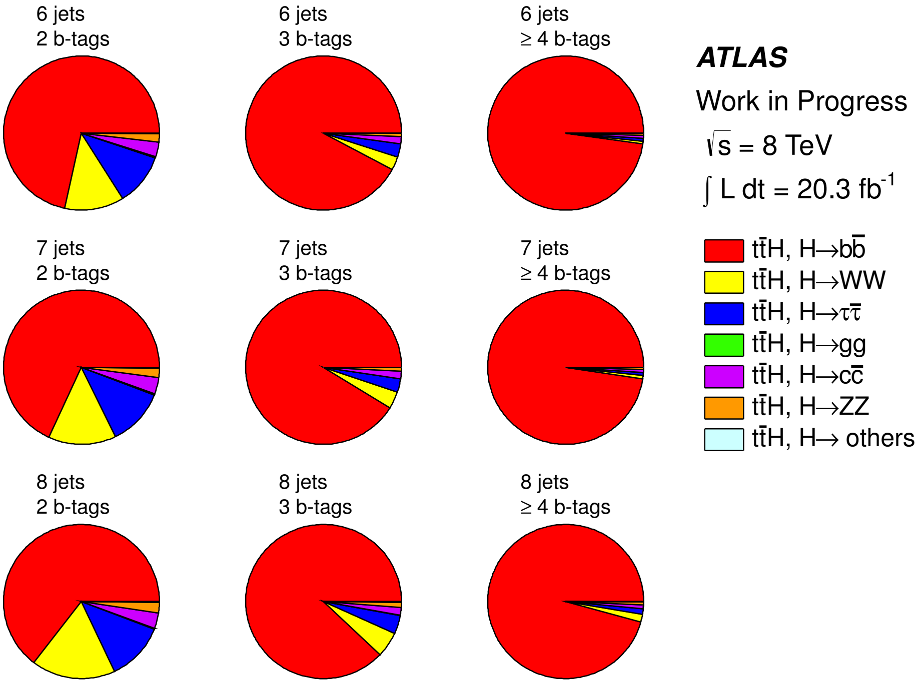


Figure 6.19: Pie-chart of the predicted $t\bar{t}H$ events simulated by MC after preselection, classified by Higgs decay.

ranked high by the figure of merit F. This is calculated using the ratio of the signal and background histograms of the BDT response, with the relative normalization factors:

$$F = \sqrt{\sum_{i=1}^{n} \frac{S_i^2}{B_i}} \tag{6.17}$$

where S_i and B_i are, respectively, the number of signal events and the number of background events in the i^{th} bin. The bins for the evaluation of F are defined in order to have no empty bins for the background distribution and at least 10% of signal + background events. Variables are added successively to the list of inputs, until the figure of merit F reaches a plateau with less than 1% variation: in most regions this occurs after adding the 11^{th} variable.

The final list of variables used by the BDT in each region is shown in Table 6.9 and a brief description of the selected variables is presented in Table 6.10. The separation between signal and background is shown in Figure 6.20 for the signal regions entering in the fit.

LogLikelihood Ratio (LLR) variable definition

The LLR variable measures the probability for an event to be a signal candidate. The LLR variable for the W resonance is built as $LLR(M_{jj}) = \ln \frac{S(M_{jj})}{B(M_{jj})}$.

$$S(M_{jj}) = \begin{cases} s \cdot G(M_{jj}|M_W, \sigma_W), & \text{for} |M_{jj} - M_W| \leq 30\text{GeV}, \\ 1 - s, & \text{for } |M_{jj} - M_W| > 30\text{GeV}. \end{cases} \tag{6.18}$$

$$B(M_{jj}) = \begin{cases} b \cdot \text{Rect}(M_W, 30\text{ GeV}), & \text{for} |M_{jj} - M_W| \leq 30\text{GeV} \\ 1 - b, & \text{for } |M_{jj} - M_W| > 30\text{GeV}. \end{cases} \tag{6.19}$$

where s and b are the probabilities in signal and background respectively to find a jet pair with an invariant mass compatible with the W boson mass. They are calculated from MC for signal and from multijet background prediction. $G(M_{jj}|M_W, \sigma_W)$ is a Gaussian distribution which models the W resonance[5] and $\text{Rect}(M_W, 30\text{GeV})$ is a uniform distribution in a mass window of ±30 GeV centered on the W boson mass. $S(M_{jj})$ and $B(M_{jj})$ are normalized to unity.

The LLR is built extending the formulation described above to top quark and Higgs boson resonances:

$$\begin{aligned} LLR(M_{jj}, M_{jjb}, M_{bb}) = LLR(M_{jj}|M_W, \sigma_W) + LLR(M_{jjb}|M_{top}, \sigma_{top}) + \\ + LLR(p_{\mathrm{T\,bb}}, M_{bb}|M_H, \sigma_H) \end{aligned} \tag{6.20}$$

For the Higgs boson the LLR is function also of the p_T of the system of two b-tagged jets assumed to be originating from its decay. σ_H is the expected experimental width of a hadronically decaying Higgs boson without combinatorial background, it is assumed to be equal to 18 GeV. To obtain the maximum LLR of the event, all possible jet combinations are considered.

---

[5] Gaussian distribution is used as approximation for the Breit-Wigner distribution.

Variable	6jet 3btag	6jet 4btag	7jet 3btag	7jet 4btag	8jet 3btag	8jet 4btag
LLR	4	4	3	3	1	1
S_{T}	2	2	2	4	2	2
Aplanarity	-	11	-	-	6	-
Centrality Mass	1	1	1	1	9	6
Min pair mass (all jets)	9	-	6	10	11	12
Minimum ΔR	6	5	9	-	8	4
$\Delta R(b,b)$ max p_{T}	11	-	7	5	5	3
$M_{bb} min \Delta R(b,b)$	3	3	8	9	3	9
$m_{2\,\text{b-jets}}$	12	7	-	6	-	8
M_{jj}	10	-	-	8	-	-
$(E_{\mathrm{T}\,1} + E_{\mathrm{T}\,2}) / \sum(\text{jets } E_{\mathrm{T}})$	5	8	4	2	7	5
$m_{top,1}$	13	10	-	-	4	11
$m_{top,2}$	7	9	5	-	10	7
$H_T 5$	8	-	-	7	-	-
softest p_{T}	-	6	10	-	-	10

Table 6.9: List of variables used in the BDT of each analysis region. The number corresponds to the ranking of that variable, in terms of discriminating power. Variables without value in few regions are variables which are not used in the BDT in the considered region.

Variable	Definition
LLR	LogLikelihood Ratio
S_{T}	Modulus of the vector sum of jet p_T after quality and JVF cuts
Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with all jets
Centrality Mass	Sum of the p_T divided by the invariant mass of the jets
Min pair mass (all jets)	Smallest mass of any combination of two jets
Minimum ΔR	Minimum ΔR between two jets
$\Delta R(b,b)$ max p_{T}	ΔR between two b-tagged jets with the largest vector sum p_T
$M_{bb} min \Delta R(b,b)$	Mass of the combination of two b-tagged jets with the smallest ΔR
$m_{2\,\text{b-jets}}$	Invariant mass of the two b-tagged jets which are selected by requiring that the invariant mass of all the remaining jets is maximal.
M_{jj}	Computed in two steps: select iteratively a 2 light-jets pair + a b-jet and compute M_{jj} and M_{jjb}, pick up the triplet with the highest $p_{T,\,jjb}$
$(E_{\mathrm{T}\,1}+E_{\mathrm{T}\,2})/\sum(\text{jets } E_{\mathrm{T}})$	Sum of the two jets with leading E_T over the sum of all jets E_T
$m_{top,\,1}$	Invariant mass of three jets that better approaches the top mass
$m_{top,\,2}$	Invariant mass of three jets that better approaches the top mass calculated excluding jets entering $m_{top,\,1}$.
$H_T\,5$	Scalar sum of jet p_T starting from the 5^{th} jet
Softest p_{T}	p_{T} of the softest jet

Table 6.10: The variables considered for use within the BDT analysis.

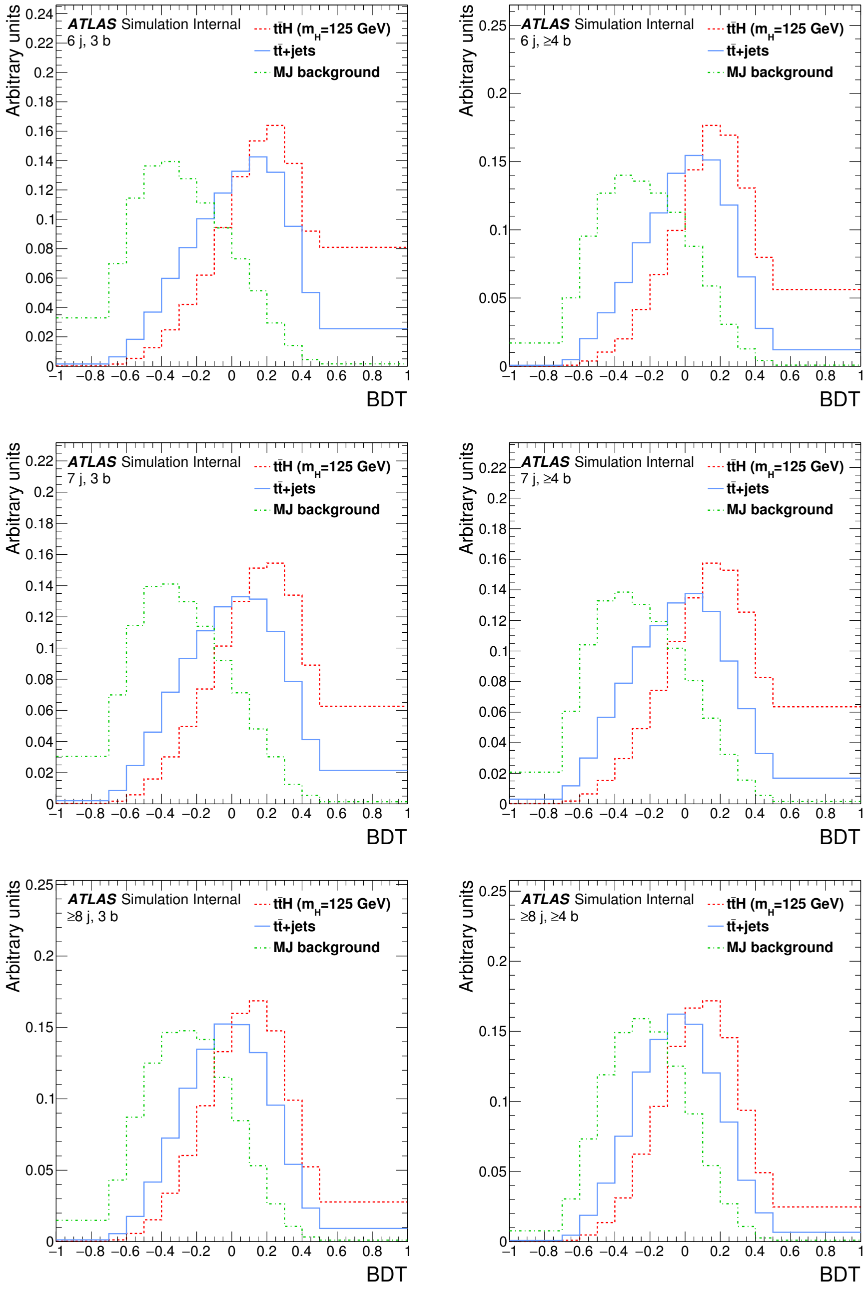


Figure 6.20: Response of the BDTs in 3 b-tags regions (*left*) (6, 7 and $\geq$ 8 jets respectively) and $\geq$ 4 b-tags regions (*right*). In the plots the red histogram refers to the signal, the blue one to the $t\bar{t}$ + jets background and the green to the multijet background.

Blinding strategy: the $\mathrm{BDT}_{\mathrm{anti}}$ cut

A blinded region in the BDT spectra is defined. This to check the agreement between data and the background prediction in a region where the presence of $t\bar{t}H$ signal is minimized. To do so $\mathrm{BDT}_{\mathrm{anti}}$ is defined such that for BDT values larger than $\mathrm{BDT}_{\mathrm{anti}}$ the signal contamination is larger than 2% in each BDT bin of size 0.1, as shown in Figure 6.21 and in Table 6.11.

The blinded region is defined for values of the BDT smaller than $\mathrm{BDT}_{\mathrm{anti}}$. This blinded region is used in the optimization steps of the analysis. In particular, it is used for the selection of the nominal $\mathrm{TRF}_{\mathrm{MJ}}$ parametrization, described in Section 6.8, and for the preliminary test of the fit.

Region	BDT_{anti}
(6j, 3b)	Unblinded
(6j, ≥4b)	Unblinded
(7j, 3b)	0.5
(7j, ≥4b)	0.2
(≥8j, 3b)	0.3
(≥8j, ≥4b)	0.0

Table 6.11: Value of the $\mathrm{BDT}_{\mathrm{anti}}$ in each region. Unblinded means that the full BDT spectrum is used.

6.8 $\mathrm{TRF}_{\mathrm{MJ}}$ parametrization selection

This section presents the selection of the nominal $\varepsilon_{\mathrm{MJ}}$ parametrization performed in the analysis regions. For the $\mathrm{TRF}_{\mathrm{MJ}}$ method to work, $\varepsilon_{\mathrm{MJ}}$ should be able to take into account the dependency on the flavor of the jet, i.e. it needs to be as sensitive as possible to heavy flavor production. For this reason, $\varepsilon_{\mathrm{MJ}}$ is described as function of the distance to already b-tagged jets to be able to exploit correlations between heavy flavored jets. In principle, with a large enough sample where to estimate the $\mathrm{TRF}_{\mathrm{MJ}}$ it should be possible to describe $\varepsilon_{\mathrm{MJ}}$ with respect to a large enough number of variables and maximize the sensitivity to flavor production. Unfortunately the region allowed for such study is relatively limited in statistics which make possible to use a maximum of 4-dimensional $\varepsilon_{\mathrm{MJ}}$. The b-tagging efficiency has been derived as function of six observables:

- Jet p_T and $|\eta|$,
- Min $\Delta R_{(j,j)}$: Minimum ΔR between the jet and any other jet in the event,
- Min $\Delta R_{(j,\mathrm{hMV1})}$: Minimum ΔR between the jet and the two with highest MV1 weight,

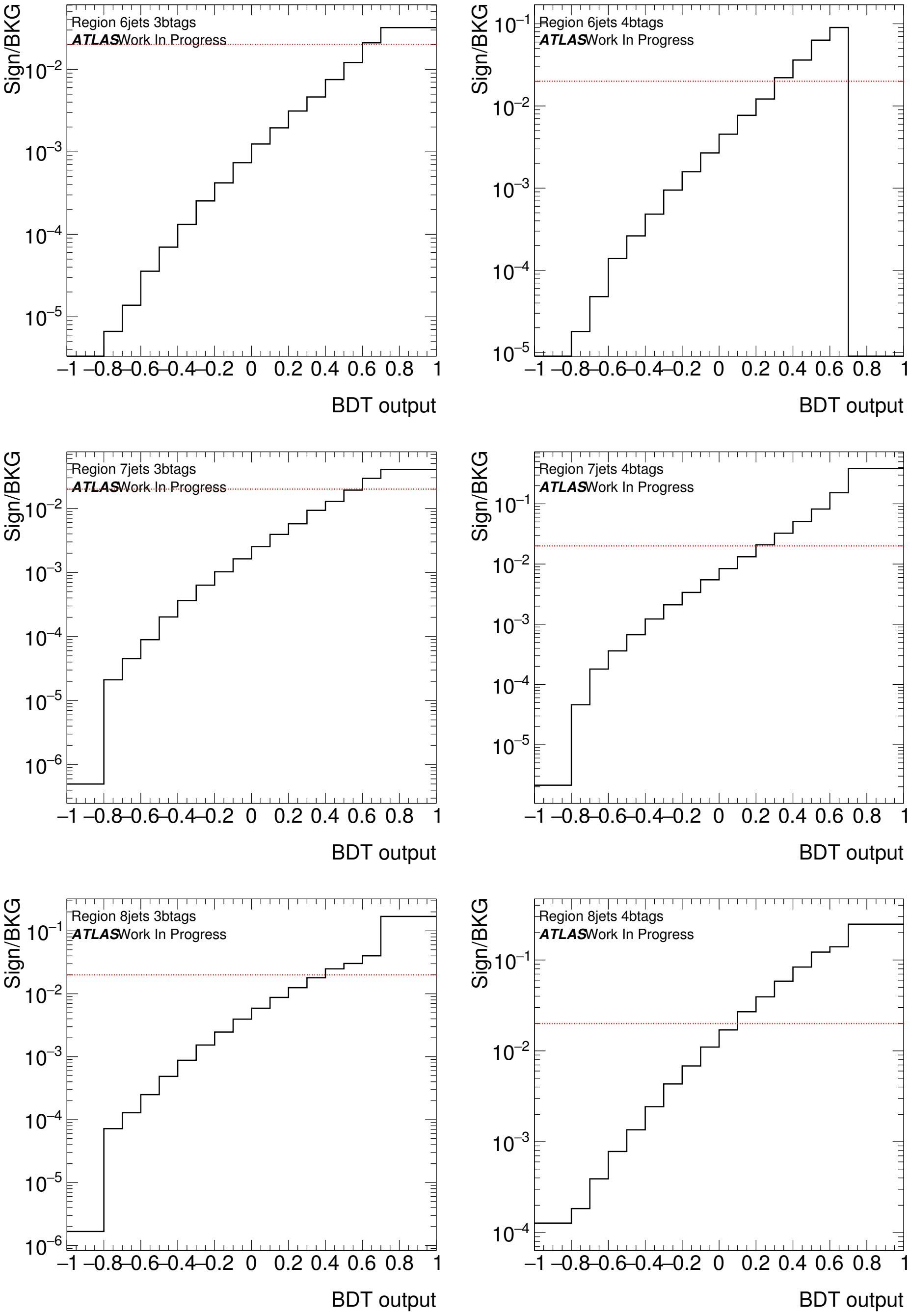


Figure 6.21: Ratio of signal over background in each bin of the BDT output.

- $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$: Average ΔR between the jet and the two with highest MV1 weight,
- MV1 ΔR: ΔR between the two jets with the highest MV1,

and the different parametrizations are:

- p_T, $|\eta|$, $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$
- p_T, $|\eta|$, Min$\Delta R_{(j,\mathrm{hMV1})}$
- p_T, $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$, MV1 ΔR
- p_T, Min $\Delta R_{(j,j)}$, MV1 ΔR
- p_T, Min $\Delta R_{(j,\mathrm{hMV1})}$, MV1 ΔR
- p_T, $|\eta|$, Min $\Delta R_{(j,\mathrm{hMV1})}$, MV1 ΔR

The efficiencies obtained are presented in Figures 6.9, 6.22, 6.23, 6.24, 6.25, and 6.26. One can notice the strong dependency of the $\mathrm{TRF}_{\mathrm{MJ}}$ upon the different b-tagging correlation variables, even though these are strongly correlated using only one of such variables in the parametrization is not enough to unfold the flavor dependency, as it can be seen in Figure 6.23 and 6.24 where Min $\Delta R_{(j,\mathrm{hMV1})}$ and Min $\Delta R_{(j,j)}$ are used together with MV1 ΔR.

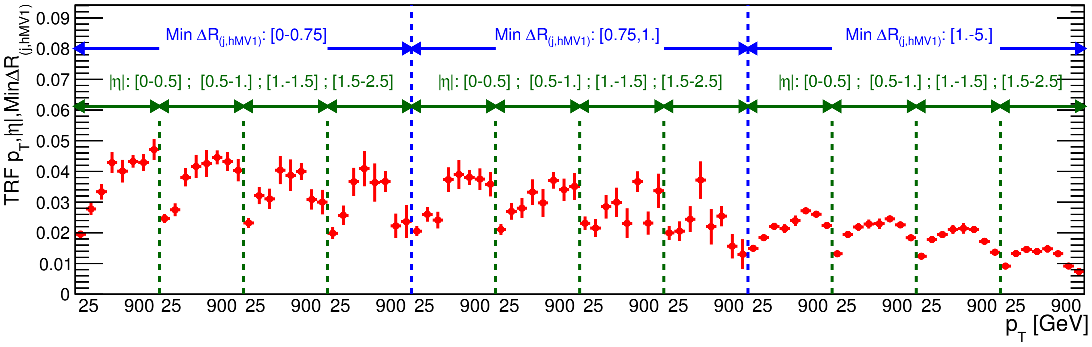


Figure 6.22: One dimensional view of the dependence of $\varepsilon_{\mathrm{MJ}}$ as a function of the p_T and $|\eta|$ of the jet and the Min $\Delta R_{(j,\mathrm{hMV1})}$.

These different sets of parameterisations have been tested in order to determine the baseline for the analysis. The selection is performed through a χ^2 minimisation on the shape modeling with respect to data using all BDT input variables in all regions. This ministration is performed after application of the blinding procedure described in Section 6.7. For this procedure the comparison is performed between $\mathrm{TRF}_{\mathrm{MJ}}$ prediction and the expected multijet background defined as the data minus the MC contribution in each region.

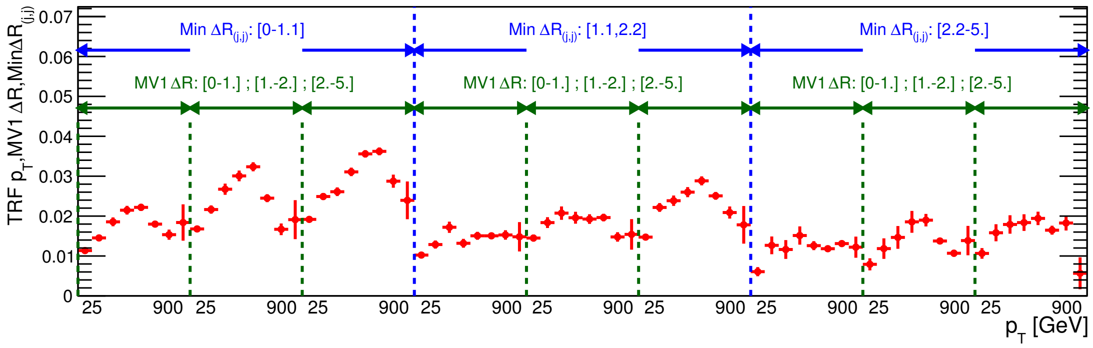


Figure 6.23: One dimensional view of the dependence of $\varepsilon_{\mathrm{MJ}}$ as a function of the p_T of the jet, MV1 ΔR and Min$\Delta R_{(j,j)}$.

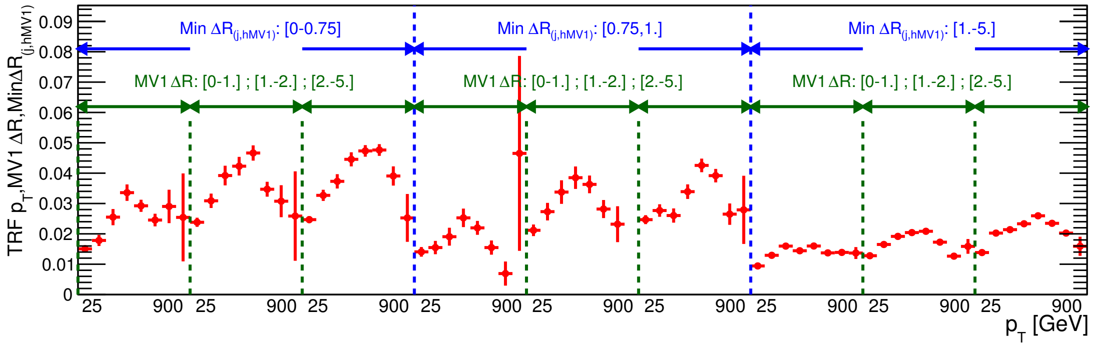


Figure 6.24: One dimensional view of the dependence of $\varepsilon_{\mathrm{MJ}}$ as a function of the p_T of the jet, MV1 ΔR and Min$\Delta R_{(j,\mathrm{hMV1})}$.

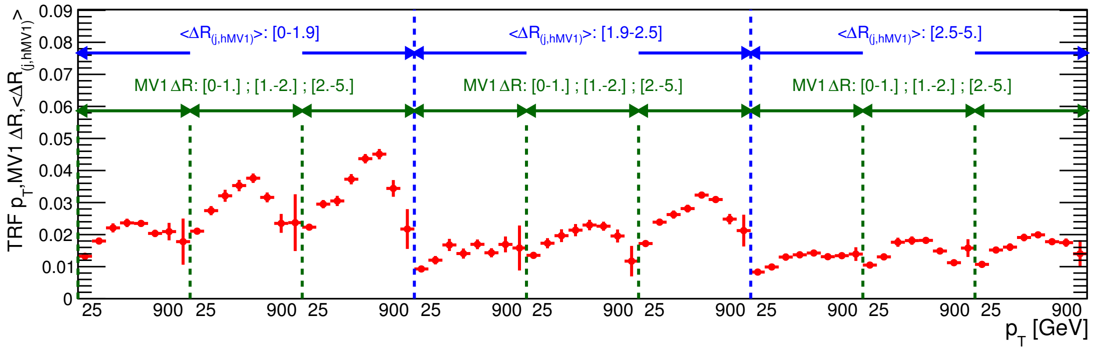


Figure 6.25: One dimensional view of the dependence of $\varepsilon_{\mathrm{MJ}}$ as a function of the p_T of the jet, MV1 ΔR and $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$.

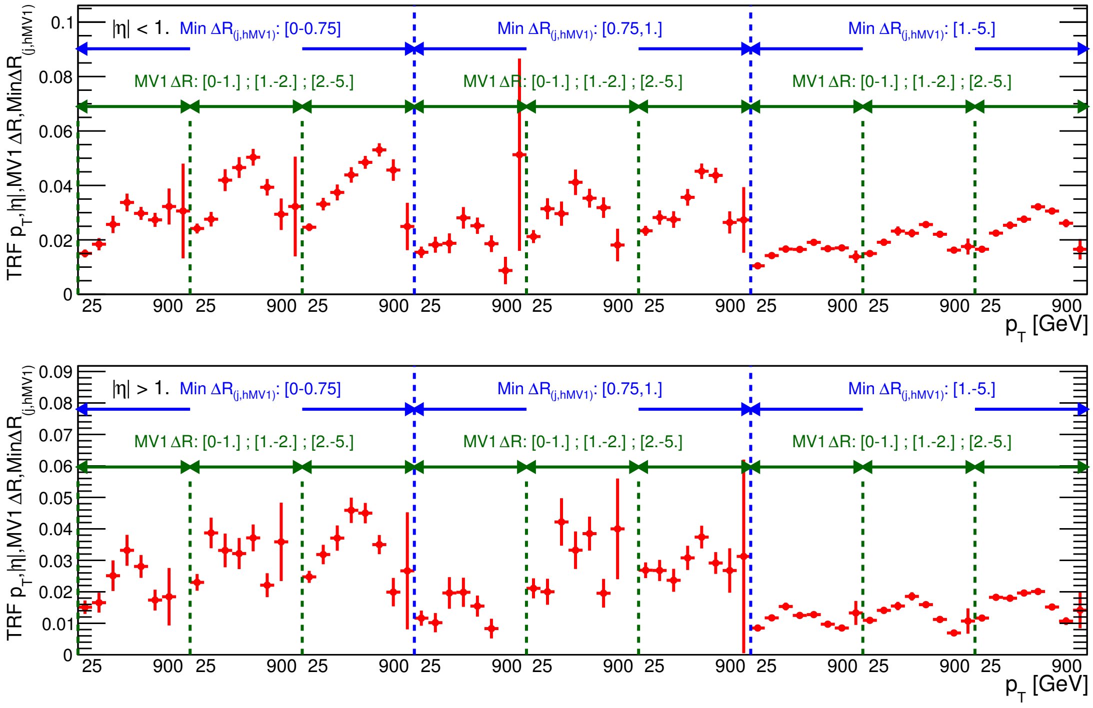


Figure 6.26: One dimensional view of the dependence of $\varepsilon_{\mathrm{MJ}}$ as a function of the p_T and $|\eta|$ of the jet, MV1 ΔR and Min$\Delta R_{(j,\mathrm{hMV1})}$. The top (bottom) plot represents the TRF$_{\mathrm{MJ}}$ for jets having $|\eta| < (>) 1$.

Table 6.12 shows the normalization predictions from the tested parameterisations compared to the direct b-tagging in the TRF$_{\mathrm{MJ}}$ estimation region, defined in Section 6.5.1. Table 6.13 shows the normalization predictions of the tested parameterisations compared to the subtraction of MC contributions from data in the regions entering in the fit. This table does not present any BDT$_{\mathrm{anti}}$ cut. The predictions have an overall spread of 15%, with maximum difference from the expected multijet background of $\sim$ 20%. Since the low BDT regions are dominated by multijet background its normalization in each region is left as a free parameter in the fit.

For the set of variables which are most sensitive to angular correlations, such as the M_{bb} for the b-tagged jets with minimum ΔR one can see a large spread in the predictions from the different TRF$_{\mathrm{MJ}}$. Whereas other variables are completely unaffected by switching from one TRF$_{\mathrm{MJ}}$ to the other, such for example the LLR output. For another class of variables the agreement reached by this class of TRF$_{\mathrm{MJ}}$s is not considered sufficient to simulate the correct multijet behaviour, nor the spread produced by using different multijet predictions enough to cover the disagreement. This is the case for S_t and H_T. The distributions for these observables are presented in Figures 6.27, 6.28, 6.29, 6.30, 6.31 and 6.32 for the regions (6j, 3b), (6j, $\geq$ 4b), (7j, 3b), (7j, $\geq$ 4b), (8j, 3b), (8j, $\geq$ 4b) respectively. The non closure of S_t and H_T is treated via ad hoc systematic as will be explained in Section 6.9.3. Figures 6.33 and 6.34 show the spectra of the BDT output as it is reproduced by all different parametrization of TRF$_{\mathrm{MJ}}$.

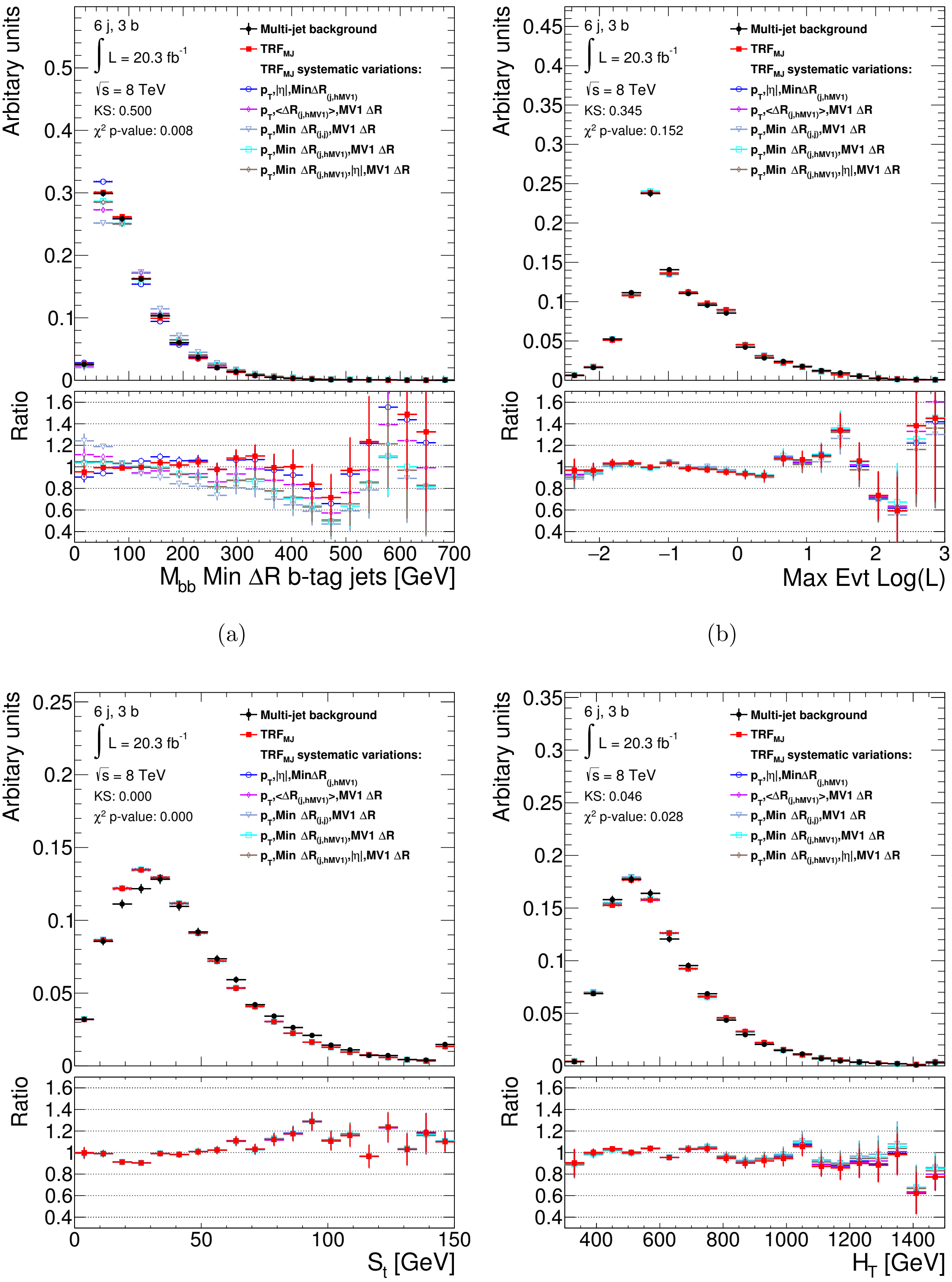
Table 6.14 summarizes the results for the χ^2 computed using all the $\varepsilon_{\mathrm{MJ}}$ parameterisations in all regions. The nominal set of variables to describe $\varepsilon_{\mathrm{MJ}}$ is then selected as the best ranked one in most of the regions, specifically (p_T,$|\eta|$,$\langle\Delta R_{(j,\mathrm{hMV1})}\rangle$) which is ranked first in 3 regions and second in the 3 others.

