Corso di dottorato di ricerca in:

"Informatica e Scienze Matematiche e Fisiche"

Ciclo XXXIII

Titolo della tesi:

"Search for four top quark production in the
single lepton final state in pp collisions at
$\sqrt{s} = 13$ TeV with the ATLAS detector"

Dottorando:
Jacopo Magro

Supervisore:
Prof. Andrea Vacchi

Co-supervisore:
Prof.ssa Marina Cobal

Anno 2021

精力善用

自他共栄

Seiryoku-zen'yō

Jita-kyō'ei

Jigoro Kano

Abstract

The Standard Model of particle physics is a fundamental theory of the modern physics, describing the elementary particles and their interactions. In the last 50 years it has been subject of several tests in order to validate it and put constraints on its parameters. However, many aspects of the world around us are unclear and they are not explained in the Standard Model. In order to improve our understanding of some of these phenomena, the Large Hadron Collider carries on its work after the Higgs boson discovery, searching for new physics beyond the Standard Model and improving the model itself. A key role in this framework is carried out by the top quark: it has a very large coupling to the Higgs boson and it is also predicted to have large couplings to beyond Standard Model particles. In this contest the production of $t\bar{t}t\bar{t}$ in proton-proton collisions has a special part, since it is a rare process predicted by the Standard Model and many beyond Standard Model scenarios lead to an enhancement of its cross section. In this thesis a particular channel of decay is considered, the single lepton channel, analysing the data of the proton-proton collisions at a centre of mass energy of 13 TeV collected in the period 2015-2018 by the ATLAS detector at LHC. In order to improve the results already obtained in the past, a multivariate analysis is performed to separate the SM $t\bar{t}t\bar{t}$ signal from the main $t\bar{t}$+jets background. Moreover, a purely Monte Carlo simulation-based method is not expected to model well the $t\bar{t}$+jets background in the high jet and b-jet multiplicities regions considered in the thesis, and therefore a data-driven, MC-assisted, approach is adopted to improve the prediction for this background: the $TRF_{t\bar{t}}$ method. A profile likelihood fit is used for the measurement of the $t\bar{t}t\bar{t}$ signal strength $\mu = \sigma_{t\bar{t}t\bar{t}}/\sigma_{t\bar{t}t\bar{t}}^{SM}$. The result is $\mu = 3.5_{-0.7}^{+0.7}(\text{stat.})_{-1.8}^{+1.8}(\text{syst.}) = 3.5_{-1.9}^{+1.9}$, which implies an observed (expected) significance of the signal over the background-only hypothesis of 1.8σ (0.5σ). The corresponding measured cross section for the $t\bar{t}t\bar{t}$ process is 42_{-23}^{+23} fb, compatible with the SM value within 1.3σ. The result is also compatible within 1σ with the recent dilepton same-sign and multilepton channel and the previous single lepton and dilepton opposite-sign channel analyses, which fitted a signal strength of $\mu = 2.0_{-0.6}^{+0.8}$ and $\mu = 1.7_{-1.7}^{+1.9}$, respectively. At the time of the submission of this thesis the current official ATLAS result for the single lepton and dilepton opposite-sign channel is not public; thence no comparison can be shown in this thesis.

Contents

V

Chapter 1

The Standard Model

The Standard Model of particle physics (SM) is a gauge quantum theory which describes the elementary particles and their interactions in terms of the quantum field dynamics. The SM was originally elaborated between 1960 and 1967 by S. Glashow [1], S. Weinberg [2] and A. Salam [3] combining electroweak and strong interactions. Moreover, in the following years, the SM has been improved in order to include the discoveries which were observed in those years, and the Higgs mechanism, able to justify the mass of particles. Several experimental confirmations of the SM were found starting from the 80's, establishing this theory as the reference point for particle physics. The most important confirmations, among several others, were the discovery of the W and Z bosons at CERN by Rubbia and Van Der Meer in 1983 [4] and the discovery of the Higgs boson at CERN by ATLAS [5] and CMS [6] collaborations in 2012. In this chapter the Standard Model will be briefly introduced, discussing in particular the aspects that are more relevant for this thesis.