N_{b-tags}		$N_{jets} == 3$	$N_{jets} \geq 4$
3	data	641	7585
	p_T,$\lvert\eta\rvert$,$\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$	632 ± 4	7952 ± 25
	p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$	692 ± 4	8051 ± 25
	p_T,$\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$,MV1 ΔR	680 ± 4	8051 ± 25
	p_T,Min $\Delta R_{(j,j)}$,MV1 ΔR	717 ± 4	7814 ± 25
	p_T,Min $\Delta R_{(j,\mathrm{hMV1})}$,MV1 ΔR	644 ± 4	7873 ± 25
	p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$, MV1 ΔR	740 ± 4	7964 ± 25
≥ 4	data		425
	p_T,$\lvert\eta\rvert$,$\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$		452 ± 2
	p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$		485 ± 2
	p_T,$\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$,MV1 ΔR		454 ± 2
	p_T,Min $\Delta R_{(j,j)}$,MV1 ΔR		470 ± 2
	p_T,Min $\Delta R_{(j,\mathrm{hMV1})}$,MV1 ΔR		476 ± 2
	p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$, MV1 ΔR		494 ± 2

Table 6.12: Normalization prediction from the sets of $\mathrm{TRF}_{\mathrm{MJ}}$ parameterisations for different event categories compared to expected multijet background, defined as data minus MC background contribution. Test is performed in the $\mathrm{TRF}_{\mathrm{MJ}}$ extraction region.

	(6j, 3b)	(6j, $\geq$ 4b)	(7j, 3b)	(7j, $\geq$ 4b)	(8j, 3b)	(8j, $\geq$ 4b)
p_T,$\lvert\eta\rvert$,$\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$	16376 ± 42	1113 ± 3	12529 ± 41	1123 ± 4	10665 ± 45	1323 ± 6
p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$	16463 ± 42	1199 ± 3	12522 ± 41	1193 ± 4	10588 ± 44	1381 ± 6
p_T,$\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$,MV1 ΔR	16344 ± 43	1126 ± 4	12417 ± 43	1126 ± 4	10543 ± 46	1325 ± 7
p_T,Min $\Delta R_{(j,j)}$,MV1 ΔR	17523 ± 45	1306 ± 4	13449 ± 45	1329 ± 5	11515 ± 49	1591 ± 8
p_T,Min $\Delta R_{(j,\mathrm{hMV1})}$,MV1 ΔR	16421 ± 43	1239 ± 4	12409 ± 42	1218 ± 5	10457 ± 45	1403 ± 7
p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$, MV1 ΔR	16683 ± 44	1274 ± 4	12612 ± 43	1255 ± 5	10634 ± 46	1449 ± 7
Multijet	16063	1427	12219	1229	10327	1285

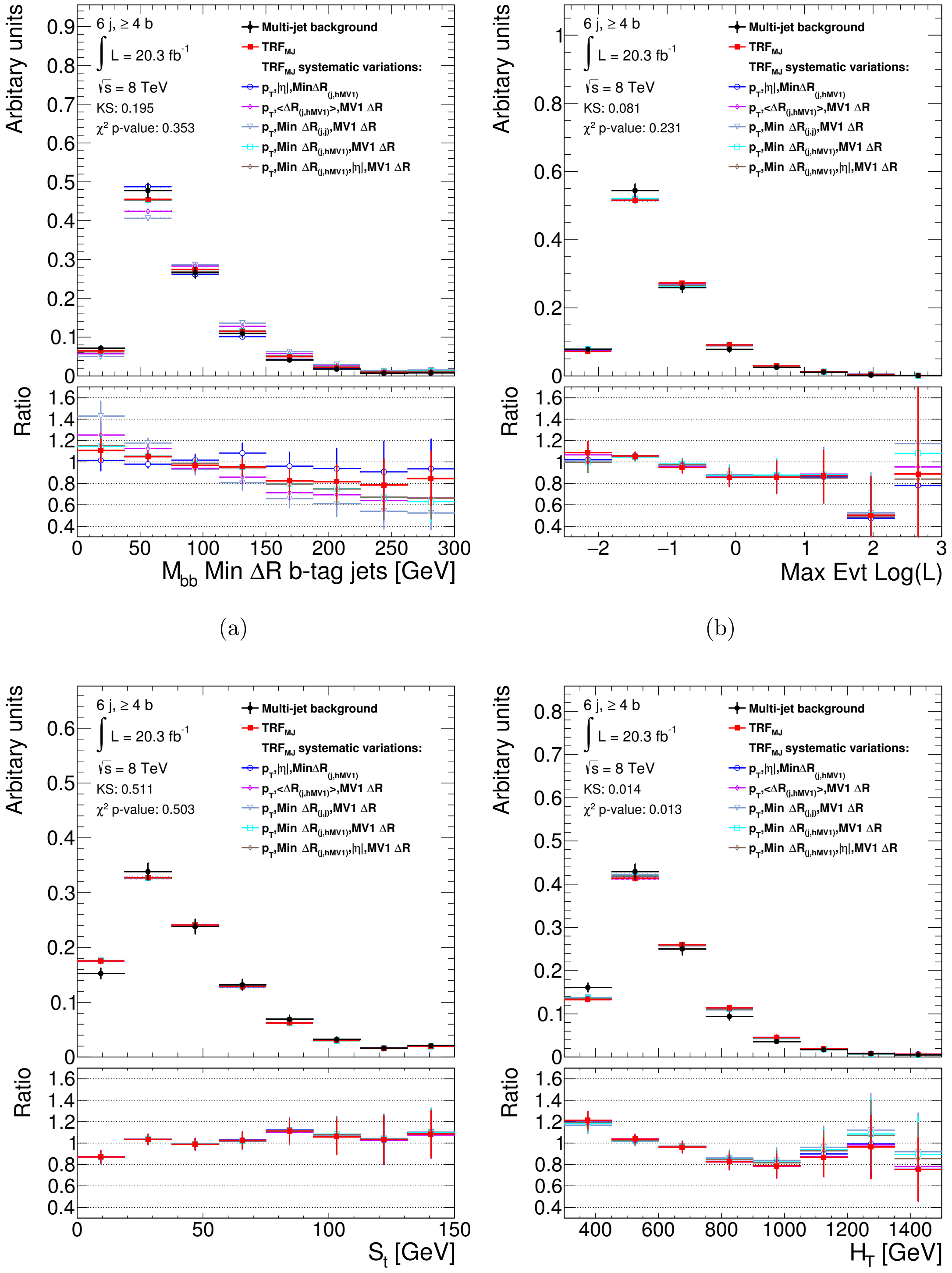
Table 6.13: Normalization prediction from the sets of $\mathrm{TRF}_{\mathrm{MJ}}$ parameterisations for the analysis' event categories compared to expected multijet background, defined as data minus MC background contribution.



(a) (b)

(c) (d)

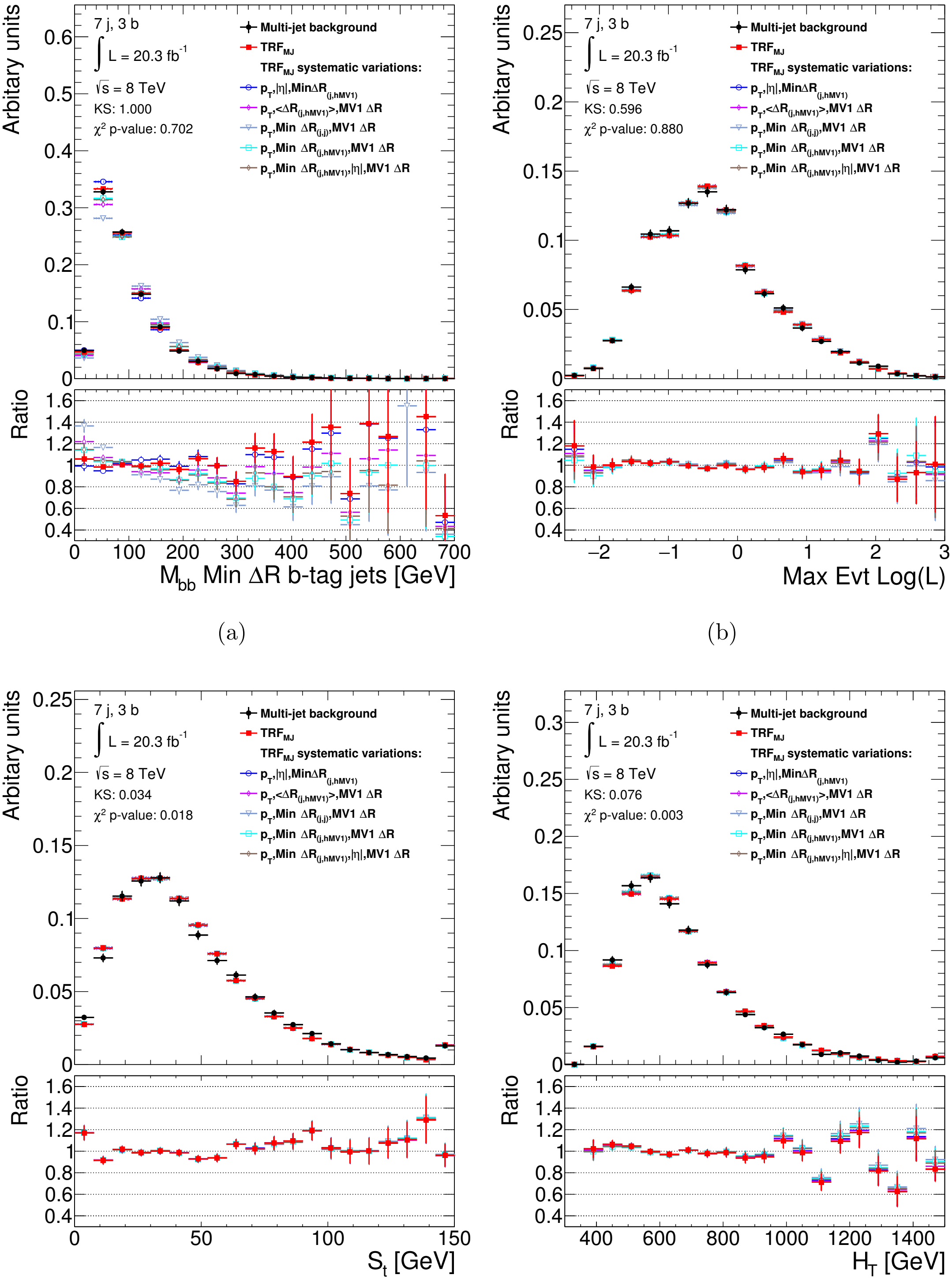
Figure 6.27: Comparison of the modeling for the M_{bb} for the b-tagged jets with minimum ΔR (a), LLR (b), S_t (c) and H_T (d) between the different sets of TRF$_{\mathrm{MJ}}$ parameterisations and expected multijet background, defined as data minus MC background contribution, for the region (6j, 3b). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF$_{\mathrm{MJ}}$ for each distribution.



(a) (b)

(c) (d)

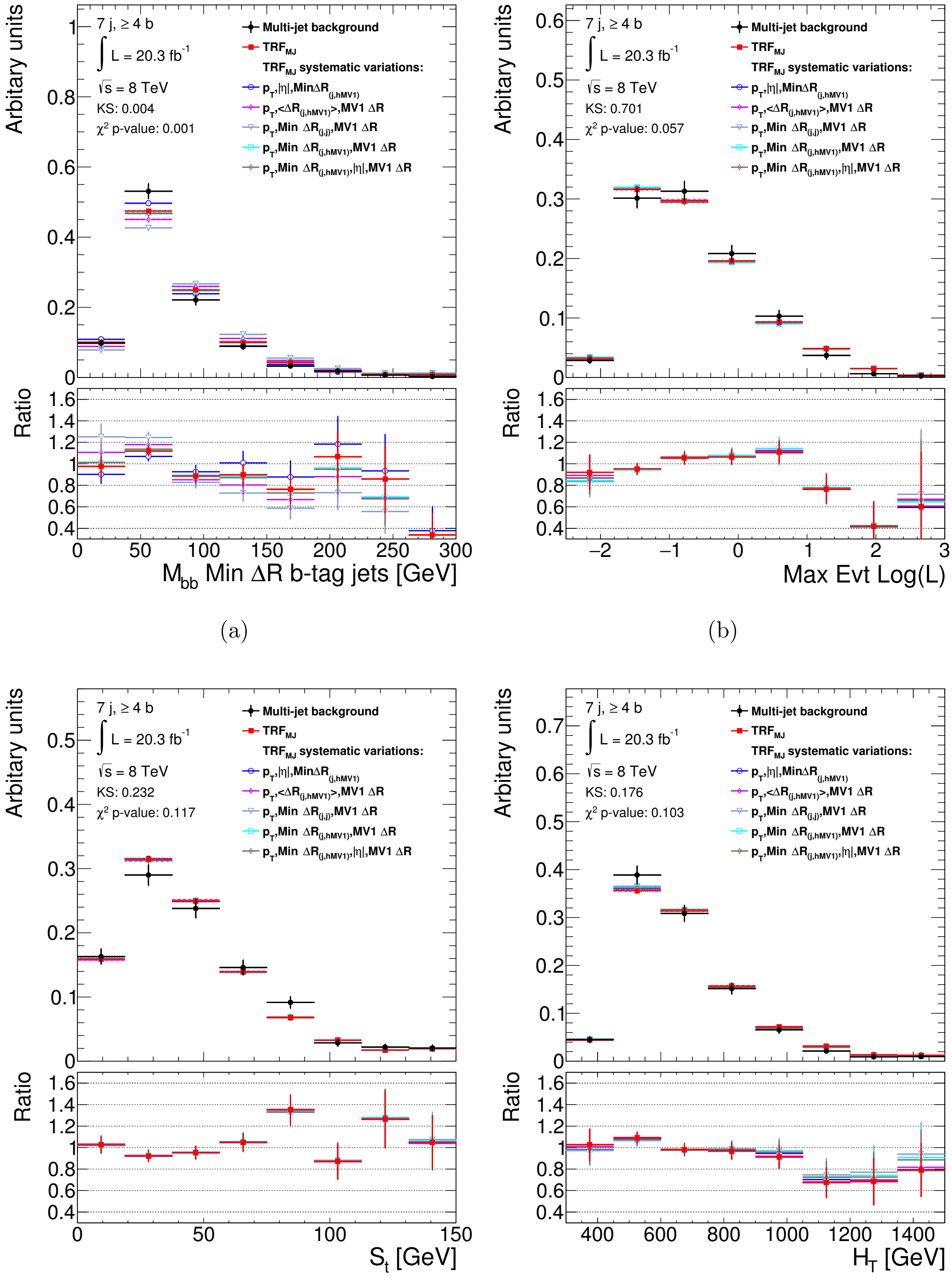
Figure 6.28: Comparison of the modeling for the M_{bb} for the b-tagged jets with minimum ΔR (a), LLR (b), S_t (c) and H_T (d) between the different sets of TRF$_{\mathrm{MJ}}$ parameterisations and expected multijet background, defined as data minus MC background contribution, for the region (6j, $\geq$ 4b). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF$_{\mathrm{MJ}}$ for each distribution.



(a) (b)

(c) (d)

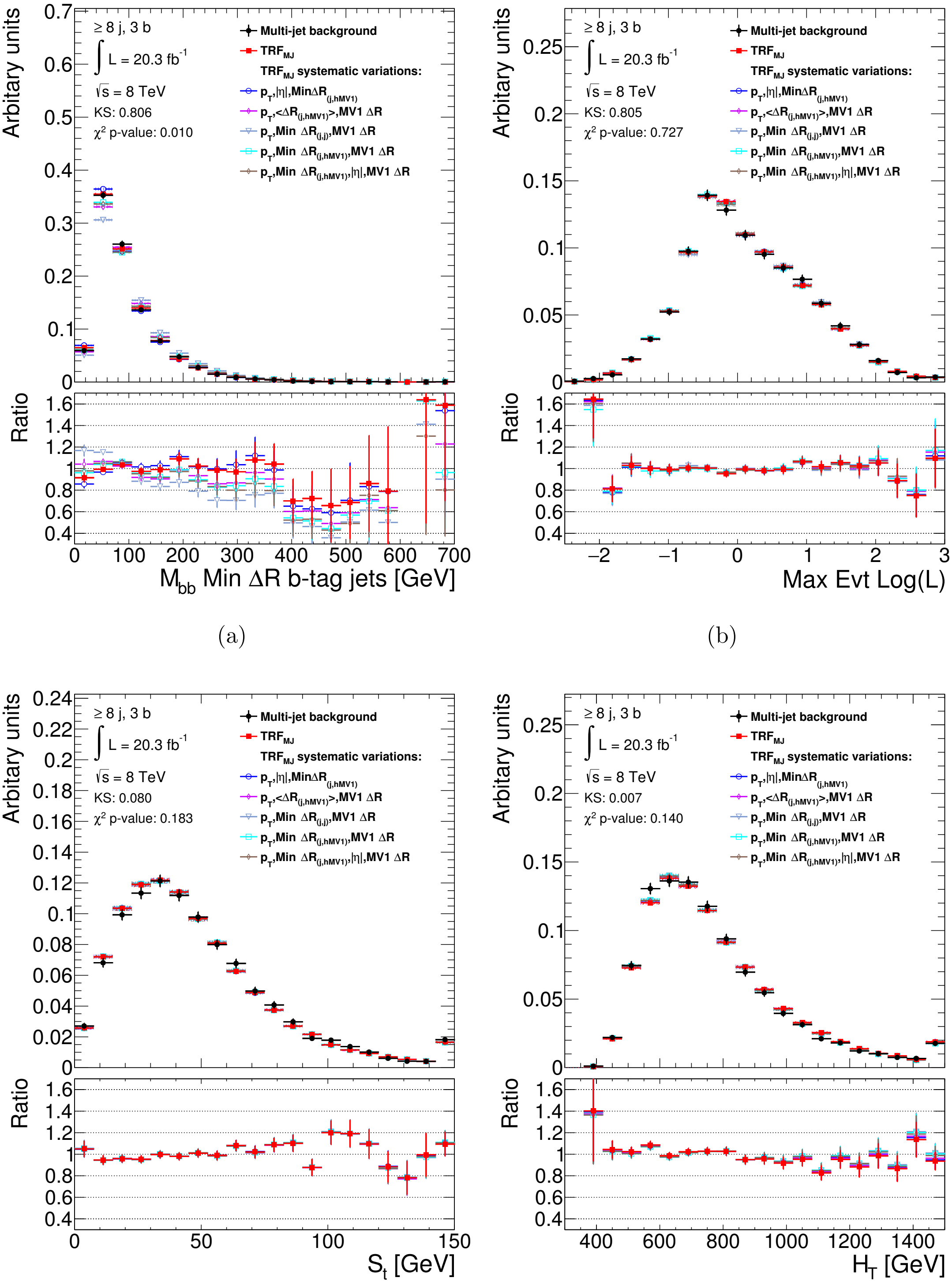
Figure 6.29: Comparison of the modeling for the M_{bb} for the b-tagged jets with minimum ΔR (a), LLR (b), S_t (c) and H_T (d) between the different sets of TRF$_{\mathrm{MJ}}$ parameterisations and expected multijet background, defined as data minus MC background contribution, for the region (7j, 3b). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF$_{\mathrm{MJ}}$ for each distribution.



(a) (b)

(c) (d)

Figure 6.30: Comparison of the modeling for the M_{bb} for the b-tagged jets with minimum ΔR (a), LLR (b), S_t (c) and H_T (d) between the different sets of TRF$_{\rm MJ}$ parameterisations and expected multijet background, defined as data minus MC background contribution, for the region (7j, $\geq$ 4b). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF$_{\rm MJ}$ for each distribution.



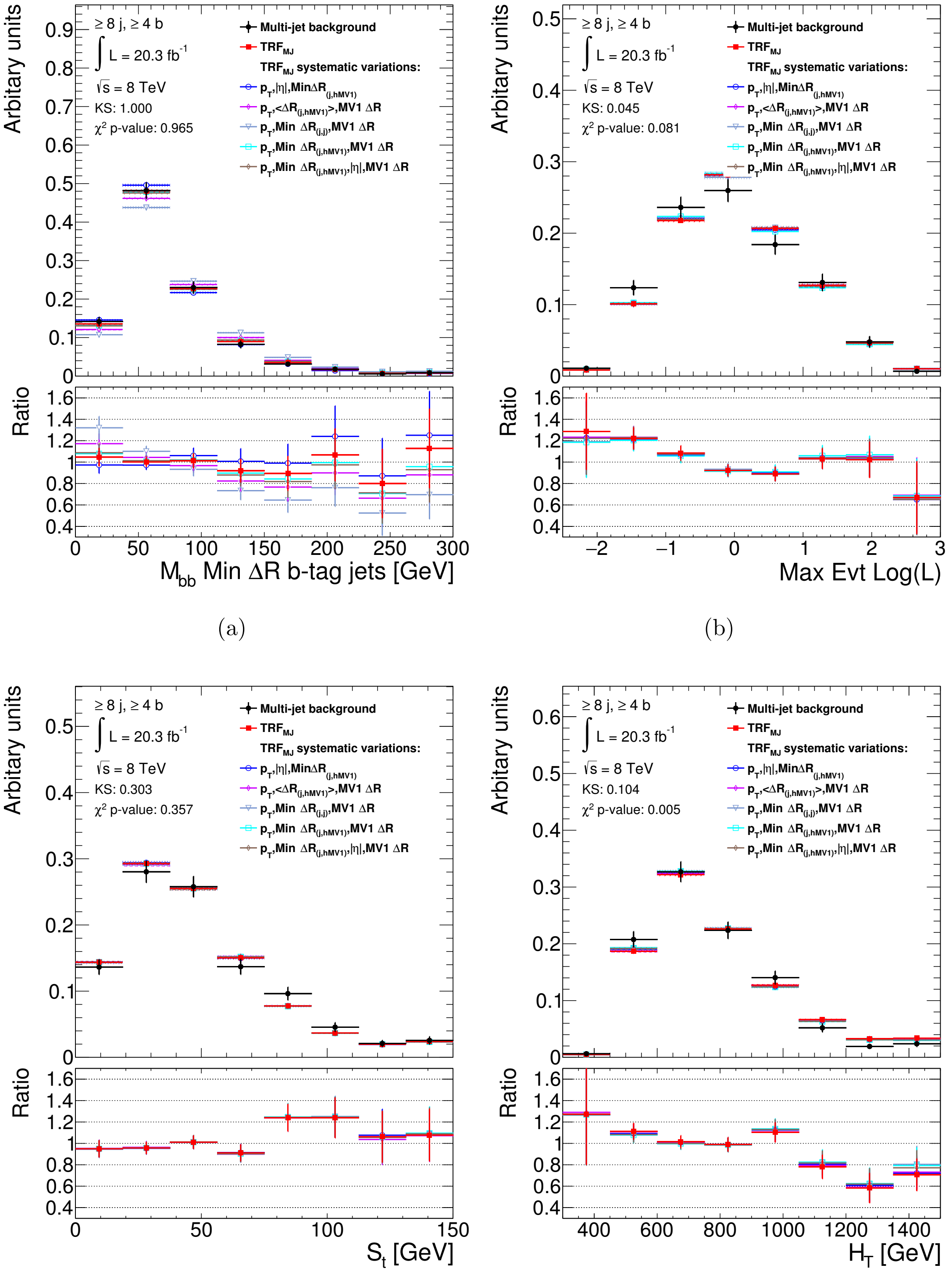
(a)

(b)

(c)

(d)

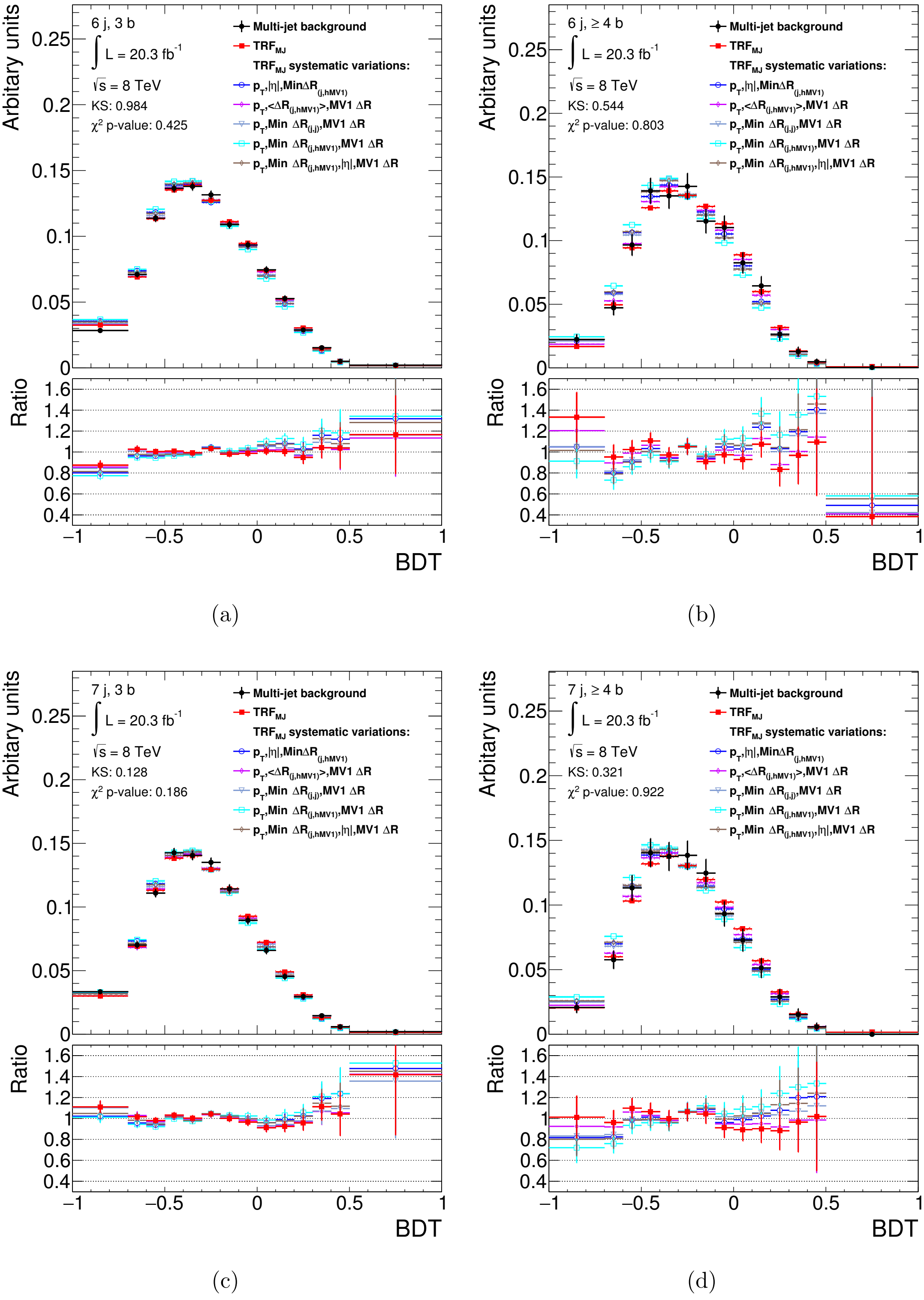
Figure 6.31: Comparison of the modeling for the M_{bb} for the b-tagged jets with minimum ΔR (a), LLR (b), S_t (c) and H_T (d) between the different sets of TRF$_{\mathrm{MJ}}$ parameterisations and expected multijet background, defined as data minus MC background contribution, for the region ($\geq$ 8j, 3b). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF$_{\mathrm{MJ}}$ for each distribution.



(a) (b)

(c) (d)

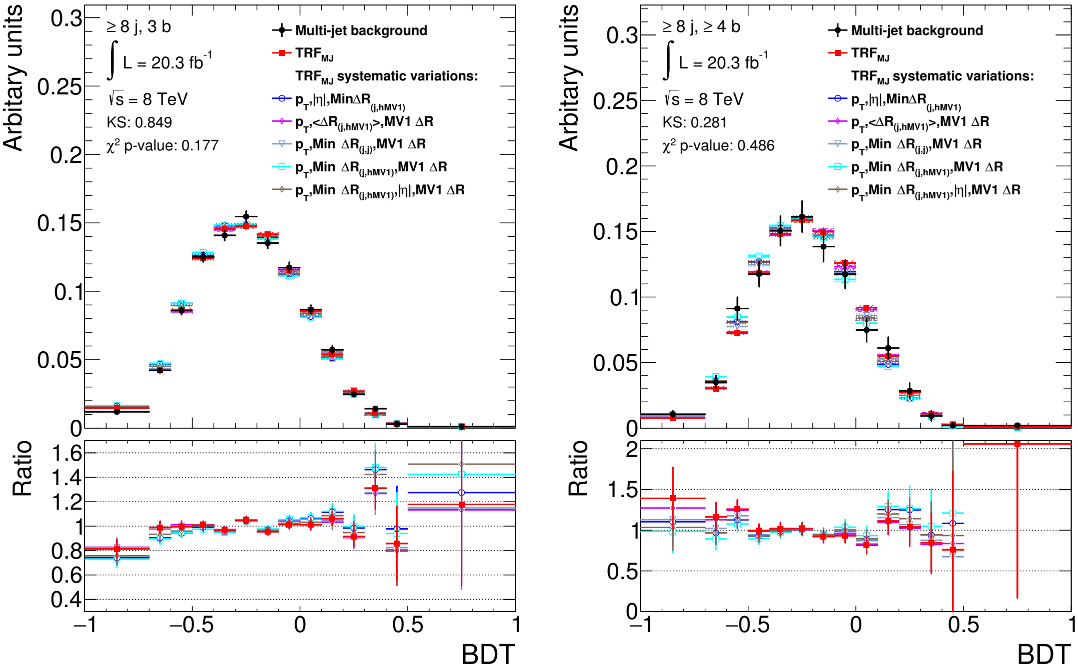
Figure 6.32: Comparison of the modeling for the M_{bb} for the b-tagged jets with minimum ΔR (a), LLR (b), S_t (c) and H_T (d) between the different sets of TRF_{MJ} parameterisations and expected multijet background, defined as data minus MC background contribution, for the region ($\geq$ 8j, $\geq$ 4b). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF_{MJ} for each distribution.



(a) (b)

(c) (d)

Figure 6.33: Comparison of the modeling for the BDT output between the different sets of TRF$_{\mathrm{MJ}}$ parameterisations and expected multijet background, defined as data minus MC background contribution, for the regions (6j, 3b) (a), (6j, $\geq$ 4b) (b), (7j, 3b) (c), (7j, $\geq$ 4b) (d). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF$_{\mathrm{MJ}}$ for each distribution.



(a)

(b)

Figure 6.34: Comparison of the modeling for the BDT between the different sets of $\mathrm{TRF_{MJ}}$ parameterisations and expected multijet background, defined as data minus MC background contribution, for the regions ($\geq$ 8j, 3b) (a), ($\geq$ 8j, $\geq$ 4b) (b). Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal $\mathrm{TRF_{MJ}}$ for each distribution.

	(6j, 3b)	(6j, $\geq$ 4b)	(7j, 3b)
p_T,$\lvert\eta\rvert$,Min$\Delta R_{(j,\mathrm{hMV1})}$	2.05	**2.40**	1.53
p_T,$\lvert\eta\rvert$,$\langle\Delta R_{(j,\mathrm{hMV1})}\rangle$	**1.65**	2.57	**1.49**
p_T,$\langle\Delta R_{(j,\mathrm{hMV1})}\rangle$,MV1 ΔR	2.55	3.28	1.93
p_T,Min$\Delta R_{(j,j)}$,MV1 ΔR	4.66	4.46	3.46
p_T,Min$\Delta R_{(j,\mathrm{hMV1})}$,MV1 ΔR	2.94	4.62	1.89
p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$,MV1 ΔR	2.51	3.30	1.91

	(7j, $\geq$ 4b)	(8j, 3b)	(8j, $\geq$ 4b)
p_T,$\lvert\eta\rvert$,Min$\Delta R_{(j,\mathrm{hMV1})}$	**3.45**	1.60	**3.60**
p_T,$\lvert\eta\rvert$,$\langle\Delta R_{(j,\mathrm{hMV1})}\rangle$	3.52	**1.34**	3.73
p_T,$\langle\Delta R_{(j,\mathrm{hMV1})}\rangle$,MV1 ΔR	3.99	1.56	3.80
p_T,Min$\Delta R_{(j,j)}$,MV1 ΔR	5.49	3.00	5.48
p_T,Min$\Delta R_{(j,\mathrm{hMV1})}$,MV1 ΔR	5.10	1.78	4.39
p_T,$\lvert\eta\rvert$,Min $\Delta R_{(j,\mathrm{hMV1})}$,MV1 ΔR	3.93	1.55	3.81

Table 6.14: χ^2 over number of degrees of freedom in the different event categories obtained using different sets of variables to describe $\varepsilon_{\mathrm{MJ}}$. The smaller value in each region is shown in bold.

6.9 Systematic uncertainties

Several sources of systematic uncertainties, that can affect the normalization of signal and background and/or the shape of their corresponding final discriminant distributions, are considered. Individual sources of systematic uncertainty are considered uncorrelated among them and correlated across analysis regions. Table 6.15 presents a summary of the systematic uncertainties considered in the analysis indicating whether they are taken to be normalization-only ("N"), shape-only ("S"), or to affect both shape and normalization ("SN").

6.9.1 Luminosity

The luminosity estimate has an uncertainty of 2.8% [120] in $\sqrt{s} = 8$ TeV analyses. This systematic uncertainty is applied to all MC simulation processes.

6.9.2 Physics Objects

In this section uncertainties in the reconstruction of jets, and b-, c-, and light flavour-tagging are considered.

Trigger efficiency

After applying data/MC Scale Factor, described in Section 4.2, to the single jet trigger efficiency, the residual difference between the efficiency in $t\bar{t}H$ events and on PYTHIA8 di-jet MC events is taken as systematic uncertainty. The effect on the shape in $t\bar{t}H$ events is shown in Figure 4.14.

Jet Reconstruction Efficiency

The jet reconstruction efficiency is found to be about 0.2% lower in the simulation than in data for jets below 30 GeV and it is consistent with data for higher jet p_T. To evaluate systematic uncertainty due to this small inefficiency, 0.2% of the jets with p_T below 30 GeV are removed randomly and all jet-related kinematic variables are recomputed. The event selection is repeated using the modified selected jet list. Given the small uncertainty on the jet reconstruction efficiency measured in this analysis, this systematic uncertainty is neglected.

Jet Vertex Fraction Efficiency

The per-jet efficiency associated to the jet vertex fraction requirement is measured in $Z(\rightarrow \ell^+\ell^-)$+1-jet events in data and simulation, selecting separately events enriched in hard-scatter jets and events enriched in jets from other proton interactions in the same bunch crossing (pileup). The corresponding uncertainty is evaluated in the analysis by changing the nominal JVF cut value by 0.03 up and down and repeating the analysis using the modified cut value.

Systematic uncertainty	Type	Components
Luminosity	N	1
Physics Objects		
Jet energy scale	SN	22
Jet vertex fraction	SN	1
Jet trigger	SN	1
Jet energy resolution	SN	1
Jet reconstruction efficiency	SN	1
b-tagging efficiency	SN	6
c-tagging efficiency	SN	4
Light jet-tagging efficiency	SN	12
Background Model		
$t\bar{t}$ cross section	N	1
$t\bar{t}$ modeling: p_T reweighting	SN	9
$t\bar{t}$ modeling: parton shower	SN	3
$t\bar{t}$+heavy-flavour: normalization	N	2
$t\bar{t}+c\bar{c}$: HF 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
Single top cross section	N	1
Multijet normalization	N	6
Multijet $\mathrm{TRF}_{\mathrm{MJ}}$ parametrization	S	6
Multijet H_T correction	S	1
Multijet S_T correction	S	1
Signal Model		
$t\bar{t}H$ modeling	SN	2
$t\bar{t}H$ hadronization	SN	1
$t\bar{t}H$ Parton Shower	SN	1

Table 6.15: List of systematic uncertainties considered. “N” means that the uncertainty is taken as normalization-only for all processes and channels affected, whereas “S” denotes systematics that are considered shape-only in all processes and channels. “SN” means that the uncertainty is taken as both shape and normalization. Some of the systematic uncertainties are split into several different components for a more accurate treatment (number given in column “Components”).

Jet Energy Scale

The jet energy scale (JES) and its uncertainty have been derived combining information from test-beam data, LHC collision data and simulation [88]. The jet energy scale uncertainty is split into 22 uncorrelated sources which can have different jet p_T and η dependencies and are treated independently in this analysis.

Jet Energy Resolution

The jet energy resolution has been measured separately for data and simulation using two *in-situ* techniques [88]. A systematic uncertainty is defined as the quadratic difference between the jet energy resolutions for data and simulation. To estimate the corresponding systematic uncertainty in the analysis, the energy of jets in the simulation is smeared by this difference, and the changes in the normalization and shape of the final discriminant are compared to the default prediction.

In order to propagate the uncertainty in the p_T resolution, for each jet in the simulation, a Gaussian distributed random number r is used with mean 0 and sigma equal to the difference in quadrature between the fractional p_T resolution, measured as a function of jet's p_T and rapidity, and the nominal one. The jet 4-momentum is then scaled by a factor $1 + r$. Resulting uncertainty on the normalization and shape of the final discriminant is one-sided, since jets in the simulation cannot be under-smeared. This uncertainty is then symmetrised.

Heavy- and Light-Flavour Tagging

The effects of uncertainties in efficiencies for the heavy flavour identification of jets by the b-tagging algorithm have been evaluated. These efficiencies are measured from data and depend on the jet flavour. Efficiencies for b and c quarks in the simulation have to be corrected by p_T-dependent factors [91]. The scale factors and their uncertainties are applied to each jet in the simulation depending on its flavour and p_T. In the case of light-flavour jets, the corrections also depend on jet η.

A total of six independent sources of uncertainty affecting the b-tagging efficiency and four affecting the c-tagging efficiency are considered. Each of these uncertainties correspond to a resulting eigenvector after diagonalising the matrix containing the information of total uncertainty per p_T bin and the bin-to-bin correlations. Twelve uncertainties are considered for the light jets tagging which depend on jet p_T and η region. These systematic uncertainties are taken as uncorrelated between b, c jets, and light flavour jets.

A per-jet weighting procedure is applied to simulated events to propagate the calibration of b-tagging and the related uncertainties.

An additional uncertainty is assigned to b-tagging efficiency for jets with $p_T >$ 300 GeV. This uncertainty is related to the fact that b-tagging efficiency map used in the $\mathrm{TRF}_{\mathrm{MC}}$ method, see Section 6.4.6, are defined up to jet $p_T = 300$ GeV and the value of the last bin is used for jets with p_T above this limit.

6.9.3 Uncertainties on the Multijet background

The systematic uncertainties considered for the multijet background involves only uncertainties on the $\mathrm{TRF_{MJ}}$ shape, indeed since the statistical uncertainties on the normalization from the $\mathrm{TRF_{MJ}}$ prediction, do not give reason of the spread between prediction and observation in the analysis' regions (Table 6.13), one can use the lower part of the BDT, dominated by multijet contribution, to normalize this background to a statistical precision of at most 5% as it is described, in Section 6.10. For the shape uncertainties two kinds of systematics are considered: one "a priori" derived from the choice of parametrization used for this analysis, and another kind "a posteriori" derived from shape unclosure in the validation region.

Multijet background systematics from the choice of parametrization

As presented in Section 6.8, six different sets of variables have been studied to describe the $\varepsilon_{\mathrm{MJ}}$ used in the multijet modeling. The nominal set has been selected through a χ^2 minimisation using all the observable inputs of the BDT. Each of the $\varepsilon_{\mathrm{MJ}}$ is used as input in the $\mathrm{TRF_{MJ}}$ method to estimate the multijet background. The difference between all different distributions and the nominal one is taken as systematic uncertainty.

In addition, the dependency on the choice of b-tagged jets excluded from the $\varepsilon_{\mathrm{MJ}}$ evaluation has also been tested. As previously discussed, the sample where the $\mathrm{TRF_{MJ}}$ method is applied contains exactly 2 b-tagged jets. Probability $\varepsilon_{\mathrm{MJ}}$ is evaluated in a sample with at least 2 b-tagged jets where the two with highest MV1 weight are excluded. This reflects the fact that the two jets excluded when applying $\mathrm{TRF_{MJ}}$ are the only two b-tagged jets in the event, a fortiori the ones with the highest MV1 weight. In order to have an assessment of the systematics coming from this assumption, different $\varepsilon_{\mathrm{MJ}}$ probabilities are derived modifying the choice of jets to be excluded, in particular 2 b-tagged jets are excluded randomly, or, by inverting the logic, the ones with lowest MV1 value are excluded. This two resulting $\varepsilon_{\mathrm{MJ}}$ are shown in Figure 6.35. The result obtained by using these modified $\varepsilon_{\mathrm{MJ}}$ is assigned as systematic uncertainty.

Figures 6.38 and 6.38 show the comparison of the shape of the multijet background prediction obtained with the different $\varepsilon_{\mathrm{MJ}}$ considered. Figures 6.36 and 6.36 show the total systematic uncertainty induced on the BDT. Plots are made for the different jet multiplicities and for both 3 b tags and ≥ 4 b-tagged jets categories.

Multijet background systematics from mismodelling found in the validation region for S_T and H_T

As shown in Section 6.5.3, Figure 6.12, the validation region shows residual mismodelling even after the envelop of different $\mathrm{TRF_{MJ}}$ are considered. Among the different mismodelling two classes are considered for systematics studies:

- H_T is not currently used in the BDT, but some variables which are highly correlated with it are, i.e. H_{T5} and $(E_{\mathrm{T}1} + E_{\mathrm{T}2})/\sum(\text{jets } E_{\mathrm{T}})$. While this

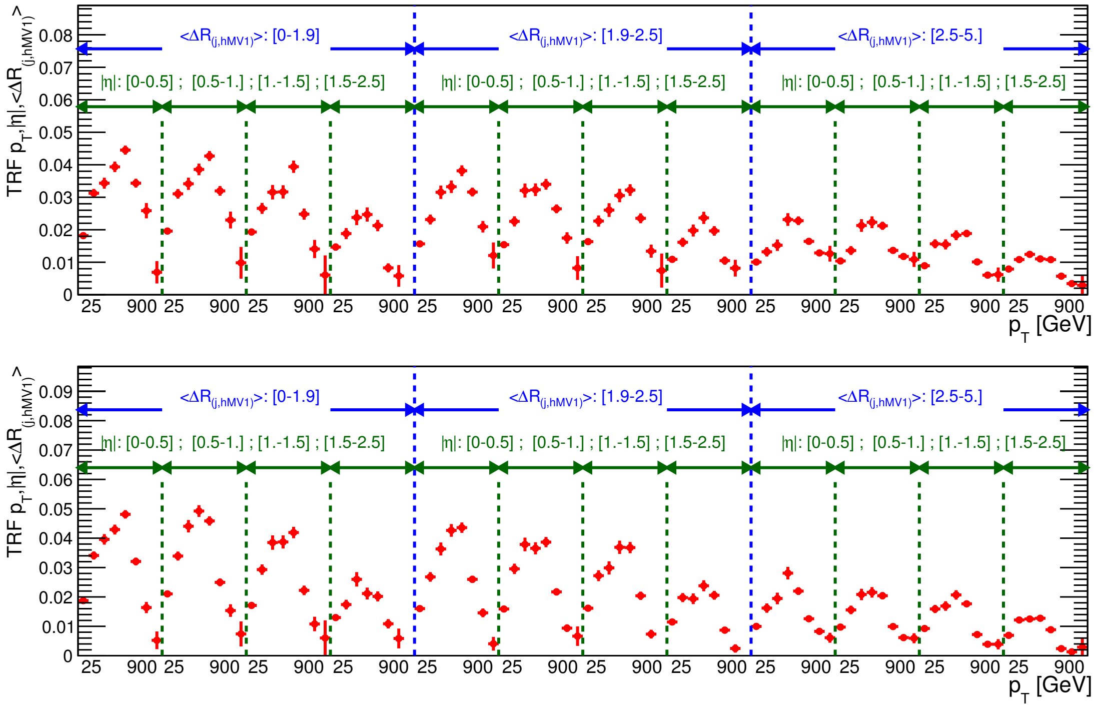
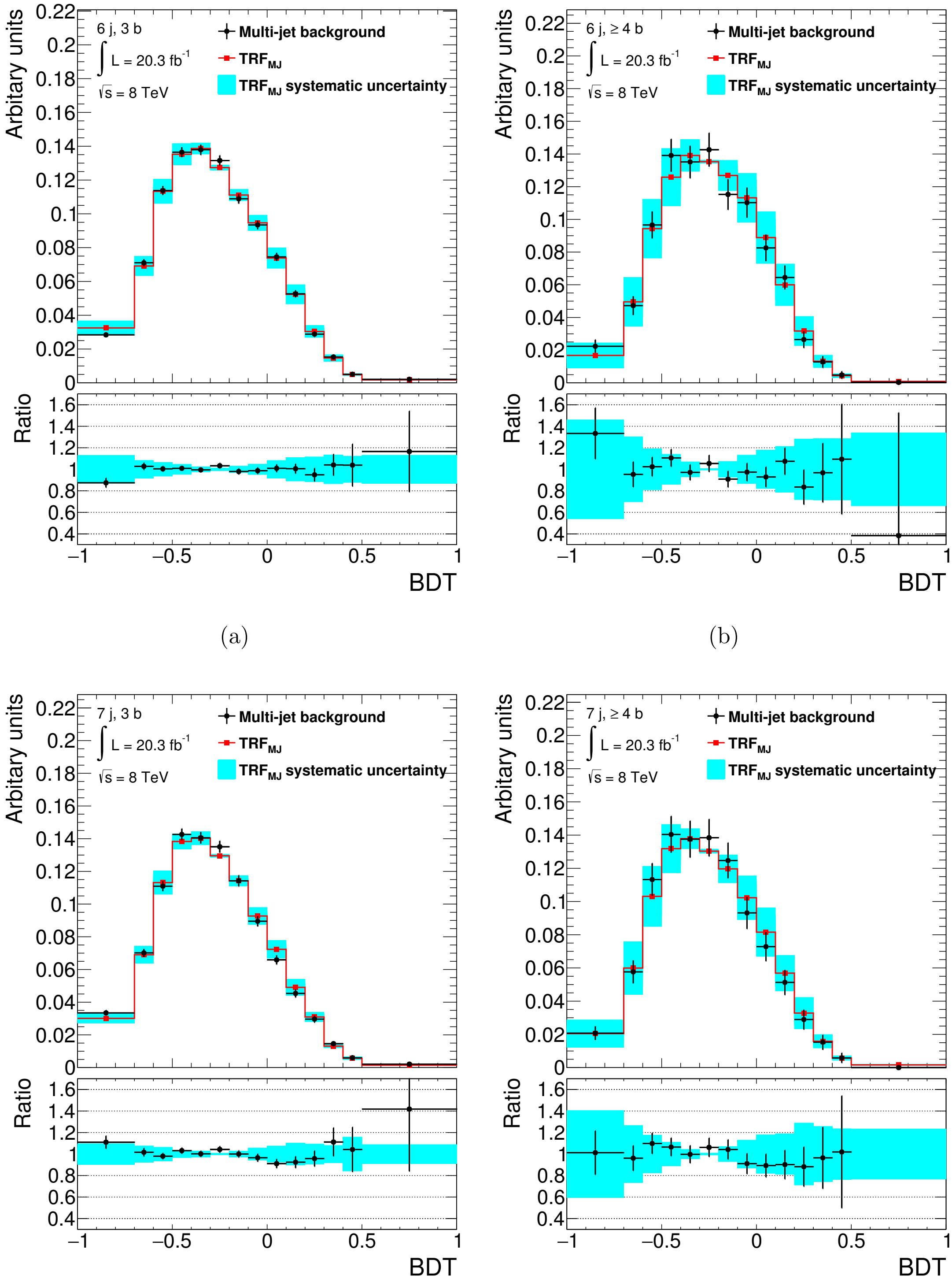


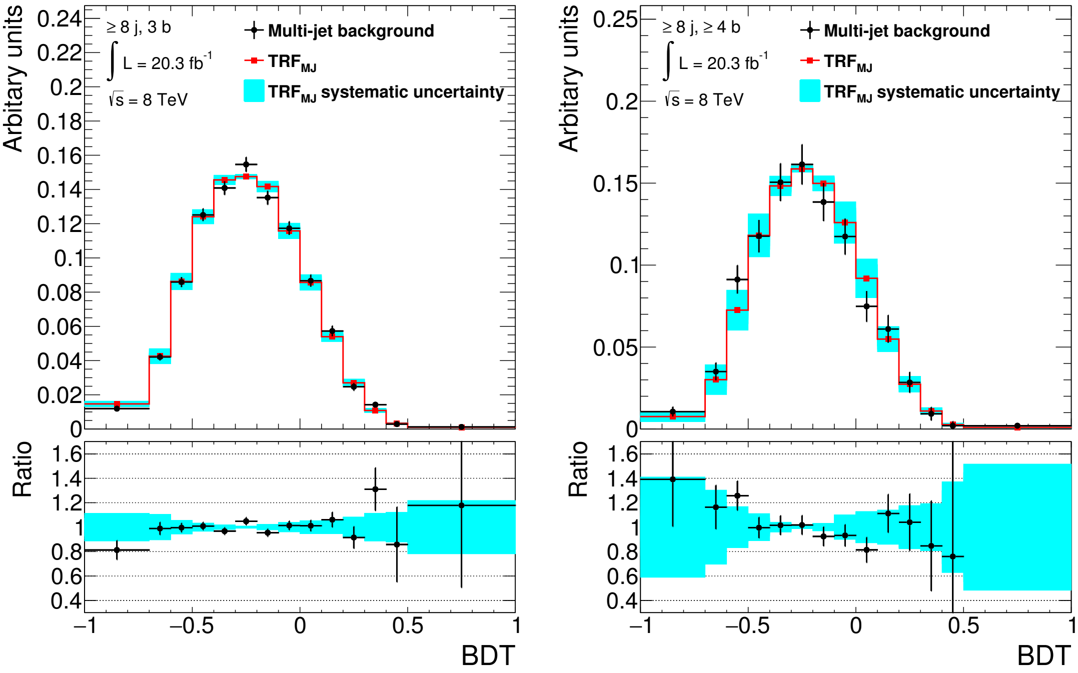
Figure 6.35: One dimensional view of $\varepsilon_{\mathrm{MJ}}$ as a function of the p_T and $|\eta|$ of the jet and $\left\langle \Delta R_{(j,\mathrm{hMV1})} \right\rangle$. In the top (bottom) plot $\varepsilon_{\mathrm{MJ}}$ is built by removing randomly (the lowest MV1) two jets from the list of b-tagged jets in the event.



(a) (b)

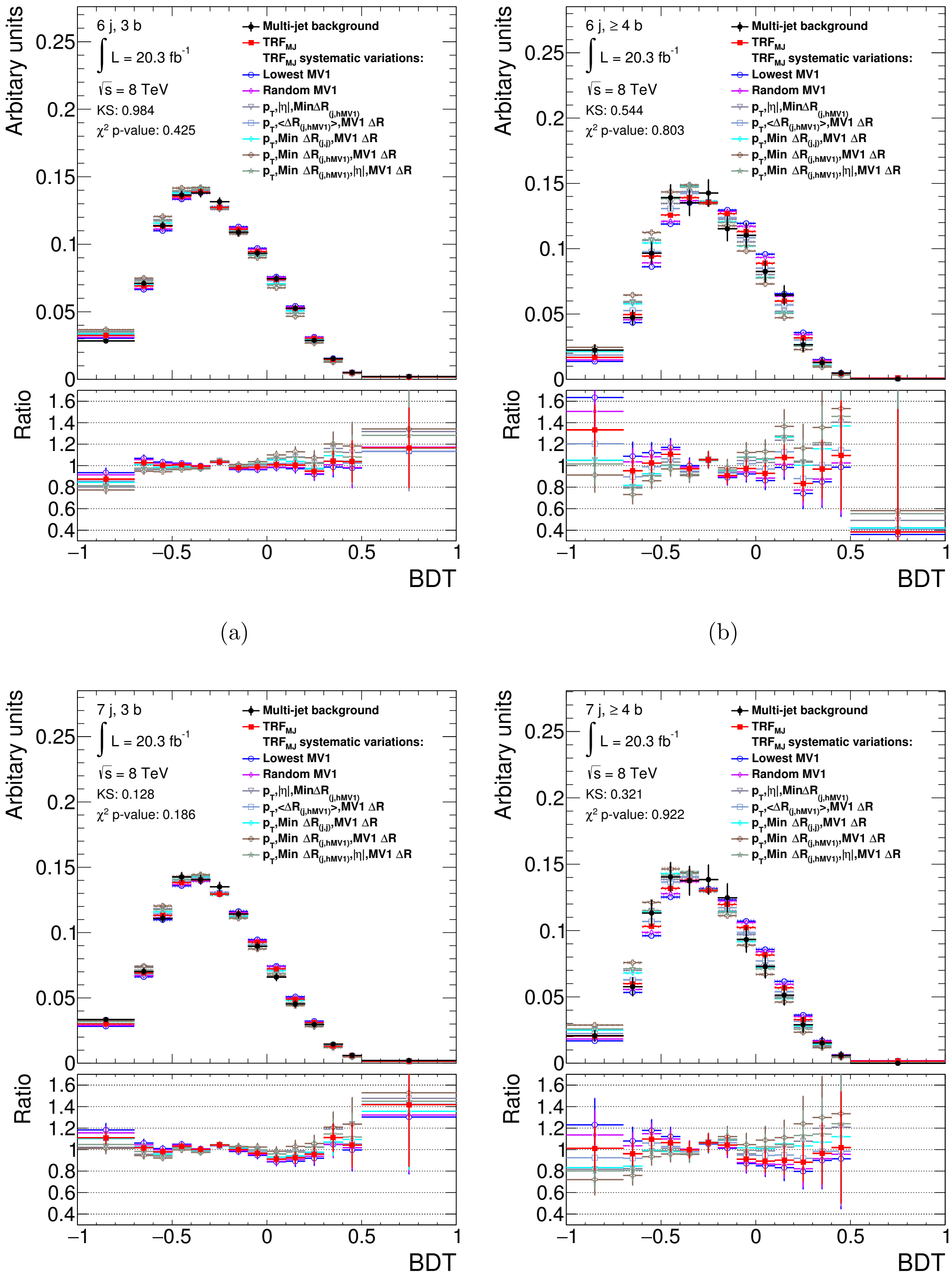
(c) (d)

Figure 6.36: Shape systematic variations induced by the systematic $\varepsilon_{\mathrm{MJ}}$ probabilities on the BDT. Regions (6j, 3b) (a), (6j, $\geq$ 4b) (b), (7j, 3b) (b) and (7j, $\geq$ 4b) (c) are shown. Black points represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the $\mathrm{TRF}_{\mathrm{MJ}}$ method. The blue area represents the total systematic uncertainty related the choice of $\varepsilon_{\mathrm{MJ}}$ description variables and b-tagged jets excluded in $\varepsilon_{\mathrm{MJ}}$ evaluation.



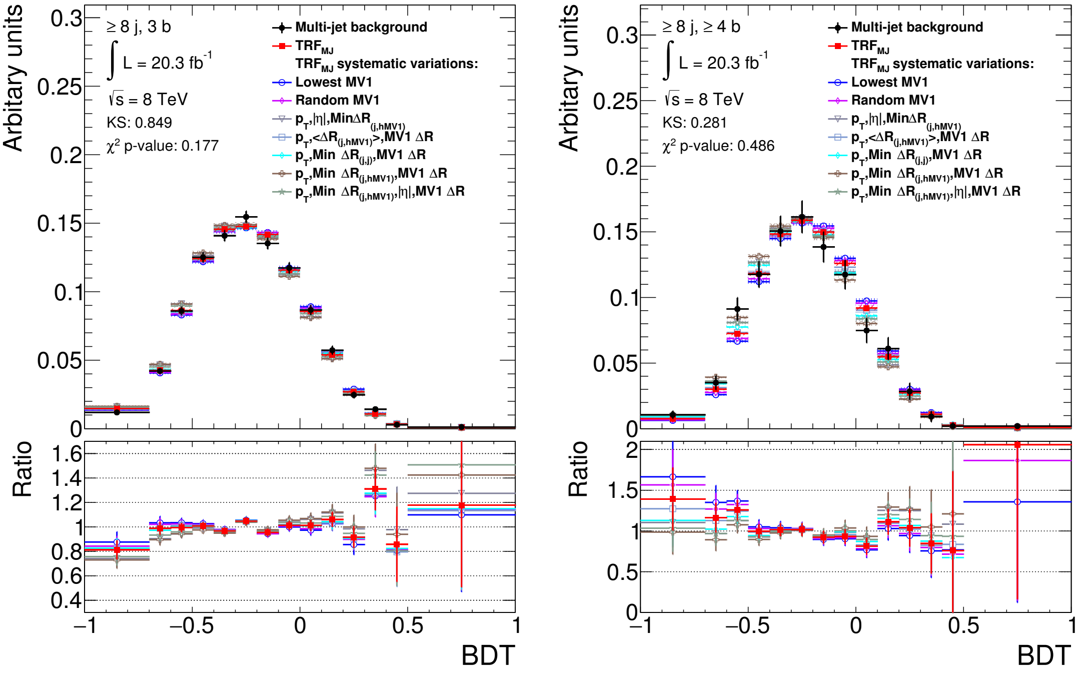
(a) (b)

Figure 6.37: Shape systematic variations induced by the systematic $\varepsilon_{\mathrm{MJ}}$ probabilities on the BDT. Regions (≥ 8j, 3b) (a), and (≥ 8j, ≥ 4b) (b) are shown. Black points represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the TRF$_{\mathrm{MJ}}$ method. The blue area represents the total systematic uncertainty related the choice of $\varepsilon_{\mathrm{MJ}}$ description variables and b-tagged jets excluded in $\varepsilon_{\mathrm{MJ}}$ evaluation.



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

Figure 6.38: Comparison of BDT shape systematic variations between the different $\varepsilon_{\mathrm{MJ}}$ and expected multijet background, defined as data minus MC background contribution. Regions (6j, 3b) (a), (6j, $\geq$ 4b) (b), (7j, 3b) (b) and (7j, $\geq$ 4b) (c) are shown. Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal TRF$_{\mathrm{MJ}}$ for each distribution.



(a) (b)

Figure 6.39: Comparison of BDT shape systematic variations between the different $\varepsilon_{\mathrm{MJ}}$ and expected multijet background, defined as data minus MC background contribution. Regions ($\geq$ 8j, 3b) (a), and ($\geq$ 8j, $\geq$ 4b) (b) are shown. Also shown the results of the Kolmogorov-Smirnov and χ^2 tests for the nominal $\mathrm{TRF}_{\mathrm{MJ}}$ for each distribution.

variable is well reproduced in the 3 b-tag region, it seems that a systematic tilt is present in the ≥ 4 b-tagged jets region, as shown in Figures 6.12, 6.28, 6.30 and 6.32.

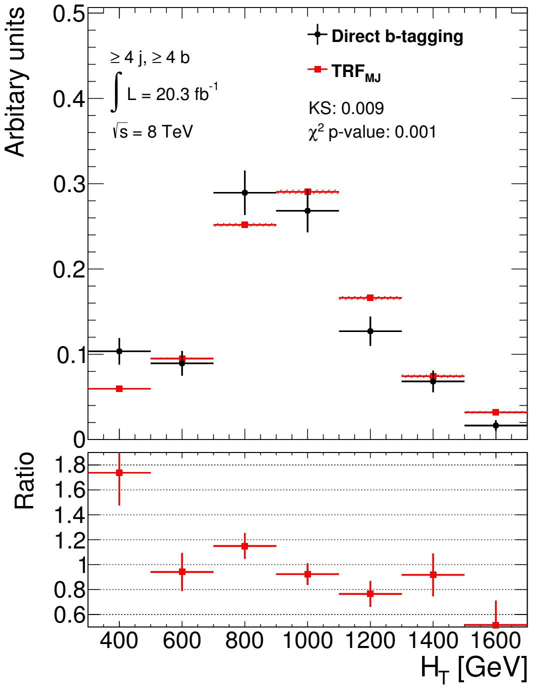
- S_T, the object based missing energy, ranks very high in several signal regions (see Table 6.9). This variable shows a same mismodelling behaviour in all different jet and b-jet multiplicities in the TRF_{MJ} extraction region. This is also confirmed when the same quantities are looked at in the signal regions, see Section 6.8, Figures 6.27, 6.28 and 6.29, 6.30, 6.31 and 6.32.

To estimate the impact of such mismodelling, reweights are derived in the TRF_{MJ} evaluation region, defined in Section 6.5.1 and applied in the signal region for both H_T and S_T.

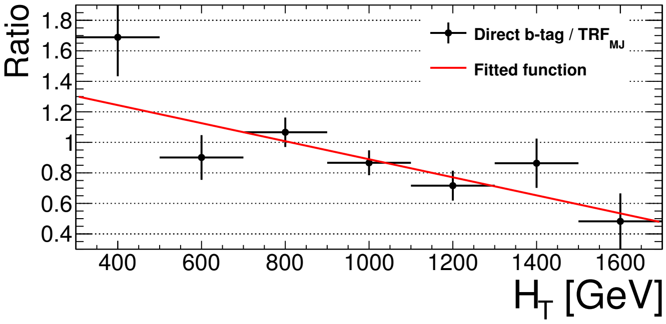
For H_T the reweight is applied (and derived) only in the inclusive 4 b-tagged jet regions. The ratio between data and prediction is fitted with a linear function in the TRF_{MJ} evaluation region, and then applied as a reweight in the signal region. Figure 6.40(a) shows the region ($\geq$ 4j, $\geq$ 4b) of the TRF_{MJ} evaluation region, where the reweight is extracted. Figure 6.40(b) shows the ratio points and the fitted function, of the form $y = x \times p_1 + p_0$. The values obtained form the fit are: $p_0 = 1.48 \pm 0.17$ and $p_1 = (-5.9 \pm 1.6) \times 10^{-4}$. Figures 6.41 show the effects of the reweighting on H_T itself for the signal region. In the ratio plot, the blue markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of multijet background and the predicted nominal one. The fact that the blue and the red markers overlap means that the mismodelling is completely recovered.