1.1 Particles and interactions

The fundamental constituents of matter interact through four types of forces: the electromagnetic, the weak, the strong and the gravitational one. The SM describes all these interactions except for the gravitational force, which is, however, completely negligible in particle physics experiments, and introducing another interaction (the Higgs mechanism) in order to justify the mass of particles. The elementary particles described by the SM are collected in Figure 1.1. There are two types of particles: *fermions*, with semi-integer spin obeying Fermi-Dirac statistics, and *bosons*, with integer spin obeying Bose-Einstein statistics. Fermions can be divided into *leptons* and *quarks* according to their interactions, and each of them has a corresponding antiparticle with the same mass, but opposite quantum numbers. Leptons are grouped into three families, each formed by a charged lepton and the corresponding neutrino: electron (e) and ν_e, muon (μ) and ν_μ, tau (τ) and ν_τ. Charged leptons can interact through the weak and the electromagnetic force, while neutrinos can interact only through the weak force. Quarks can interact through the electromagnetic, the weak and the strong force, being the latter the dominant one; there are six different *flavours* of quarks, grouped into three families: up and down, charm and strange, top and bottom. They carry fractional electric charge and "live" only in bound states, called *hadrons*, which can be divided in: *mesons*, made up of a quark and an

Standard Model of Elementary Particles

	three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III		
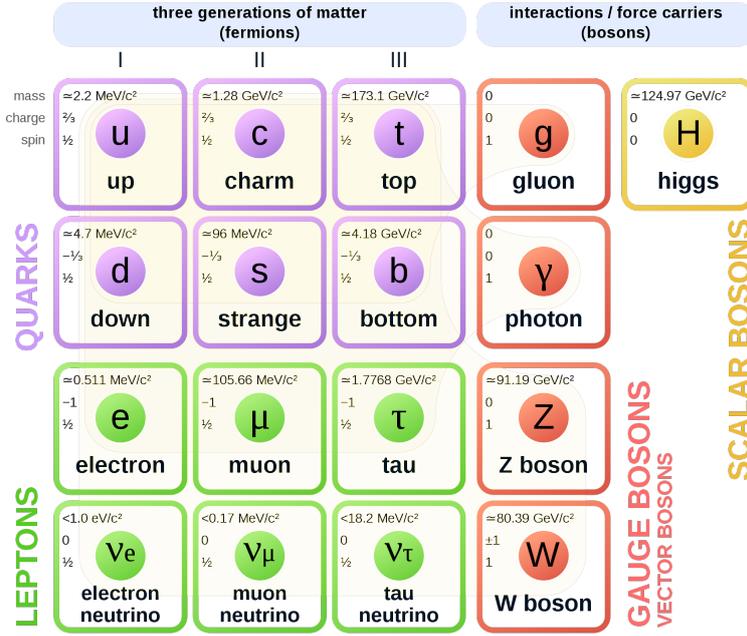
Figure 1.1: Scheme of the elementary particles described in the SM.

antiquark, and *baryons*, formed by three quarks. Quarks have an additional quantum number called *colour*, the charge of the strong interaction, which was originally introduced to accommodate for the Pauli exclusion principle. The three interactions considered in the SM are mediated by the force-carrier gauge bosons: the electromagnetic force by the photon (γ), the weak force by $W^\pm$ and Z^0, and the strong force by gluons. Gluons and γ are massless bosons, while W and Z are massive: $M_W = 80.379 \pm 0.012$ GeV and $M_Z = 91.1876 \pm 0.0021$ GeV [7]. Each interaction is described by a quantum field theory (QFT) based on a gauge group with a specified Lie algebra, and the SM is itself a particular QFT based on the gauge group

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y, \qquad (1.1)$$

which will be explained in the following sections.

1.2 Quantum Electrodynamics

The electromagnetic (EM) 3-vector potential $\vec{A}$ and the scalar potential φ are not unique for given fields $\vec{E}$ and $\vec{B}$. So then, the transformations of the potentials which preserve the fields are called *gauge transformations* and the associated invariance of the Maxwell equations is called *gauge invariance*. The Maxwell equations for the free EM field can be written in a Lorentz covariant form:

$$\partial^\mu F_{\mu\nu} = 0, \qquad (1.2)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the EM field strength tensor, being $A_\mu = (\varphi, -\vec{A})$ the 4-vector potential and $\partial_\mu = (\partial/\partial t, \vec{\nabla})$ the 4-vector differential operator. For

this potential the gauge transformation is simply $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu\alpha$ (e is the electric charge). The Lagrangian density of the free Maxwell field is:

$$\mathcal{L}_{EM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.3}$$

Consider then the Dirac Lagrangian for a fermion of mass m:

$$\mathcal{L}_D = \overline{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x), \tag{1.4}$$

where $\psi(x) = (\psi_1(x), \psi_2(x), \psi_3(x), \psi_4(x))^T$ is the Dirac spinor, $\overline{\psi}(x) = \psi^\dagger(x)\gamma^0$ and γ^μ are the Dirac matrices. Such Lagrangian is invariant under the global symmetry of the unitary Abelian group $U(1)$:

$$\psi(x) \to \psi'(x) = e^{i\alpha}\psi(x). \tag{1.5}$$

If α is not a constant, but a function $\alpha(x)$ of the space-time, $\mathcal{L}_D$ is not anymore invariant under this continuous local transformation. However the invariance can be recovered by coupling the Dirac field to the Maxwell field:

$$\mathcal{L}_{D,EM} = \overline{\psi}(x)(i\gamma^\mu(\partial_\mu - ieA_\mu) - m)\psi(x), \tag{1.6}$$

where $(\partial_\mu - ieA_\mu)$ is the gauge covariant derivative D_μ. The complete gauge invariant Lagrangian for the QED $U(1)$ is then:

$$\mathcal{L}_{QED} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \overline{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) = \mathcal{L}_{EM} + \mathcal{L}_D + \mathcal{L}_{int}, \tag{1.7}$$

where the interaction term $\mathcal{L}_{int} = e\overline{\psi}\gamma^\mu A_\mu\psi$ represents the interaction of the Dirac field with the EM field.

1.3 Electroweak theory

The Electroweak theory (EW) is based on the same principle of gauge invariance as QED and unifies the EM and weak interactions as two different manifestations of the same fundamental interaction: $SU(2)_L \times U(1)_Y$, where $SU(2)_L$ is the weak isospin group which is non-Abelian and has three generators $T_{1,2,3} = \frac{1}{2}\sigma_{1,2,3}$, that are the Pauli matrices, and $U(1)_Y$ is the weak hypercharge group which is Abelian and has one generator: $Y/2$. The electromagnetic interaction is a subgroup of the electroweak group and its generator is a linear combination of the third component of the weak isospin with the weak hypercharge:

$$Q = T_3 + \frac{Y}{2}. \tag{1.8}$$

The fermion spinors ψ are split up into right-handed and left-handed:

$$\psi_R = \frac{1}{2}(1 + \gamma^5)\psi$$
$$\psi_L = \frac{1}{2}(1 - \gamma^5)\psi, \tag{1.9}$$

where γ^5 is proportional to the product of the four Dirac matrices. The weak isospin current couples exclusively to left-handed fermions, implying a non-conservation of Parity in weak interactions. Therefore fermions appear as families with left-handed weak isospin $T = 1/2$ doublets and right-handed $T = 0$ singlets:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, u_R, d_R; \begin{pmatrix} c \\ s \end{pmatrix}_L, c_R, s_R; \begin{pmatrix} t \\ b \end{pmatrix}_L, t_R, b_R;$$

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, e_R; \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \mu_R; \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \tau_R.$$

In the doublets, neutrinos and up-type quarks (u, c, t) have the third component of the weak isospin $T_3 = +1/2$, while the charged leptons and down-type quarks (d, s, b) have $T_3 = -1/2$. Thence, members within a doublet carry the same hypercharge: $Y = -1$ for leptons and $Y = 1/3$ for quarks. The $SU(2)_L \times U(1)_Y$ gauge group requires the existence of four massless gauge bosons, of which two are charged under the EM interaction and two are neutral. However, the short range of the weak force suggests that the corresponding gauge bosons must be massive particles. In order to include these observed masses, a spontaneous EW symmetry breaking mechanism is implemented, giving mass to the three electroweak gauge bosons, $W^\pm$ and Z^0, but not to the photon, leaving the electromagnetic symmetry unbroken.