For S_T, the reweight is based on the actual mismodelling as seen in the TRF_{MJ} evaluation region since the mismodelling itself does not seem to follow a particular functional form. In Figures 6.42 and 6.43 the reweighted S_T is shown. To make sure that the two reweighting processes are uncorrelated, the S_T correction is applied to H_T as can be seen in Figures 6.44 and 6.45, and vice versa in Figure 6.46. In this case the blue markers in the ratio plot are flat. This means that no distortion of the original shape is seen after reweight.

The effect on the BDT output is shown in Figures 6.47, 6.48 and 6.49. Both the indicators, KS and χ^2 test, indicates a better agreement of the reweighted prediction with data. The difference in shape is so small that the unweighted distribution is kept as nominal and the difference between the nominal and the reweighted BDT is taken as a systematic. The magnitude of systeamtic uncertainty is the 100% of the reweight. All the multijet shape NP are listed in Table 6.16.

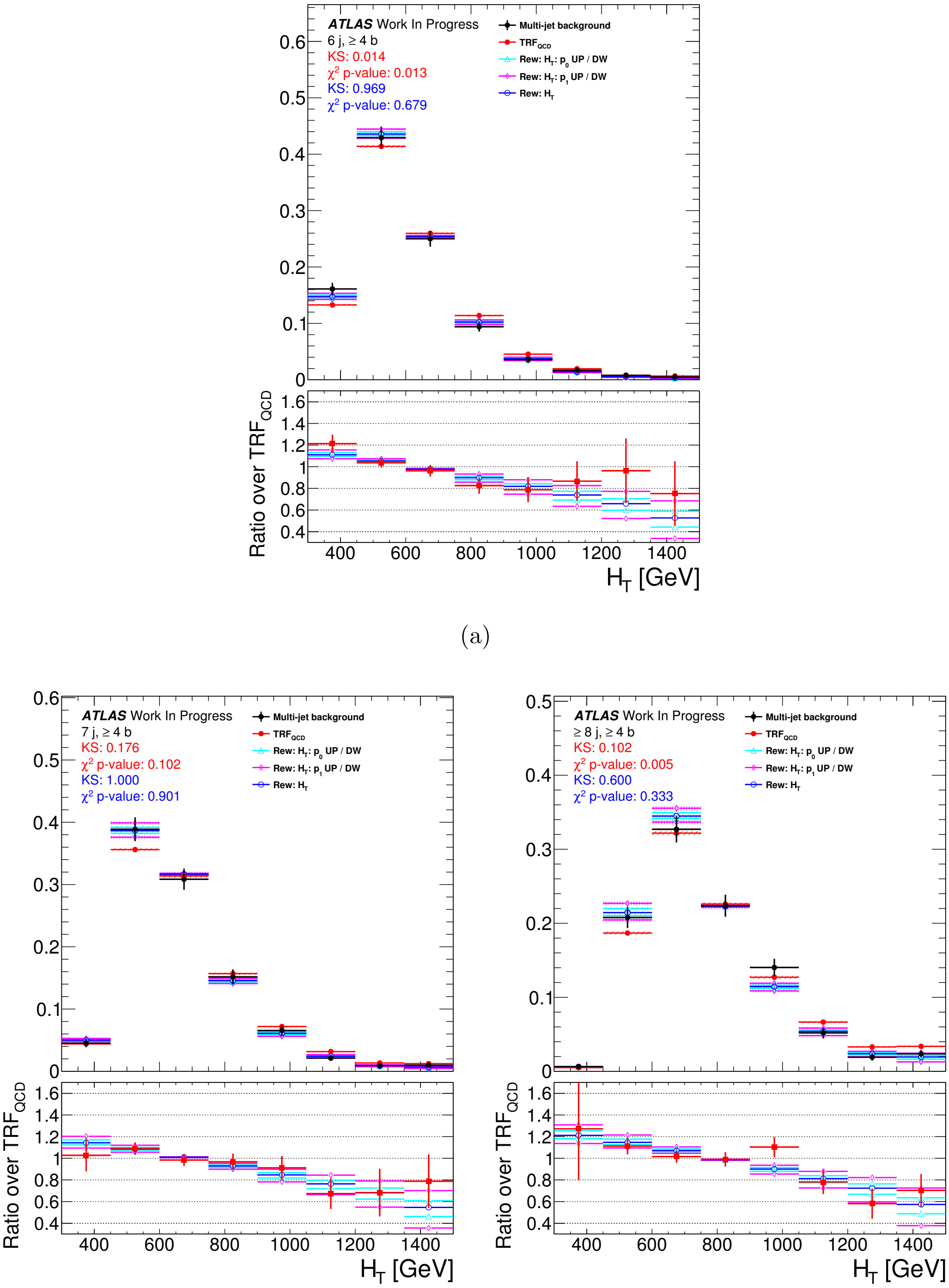


(a)



(b)

Figure 6.40: (a) Comparison of shapes for H_T in events obtained with direct b-tagging (black points) and prediction with $\mathrm{TRF_{MJ}}$ method. Bottom panel shows the ratio of the direct b-tag over the $\mathrm{TRF_{MJ}}$ prediction (red squares). (b) Ratio of the direct b-tag over the $\mathrm{TRF_{MJ}}$ prediction (black circles) and fit of a linear function of the form $y = x \times p_1 + p_0$, where best-fit values are $p_0 = 1.48 \pm 0.17$ and $p_1 = (-5.9 \pm 1.6) \times 10^{-4}$ (red line). Both plots refer to the category ($\geq$ 4j, $\geq$ 4b) of the $\mathrm{TRF_{MJ}}$ evaluation region.

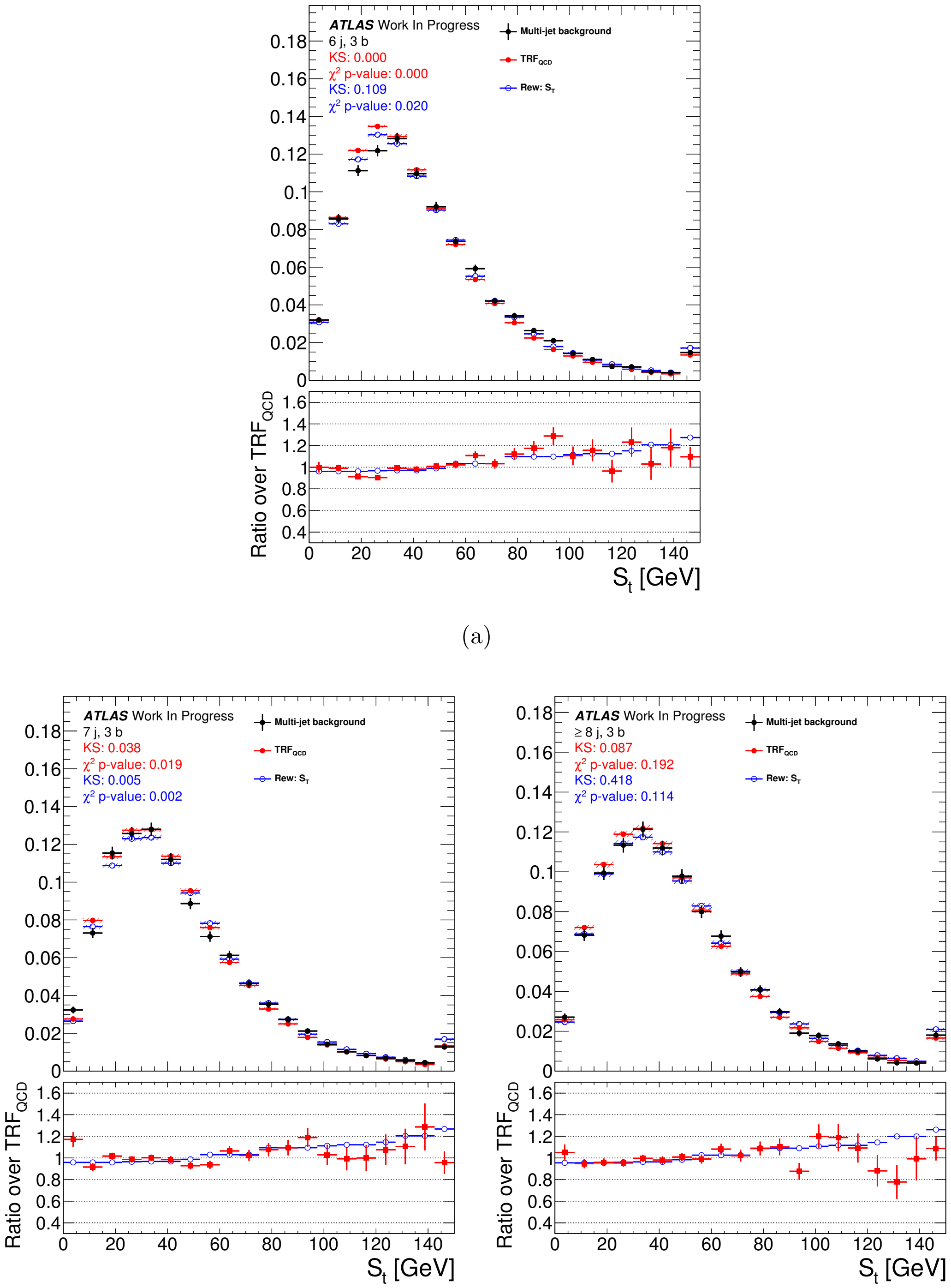


(a)

(b)

(c)

Figure 6.41: Multijet shape systematic variations on H_T induced by the reweight on H_T. Regions (6j, $\geq$ 4b) (a), (7j, $\geq$ 4b) (b) and ($\geq$ 8j, $\geq$ 4b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represent the prediction using the TRF_{MJ} method, blue markers are the reweighted prediction together with the variation on the two parameters used in the fit to evaluate the H_T reweight. In the ratio plots, the blue, cyan and purple markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.

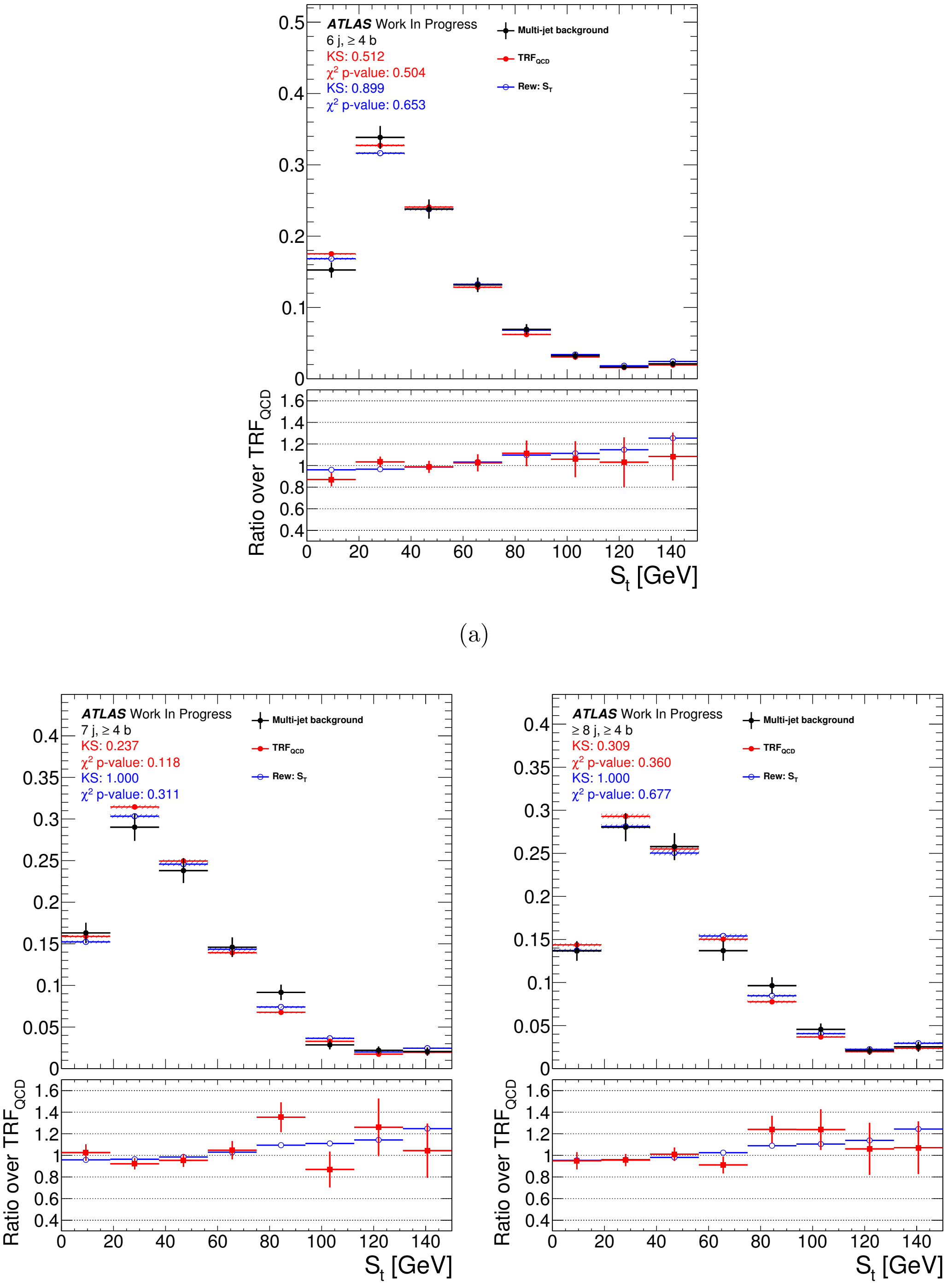


(a)

(b)

(c)

Figure 6.42: Multijet shape systematic variations on S_T induced by the reweight on S_T. Regions (6j, 3b) (a), (7j, 3b) (b) and (≥ 8j, 3b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the TRF_{MJ} method, blue markers are the reweighted prediction. In the ratio plots, the blue markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.

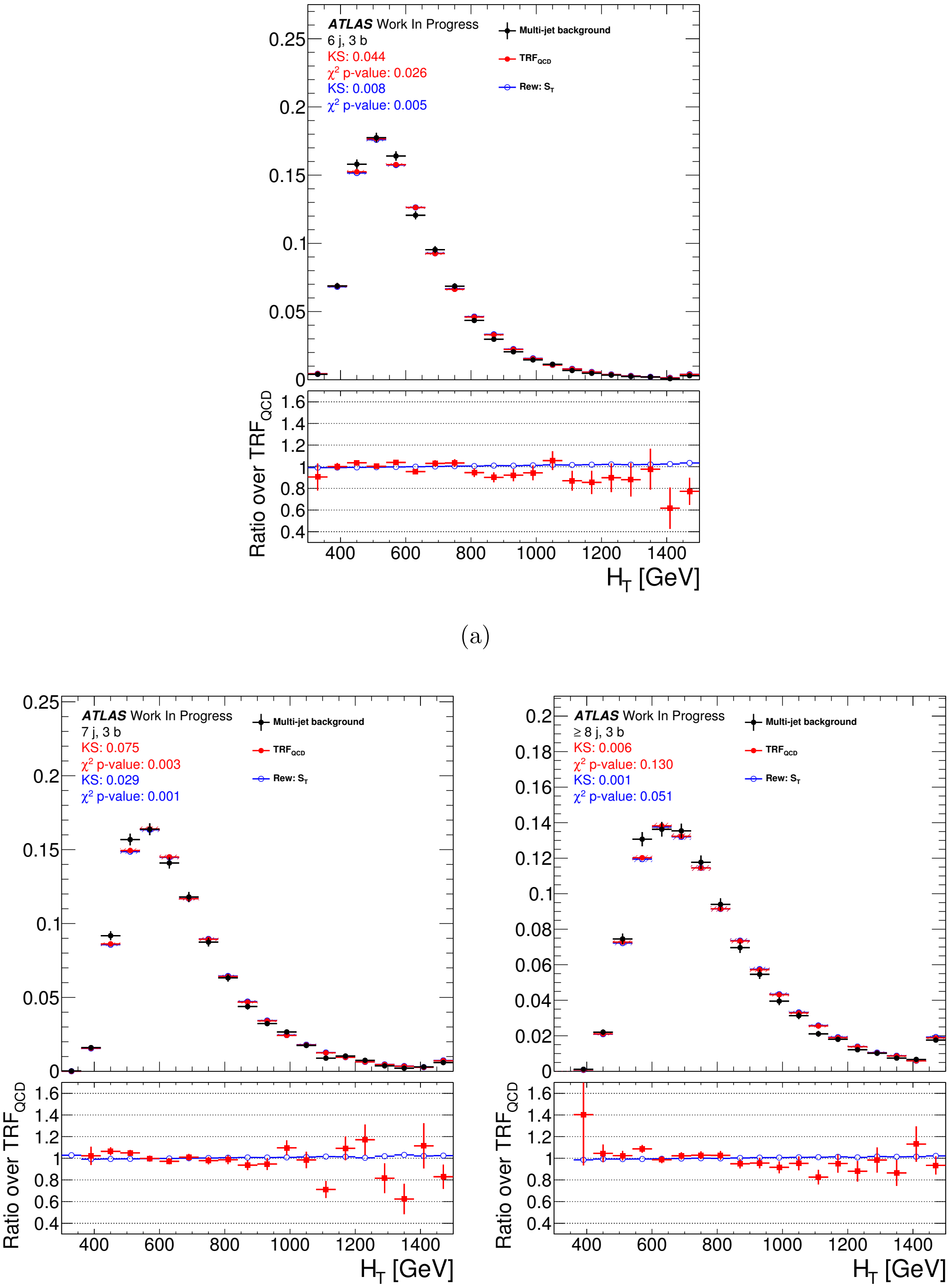


(a)

(b)

(c)

Figure 6.43: Multijet shape systematic variations on S_T induced by the reweight on S_T. Regions (6j, $\geq$ 4b) (a), (7j, $\geq$ 4b) (b) and ($\geq$ 8j, $\geq$ 4b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the $\mathrm{TRF}_{\mathrm{MJ}}$ method, blue markers are the reweighted prediction. In the ratio plots, the blue markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.

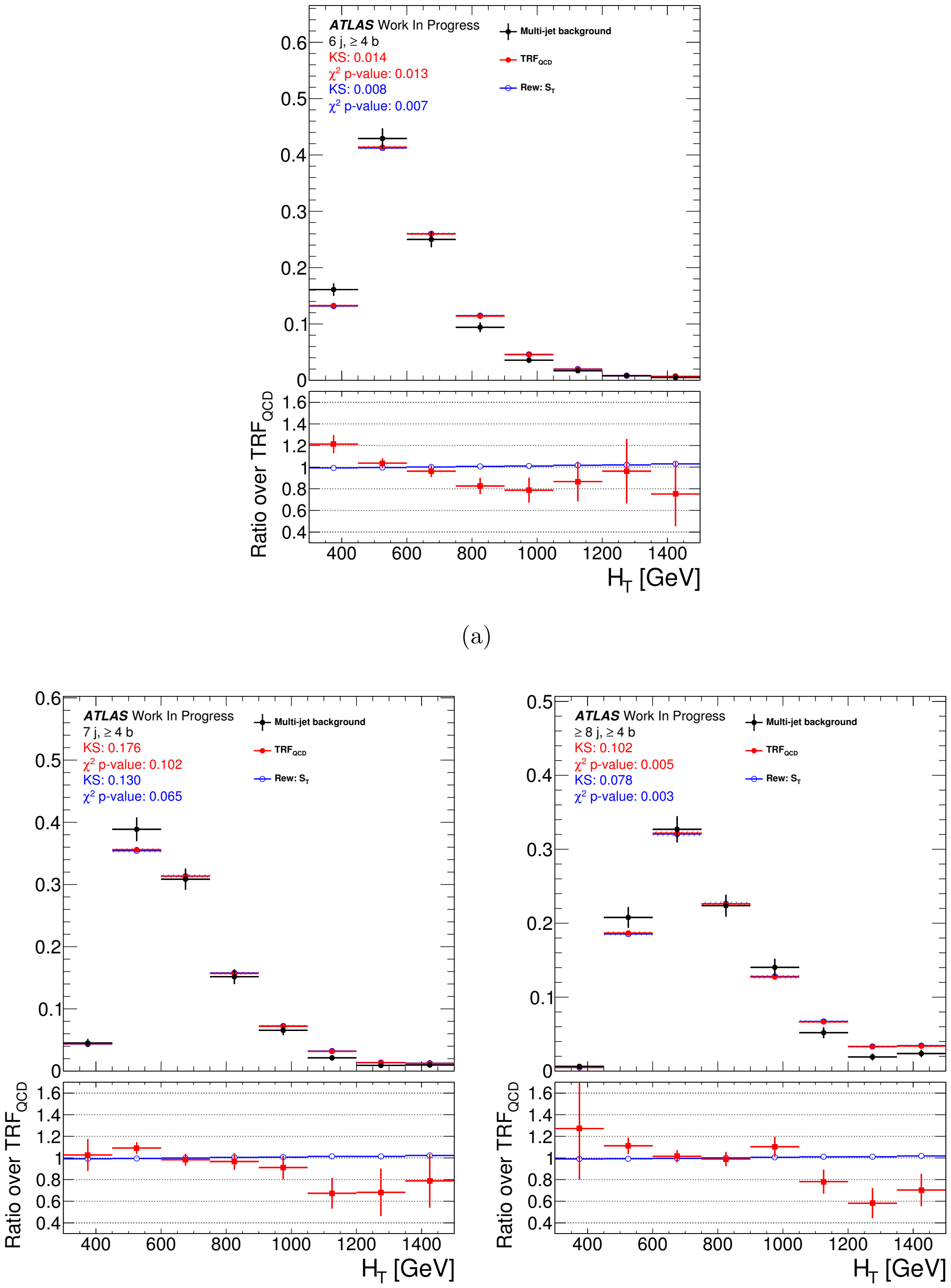


(a)

(b)

(c)

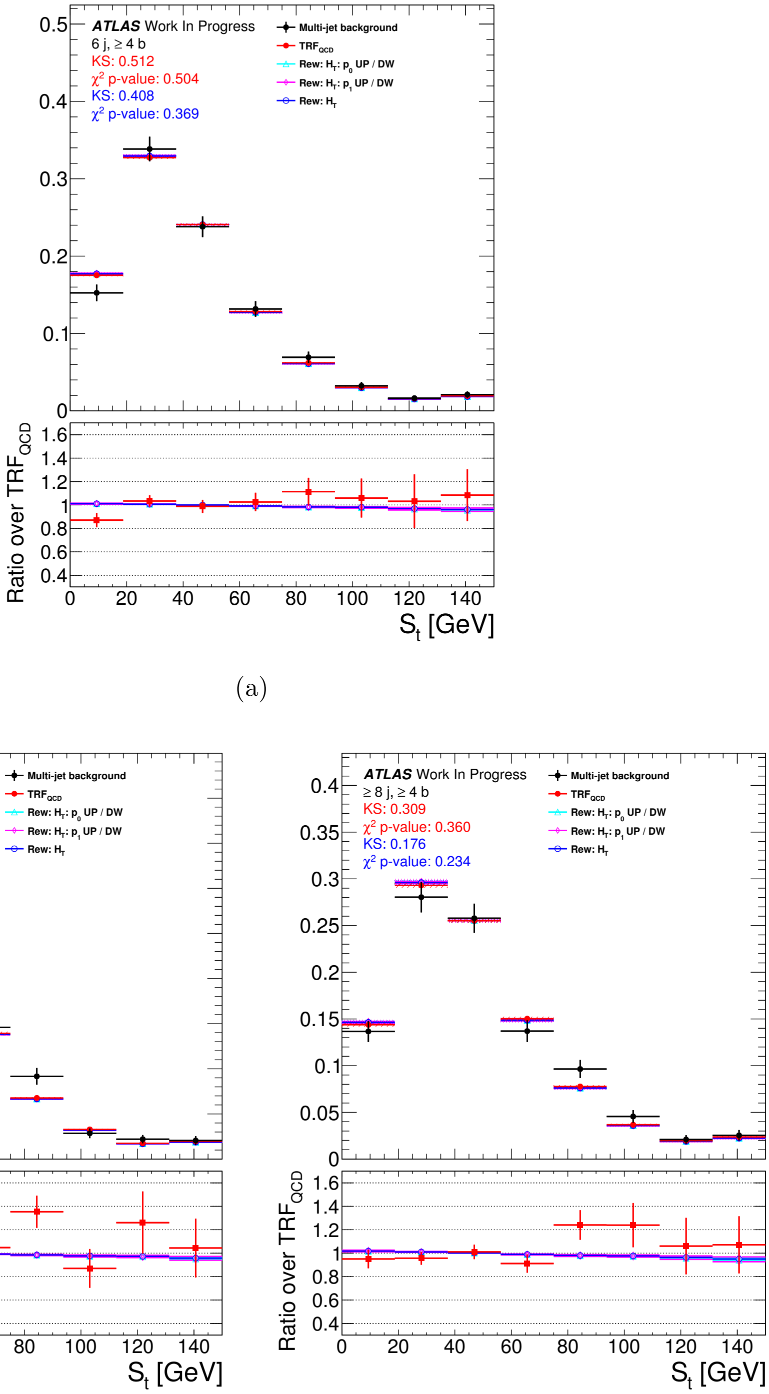
Figure 6.44: Multijet shape systematic variations on H_T induced by the reweight on S_T. Regions (6j, 3b) (a), (7j, 3b) (b) and ($\geq$ 8j, 3b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the $\mathrm{TRF}_{\mathrm{MJ}}$ method, blue markers are the reweighted prediction. In the ratio plots, the blue markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.



(a)

(b) (c)

Figure 6.45: Multijet shape systematic variations on H_T induced by the reweight on S_T. Regions (6j, $\geq$ 4b) (a), (7j, $\geq$ 4b) (b) and ($\geq$ 8j, $\geq$ 4b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the TRF_{MJ} method, blue markers are the reweighted prediction. In the ratio plots, the blue markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.

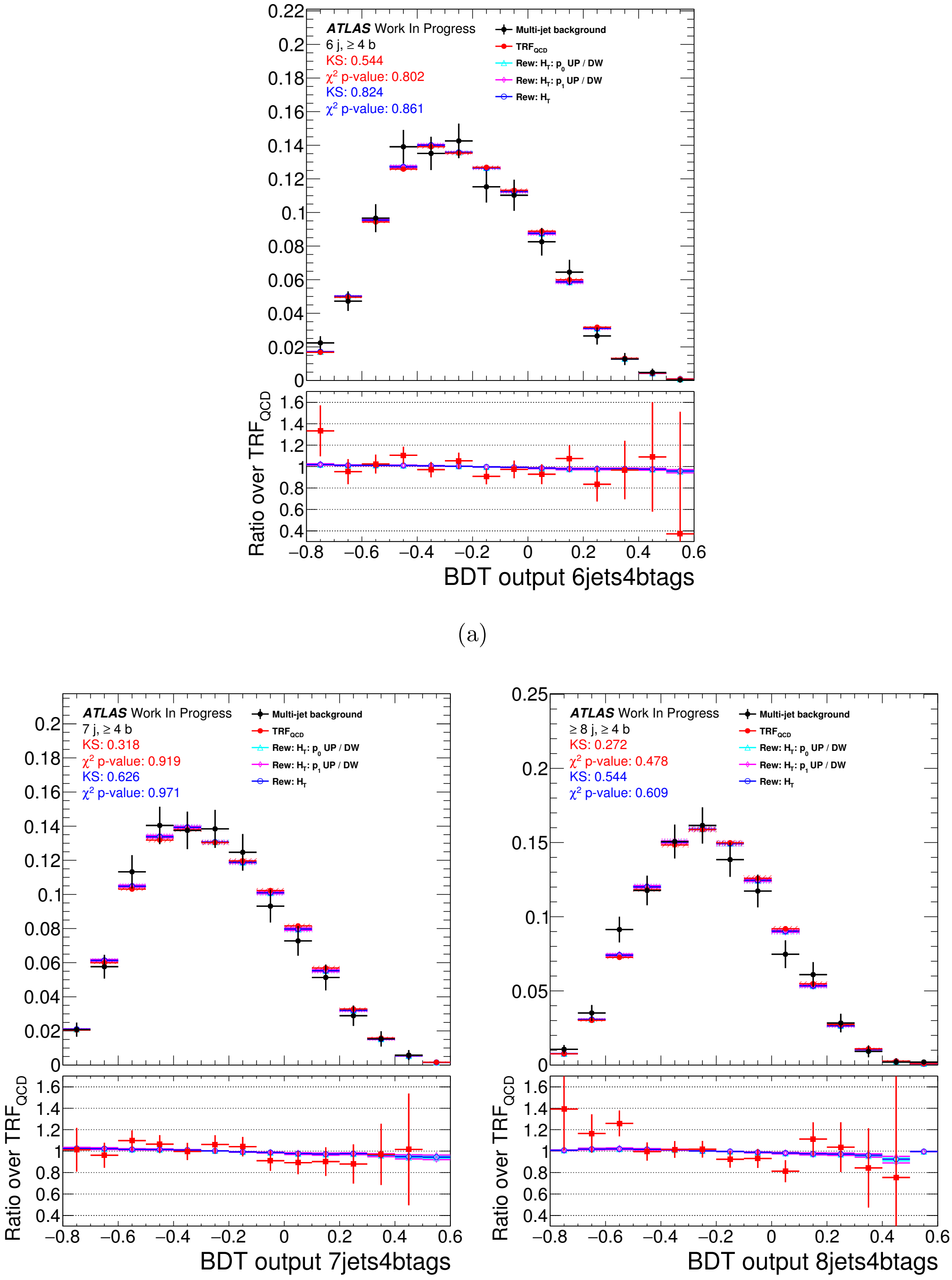


(a)

(b)

(c)

Figure 6.46: Multijet shape systematic variations on S_T induced by the reweight on H_T. Regions (6j, ≥ 4b) (a), (7j, ≥ 4b) (b) and (≥ 8j, ≥ 4b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the $\mathrm{TRF}_{\mathrm{MJ}}$ method, blue markers are the reweighted prediction together with the variation on the two parameters used in the fit to evaluate the H_T reweight. In the ratio plots, the blue, cyan and purple markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the supposed QCD and the predicted nominal one.