1.4 Spontaneous symmetry breaking: the Higgs mechanism

The spontaneous symmetry breaking mechanism was introduced in the works of Higgs, Englert, Brout, Guralnik, Hagen and Kibble [8]. Consider a complex scalar field $\varphi(x) = (\varphi_1 + i\varphi_2)$ interacting with a real vector field $A_\mu(x)$ via the following Lagrangian, invariant under a local $U(1)$ transformation:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\varphi)^* D^\mu\varphi - V(|\varphi|^2), \tag{1.10}$$

where $F_{\mu\nu}$ is the field strength tensor, D_μ is the already defined covariant derivative and $V(|\varphi|^2) = -\mu^2|\varphi|^2 + \lambda|\varphi|^4$ is the potential of a scalar field with real constants λ and μ. If λ is negative, V has no stable minima. In case $-\mu^2 > 0$ and $\lambda > 0$, V has a unique stable minimum. When $-\mu^2 < 0$ and $\lambda > 0$, V has two different minima at $|\varphi| = \mu/\sqrt{2\lambda}$, as can be seen in Figure 1.2. The minimum of the potential will be at:

$$\langle\varphi\rangle_0 = \langle 0|\varphi|0\rangle = \varphi_0/\sqrt{2}, \tag{1.11}$$

and the Lagrangian in Equation 1.10 can be expanded around the vacuum state:

$$\varphi(x) = \frac{1}{\sqrt{2}}[\varphi_0 + \chi(x) + i\theta(x)] \tag{1.12}$$

$$\rightarrow \mathcal{L} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \frac{1}{2}(e\varphi_0)^2 B^\mu B_\mu + \frac{1}{2}(\partial_\mu\chi\partial^\mu\chi) - \mu^2\chi^2, \tag{1.13}$$

where $B_\mu = A_\mu - \partial_\mu\theta/(e\varphi_0)$ and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ stands for a simplified version of the weak field strength tensor. This Lagrangian includes a massive vector field B_μ with mass $e\varphi_0$ and a massive scalar field χ with mass $\mu\sqrt{2}$: the vector field has "absorbed" the field $\theta(x)$, becoming massive. The scalar field that appears is the so-called *Higgs field* and the particle associated is the *Higgs boson*. The $U(1)$ gauge symmetry is no more apparent and it is spontaneously broken.
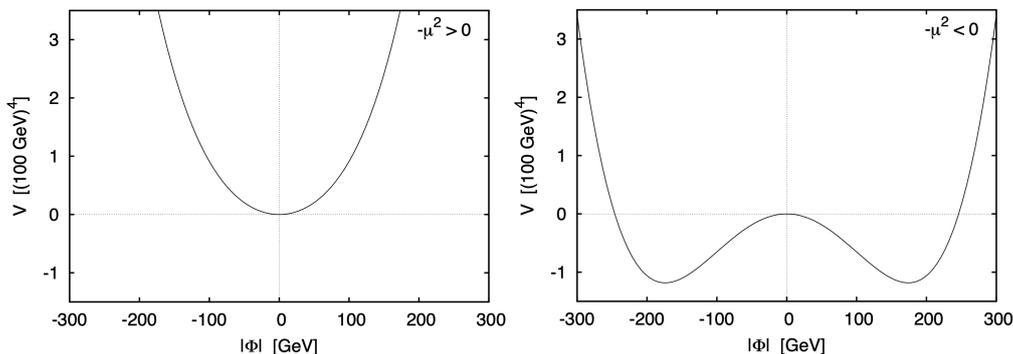
Figure 1.2: Examples of a potential of the type $V(|\varphi|^2) = -\mu^2|\varphi|^2 + \lambda|\varphi|^4$, with $\lambda > 0$ and two different choices for $-\mu^2$ [9].

1.4.1 Gauge boson masses

Such mechanism of spontaneous symmetry breaking is introduced in the EW theory in order to give mass to the electroweak gauge bosons. Before the symmetry breaking the Lagrangian is:

$$\mathcal{L} = -\frac{1}{4}F^a_{\mu\nu}(F^a)^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + (D_\mu\varphi)^\dagger D^\mu\varphi - V(\varphi^\dagger, \varphi), \qquad (1.14)$$

where $F^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g_2\varepsilon^{abc}W_{\mu b}W_{\nu c}$ is the $SU(2)_L$ field strength tensor, being W^a_μ the corresponding gauge field; $B_{\mu\nu}$ is the field strength tensor of $U(1)_Y$, being B_μ the corresponding gauge field; ε^{abc}, with $a, b, c = 1, 2, 3$, are the structure constants of $SU(2)$, and $V(\varphi^\dagger, \varphi) = \lambda(\varphi^\dagger\varphi - \frac{1}{2}v^2)^2$ is the potential of the scalar field, with $v \in \mathbb{R}$. The simplest choice for the scalar field is a doublet $\varphi(x) = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}$ with hypercharge $Y = 1$. The covariant derivative acting on this scalar field is:

$$D_\mu\varphi = (\partial_\mu - ig_2 T_a W^a_\mu - ig_1\frac{Y}{2}B_\mu)\varphi \qquad (1.15)$$

where $T^a = \frac{1}{2}\sigma^a$ are the generators of $SU(2)_L$ with coupling strength g_2, while Y is the generator of $U(1)_Y$ with coupling strength g_1. For the vacuum state a choice can be:

$$W^a_\mu = B_\mu = 0; \langle\varphi\rangle_0 = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \qquad (1.16)$$

Considering small perturbations around the ground state (see Equation 1.12) and making a gauge transformation on this field to move to the unitary gauge, $\varphi(0)$ becomes:

$$\varphi(0) = \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}, \qquad (1.17)$$

with $H(x)$ a scalar field. In this way four fields are introduced (two complex and two real):

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu); Z_\mu = \frac{g_2 W^3_\mu - g_1 B_\mu}{\sqrt{g_1^2 + g_2^2}}; A_\mu = \frac{g_2 W^3_\mu + g_1 B_\mu}{\sqrt{g_1^2 + g_2^2}}; \qquad (1.18)$$

5

chosen in order to satisfy $(W_\mu^-)^* = W_\mu^+$ and $Z_\mu^2 + A_\mu^2 = (W_\mu^3)^2 + B_\mu^2$. The covariant derivative of Equation 1.15 can be written in terms of these fields, so then the Lagrangian of the EW model can be rewritten after the spontaneously symmetry breaking as:

$$
\begin{aligned}
\mathcal{L} = &-\frac{1}{4}F_{\mu\nu}^a(F^a)^{\mu\nu} - \frac{1}{2}\mathcal{W}_{\mu\nu}^+\mathcal{W}^{-,\mu\nu} - \frac{1}{4}\mathcal{Z}_{\mu\nu}\mathcal{Z}^{\mu\nu} \\
&+ m_W^2 W_\mu^+ W^{-,\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu + \frac{1}{2}(\partial_\mu H \partial^\mu H) - \frac{1}{2}m_H^2 H^2,
\end{aligned}
\tag{1.19}
$$

where $\mathcal{W}_{\mu\nu}^\pm = \partial_\mu W_\nu^\pm - \partial_\nu W_\mu^\pm$ and $\mathcal{Z}_{\mu\nu} = \partial_\mu Z_\nu - \partial_\nu Z_\mu$. Such Lagrangian describes the massless EM field A_μ, the massive $W_\mu^\pm$ fields associated to the W bosons with mass $m_W = \frac{g_2 v}{2}$, the massive Z_μ field associated to the Z boson with mass $m_Z = \frac{v\sqrt{g_1^2 + g_2^2}}{2}$, and the massive H field associated to the Higgs boson with mass $m_H = v\sqrt{2\lambda}$. Since the Higgs boson mass depends from the free parameter λ (the Higgs self-coupling), it can not be predicted by the theory. Instead, the masses of the W and Z bosons can be connected one to each other via the weak mixing angle:

$$
\begin{aligned}
\cos\theta_W &= \frac{g_2}{\sqrt{g_1^2 + g_2^2}}; \\
m_Z &= \frac{M_W}{\cos\theta_W}.
\end{aligned}
\tag{1.20}
$$

1.4.2 Fermion masses

The fermion masses can be generated via the *Yukawa interaction* which describes the coupling between the Higgs field and the massless fermion fields. For example, consider the coupling of the electron, which involves the left-hand electron doublet $L_L^e = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$, the Higgs doublet φ and the right-component of the electron field e_R. In this case the effective Yukawa Lagrangian is:

$$
\mathcal{L}_{Yu}^e = -g_e \left(\bar{L}_L^e \varphi e_R + \bar{e}_R \varphi^\dagger L_L^e \right).
\tag{1.21}
$$