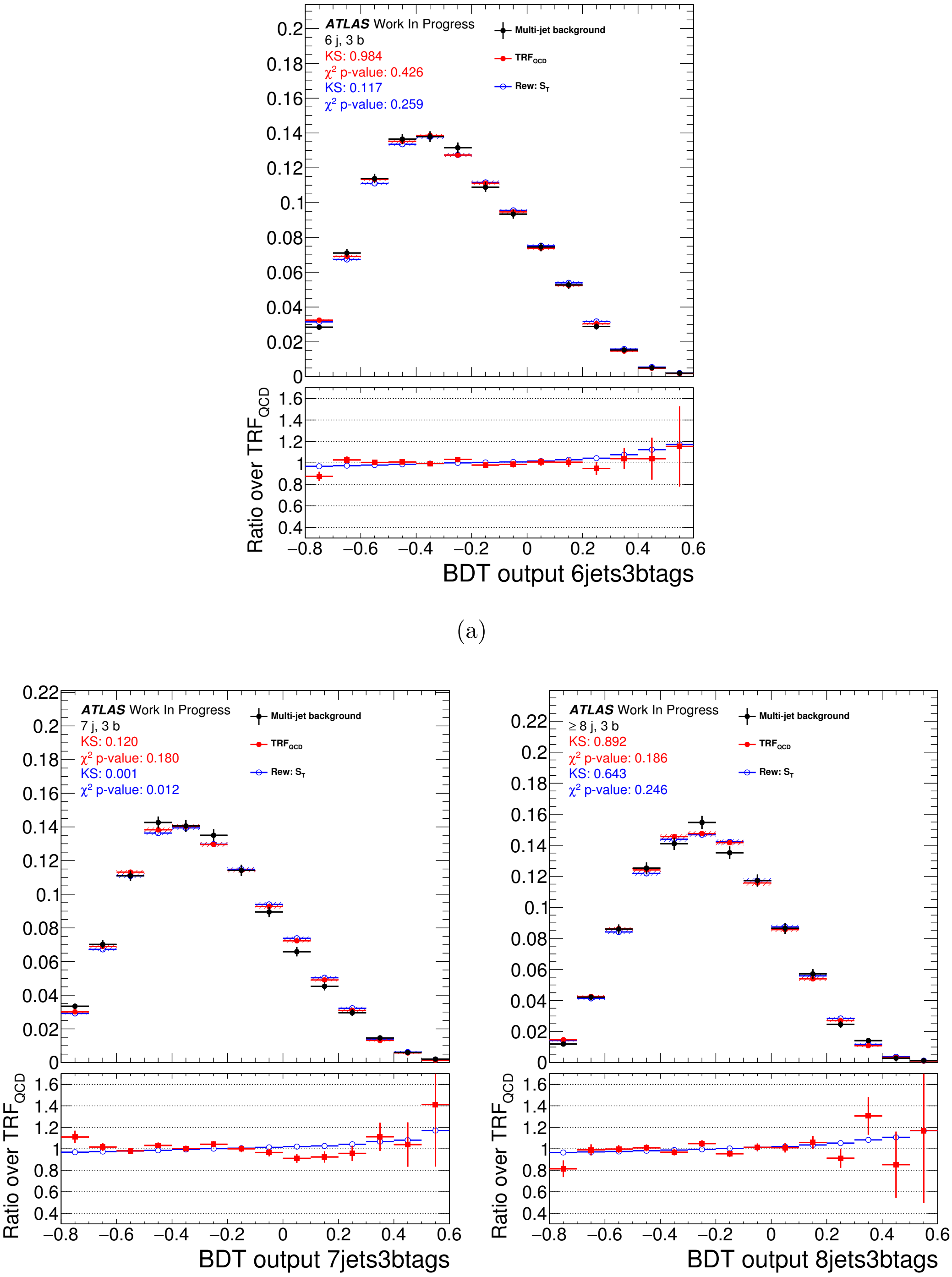


(a)

(b)

(c)

Figure 6.47: Multijet shape systematic variations on the BDT output induced by the reweight on H_T. Regions (6j, $\geq$ 4b) (a), (7j, $\geq$ 4b) (b) and ($\geq$ 8j, $\geq$ 4b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the $\mathrm{TRF_{MJ}}$ method, blue markers are the reweighted prediction together with the variation on the two parameters used in the fit to evaluate the H_T reweight In the ratio plots, the blue, cyan and purple markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.

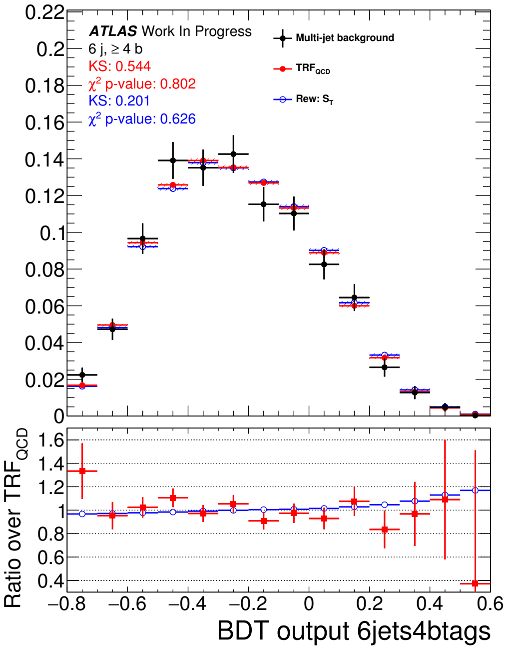


(a)

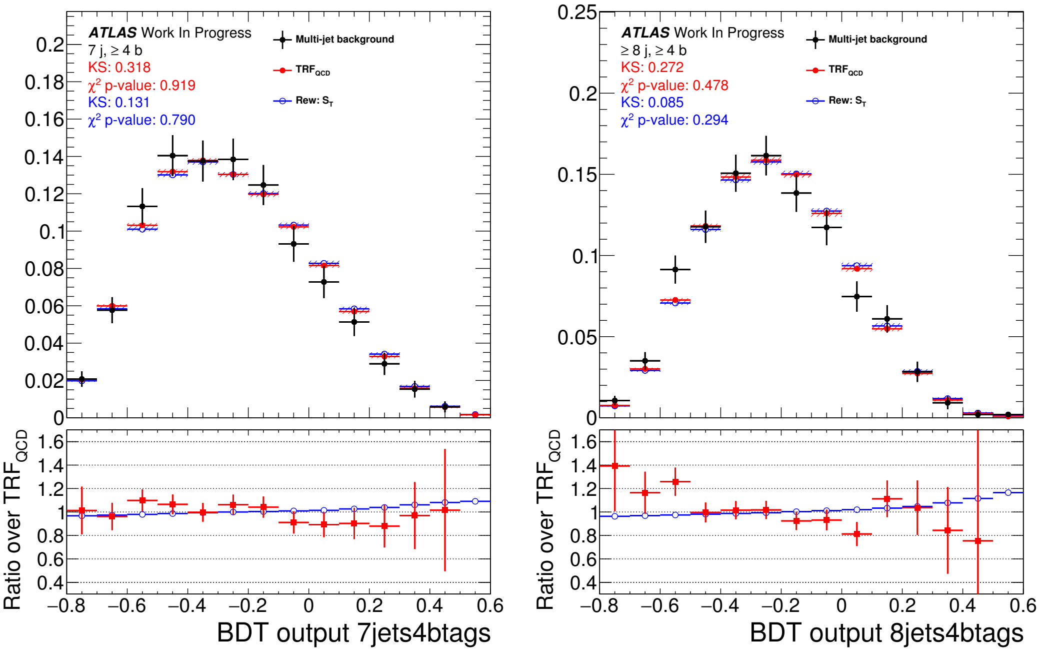
(b)

(c)

Figure 6.48: Multijet shape systematic variations BDT output induced by the reweight on S_T. Regions (6j, 3b) (a), (7j, 3b) (b) and ($\geq$ 8j, 3b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the $\mathrm{TRF}_{\mathrm{MJ}}$ method, blue markers are the reweighted prediction. In the ratio plots, the blue markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.



(a)



(b) (c)

Figure 6.49: Multijet shape systematic variations BDT output induced by the reweight on S_T. Regions (6j, ≥ 4b) (a), (7j, ≥ 4b) (b) and (≥ 8j, ≥ 4b) (c) are shown. In the upper part of the plots the black markers represent the expected multijet background, defined as the data minus the MC contribution, while red markers represents the prediction using the TRF_{MJ} method, blue markers are the reweighted prediction. In the ratio plots, the blue markers correspond to the ratio between the reweighted distribution and the nominal one, while the red markers correspond to the ratio of the expected multijet background and the predicted nominal one.

Nuisance parameter	Set of variables
Multijet set 1	p_T, MV1 ΔR, Min$\Delta R_{(j,\mathrm{hMV1})}$
Multijet set 2	p_T, MV1 ΔR, Min$\Delta R_{(j,j)}$
Multijet set 3	p_T, $\lvert\eta\rvert$, Min$\Delta R_{(j,\mathrm{hMV1})}$
Multijet set 4	p_T, $\lvert\eta\rvert$, MV1 ΔR, Min$\Delta R_{(j,\mathrm{hMV1})}$
Multijet set 5	p_T, MV1 ΔR, $\langle \Delta R_{(j,\mathrm{hMV1})} \rangle$
Multijet Lowest MV1	Nominal set removing the two lowest MV1 jets from computation
Multijet Random MV1	Nominal set removing randomly two MV1 jets from computation
Multijet HT RW	Nominal set with H_T reweighting
Multijet ST RW	Nominal set with S_t reweighting

Table 6.16: Description of the multijet shape nuisance parameters. The five first lines correspond to variations of the nominal set of variables describing $\mathrm{TRF}_{\mathrm{MJ}}$. The two following present variations of the computation of $\varepsilon_{\mathrm{MJ}}$ based on the selection of the 2 b-tagged jets to remove. The two last lines include the impact of the residual mismodelling on H_T and S_t.

6.9.4 Electroweak backgrounds Modeling

Uncertainty of ±7% is assumed for the theoretical cross sections of the single top production [141, 142]. This corresponds to the theoretical uncertainty on t- and Wt-channel production.

$t\bar{t}V$ systematics

An uncertainty of ±30% is assumed for the theoretical cross sections of the $t\bar{t}V$ [144, 145] backgrounds. Studies have been performed at the truth level to test the modeling of this background modifying the parameters of the Madgraph+Pythia6 generator related to ISR, FSR, QCD emission scale, renormalization and factorization scale. The biggest effect is given by the variation of the QCD emission scale with shape differences of the order of 10%. To account for this systematic effect $t\bar{t}V$ samples are reweighted to match the distributions with QCD emission scale varied up and down. The size of this systematic uncertainty is sufficiently big to cover the effect of the variation of all the parameters, therefore it is the only one considered.

6.9.5 $t\bar{t}$ +jets Background Modeling

A number of systematic uncertainties affecting the modeling of $t\bar{t}$+jets are considered in this analysis. In particular, systematics due to the uncertainty on the theoretical cross section, related to the choice of the parton shower and hadronization model as well as several uncertainties arising from the reweighting procedure applied to correct $t\bar{t}$ MC model are taken into account. Additional uncertainties are assigned to account for the limited knowledge of the $t\bar{t}$+heavy-flavour jets production.

Theoretical cross section

Uncertainties of +6.5%/-6% are assumed for the inclusive $t\bar{t}$ production cross section, including components from PDF and α_S uncertainties. These were calculated using the PDF4LHC prescription [155] with the MSTW2008 68% CL NNLO [105, 106], CT10 NNLO [130, 156] and NNPDF2.3 5f FFN [157] PDF sets, added in quadrature to the scale uncertainty.

Top quark p_T and $t\bar{t}$ system p_T reweighting

To reach an agreement between data and $t\bar{t}$ MC model a reweighting procedure based on the difference between top quark p_T and $t\bar{t}$ p_T distributions measured in data and in the simulation is applied to $t\bar{t}$ MC events. Nine largest uncertainties associated with the experimental measurement [116] of top quark and $t\bar{t}$ p_T are applied changing the size of the correction. This represents approximately 95% of the total experimental uncertainty. Each source is represented by a separate nuisance parameter in the fit thus making 9 nuisance parameters in total.

Given that the measurement is performed for the inclusive $t\bar{t}$ sample and the size of the uncertainties to be applied to the $t\bar{t} + c\bar{c}$ component is not known, two additional uncertainties are assigned to $t\bar{t}+c\bar{c}$ events corresponding to the $t\bar{t}$ p_T and top quark p_T corrections being turned off.

Parton shower

An uncertainty due to the choice of the parton shower and hadronization model is derived by comparing events produced by POWHEG interfaced with PYTHIA or HERWIG. Effects on the shapes are compared, symmetrised and applied to the shapes predicted by the default model after correcting both samples to match top quark p_T and $t\bar{t}$ p_T distributions in data. Given that the change of the parton shower model leads to two separate effects - a change of the number of jets distribution and a change of the heavy flavor content - parton shower uncertainty is represented by two parameters, one acting on $t\bar{t}$ +light and another on $t\bar{t}$ +HF contribution which are treated as uncorrelated in the fit. These uncertainties have a significant impact on the fitted signal strength, as can be seen in Figure 6.58.

$t\bar{t}$+jets Heavy-Flavour Content

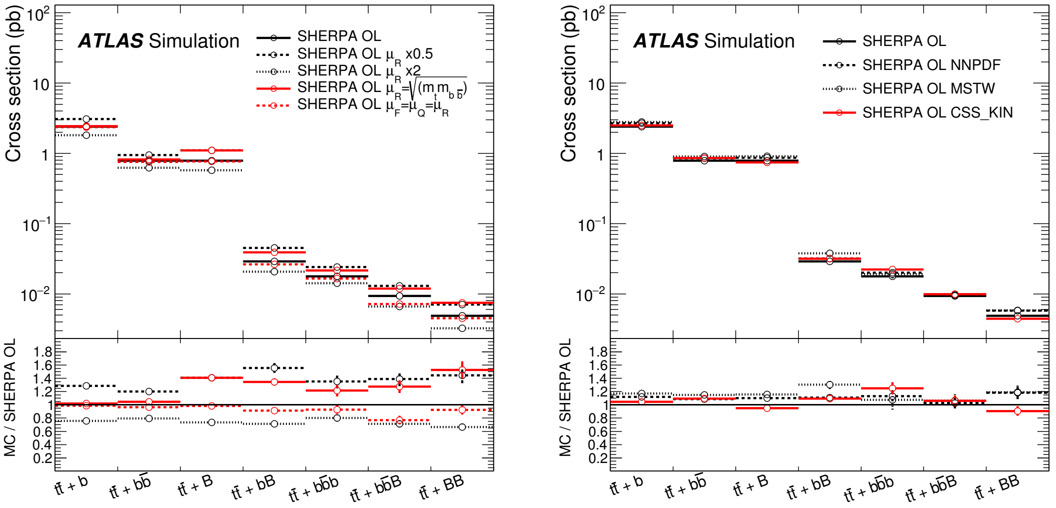
Comparison of the SHERPAOL and POWHEG + PYTHIA simulations show a difference in the cross section of the $t\bar{t} + b\bar{b}$ process of 50%. Hence a systematic of 50% is applied to the $t\bar{t} + b\bar{b}$ background normalization. Moreover, since there are no NLO predictions of $t\bar{t} + c\bar{c}$ process, the same value of 50% uncertainty is applied to the $t\bar{t} + c\bar{c}$ normalization. These two uncertainties are considered uncorrelated. Uncertainty on $t\bar{t} + b\bar{b}$ normalization is the dominant systematic of $t\bar{t}H$ ($H \to b\bar{b}$) fully hadronic analysis.

Three scale uncertainties are evaluated for the $t\bar{t} + b\bar{b}$ background. Functional form of the renormalization scale is changed to $\mu_R = (m_t m_{b\bar{b}})^{1/2}$, Functional form of the factorization μ_f and resummation μ_Q scales are changed to $\mu_F = \mu_Q = \prod_{i=t,\bar{t},b,\bar{b}} E_{T,i}^{1/4}$. Also variation of the renormalization scale μ_R by a factor of two up and down are evaluated. Shower recoil model uncertainty and two uncertainties due to the PDF choice in the SHERPAOL NLO calculation are also evaluated. The effect of these uncertainties on the $t\bar{t} + b\bar{b}$ categories defined in Section 6.4.3 is presented in Figure 6.50. Additional uncertainties are applied to take into account multi-parton interaction and final-state radiation $t\bar{t} + b\bar{b}$ production, not included in the SHERPAOL NLO calculation.

Systematic uncertainty in $t\bar{t}+c\bar{c}$ contribution due to the choice of the MC generator are estimated comparing simulations in MADGRAPH + PYTHIA with POWHEG + PYTHIA, since the former has the $t\bar{t}+c\bar{c}$ process in the matrix element calculation. In addition also variations of factorization and renormalization scales, matching threshold and c-quark mass are taken into account. These uncertainties have quite small impact on the fit result.

6.9.6 Signal Modeling

Uncertainties related to the $t\bar{t}H$ modeling are described in this Section.



(a) (b)

Figure 6.50: Estimation of the systematic uncertainties on the $t\bar{t} + b\bar{b}$ background across the different categories. (a) shows effects of scale variations and (b) shows effects of PDF choice and shower recoil model of SHERPAOL [11].

Factorization, renormalization and functional form scale

The impact of the choice of factorization and renormalization scale on $t\bar{t}H$ signal kinematics is estimated via dedicated NLO POWHEL samples. These events are simulated with variation of a factor of two up and down in the default scale. The nominal POWHEL $t\bar{t}H$ is reweighted with respect to truth level variables to reproduce the observed variations. Effect of the change of the functional form of the scale is evaluated similarly.

PDF

The PDF uncertainty on the $t\bar{t}H$ signal is evaluated following the recommendation of the PDF4LHC [155]. It takes into account the differences between three PDF sets - CT10 NLO [130], MSTW2008 68% CL NLO [105, 106] and NNPDF 2.3 NLO [157]. The final PDF uncertainty is an envelope of a) intra-PDF uncertainty, which evaluates the changes due to the variation of different PDF parameters within a single PDF error set and b) inter-PDF uncertainty, which evaluates differences between different PDF sets. The uncertainty is evaluated by reweighting the signal MC to the different PDF sets and evaluating the change in acceptance as a function of the BDT output and applying the PDF4LHC prescription to combine the different variations.

Parton shower

The uncertainty due to the choice of parton shower model is evaluated by comparing POWHEL samples showered by PYTHIA8 and HERWIG. For the latter corrections have been introduced to match the Higgs branching fractions in HERWIG to the NLO calculations from Ref. [104] used to generate POWHEL+PYTHIA8 sample.

Generator comparison

An additional $t\bar{t}H$ sample is generated using MADGRAPH5_AMC@NLO [158] and showered with HERWIG++[159, 160] to be compared with the nominal POWHEL + PYTHIA8 sample. Differences are observed in the distributions of the $t\bar{t}$, $t\bar{t}H$ and Higgs boson kinematics. The largest difference observed is between the p_T distributions of the Higgs bosons. Differences are recovered after reweighting of this observable, therefore this reweight is used as estimation for this systematic uncertainty.

6.10 Results

As discussed in Section 6.6 six independent event categories are considered to constrain backgrounds and search for signal simultaneously.

Hypothesis testing is performed using a modified frequentist method as implemented in ROOSTATS [161] and based on a profile likelihood which takes into account the systematic uncertainties as Nuisance Parameters (NP) which are fitted to the data. This procedure allows to take into account the impact of systematic uncertainties on the search sensitivity to be minimized, by taking advantage of the high-statistics background-dominated control channels included in the likelihood fit. The process of including these nuisance parameters in this likelihood fit is referred to as profiling. The distributions of the final discriminants from each of the channels considered are combined in the statistical analysis to test for the presence of a signal for a hypothesized Higgs boson mass of $M_H = 125$ GeV.

To obtain the final result a simultaneous fit of all nuisance parameters is performed to the data using the histograms of the final variable considered in each of the six analysis regions of the all hadronic channel. Fits are performed under the signal-plus-background and background-only hypotheses.

The analysis is prepared using a blinding procedure in order to determine that the nuisance parameters are consistently describing the uncertainties in all the regions. The blinded regions are defined as those, for each analysis region, with an expected signal over background higher than 2%, as described in Section 6.7. In a first step, referred as "blinded" analysis, the checks are done by excluding these regions. This allows to use sufficient amount of data to test the description of the systematic uncertainties. In this step also the multijet background predictions with the TRF_{MJ} method have been validated. After this step, if everything is behaving in the expected way, the analysis is unblinded, allowing to use the complete datatset and present the results. The unblinded result is presented in the following.

6.10.1 Expected performance of the fit

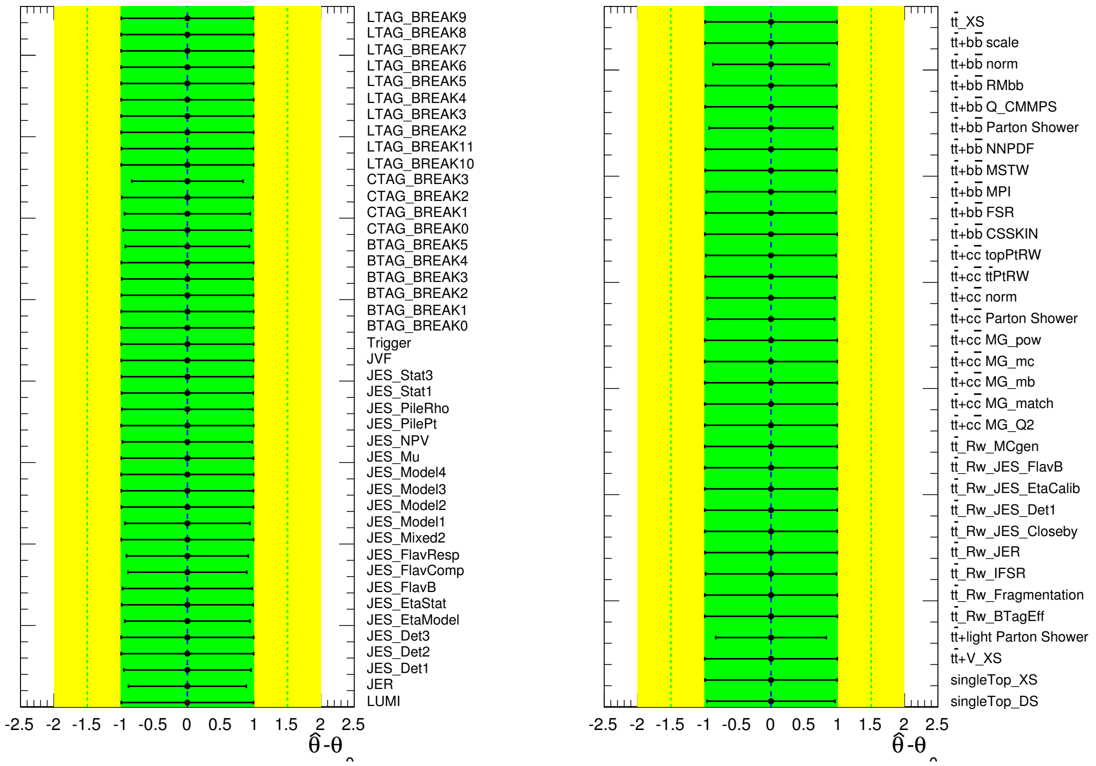
Expected performance of the fit has been studied using so-called Asimov dataset approach [162]; this procedure replaces the ensemble testing performed with Monte Carlo pseudo-experiments with a single "representative" dataset which returns the true value for each estimated parameter. The Asimov dataset 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 unexpected over-constraints from data in the fit.

The result of the fit to the Asimov dataset under the signal plus background hypothesis is shown in Figure 6.51. Both nuisance parameters and multi-jet background normalization scale factors (`SF_QCD_xxx`) are reported on Figure 6.52. As expected nuisance parameters corresponding to systematic uncertainties are all centered on zero and the normalization scale factors are all centered around one. The majority of the parameters (especially the ones related to the detector performance) are not constrained with respect to their prior uncertainties. The constrains on the parameters related to the shape of multi-jet background can be 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 normalization of the multi-jet background in each region can be known 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 7 jets.

The correlation plot of the fitted nuisance parameters and normalization scale factors, under the signal plus background hypothesis is found in Figure 6.53. Only nuisance parameters with at least one correlation higher than 25% are shown in the plot.

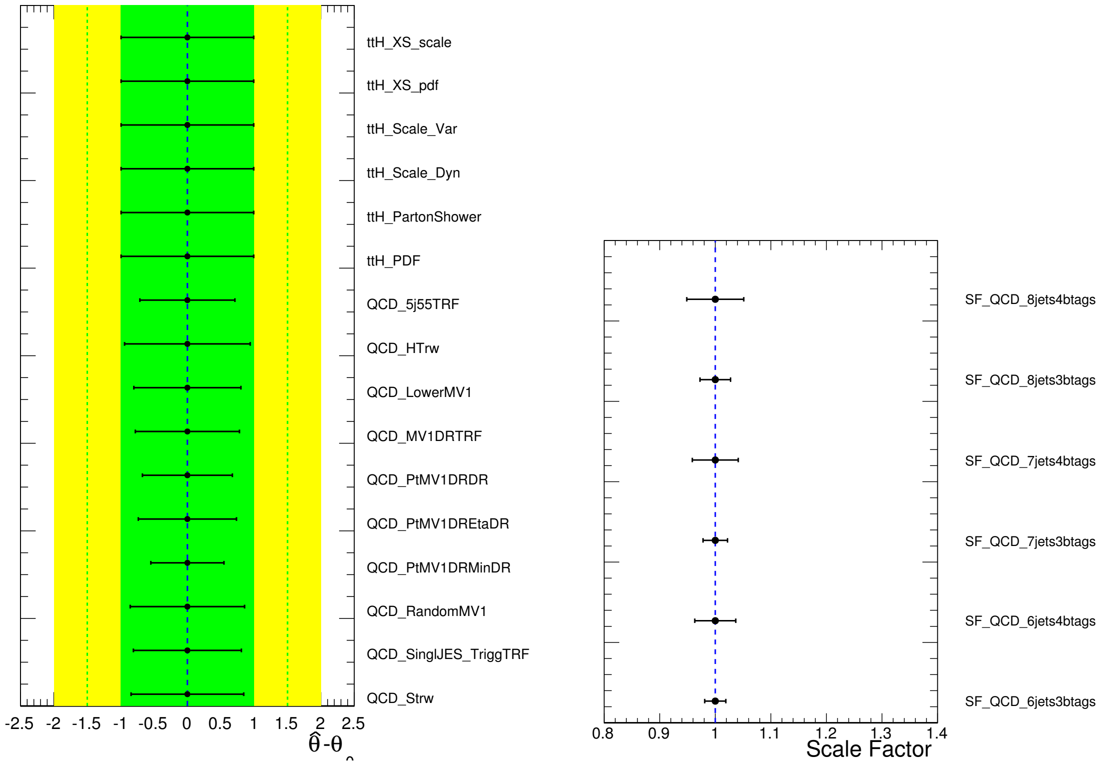
For what concerns the analysis sensitivity, the leading source of systematics uncertainties are the ones which show a high level of correlation with the signal strength. These are the uncertainties on the normalization of the $t\bar{t}+b\bar{b}$ background as well as its shape and the uncertainty on the multijet process normalization. This is also confirmed by the ranking plot on Figure 6.54; In this plot the post-fit effect on μ is calculated by fixing the corresponding NP at $\hat{\theta} \pm \sigma_\theta$, where $\hat{\theta}$ is the fitted value of the nuisance parameter and σ_θ is its post-fit uncertainty, and performing the fit again. The difference between the default and the modified μ, $\Delta\mu$, represents the effect on μ of this particular systematic uncertainty. The parameters with the largest impact are the normalization of the multi-jet background followed by the uncertainty on the $t\bar{t}+b\bar{b}$ process.

In the presence of a signal compatible with the one predicted by the SM, the expected uncertainty on the signal strength extraction in units of σ/σ_{SM} is 2.79 in case of all sources of systematics are considered, and 0.92 in case only the data statistical uncertainty and the template statistics are taken into account. This corresponds to an expected signal significance of 0.37, in case of a SM signal. Upper limits on the signal production cross section are reported in Table 6.17.



(a) (b)

Figure 6.51: Fitted nuisance parameters and normalization scale factors to Asimov data set under the signal plus background hypothesis (assuming $M_H = 125$ GeV). Figure on the *left* shows detector-related uncertainties while the one on the *right* displays uncertainties related to simulated background modeling and normalization.



(a) (b)

Figure 6.52: Fitted nuisance parameters and normalization scale factors to Asimov data set under the signal plus background hypothesis (assuming $M_H = 125$ GeV). (a) Shows uncertainties related to multijet background while (b) displays multijet background normalization scale factors (`SF_QCD_xxx`).

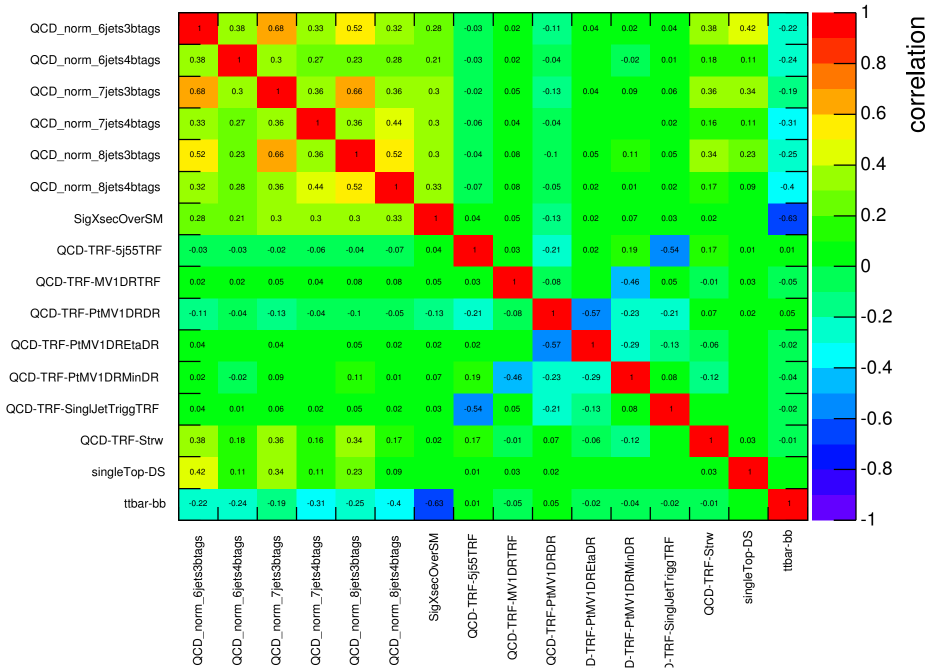


Figure 6.53: Correlation of nuisance parameters under the signal plus background hypothesis in full fit to the Asimov data set. Only nuisance parameters with at least one correlation above 25% are plotted.

Channel	Stat. Only	All Systematics
All hadronic	1.73	5.42

Table 6.17: Expected (median, for the background-only hypothesis) 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, σ/σ_{SM}.

6.10.2 Fit to data

The regions of the analysis considered in the fit are (6j, 3b), (7j, 3b), (≥8j, 3b). (6j, ≥4b), (7j, ≥4b), (≥8j, ≥4b). The region (≥8j, ≥4b) is the most sensitive one. The fit is performed with the signal plus background hypothesis using the full BDT discriminant in each region.

Figure 6.55 and Figure 6.56 show the result of the fit to data. The behaviour of the nuisance parameters is consistent with the one observed in the Asimov dataset, showed in Section 6.10.1. The correlation plot of the fitted nuisance parameters and normalization scale factors on unblinded data is presented in Figure 6.57. Only nuisance parameters with at least one correlation higher than 25% are shown in the plot. Figure 6.58 shows the ranking of the parameters with the largest impact on the extracted signal strength. The post-fit effect on μ is calculated by fixing the corresponding nuisance parameter at $\hat{\theta} \pm \sigma_\theta$, where $\hat{\theta}$ is the fitted value of the nuisance parameter and σ_θ is its post-fit uncertainty, and performing the fit again.

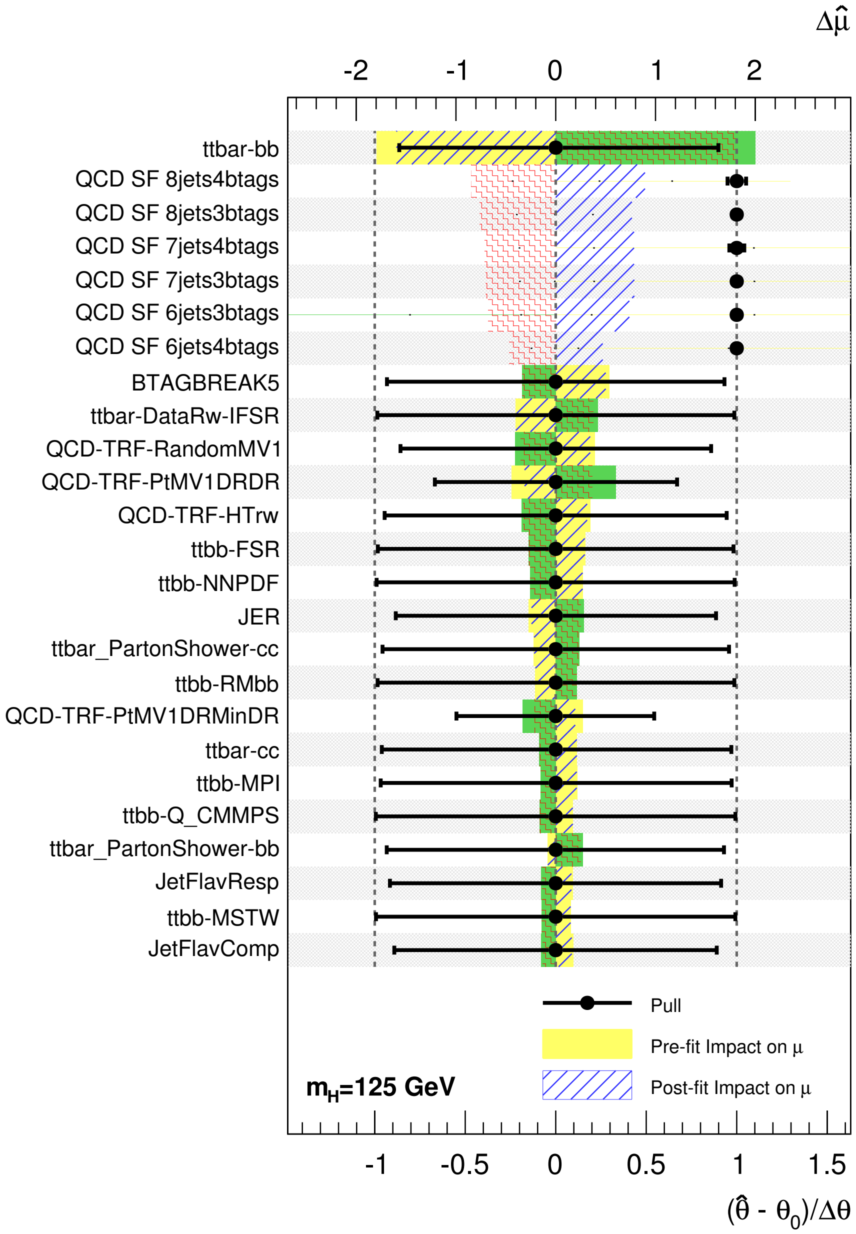
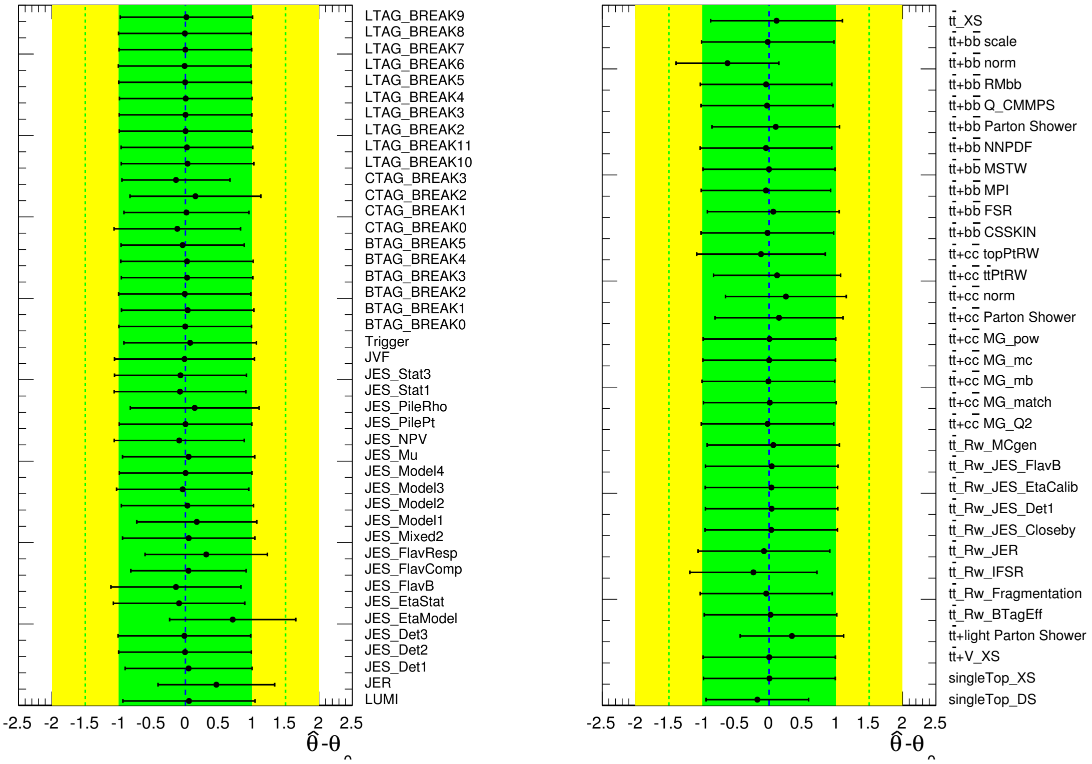


Figure 6.54: Best fit values of the nuisance parameters with the largest impact on the signal strength μ in the Asimov dataset. Black points are drown according to the bottom axis and represent the deviation of the fitted nuisance parameters $\hat{\theta}$ from the pre-fit value θ_0 in units of the pre-fit standard deviation $\Delta\theta$. Error bars represent the post-fit uncertainty σ_θ. i.e. if data are constraining the uncertainties, the bars are smaller than 1, otherwise these are close to 1. Yellow and dashed blue bands represent the pre- and post-fit impact on μ of the nuisance parameter. Bands are drown according to the top axis.

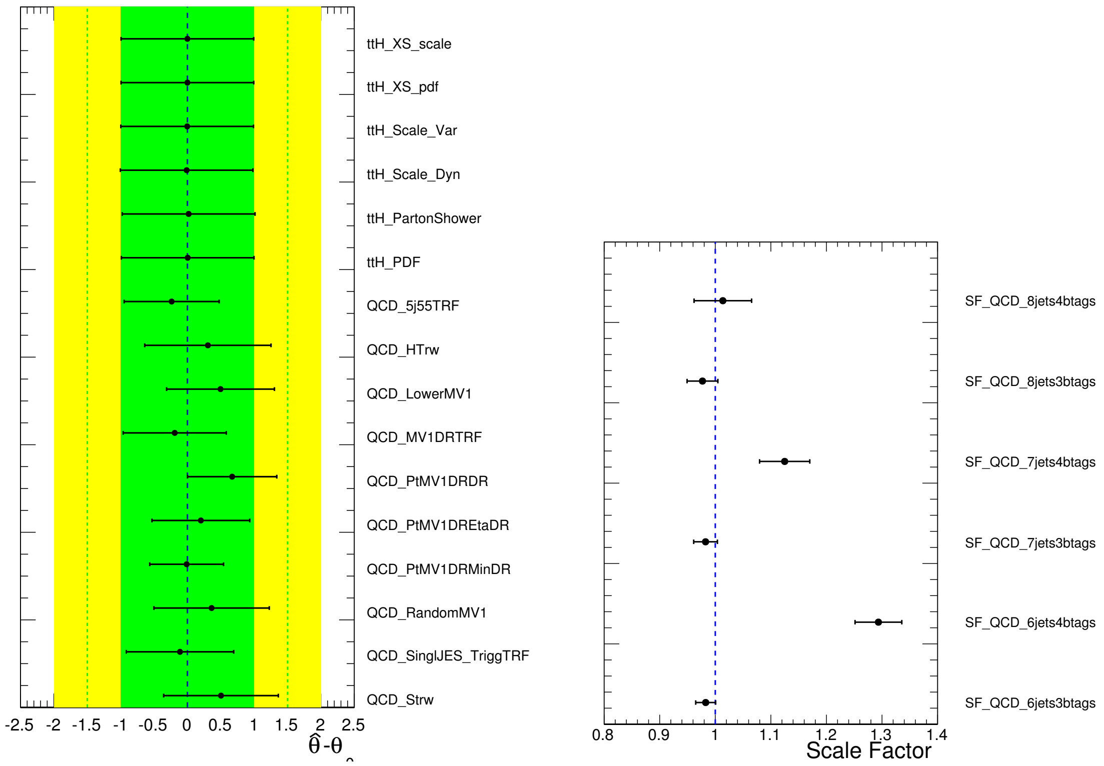
The difference between the default and the modified μ, $\Delta\mu$, represents the effect on μ of this particular systematic uncertainty. The systematics with the largest impact are uncertainty on the $t\bar{t}+b\bar{b}$ process followed by the normalizations of the multi-jet background. Two of the multijet background shape uncertainties are ranked fourth and fifth and their pulls are slightly positive. This ranking is expected as the multijet background is the main background of this analysis. The small constraint of these systematics is expected since these are defined in a conservative way. Other leading uncertainties include b-tagging and some components of the JES uncertainty.



(a) (b)

Figure 6.55: Fitted nuisance parameters and normalization scale factors for the fit to data under the signal plus background hypothesis (assuming $m_H = 125$GeV). (a) Shows detector-related uncertainties while (b) displays uncertainties related to background modeling and normalization.

Figure 6.59 shows a comparison of data and prediction in the final discriminant for each of the regions considered, before the fit to data. Figure 6.60 shows similar distributions after the Monte Carlo predictions have been modified taking into account the result of the fit to data. As it could be seen the fit significantly improves the agreement especially in regions with 4 b-tags, mainly thank to the adjustment of the multi-jet normalization SF. The uncertainty band is significantly reduced mainly as a result of the correlation among the nuisance parameters built by the fit. The comparison of pre- and post-fit distributions for some of the observables entering



(a) (b)

Figure 6.56: Fitted nuisance parameters and normalization scale factors for the fit to data under the signal plus background hypothesis (assuming $m_H = 125\text{GeV}$). (a) Shows uncertainties related to multijet background while (b) displays multijet background normalization scale factors. uncertainties related to background modeling and normalization (`SF_QCD_xxx`).

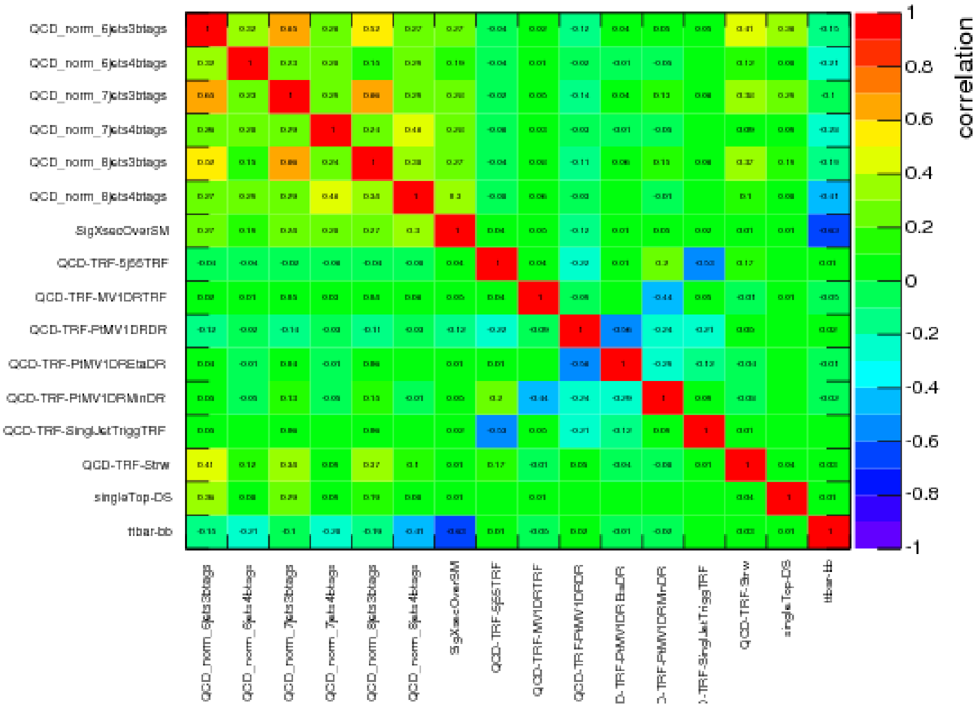


Figure 6.57: Correlation of nuisance parameters in full fit to the unblinded data set. Only nuisance parameters with at least one correlation above 25% are plotted.

the BDTs is presented in Figures 6.61, 6.62, 6.63, 6.64, 6.65 and 6.66 for the regions (6j, 3b), (6j, ≥ 4b), (7j, 3b), (7j, ≥ 4b), (8j, 3b), (8j, ≥ 4b) respectively.

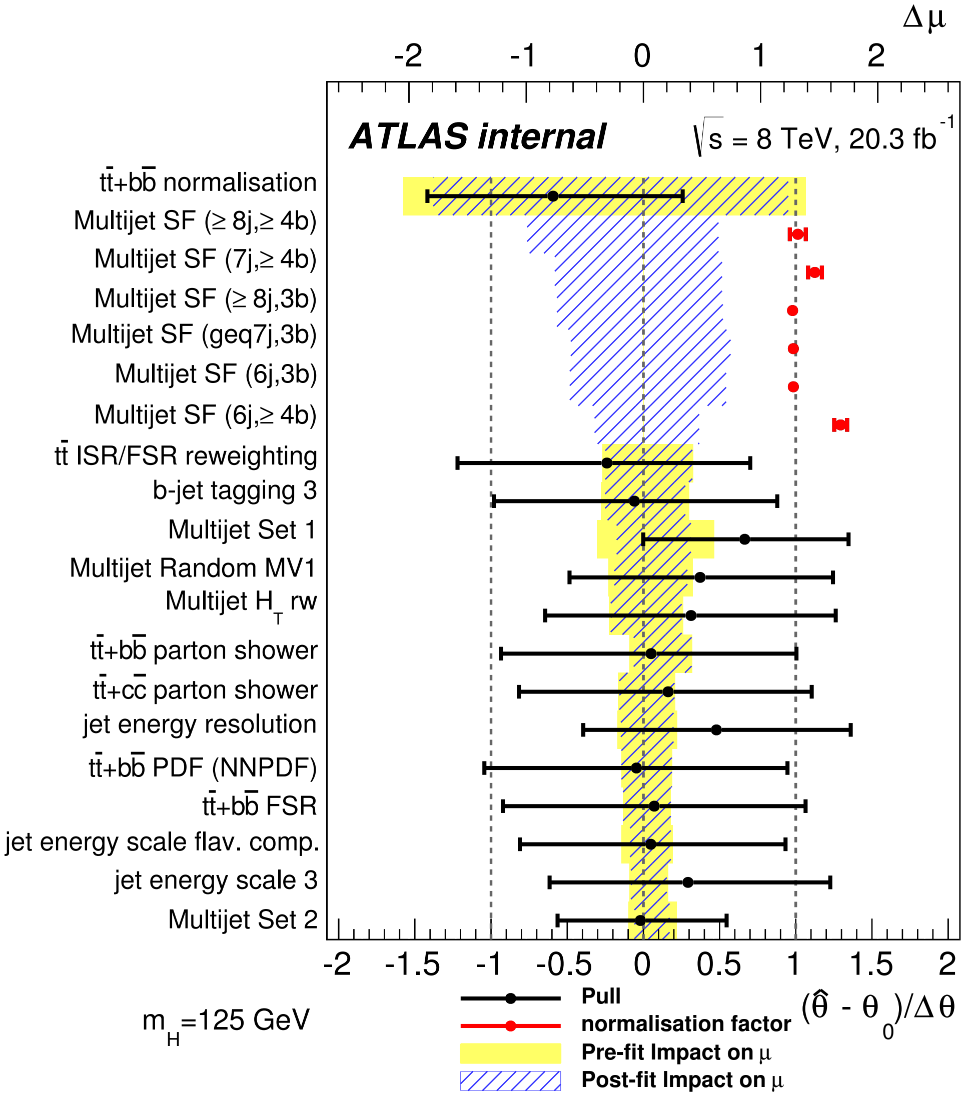


Figure 6.58: Best fit values of the nuisance parameters with the largest impact on the signal strength μ in the Asimov dataset. Points are drown according to the bottom axis and represent the deviation of the fitted nuisance parameters $\hat{\theta}$ from the pre-fit value θ_0 in units of the pre-fit standard deviation $\Delta\theta$. Error bars represent the post-fit uncertainty σ_θ. i.e. if data are constraining the uncertainties, the bars are smaller than one, otherwise these are close to one. Red points show the multi-jet background normalization parameters, these are freely floating in the fit and have a core-fit value of one. Yellow and dashed blue bands represent the pre- and post-fit impact on μ of the nuisance parameter. Bands are drown according to the top axis.

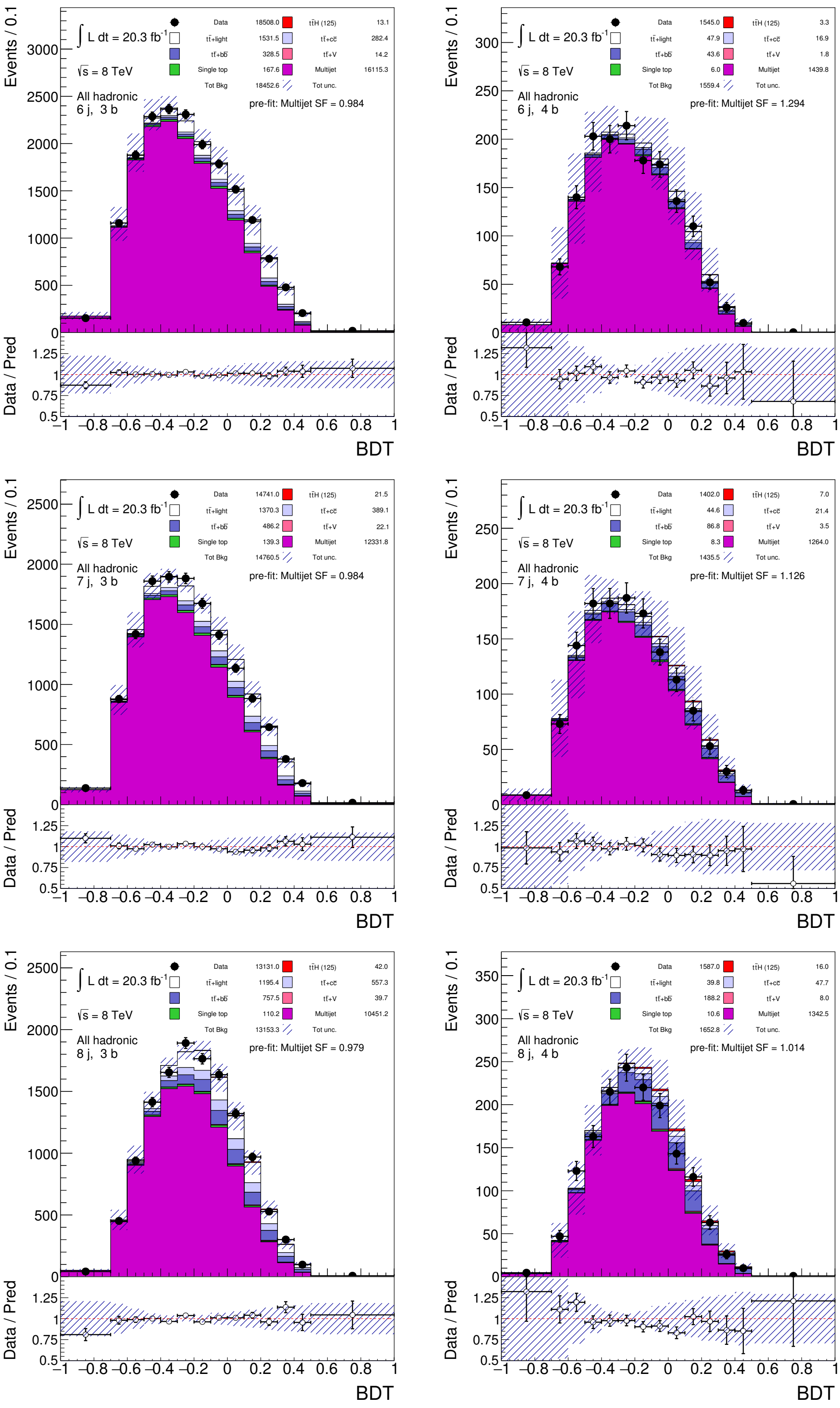


Figure 6.59: Comparison between data and predictions for the final discriminant in the analysis regions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties. The signal contribution is not included in the stack plot. Left (resp. right) plots concern the regions with exactly 3 (resp. ≥ 4) b-tagged jets.

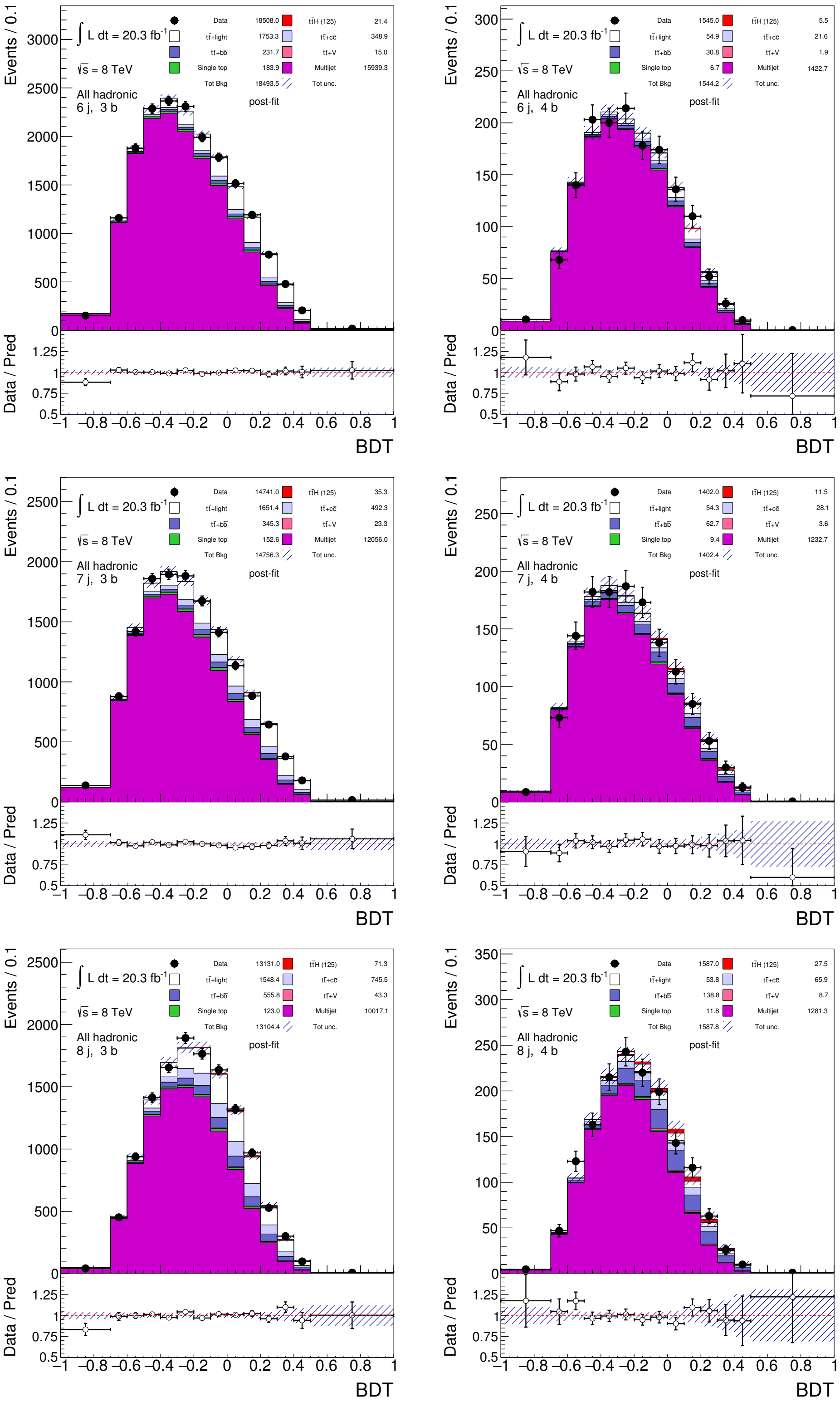


Figure 6.60: Comparison between data and predictions for the final discriminant in the analysis regions; signal and background predictions have been corrected with the result of the fit to data. The uncertainty band contains statistical and systematical uncertainties and takes into account the correlations among the nuisance parameters induced by the fit. Signal contribution is included in the stack plot. Left (resp. right) plots concern the regions with exactly 3 (resp. ≥ 4) b-tagged jets.

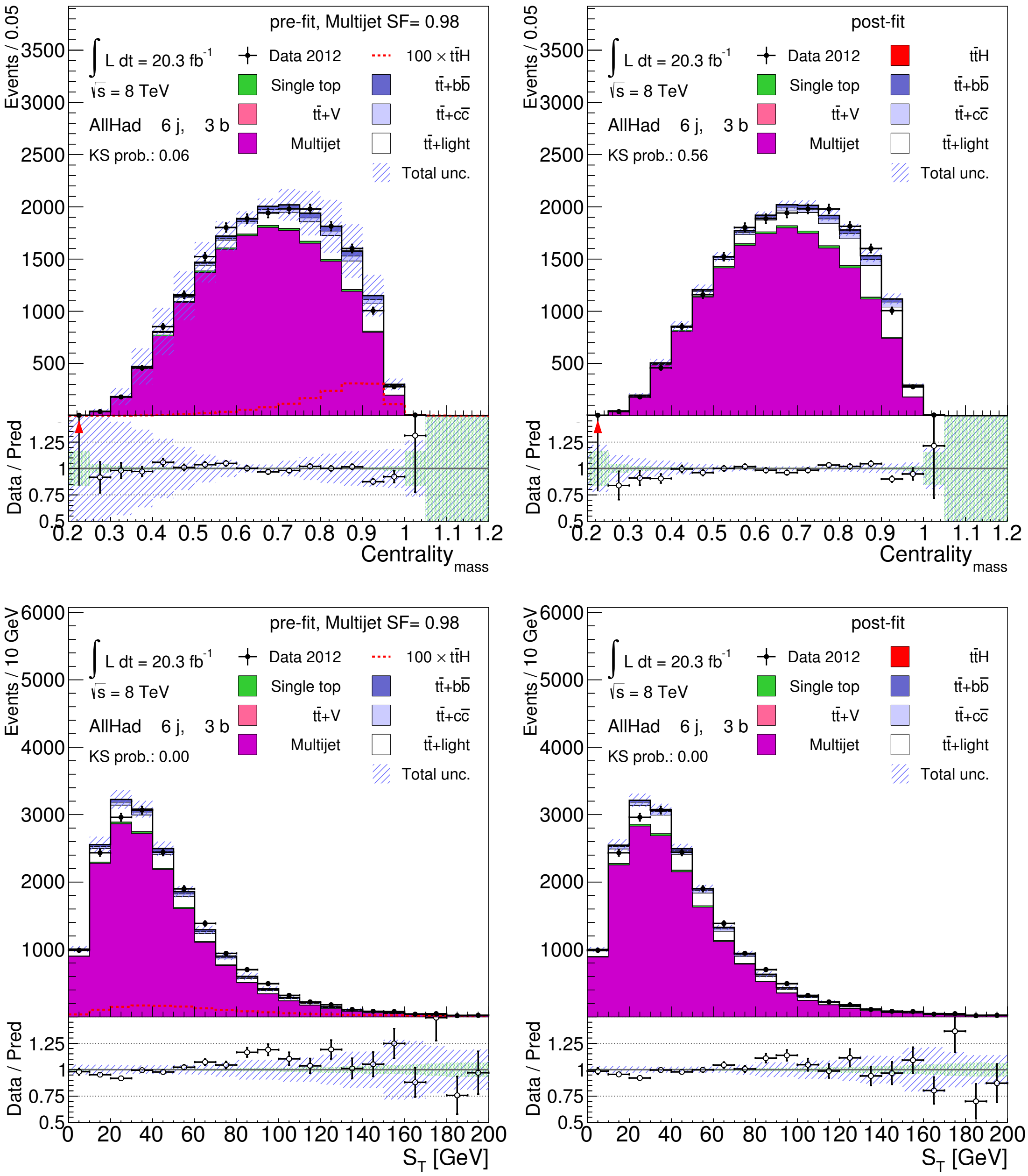


Figure 6.61: Comparison between data and predictions for the Centrality mass (top) and for the object missing energy S_T (bottom) in the (6j, 3b) region; *Left*: background predictions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties. The signal contribution is not included in the stack plot. *Right*: signal and background predictions 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. Signal contribution is included in the stack plot.

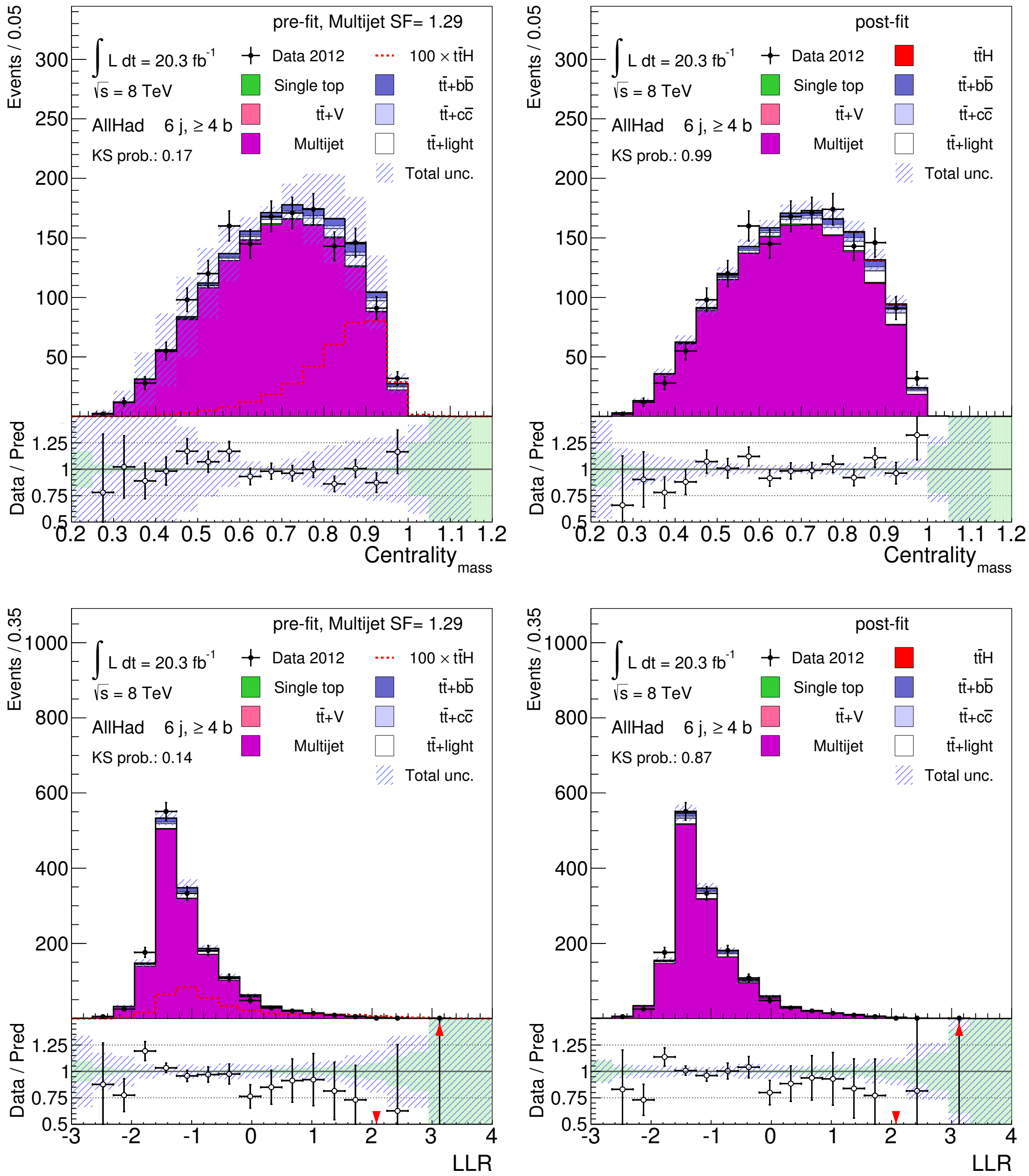


Figure 6.62: Comparison between data and predictions for the Centrality mass (top) and for the LLR (bottom) in the (6j, ≥ 4b) region; *Left*: background predictions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties. The signal contribution is not included in the stack plot. *Right*: signal and background predictions 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. Signal contribution is included in the stack plot.

Table 6.18 summarizes the expected and observed limits and Table 6.19 shows the observed $\hat{\mu}$ value. Observed signal strength for the $m_H = 125$ GeV is $\hat{\mu} = 1.56 \pm 2.55$. This corresponds to an observed signal significance of 0.59.

Channel	observed	-2 s.d.	-1 s.d.	expected	+1 s.d.	+2 s.d.
All hadronic	6.45	2.91	3.90	5.42	7.55	10.12

Table 6.18: Observed and expected (median, for the background-only hypothesis) 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, σ/σ_{SM} assuming $m_H = 125$ GeV, obtained from the all hadronic $t\bar{t}H$ $(H \to b\bar{b})$ channel.