Considering then a spontaneously broken symmetry and small perturbations around the vacuum state, such Lagrangian becomes:

$$
\mathcal{L}_{Yu}^e = -\frac{g_e v}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) - \frac{g_e}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L)H = -m_e(\bar{e}e) - \frac{g_e}{\sqrt{2}}(\bar{e}e)H, \tag{1.22}
$$

where the first term contains the electron mass $(m_e = g_e v/\sqrt{2})$ and the second one describes the interaction between the fields. Including also the quarks interaction, the Yukawa Lagrangian is:

$$
\mathcal{L}_{Yu}^e = -\Gamma_u^{ij}\bar{Q}_L^i(i\sigma_2)\varphi^* u_R^j - \Gamma_d^{ij}\bar{Q}_L^i\varphi d_R^j - \Gamma_e^{ij}\bar{L}_L^i\varphi e_R^j + [h.c.],
\tag{1.23}
$$

where $[h.c.]$ means hermitian conjugate and $\Gamma_{u,d,e}$ are 3×3 complex matrices (Yukawa matrices). They are not necessarily diagonal, so that in general a mixing between different fermion generations is allowed. After the Higgs field acquires a vacuum expectation value, the fermions become massive:

$$
m_f = \frac{g_f v}{\sqrt{2}},
\tag{1.24}
$$

6

where g_f is the Yukawa coupling, obtained by diagonalizing the matrices Γ. It describes at the same time the mass of the fermion and its coupling with the Higgs boson. The mass eigenstates for the left-handed doublets do not coincide with the eigenstates of the weak interaction of quarks; in fact, the quark fields can form mixed combinations. The transformation of the mass eigenstates (d, s, and b) into weak eigenstates (d', s' and b') is regulated by the CKM (Cabibbo-Kobayashi-Maskawa) matrix [10]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

a 3×3 unitary matrix, that can be parameterised by three mixing angles and a CP-violating phase. The transition probability between different quark flavours is described by the matrix elements, and it is proportional to $|V_{qq'}|^2$. Currently, the best determination of the magnitudes of the CKM matrix elements is [7]:

$$V_{CKM} \simeq \begin{pmatrix} 0.97401 & 0.22650 & 0.00361 \\ 0.22636 & 0.97320 & 0.04053 \\ 0.00854 & 0.03978 & 0.999172 \end{pmatrix}$$

and so the CKM matrix is not diagonal, although the off-diagonal elements are suppressed. By convention, the mixing takes place only between down-type quarks, while the up-type mass matrix is diagonal. In the SM, neutrinos are assumed to be massless, and so the same mixing for leptons does not happen. The experimental evidence of massive neutrinos has risen the need of theories beyond the Standard Model, introducing right-handed neutrinos.

1.5 Quantum Chromodynamics

The last piece of the Standard Model is the interaction between quarks and gluons, a gauge theory called Quantum Chromodynamics (QCD), based on the unbroken colour group $SU(3)_C$ [11]-[12]. The colour symmetry implies eight massless gauge bosons called gluons (related to the generators of $SU(3)_C$), which mediate the strong interaction and couple only to colour charges. Quantum Chromodynamic has a non-Abelian structure, and therefore the gluons also carry colour charge themselves and can couple to each other. The Deep Inelastic Scattering (DIS) experiments involving leptons and nuclei gave the first confirmation of the existence of a sub-structure in nucleons. In particular, some aspects were underlined: the quarks are fermions with 1/2 spin and carry colour charge and fractional electric charge, gluons exist and do not interact with electromagnetic or weak force. The QCD Lagrangian is:

$$\begin{aligned} \mathcal{L}_{QCD} &= -\frac{1}{4} G_{\mu\nu}^a (G^a)^{\mu\nu} + \sum_k^{n_f} \bar{\psi}^k (i\gamma^\mu D_\mu - m_k)\psi^k \\ &= -\frac{1}{4} G_{\mu\nu}^a (G^a)^{\mu\nu} + \sum_k^{n_f} \bar{\psi}^k (i\gamma^\mu \partial_\mu - m_k)\psi^k + g_s \sum_k^{n_f} \bar{\psi}^k (\gamma^\mu T_a G_\mu^a)\psi^k \\ &= \mathcal{L}_g + \mathcal{L}_q + \mathcal{L}_{int}, \end{aligned} \qquad (1.25)$$