Channel	$\hat{\mu}$	error
All hadronic	1.56	2.55

Table 6.19: Observed signal strength and its uncertainty assuming $m_H = 125$ GeV in the all hadronic $t\bar{t}H$ $(H \to b\bar{b})$ channel.

Figure 6.67 shows a summary of the 95% CL upper limits of $\sigma(t\bar{t}H)$ relative to the SM prediction and of the observed signal strength. The expected significance of the combination of semi- and di-leptonic $t\bar{t}H$ $(H \to b\bar{b})$ analyses passes from 1.0 σ to 1.1 σ with the addition of the fully hadronic channel. Tables 6.20 and 6.21 show the event yield pre- and post-fit respectively for all the regions considered. Figure 6.68 summaries post-fit event yields as function of $\log_{10}(S/B)$ for all bins of the distributions used in the fit. The value of $\log_{10}(S/B)$ is calculated according to the post-fit yields for both signal ($\mu = 1.6$) and background. The total number of background and signal events is displayed in bins of $\log_{10}(S/B)$. A signal strength 6.4 times larger than predicted by the SM, which is excluded at 95% CL by this analysis, is also shown. The aim of this plots is to show the overall good agreement between background modelization and observed data and visualize the regions where the signal is supposed to accumulate.

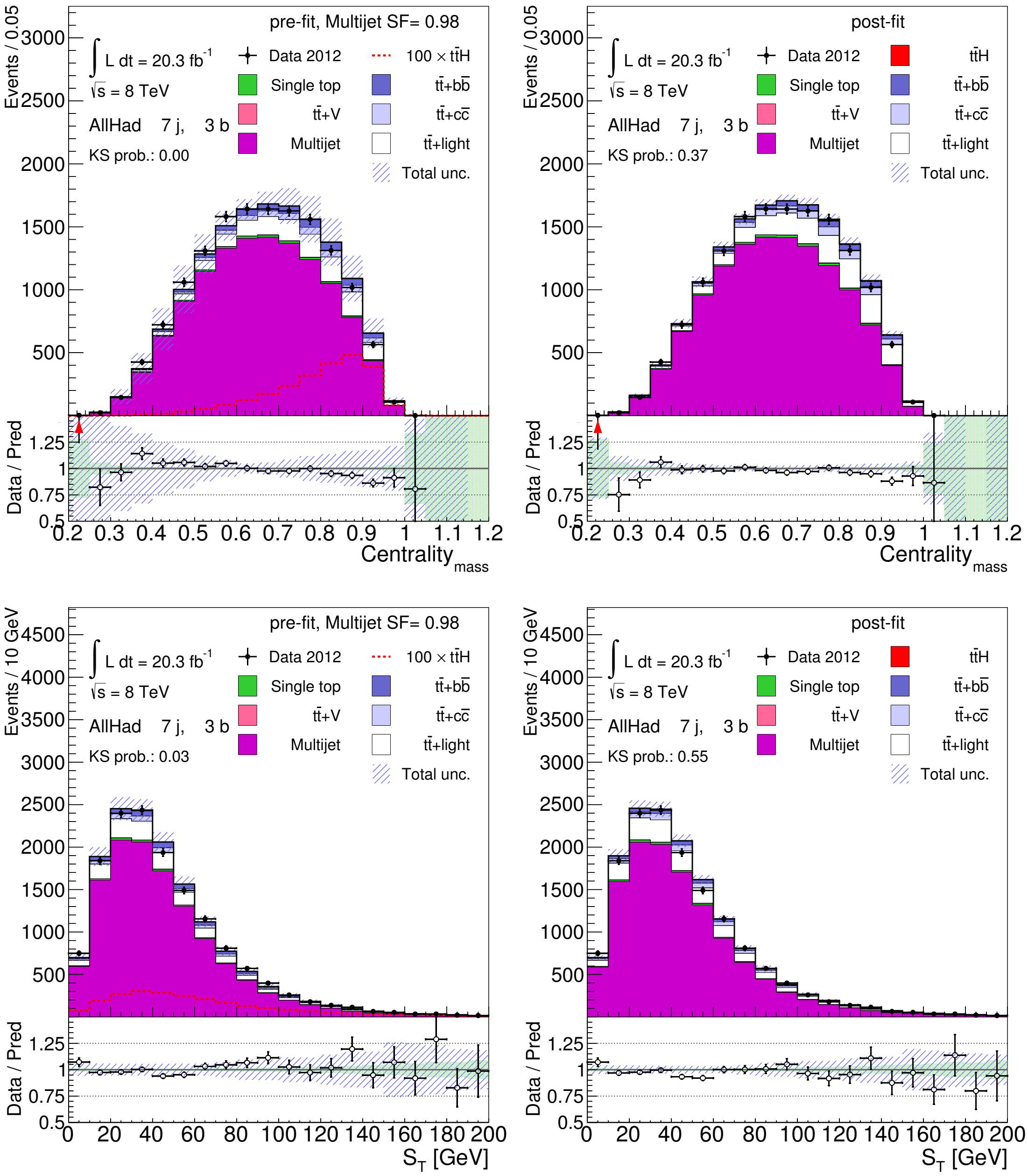


Figure 6.63: Comparison between data and predictions for the Centrality mass (top) and for the object missing energy S_T (bottom) in the (7j, 3b) region; *Left*: background predictions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties. The signal contribution is not included in the stack plot. *Right*: signal and background predictions 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. Signal contribution is included in the stack plot.

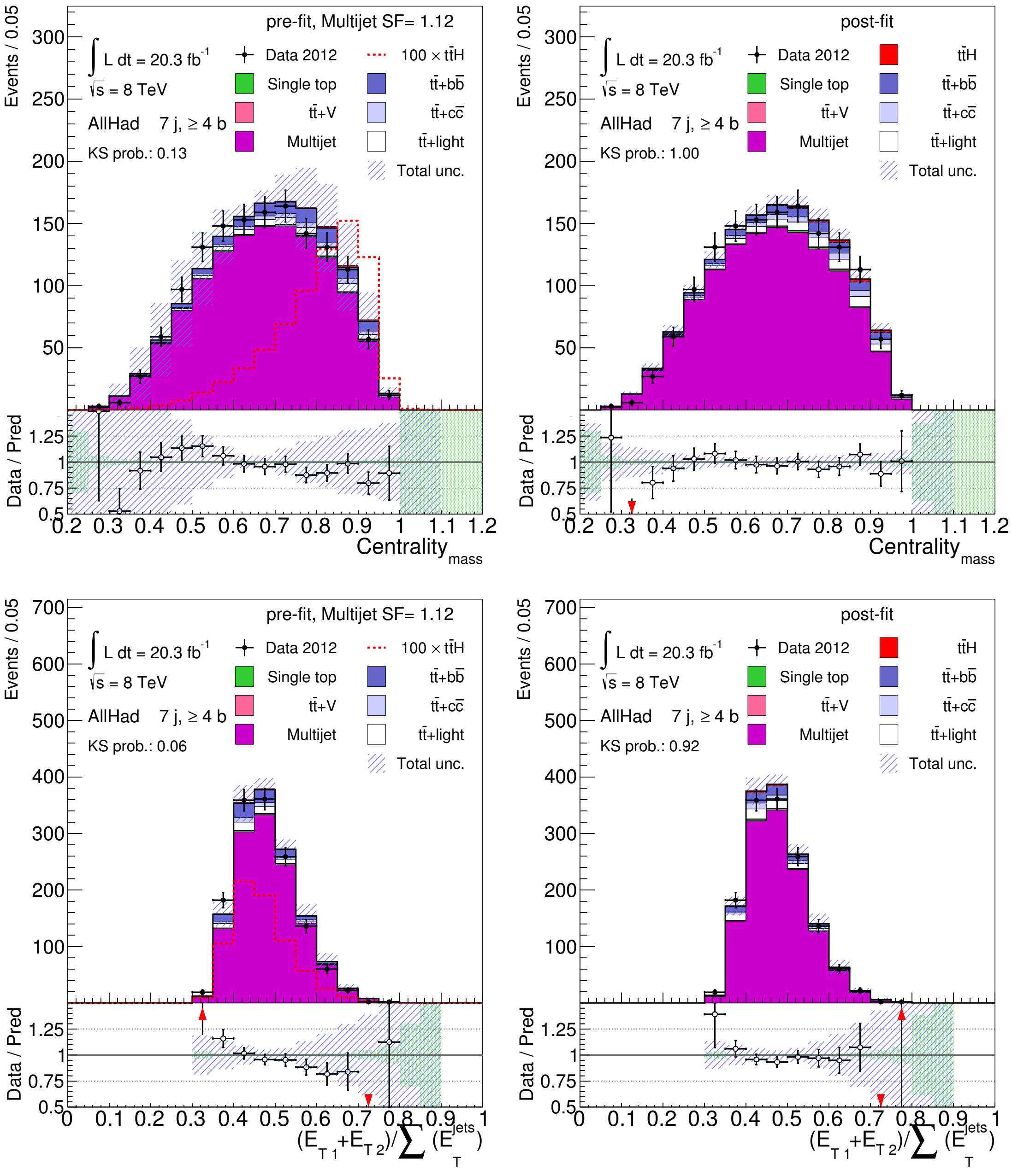


Figure 6.64: Comparison between data and predictions for the Centrality mass (top) and for the $(E_{\mathrm{T}\,1} + E_{\mathrm{T}\,2})/\sum(\text{jets } E_{\mathrm{T}})$ (bottom) in the (7j, $\geq$ 4b) region; *Left*: background predictions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties. The signal contribution is not included in the stack plot. *Right*: signal and background predictions 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. Signal contribution is included in the stack plot.

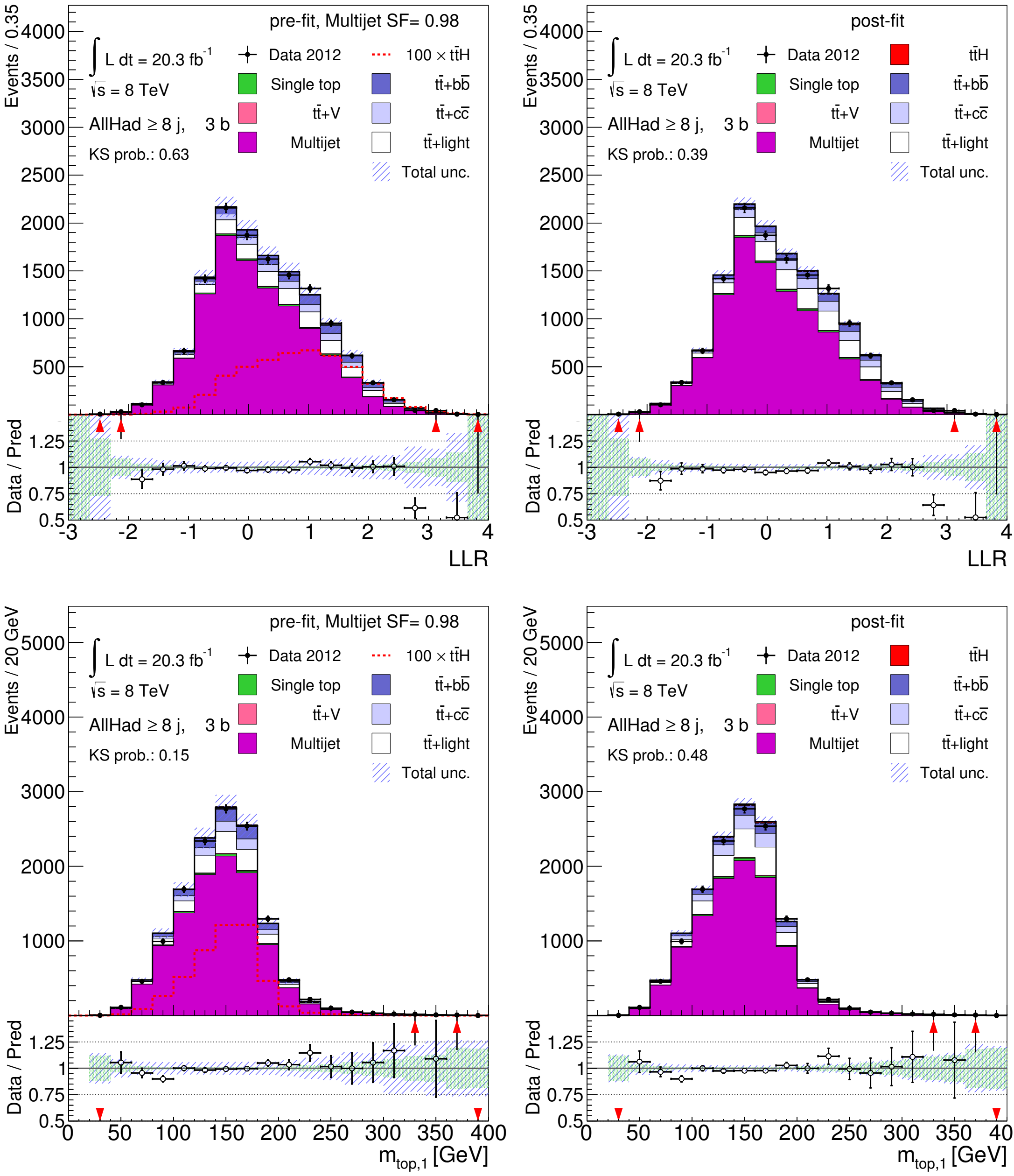


Figure 6.65: Comparison between data and predictions for the LLR (top) and for the $m_{top,1}$ (bottom) in the ($\geq$ 8j, 3b) region; *Left*: background predictions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties. The signal contribution is not included in the stack plot. *Right*: signal and background predictions 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. Signal contribution is included in the stack plot.

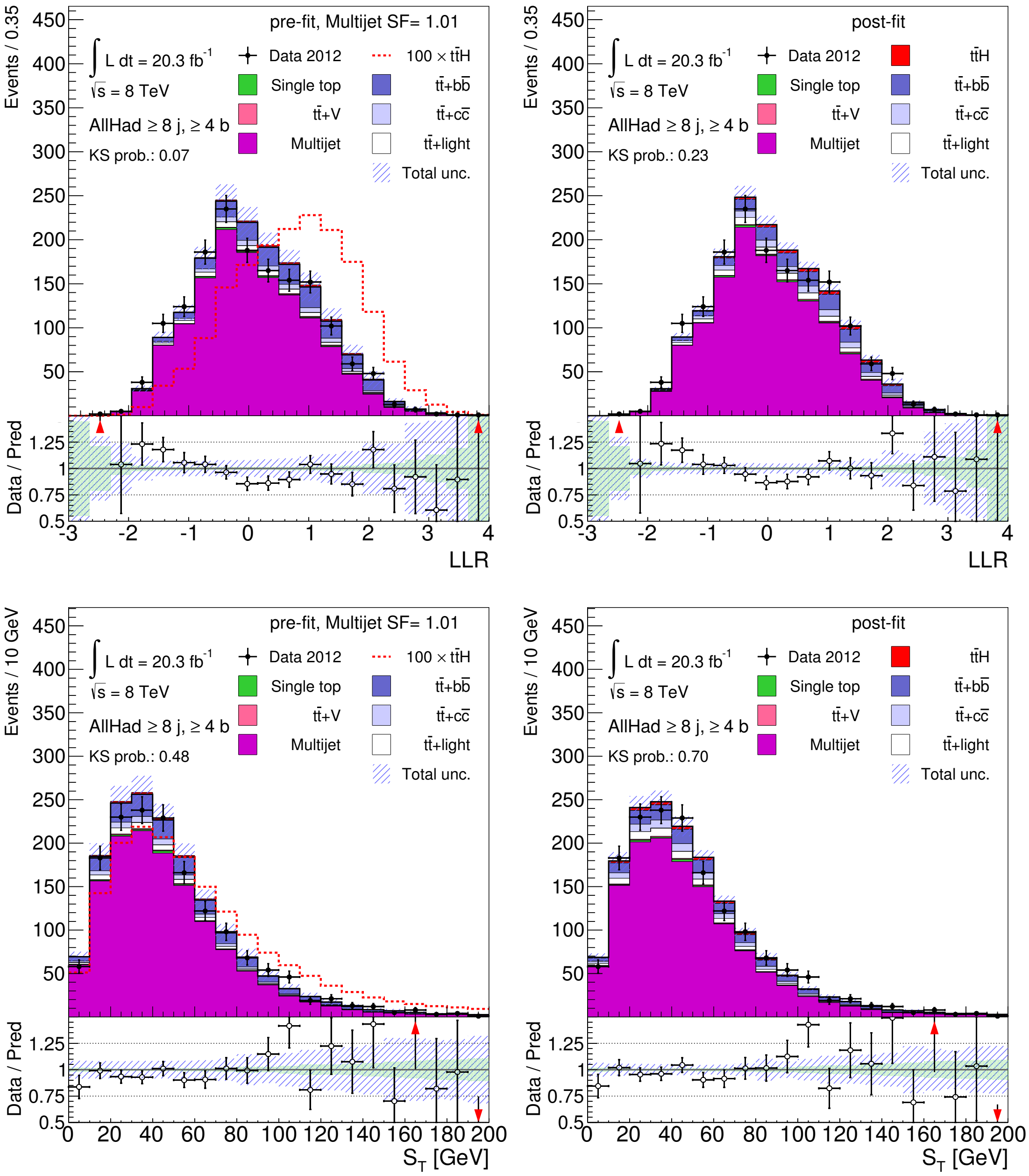


Figure 6.66: Comparison between data and predictions for the LLR (top) and for the object missing energy S_T (bottom) in the ($\geq$ 8j, $\geq$ 4b) region; *Left*: background predictions before performing the fit to data. The uncertainty band contains the full statistical and systematical uncertainties. The signal contribution is not included in the stack plot. *Right*: signal and background predictions 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. Signal contribution is included in the stack plot.

	(6j, 3b)	(6j, ≥4b)	(7j, 3b)	(7j, ≥4b)	(8j, 3b)	(8j, ≥4b)
Multijet	16400 ± 130	1100 ± 33	12500 ± 12	1100 ± 33	10600 ± 100	1300 ± 36
single top	170 ± 63	6.0 ± 3.7	140 ± 55	8.3 ± 4.6	110 ± 50	11 ± 5.9
$t\bar{t}+V$	14 ± 6.3	1.8 ± 1.5	22 ± 9.0	3.5 ± 2.3	40 ± 15	8.0 ± 4.2
$t\bar{t}+b\bar{b}$	330 ± 180	44 ± 26	490 ± 270	87 ± 51	760 ± 450	190 ± 110
$t\bar{t}+c\bar{c}$	280 ± 170	17 ± 12	390 ± 240	21± 15	560 ± 350	48 ± 33
$t\bar{t}+light$	1500 ± 400	48 ± 18	1370 ± 400	45 ± 18	1200 ± 500	40 ± 23
$t\bar{t}H$ (125)	13 ± 4.5	3.3 ± 2.1	21 ± 6.2	7.0 ± 3.2	42 ± 11	16 ± 6.1
Total bkg.	18700 ± 500	1200 ± 50	14960 ± 580	1300 ± 65	13380 ± 77	1650 ± 130
Data	18508	1545	14741	1402	13131	1587

Table 6.20: Pre-fit event yields for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields for all samples but the multijet background. As multijet normalization has no pre-fit value, only statistical uncertainty is quoted. As numbers were rounded, the sum of all contributions may not equal the total value.

	(6j, 3b)	(6j, ≥4b)	(7j, 3b)	(7j, ≥4b)	(8j, 3b)	(8j, ≥4b)
Multijet	16000 ± 320	1400 ± 66	12000 ± 350	1230 ± 78	10000 ± 490	1300 ± 100
single top	180 ± 59	6.7 ± 3.6	153 ± 12	9.4 ± 4.4	120 ± 47	12 ± 5.7
$t\bar{t}+V$	15 ± 6.2	1.9 ± 1.5	23 ± 8.9	3.6 ± 2.1	43 ± 15	8.7 ± 4.2
$t\bar{t}+b\bar{b}$	230 ± 120	31 ± 17	340 ± 190	63 ± 34	560 ± 320	140 ± 75
$t\bar{t}+c\bar{c}$	350 ± 170	22 ± 11	490 ± 240	28 ± 15	740 ± 360	66 ± 32
$t\bar{t}+light$	1750 ± 270	55 ± 13	1650 ± 340	54 ± 19	1500 ± 450	54 ± 21
$t\bar{t}H$ (125)	21 ± 6.1	5.5 ± 2.7	35 ± 8.6	11 ± 4.4	71 ± 15	27 ± 8.4
Total bkg.	18500 ± 310	1540 ± 61	14700 ± 300	1400 ± 69	13100 ± 340	1590 ± 72
Data	18508	1545	14741	1402	13131	1587

Table 6.21: Post-fit event yields for signal, backgrounds and data in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields. As numbers were rounded, the sum of all contributions may not equal the total value.

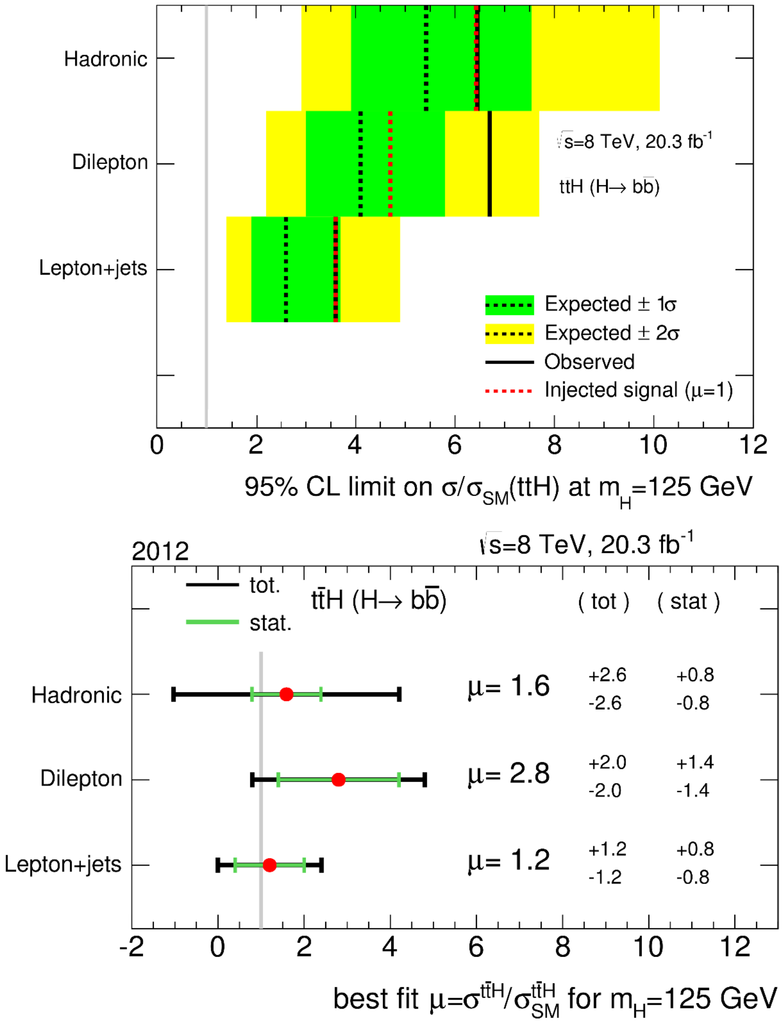


Figure 6.67: *Top*: summary of 95% CL upper limits on $\sigma(t\bar{t}H)$ relative to the SM prediction, for the $t\bar{t}H$ $(H \rightarrow b\bar{b})$ channels. The observed limits (solid lines) are compared to the expected (median) limits under the background-only hypothesis (black dashed lines) and under the signal-plus-background hypothesis assuming the SM prediction for $\sigma(t\bar{t}H)$ (light dashed lines). Bands correspond to the $\pm1\sigma$ and $\pm2\sigma$ ranges around the expected limits under the background-only hypothesis. *Bottom*: summary of the signal strength measurement for all $t\bar{t}H$ $(H \rightarrow b\bar{b})$ channels. Values for semi- and di-leptonic channels are from [11]. Higgs boson mass of 125 GeV is assumed.

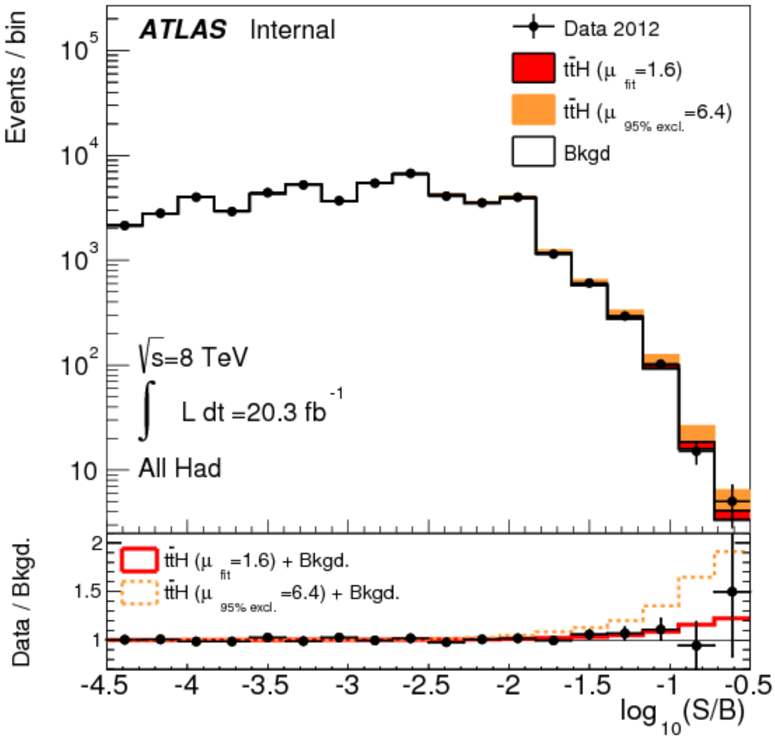


Figure 6.68: Event yields as function of decimal logarithm of signal to background ratio: $\log(S/B)$. All the regions entering the analysis are considered. Background and signal yields shown are obtained from the fit to data. $t\bar{t}H$ is shown as red area. The yield for the 95% CL excluded cross section is also shown.

6.11 Prospects for Run 2 of LHC

The LHC Run 2 data taking period has started in spring 2015 with increased center of mass energy of 13 TeV. The $t\bar{t}H$ total cross section ($\sigma_{t\bar{t}H}$) calculated at NLO in QCD assuming a SM Higgs boson with mass $M_H = 125$ GeV increases from $\sigma_{t\bar{t}H} = 0.1290$ pb $^{+3.9\%}_{-9.3\%}$ (QCDScale) $\pm 8.1\%(\text{PDF} + \alpha_s)$ at $\sqrt{s} = 8$ TeV [30] to $\sigma_{t\bar{t}H} = 0.5085$ pb $^{+5.7\%}_{-9.3\%}$ (QCDScale) $\pm 8.8\%(\text{PDF} + \alpha_s)$ at $\sqrt{s} = 13$ TeV [163] which corresponds to a factor of ~ 4 while the inclusive $t\bar{t}$ cross section scales only by a factor of 3.3. Another benefit for all Run 2 analyses comes from the increase in instantaneous luminosity which has as a result that ATLAS is expected to collect ~ 200 fb^{-1} by the end of Run 2.

An important improvement is also expected from the upgrade of the trigger in ATLAS. In Run 2 it is possible for LVL1 multi-jet triggers item to make decisions based on cuts on E_T as well as η of the online objects. In particular a LVL1 trigger item requiring three jets with $E_T > 25$ GeV and $|\eta| < 2.3$ is already implemented in the Run 2 and is collecting data with a sustainable rate. This allows to have less tight p_T requirement in the analysis event selection, passing from requiring at least five jets with $p_T > 55$ GeV (and any other jets with $p_T > 25$ GeV) to at least three jets with $p_T > 60$ GeV (and any other jets with $p_T > 25$ GeV). With this analysis requirements the gain in signal efficiency is of the order of 90%. The analysis can take advantage also from the amelioration brought to online b-tagging. This has now light- and c-jet rejection close to the one of offline algorithms, see Figure 6.69, allowing a more sustainable trigger rate than a standard jet or multi-jet triggers. Notice that contrary to fully hadronic state, all lepton based signature decrease their trigger efficiency due to the tighter requirements on lepton triggers in Run 2.

For Run 2 a new b-tagging algorithm has been implemented [164]. This uses similar multivariate techniques to the one used in Run 1, described in Section 3.2. It benefits from the improved performance of the tracking algorithms and of the upgrade of the ID with the insertion of the Insertable B-Layer (IBL) [16], between the new beam pipe with a smaller radius and the previously existing pixel detector, described in Section 2.2.3. It is expected to have a major impact on the b-tagging performance due to the significantly improved tracks' impact parameter resolution. The b-tagging efficiency is improved by $\sim$10% for the same light and c-quark rejection obtained with MV1. Since $t\bar{t}H$ ($H \to b\bar{b}$) contains four b-jets in the final state, passing from a 60% to a 70% operating point corresponds to an increase of signal efficiency of the order of 90% without loss of signal purity.

Several improvements can also be done on the analysis level. It would be possible to introduce a BDT to discriminate $t\bar{t} + b\bar{b}$ from $t\bar{t}H$. As $t\bar{t} + b\bar{b}$ is the dominant irreducible background for $t\bar{t}H$ ($H \to b\bar{b}$), this will help to reduce the impact of the systematic uncertainty related to the $t\bar{t} + b\bar{b}$ cross section uncertainty, which is the dominant uncertainty of the analysis as can be seen in Figure 6.58. It will also be possible to study boosted regimes where the decay products the top-quark or Higgs boson are combined in a single jet. In these case rejection of multijet background against events with a $t\bar{t}$ pair can be improved using discriminants like

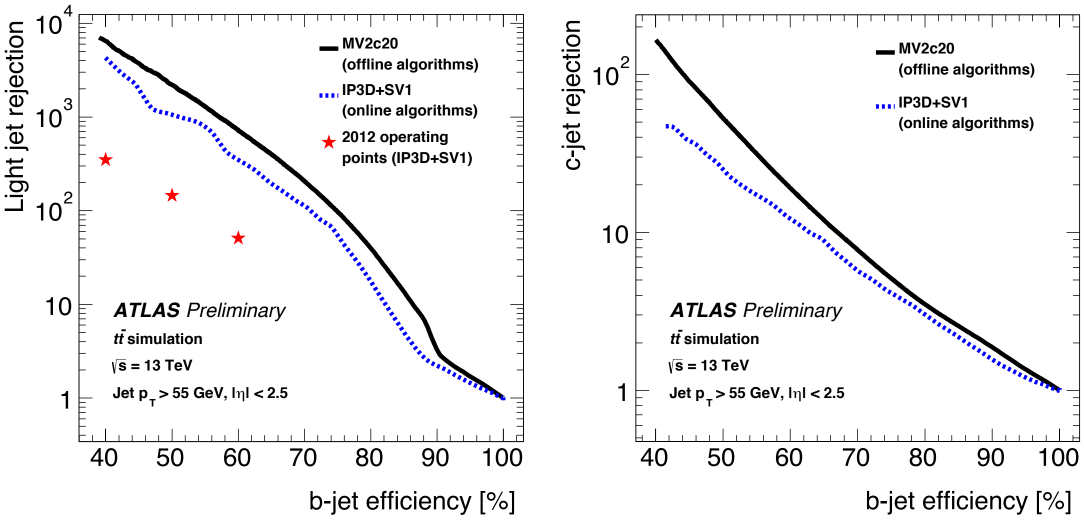


Figure 6.69: The expected online performance in terms of light-jet rejection (left) and c-jet rejection (right) of the MV2c20 tagger, ATLAS b-tagging algorithm for Run 2, (solid black line) is shown together with the expected performance of the IP3D+SV1 tagger in Run 2 (dashed purple line). For the light-jet rejection the performance of the IP3D+SV1 tagger that was achieved during Run 1 (red stars) is also shown. The tuning is performed on $t\bar{t}$ simulation with $\sqrt{s} = 13$ TeV. Jets used are required to have $p_T > 55$ GeV and $|\eta| < 2.5$. The points illustrating the Run 1 performance were derived using $t\bar{t}$ simulation with $\sqrt{s} = 8$ TeV [100].

top-taggers [165] or general jet substructure observables [166].

Improvements in the multijet background description can be obtained with a refined definition of the $\mathrm{TRF}_{\mathrm{MJ}}$ extraction region. Figure 6.70 shows values of $\varepsilon_{\mathrm{MJ}}$ obtained requiring additional selections on jet and b-tagged jets multiplicity in the $\mathrm{TRF}_{\mathrm{MJ}}$ extraction region, defined in Section 6.5.1. It can be noticed a dependency on the jet and b-tagged jet multiplicity which is not exploited in the current analysis. This difference in behaviour can be linked to the mismodelling observed in the H_T and S_t variables and will be scrutinized further in the next generations of fully hadronic analyses.

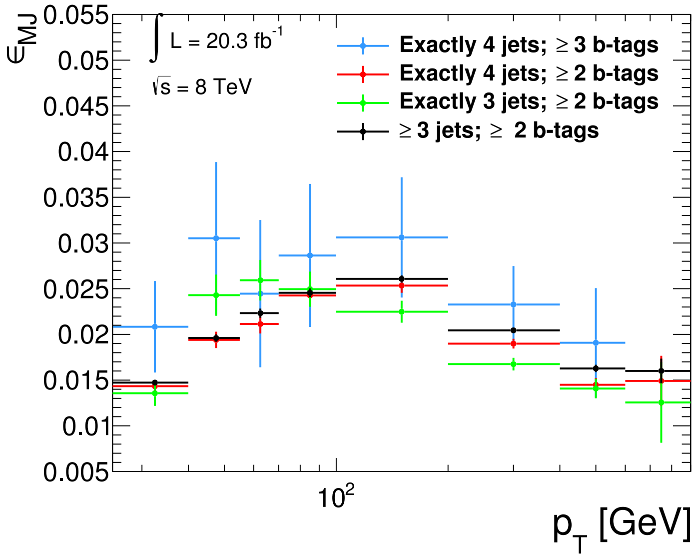


Figure 6.70: One dimensional dependence of $\varepsilon_{\mathrm{MJ}}$ as a function of jet p_T evaluated in data. Different efficiencies are derived with different requirements on the $\mathrm{TRF}_{\mathrm{MJ}}$ evaluation region. The range covered in p_T goes from 25 GeV to 900 GeV.

Conclusion

The discovery of a new particle in 2012 by the ATLAS and CMS collaborations at the LHC has been a historical moment in the history of particle physics. This particle is likely to be identified as the Brout-Englert-Higgs boson, which is the last particle predicted but not yet observed by the SM of particle physics. This is a great success in particle physics and by no means the last challenge for the SM. Indeed, the SM is not able to account for a series of observations that could lead to the discovery of new interactions and new particles. To assess the scale where the physics Beyond the SM lies, one of the paths follows the precise measurement of the properties of the recently discovered particle. In particular, the strength of its interaction with the top quark can be directly constrained at the LHC measuring the associated production of a Higgs boson with a top-antitop quark pair ($t\bar{t}H$).

To be able to have the best possible measure of the $t\bar{t}H$ cross section, it is necessary to analyze the largest amount of data. Several measures have been made exploiting different Higgs boson decays. The biggest branching ratio is in $b\bar{b}$ quark pairs ($H \to b\bar{b}$), ATLAS and CMS have performed this search looking for topologies where one or both top quark decays in a leptonic final states. The search in the fully hadronic final state has never been performed up to now. This doctoral thesis describes the first attempt ever performed of this search.

In this thesis great detail has been given to the description of the $\mathrm{TRF}_{\mathrm{MJ}}$ method, the data-driven technique that allows to estimate the overwhelming multijet background. The $\mathrm{TRF}_{\mathrm{MJ}}$ method is based on the probability for a multijet event to be b-tagged, $\varepsilon_{\mathrm{MJ}}$, which is evaluated in a dedicated data sample selected with a combination of high E_T single jet triggers (145 GeV $\leq E_T \leq$ 360 GeV) and multi-jet trigger (E_T : 3 × 55 GeV). The method allows to determine the normalization of the multijet background with an accuracy of 5% in the $\varepsilon_{\mathrm{MJ}}$ evaluation region and in a Pythia8 di-jet MC sample. The $\mathrm{TRF}_{\mathrm{MJ}}$ method allows to describe observables depending on properties of b-tagged jets. Systematic uncertainties on the method have been assessed using five different sets of variables to describe the b-tagging efficiency $\varepsilon_{\mathrm{MJ}}$. Two additional systematic uncertainties have been evaluated changing the criterion used for the evaluation of $\varepsilon_{\mathrm{MJ}}$. Finally, two reweightings have been evaluated to take into account a residual mismodelling in H_T and S_t and were used as systematic uncertainties.

A per-jet efficiency has been evaluated for a multi-jet trigger, and it has been proven that with this type of parametrization it is possible to describe the trigger behaviour using only offline jets information. Scale Factors (SF) have been evaluated to allow to include events outside the trigger fully efficient region to increase the signal efficiency.

A BDT is used to discriminate the $t\bar{t}H$ signal against the background and its distribution is fitted to the data. The result obtained, for a Higgs Boson mass of 125 GeV, is the measurement of a 95% CL observed upper limit of 6.5 times the SM cross section, while the expected limit is 5.4. A $t\bar{t}H$ signal strength of 1.6 ± 2.6 times the SM value is obtained. This result takes into account theoretical systematics uncertainties, systematic uncertainties on the physics objects used and dedicated systematic uncertainties on the trigger SF and the $\mathrm{TRF_{MJ}}$ method. The great impact of the lack of knowledge on the irreducible background $t\bar{t} + b\bar{b}$ for this measurement indicates the necessity of its better understanding to be able to perform a $t\bar{t}H$ ($H \rightarrow b\bar{b}$) measurement.

The result presented in this work demonstrated that this analysis is within reach for LHC experiments and that it could give a non negligible contribution in a future combination with other $t\bar{t}H$ searches, where the dependency from $t\bar{t}$+$b\bar{b}$ uncertainties are mitigated. In addition, the development of methods to estimate the multijet background is of interest for future searches of new physics in fully hadronic channels.

A description of the prospects for Run 2 of LHC has also been given. The analysis will benefit from the update of the detector, with better performing trigger and b-tagging algorithms with respect to Run 1. This can reflect in an increase of signal yield of a factor ~ 2. Moreover, increased luminosity can allow for more precise description of the multijet background refining the $\mathrm{TRF_{MJ}}$ method and reducing the systematic uncertainties related to it.

My work started with the trigger studies described in this thesis, and has continued in the $t\bar{t}H$ ($H \rightarrow b\bar{b}$) fully hadronic analysis. I took care of the preparation of the data and MC samples used, and developed the code for the trigger efficiency and SF evaluation and for the application of the latter in the analysis. I contributed to the update of the code for the skimming and slimming of the data and MC samples. I implemented the Tag Rate Function method for MC simulation ($\mathrm{TRF_{MC}}$) in the analysis code and performed its validation using POWHEG $t\bar{t}$ and POWHEL $t\bar{t}H$ samples. Great amount of work has been devoted to the Tag Rate Function method for estimation of multijet background ($\mathrm{TRF_{MJ}}$). I contributed to its design and development from the beginning to its implementation in the analysis code. I performed all the validation studies using two independent data samples and a PYTHIA8 di-jet MC sample prepared by me. I have performed the evaluation of the systematic uncertainties for the $\mathrm{TRF_{MJ}}$ method. I also contributed to the implementation of the majority of the observables used in the BDTs in the analysis code and to the running of the analysis chain from the D3PDs to the inputs for the final fit.

Bibliography

[1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) no. 4, 579 – 588. `http://www.sciencedirect.com/science/article/pii/0029558261904692`.

[2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (Nov, 1967) 1264–1266. `http://link.aps.org/doi/10.1103/PhysRevLett.19.1264`.

[3] A. Salam, *Weak and electromagnetic interactions*, . Proc. of the 8th Nobel Symposium 1968.

[4] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (Aug, 1964) 321–323. `http://link.aps.org/doi/10.1103/PhysRevLett.13.321`.

[5] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (Oct, 1964) 508–509. `http://link.aps.org/doi/10.1103/PhysRevLett.13.508`.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (Nov, 1964) 585–587. `http://link.aps.org/doi/10.1103/PhysRevLett.13.585`.

[7] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) 132–133.

[8] 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) 30–61, `arXiv:1207.7235 [hep-ex]`.

[9] 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) 1–29, `arXiv:1207.7214 [hep-ex]`.

[10] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, J. Exp. Theor. Phys. **120** (2015) 335–343, `arXiv:1411.1923 [hep-ph]`. [Zh. Eksp. Teor. Fiz.147,389(2015)].

[11] 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) no. 7, 349, `arXiv:1503.05066 [hep-ex]`.

[12] 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) 222–242, `arXiv:1409.3122 [hep-ex]`.

[13] ATLAS Collaboration, *Search for the associated production of the Higgs boson with a top quark pair in multilepton final states with the ATLAS detector*, `arXiv:1506.05988 [hep-ex]`.

[14] 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]`.

[15] 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. C **75** (2015) no. 6, 251, `arXiv:1502.02485 [hep-ex]`.

[16] M. Capeans, G. Darbo, K. Einsweiller, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet, *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Geneva, Sep, 2010. `https://cds.cern.ch/record/1291633`.

[17] C. P. Burgess and G. D. Moore, *The standard model, a primer*. 2011.

[18] A. Bettini, *Introduction to elementary particle physics*. 2008.

[19] F. Halzen and A. D. Martin, *Quark and leptons: an introductory course in modern particle physics*. 1984.

[20] S. Coleman, *Aspects of symmetry*. 1985.

[21] Y. Nambu, *Quasi-Particles and Gauge Invariance in the Theory of Superconductivity*, Phys. Rev. **117** (Feb, 1960) 648–663. `http://link.aps.org/doi/10.1103/PhysRev.117.648`.

[22] J. Goldstone, A. Salam, and S. Weinberg, *Broken Symmetries*, Phys. Rev. **127** (Aug, 1962) 965–970. `http://link.aps.org/doi/10.1103/PhysRev.127.965`.

[23] Super-Kamiokande Collaboration, *Measurements of the Solar Neutrino Flux from Super-Kamiokande's First 300 Days*, Phys. Rev. Lett. **81** (Aug, 1998) 1158–1162. `http://link.aps.org/doi/10.1103/PhysRevLett.81.1158`.

[24] G. F. Giudice, *Naturalness after LHC8*, PoS **EPS-HEP2013** (2013) 163, `arXiv:1307.7879 [hep-ph]`.

[25] K. G. Wilson, *The Renormalization Group and Strong Interactions*, Phys. Rev. D **3** (1971) 1818.

[26] E. Gildener, *Gauge-symmetry hierarchies*, Phys. Rev. D **14** (Sep, 1976) 1667. `http://link.aps.org/doi/10.1103/PhysRevD.14.1667`.

[27] S. Weinberg, *Gauge hierarchies*, Phys. Lett. B **82** (1979) no. 3-4, 387. `http://www.sciencedirect.com/science/article/pii/037026937990248X`.

[28] G. 't Hooft, *Procedings of 1979 Cargese Summer Institute: Recent Developments in Gauge Theories.* New York, 1980.

[29] S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*, `arXiv:1201.3084 [hep-ph]`.

[30] LHC Higgs Cross Section Working Group Collaboration, S. Heinemeyer et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, `arXiv:1307.1347 [hep-ph]`.

[31] R. V. Harlander and W. B. Kilgore, *Next-to-next-to-leading order Higgs production at hadron colliders*, Phys. Rev. Lett. **88** (2002) 201801, `arXiv:hep-ph/0201206 [hep-ph]`.

[32] C. Anastasiou and K. Melnikov, *Higgs boson production at hadron colliders in NNLO QCD*, Nucl. Phys. B **646** (2002) 220–256, `arXiv:hep-ph/0207004 [hep-ph]`.

[33] V. Ravindran, J. Smith, and W. L. van Neerven, *NNLO corrections to the total cross-section for Higgs boson production in hadron hadron collisions*, Nucl. Phys. B **665** (2003) 325–366, `arXiv:hep-ph/0302135 [hep-ph]`.

[34] S. Catani, D. de Florian, M. Grazzini, and P. Nason, *Soft gluon resummation for Higgs boson production at hadron colliders*, JHEP **07** (2003) 028, `arXiv:hep-ph/0306211 [hep-ph]`.

[35] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini, *Two loop light fermion contribution to Higgs production and decays*, Phys. Lett. B **595** (2004) 432–441, `arXiv:hep-ph/0404071 [hep-ph]`.

[36] G. Degrassi and F. Maltoni, *Two-loop electroweak corrections to Higgs production at hadron colliders*, Phys. Lett. B **600** (2004) 255–260, `arXiv:hep-ph/0407249 [hep-ph]`.

[37] U. Aglietti, R. Bonciani, G. Degrassi, and A. Vicini, *Two-loop electroweak corrections to Higgs production in proton-proton collisions*, in *TeV4LHC Workshop: 2nd Meeting Brookhaven, Upton, New York, February 3-5, 2005.* 2006. `arXiv:hep-ph/0610033 [hep-ph]`.

[38] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, *Higgs production via vector-boson fusion at NNLO in QCD*, Phys. Rev. Lett. **105** (2010) 011801, `arXiv:1003.4451 [hep-ph]`.

[39] P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro, *Vector boson fusion at NNLO in QCD: SM Higgs and beyond*, Phys. Rev. D **85** (2012) 035002, `arXiv:1109.3717 [hep-ph]`.

[40] M. L. Ciccolini, S. Dittmaier, and M. Kramer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys. Rev. D **68** (2003) 073003, `arXiv:hep-ph/0306234 [hep-ph]`.

[41] O. Brein, R. V. Harlander, and T. J. E. Zirke, *vh@nnlo - Higgs Strahlung at hadron colliders*, Comput. Phys. Commun. **184** (2013) 998–1003, `arXiv:1210.5347 [hep-ph]`.

[42] S. Dittmaier, M. Kramer, 1, and M. Spira, *Higgs radiation off bottom quarks at the Tevatron and the CERN LHC*, Phys. Rev. D **70** (2004) 074010, `arXiv:hep-ph/0309204 [hep-ph]`.

[43] S. Dawson, C. B. Jackson, L. Reina, and D. Wackeroth, *Exclusive Higgs boson production with bottom quarks at hadron colliders*, Phys. Rev. D **69** (2004) 074027, `arXiv:hep-ph/0311067 [hep-ph]`.

[44] R. V. Harlander and W. B. Kilgore, *Higgs boson production in bottom quark fusion at next-to-next-to leading order*, Phys. Rev. D **68** (2003) 013001, `arXiv:hep-ph/0304035 [hep-ph]`.

[45] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, and P. M. Zerwas, *Higgs radiation off top quarks at the Tevatron and the LHC*, Phys. Rev. Lett. **87** (2001) 201805, `arXiv:hep-ph/0107081 [hep-ph]`.

[46] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, and P. M. Zerwas, *NLO QCD corrections to t anti-t H production in hadron collisions*, Nucl. Phys. B **653** (2003) 151–203, `arXiv:hep-ph/0211352 [hep-ph]`.

[47] L. Reina and S. Dawson, *Next-to-leading order results for t anti-t h production at the Tevatron*, Phys. Rev. Lett. **87** (2001) 201804, `arXiv:hep-ph/0107101 [hep-ph]`.

[48] S. Dawson, L. H. Orr, L. Reina, and D. Wackeroth, *Associated top quark Higgs boson production at the LHC*, Phys. Rev. D **67** (2003) 071503, `arXiv:hep-ph/0211438 [hep-ph]`.

[49] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackeroth, *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, Phys. Rev. **D68** (2003) 034022, `arXiv:hep-ph/0305087 [hep-ph]`.

[50] 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]`.

[51] CMS Collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV*, Eur. Phys. J. C **75** (2015) no. 5, 212, `arXiv:1412.8662 [hep-ex]`.

[52] CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, Phys. Rev. D **92** (2015) no. 1, 012004, `arXiv:1411.3441 [hep-ex]`.

[53] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, `arXiv:1506.05669 [hep-ex]`.

[54] ATLAS Collaboration, *Determination of spin and parity of the Higgs boson in the* $WW^* \rightarrow e\nu\mu\nu$ *decay channel with the ATLAS detector*, Eur. Phys. J. C **75** (2015) no. 5, 231, `arXiv:1503.03643 [hep-ex]`.

[55] 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*, Tech. Rep. ATLAS-CONF-2015-044, CERN, Geneva, Sep, 2015. `http://cds.cern.ch/record/2052552`.

[56] K. Olive and P. D. Group, *Review of Particle Physics*, Chinese Physics C **38** (2014) no. 9, 090001. `http://stacks.iop.org/1674-1137/38/i=9/a=090001`.

[57] ATLAS, CDF, CMS, D0 Collaboration, *First combination of Tevatron and LHC measurements of the top-quark mass*, `arXiv:1403.4427 [hep-ex]`.

[58] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at* $\sqrt{s} = 7$ *and* 8 *TeV in the ATLAS experiment*, `arXiv:1507.04548 [hep-ex]`.

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

[60] ATLAS Collaboration, *Luminosity Public Results*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`.

[61] ALICE Collaboration, *ALICE: Physics Performance Report, Volume II*, Journal of Physics G: Nuclear and Particle Physics **32** (2006) no. 10, 1295. `http://stacks.iop.org/0954-3899/32/i=10/a=001`.

[62] LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008) no. 08, S08005. `http://stacks.iop.org/1748-0221/3/i=08/a=S08005`.

[63] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) no. 08, S08003. `http://stacks.iop.org/1748-0221/3/i=08/a=S08003`.

[64] CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) no. 08, S08004. `http://stacks.iop.org/1748-0221/3/i=08/a=S08004`.

[65] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1*, Tech. Rep. ATLAS-TDR-6, CERN-LHCC-97-018, Geneva, 1997. `https://cds.cern.ch/record/338080`.

[66] ATLAS Collaboration, *ATLAS central solenoid: Technical Design Report*, Tech. Rep. ATLAS-TDR-9, CERN-LHCC-97-021, Geneva, 1997. `https://cds.cern.ch/record/331067`.

[67] ATLAS Collaboration Collaboration, ATLAS Collaboration, *ATLAS barrel toroid: Technical Design Report*, Tech. Rep. ATLAS-TDR-7, CERN-LHCC-97-019, Geneva, 1997. `https://cds.cern.ch/record/331065`.

[68] ATLAS Collaboration, *ATLAS end-cap toroids: Technical Design Report*, Tech. Rep. ATLAS-TDR-8, CERN-LHCC-97-020, Geneva, 1997. `https://cds.cern.ch/record/331066`. Electronic version not available.

[69] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*, Tech. Rep. ATLAS-TDR-4, CERN-LHCC-97-016, Geneva, 1997. `https://cds.cern.ch/record/331063`.

[70] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 2*, Tech. Rep. ATLAS-TDR-5, CERN-LHCC-97-017, Geneva, 1997. `https://cds.cern.ch/record/331064`.

[71] ATLAS Collaboration, ATLAS Collaboration, *A study of the material in the ATLAS inner detector using secondary hadronic interactions*, JINST **7** (2012) P01013, `arXiv:1110.6191 [hep-ex]`.

[72] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.

[73] ATLAS Collaboration, *The barrel modules of the ATLAS semiconductor tracker*, Nucl. Instrum. Meth. A **568** (2006) 642–671.

[74] ATLAS Collaboration, *The ATLAS semiconductor tracker end-cap module*, Nucl. Instrum. Meth. A **575** (2007) 353–389.

[75] ATLAS TRT Collaboration, *The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: Design and performance*, JINST **3** (2008) P02013.

[76] ATLAS Collaboration, *ATLAS calorimeter performance: Technical Design Report*, Tech. Rep. ATLAS-TDR-1, CERN-LHCC-96-040, Geneva, 1996. `http://cds.cern.ch/record/331059`.

[77] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*, Tech. Rep. ATLAS-TDR-10, CERN-LHCC-97-022, Geneva, 1997. `http://cds.cern.ch/record/331068`.

[78] ATLAS Collaboration, *ATLAS Computing: technical design report*, Tech. Rep. ATLAS-TDR-17, CERN-LHCC-2005-022, Geneva, 2005. `https://cds.cern.ch/record/837738`.

[79] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, *Calorimeter Clustering Algorithms: Description and Performance*, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008. `https://cds.cern.ch/record/1099735`.

[80] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. D **48** (1993) 3160–3166, `arXiv:hep-ph/9305266 [hep-ph]`.

[81] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better jet clustering algorithms*, JHEP **08** (1997) 001, `arXiv:hep-ph/9707323 [hep-ph]`.

[82] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999.* 1998. `arXiv:hep-ph/9907280 [hep-ph]`. `http://inspirehep.net/record/484872/files/arXiv:hep-ph_9907280.pdf`.

[83] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k_t jet clustering algorithm*, JHEP **0804** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[84] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, Eur. Phys. J. C **72** (2012) 1896, `arXiv:1111.6097 [hep-ph]`.

[85] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B **641** (2006) 57–61, `arXiv:hep-ph/0512210 [hep-ph]`.

[86] G. C. Blazey et al., *Run II jet physics*, in *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pp. 47–77. 2000. `arXiv:hep-ex/0005012 [hep-ex]`. `http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-00-092`.

[87] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. C **73** (2013) no. 3, 2304, `arXiv:1112.6426 [hep-ex]`.

[88] ATLAS Collaboration, ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, Eur. Phys. J. C **75** (2015) no. 1, 17, `arXiv:1406.0076 [hep-ex]`.

[89] A. B. Galtieri, F. Margaroli, and I. Volobouev, *Precision measurements of the top quark mass from the Tevatron in the pre-LHC era*, Rept. Prog. Phys. **75** (2012) 056201, `arXiv:1109.2163 [hep-ex]`.

[90] ATLAS collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, Tech. Rep. ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014. `http://cds.cern.ch/record/1741020`.

[91] ATLAS collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2014-004, CERN, Geneva, 2014. `http://cds.cern.ch/record/1664335`.

[92] 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) no. 7, 2941, `arXiv:1404.2240 [hep-ex]`.

[93] ATLAS collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, Eur. Phys. J. C **74** (2014) no. 11, 3130, `arXiv:1407.3935 [hep-ex]`.

[94] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, Eur. Phys. J. C **72** (2012) 1844, `arXiv:1108.5602 [hep-ex]`.

[95] ATLAS Collaboration, *Measurements of the photon identification efficiency with the ATLAS detector using 4.9* $fb^{-}1$ *of pp collision data collected in 2011*, Tech. Rep. ATLAS-CONF-2012-123, CERN, Geneva, Aug, 2012. `https://cds.cern.ch/record/1473426`.

[96] ATLAS Collaboration, *Expected photon performance in the ATLAS experiment*, Tech. Rep. ATL-PHYS-PUB-2011-007, CERN, Geneva, Apr, 2011. `http://cds.cern.ch/record/1345329`.

[97] ATLAS Collaboration, *Public Jet Trigger Plots for Collision Data*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetTriggerPublicResults`.

[98] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross section in the all-hadronic channel in ATLAS with $\sqrt{s} = 7$ TeV data*, Tech. Rep. ATLAS-CONF-2012-031, CERN, Geneva, Mar, 2012. `http://cds.cern.ch/record/1432196`.

[99] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852–867, `arXiv:0710.3820 [hep-ph]`.

[100] ATLAS Collaboration, *Public b-Jet Trigger Plots for Collision Data*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults`.

[101] ATLAS Collaboration, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, JHEP **1501** (2015) 069, `arXiv:1409.6212 [hep-ex]`.

[102] CMS Collaboration, *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, Phys. Rev. D **89** (2014) no. 1, 012003, `arXiv:1310.3687 [hep-ex]`.

[103] CDF and D0 Collaborations, *Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron*, Phys. Rev. Lett. **109** (2012) 071804, `arXiv:1207.6436 [hep-ex]`.

[104] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, `arXiv:1101.0593 [hep-ph]`.

[105] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. C **63** (2009) 189–285, `arXiv:0901.0002 [hep-ph]`.

[106] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Uncertainties on α_S in global PDF analyses and implications for predicted hadronic cross sections*, Eur. Phys. J. C **64** (2009) 653–680, `arXiv:0905.3531 [hep-ph]`.

[107] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012, `arXiv:hep-ph/0201195 [hep-ph]`.

[108] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *A first unbiased global NLO determination of parton distributions and their uncertainties*, Nucl. Phys. **B838** (2010) 136–206, `arXiv:1002.4407 [hep-ph]`.

[109] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004, `arXiv:1303.6254 [hep-ph]`.

[110] CMS Collaboration, *Measurement of the cross section ratio $\sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$ in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **746** (2015) 132–153, `arXiv:1411.5621 [hep-ex]`.

[111] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, Tech. Rep. CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr, 2009. `http://cds.cern.ch/record/1194487`.

[112] CMS Collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, Tech. Rep. CMS-PAS-PFT-10-001, CERN, 2010. `http://cds.cern.ch/record/1247373`.

[113] CMS collaboration, *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*, Journal of Instrumentation **7** (2012) no. 10, P10002. `http://stacks.iop.org/1748-0221/7/i=10/a=P10002`.

[114] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois, *Electron reconstruction in CMS*, The European Physical Journal C **49** (2007) no. 4, 1099–1116. `http://dx.doi.org/10.1140/epjc/s10052-006-0175-5`.

[115] CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, JINST **8** (2013) P04013, `arXiv:1211.4462 [hep-ex]`.

[116] ATLAS Collaboration, 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. **D90** (2014) no. 7, 072004, `arXiv:1407.0371 [hep-ex]`.

[117] K. Kondo, *Dynamical Likelihood Method for Reconstruction of Events with Missing Momentum. I. Method and Toy Models*, Journal of the Physical Society of Japan **57** (1988) no. 12, 4126–4140.

[118] CMS collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, JINST **6** (2011) no. 11, P11002, `arXiv:1107.4277`.

[119] CMS Collaboration, *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update*, Tech. Rep. CMS-PAS-LUM-13-001, CERN, Geneva, 2013. `http://cds.cern.ch/record/1598864`.

[120] ATLAS Collaboration, ATLAS collaboration, *Improved luminosity determination in pp collisions at sqrts = 7 TeV using the ATLAS detector at*

the LHC, Eur. Phys. J. C **73** (2013) no. 8, 2518, `arXiv:1302.4393 [hep-ex]`.

[121] F. Cascioli, P. Maierhofer, N. Moretti, S. Pozzorini, and F. Siegert, *NLO matching for $t\bar{t}b\bar{b}$ production with massive b-quarks*, Phys. Lett. B **734** (2014) 210–214, `arXiv:1309.5912 [hep-ph]`.

[122] D0 Collaboration, *Useful Diagrams of Top Signals and Backgrounds*, `http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html`.

[123] ATLAS Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, Tech. Rep. ATLAS-CONF-2012-020, CERN, Geneva, Mar, 2012. `https://cds.cern.ch/record/1430034`.

[124] G. Bevilacqua, M. Czakon, M. V. Garzelli, A. van Hameren, A. Kardos, C. G. Papadopoulos, R. Pittau, and M. Worek, *HELAC-NLO*, Comput. Phys. Commun. **184** (2013) 986–997, `arXiv:1110.1499 [hep-ph]`.

[125] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, `arXiv:hep-ph/0409146 [hep-ph]`.

[126] 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]`.

[127] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, `arXiv:1002.2581 [hep-ph]`.

[128] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trocsanyi, *Standard Model Higgs boson production in association with a top anti-top pair at NLO with parton showering*, Europhys. Lett. **96** (2011) 11001, `arXiv:1108.0387 [hep-ph]`.

[129] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, Tech. Rep. ATL-PHYS-PUB-2012-003, CERN, Geneva, Aug, 2012. `http://cds.cern.ch/record/1474107`.

[130] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, `arXiv:1007.2241 [hep-ph]`.

[131] T. Sjöstrand, S. Mrenna, and P. Skands, *Pythia 6.4 Physics and Manual*, JHEP **05** (2006) 026, `arXiv:0603175v2 [hep-ph]`.

[132] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612–622, `arXiv:1111.5869 [hep-ph]`.

[133] P. Bärnreuther, M. Czakon, and A. Mitov, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to* $q\bar{q} \to t\bar{t} + X$, Phys. Rev. Lett. **109** (2012) 132001, `arXiv:1204.5201 [hep-ph]`.

[134] 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]`.

[135] 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]`.

[136] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, `arXiv:1112.5675 [hep-ph]`.

[137] 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) 919–985, `arXiv:hep-ph/0405247 [hep-ph]`.

[138] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan, *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, `arXiv:0802.0007 [hep-ph]`.

[139] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D **82** (2010) 074018, `arXiv:1005.3457 [hep-ph]`.

[140] S. Frixione, E. Laenen, P. Motylinski, and B. R. Webber, *Single-top production in MC@NLO*, JHEP **03** (2006) 092, `arXiv:hep-ph/0512250 [hep-ph]`.

[141] 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]`.

[142] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, Phys. Rev. D **81** (Mar, 2010) 054028. `http://link.aps.org/doi/10.1103/PhysRevD.81.054028`.

[143] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D. L. Rainwater, and T. Stelzer, *MadGraph/MadEvent v4: The New Web Generation*, JHEP **09** (2007) 028, `arXiv:0706.2334 [hep-ph]`.

[144] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^{+-}$ production and decay at NLO*, JHEP **07** (2012) 052, `arXiv:1204.5678 [hep-ph]`.

[145] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, and Z. Trocsanyi, *$t\,\bar{t}\,W^{+-}$ 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]`.

[146] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. C **72** (1996) 637–646, `arXiv:hep-ph/9601371 [hep-ph]`.

[147] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97–107, `arXiv:hep-ph/0506026 [hep-ph]`.

[148] S. Jadach, J. H. K"uhn, and Z. Wąs, *TAUOLA - a library of Monte Carlo programs to simulate decays of polarized τ leptons*, Computer Physics Communications **64** (1991) no. 2, 275 – 299. `http://www.sciencedirect.com/science/article/pii/001046559190038M`.

[149] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, European Physical Journal C **70** (Dec., 2010) 823–874, `arXiv:1005.4568 [physics.ins-det]`.

[150] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[151] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, Nucl. Instr. Meth. A **389** (1997) 81.

[152] A. Hoecker et al., *TMVA-v4 Users Guide*, `arXiv:0703039v5 [physics]`.

[153] H.-J. Yang, B. P. Roe, and J. Zhu, *Studies of boosted decision trees for MiniBooNE particle identification*, Nuclear Instruments and Methods in Physics Research A **555** (Dec., 2005) 370–385, `physics/0508045`.

[154] T. Hastie et al, *The elements of statistical learning*, Springer (2001) .

[155] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, `arXiv:1101.0538 [hep-ph]`.

[156] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li, P. Nadolsky, J. Pumplin, D. Stump, and C. P. Yuan, *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. D **89** (2014) no. 3, 033009, `arXiv:1302.6246 [hep-ph]`.

[157] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244–289, `arXiv:1207.1303 [hep-ph]`.

[158] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, `arXiv:1405.0301 [hep-ph]`.

[159] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639–707, `arXiv:0803.0883 [hep-ph]`.

[160] J. Bellm et al., *Herwig++ 2.7 Release Note*, `arXiv:1310.6877 [hep-ph]`.

[161] W. Verkerke and D. Kirkby, *RooFit Users Manual v2.91*, . `http://rootfit.sourceforge.net`.

[162] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, European Physical Journal C **71** (Feb., 2011) 1554, `arXiv:1007.1727 [physics.data-an]`.

[163] LHC Higgs Cross Section Working Group, *SM Higgs production cross sections at* $\sqrt{s} = 13\text{-}14$ *TeV*, `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt1314TeV`.

[164] ATLAS Collaboration, *Expected performance of the ATLAS b-tagging algorithms in Run-2*, Tech. Rep. ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul, 2015. `http://cds.cern.ch/record/2037697`.

[165] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at* $\sqrt{s} = 7$ *TeV using the ATLAS detector*, JHEP **09** (2013) 076, `arXiv:1306.4945 [hep-ex]`.

[166] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, *Stop Reconstruction with Tagged Tops*, JHEP **10** (2010) 078, `arXiv:1006.2833 [hep-ph]`.