where a ranges from 1 to 8, D_μ is the covariant derivative and $G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c$ indicates the field strength tensor. The $SU(3)$ structure constant is f^{abc}, $T^a = \frac{1}{2}\lambda^a$ are the group generators in terms of the Gell-Mann matrices λ^a, and n_f is the number of independent quark fields (the six flavours). The Lagrangian can be written as well as the sum of three pieces: $\mathcal{L}_g$ is the kinetic term for the gluon massless field, $\mathcal{L}_q$ the kinetic term for the massive quark field, and $\mathcal{L}_{int}$ is the interaction between gluons and quarks. Finally, g_s (or $\alpha_s = g_s^2/4\pi$) is the strong coupling constant and its size depends on the energy scale of the interaction (Q^2): it is small for large momentum transfers (*asymptotic freedom* [13]), and large for small momentum transfers (*colour confinement*). At high energy, quarks can be described as almost free particles, while the colour confinement at small energy binds quarks together, not allowing the observation of free quarks and gluons.

Considering two colliding protons, their structures can be described as a set of some point-like components (quarks and gluons), each carrying a fraction of the total proton momentum. The production cross section can be seen as the product of two terms: one at large momentum transfers (short distance) and one at small momentum transfers (long distance). The former can be analysed in the perturbative QCD frame, while for the latter it is necessary to introduce a parametrisation through the Parton Distribution Functions (PDFs). For a proton A with a momentum p_A, the PDF is a function $f_{a/A}(x, Q^2)$ of the relative momentum ($x = \frac{p_a}{p_A}$) of the parton in the direction of the proton momentum, and depends also on the energy scale of the process Q. The proton-proton collision cross section is therefore:

$$\sigma_{p_A p_B \to n} = \sum_q \int \mathrm{d}x_a \mathrm{d}x_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \sigma_{ab \to n}, \qquad (1.26)$$

where A and B denote the two protons. It is possible to rewrite the cross section in terms of the strong coupling:

$$\sigma_{p_A p_B \to n} = \sum_q \int \mathrm{d}x_a \mathrm{d}x_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) [\sigma_0 + \alpha_s(Q^2)\sigma_1 + \cdots], \qquad (1.27)$$

where σ_0 is the tree-level parton-parton cross section, σ_1 is the first order QCD correction to the parton-parton cross section, and so on. In order to have a realistic prediction, higher orders of QCD corrections have to be employed. As stated above, the partons can not be detected because of colour confinement: every parton generated in the interaction undergoes a hadronisation process that produces a jet of baryons and mesons which can be detected when they interact with matter. The PDF is the probability density for finding a parton with a momentum fraction x_a of the proton momentum at a given factorisation scale μ_F. Their determination is important in order to better understand the initial state of the colliding particles. The PDFs can not be directly calculated, and must be determined using data from different experiments: deep inelastic scattering, Drell-Yan events and jets production. The fits to various sets of experimental data are performed within the evolution scheme,

$$\frac{\partial f(x, Q^2)}{\partial \log(Q^2)} = \frac{\alpha_s}{4\pi} \int_x^1 \frac{\mathrm{d}z}{z} [P_{aa'} f(\frac{x}{z}, Q^2)], \qquad (1.28)$$

where $P_{aa'}$ are the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi splitting functions [14] and give the probability to have a transformation of the parton a with momentum

x into parton a' with momentum z, as a consequence of the emission of one or more quarks or gluons. So, the PDFs measured at one scale can be used to predict the results of experiments at other scales using Equation 1.28.

1.6   Beyond the Standard Model

The SM is a unitary and renormalisable theory, that can be used to perturbatively calculate quantities describing processes at high energies. It incorporates 18 free parameters:

- 9 Yukawa couplings for the fermion masses (six quark and three leptons);

- 4 parameters for the CKM matrix (three mixing angles and a CP-violating phase);

- 2 parameters from EW symmetry breaking (the vacuum expectation value v and the Higgs boson mass m_H);

- 3 couplings g_1, g_2 and g_s for the SM gauge groups $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$, respectively.

At the currently accessible energy scales, the SM describes successfully the particle physics processes, and its predictions have been verified with very high precision (at SPS, LEP, SLAC, HERA, Tevatron and LHC accelerators). However, the Standard Model does not explain some observations:

- the SM does not propose any proper candidate for the presence of Dark Matter in the universe. Cosmological observations reveal that only 4% of the energy in the universe is bright matter, that can be described within the SM. About 27% should be Dark Matter, a type of matter that interacts only gravitationally but not electromagnetically. Possible Dark Matter candidates are primordial black holes, supersymmetric particles and massive neutrinos (weakly interacting massive particles - WIMPs);

- neutrinos are treated as massless particles and their flavour mixing is not considered;

- the prevalence of matter over antimatter. In fact the SM predicts an almost equal amount for matter and antimatter, but there is obviousness of the opposite;

- the gravitational force is not included in the SM, although an explanation could be the graviton, an hypothetical particle, not discovered yet, that should mediate the gravitational force.

There is thence the need of new theories going beyond the Standard Model (BSM) framework. Some attempts are done in the Supersymmetry (SUSY) and in the Grand Unified Theory (GUT) frames, but there is still no experimental evidence sustaining these theories.

1.7 The top quark

With a mass of 172.76 ± 0.30 GeV [7], the top quark is the heaviest elementary particle in the SM. It is the $Q = 2/3$, $T_3 = 1/2$ component of the weak-isospin doublet containing the bottom quark. It was discovered in 1995 by the CDF and D0 Collaborations at Fermilab ([15]-[16]), almost 20 years after the b quark. Due to the large mass, close to the scale of the EW symmetry breaking, it is also characterised by a Yukawa coupling $g_t \approx 1$, playing therefore a special role in the SM and in many BSM theories. Being heavier than a W boson, the top quark is the only one that can decay weakly (actually, it is its almost exclusively decay): $t \to W + b$. It has a lifetime of about $5 \cdot 10^{-25}$ s, so it can not form hadron states before decaying.

1.7.1 Top quark production

At hadron colliders, top quarks are produced predominantly in $t\bar{t}$ pairs, through *gluon-gluon fusion* ($gg \to t\bar{t}$) or *quark-antiquark annihilation* ($q\bar{q} \to t\bar{t}$), at the leading order (LO) in QCD. The Feynman diagrams of these processes are shown in Figure 1.3. The production ratios depend from both the energy and the type of the collision: at Tevatron (proton-antiproton at $\sqrt{s} = 1.96$ TeV) the leading process was $q\bar{q}$ annihilation ($\sim 85\%$), while at LHC (proton-proton at $\sqrt{s} = 13$ TeV) gg fusion dominates the production ($\sim 90\%$). Predictions for the top quark production total cross sections are available at next-to-next-to-leading order (NNLO), also including next-to-next-to-leading-logarithmic (NNLL) soft gluon resummation: at LHC, assuming a top quark mass of 172.5 GeV, the cross section is $\sigma_{t\bar{t}} = 831.8^{+19.8+35.1}_{-29.2-35.1}$ pb ($\sqrt{s} = 13$ TeV), where the first uncertainty is from scale dependence and the second from PDFs [7]. Figure 1.4 shows the current status of $\sigma_{t\bar{t}}$ measurements, compared with the theoretical prediction. The top quark can also be produced in different ways, but with smaller cross sections: for example, it can be produced singly via the weak interaction involving the Wtb vertex, or in a system of four top quarks ($t\bar{t}t\bar{t}$).
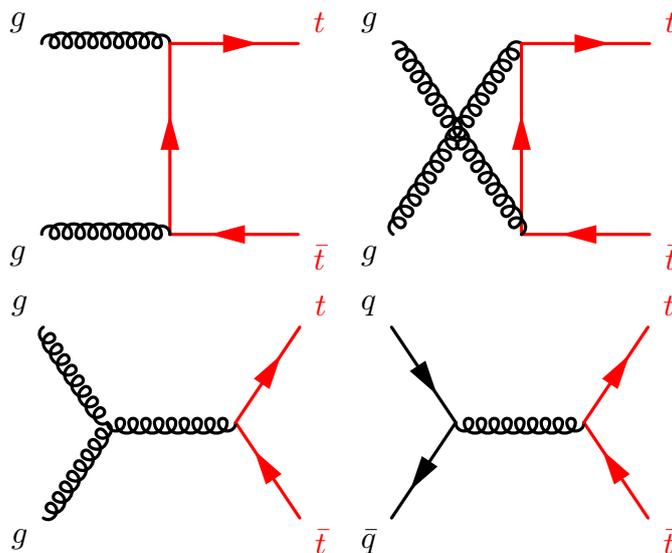
Figure 1.3: Leading order Feynman diagrams for $t\bar{t}$ production in case of gg fusion and $q\bar{q}$ annihilation.
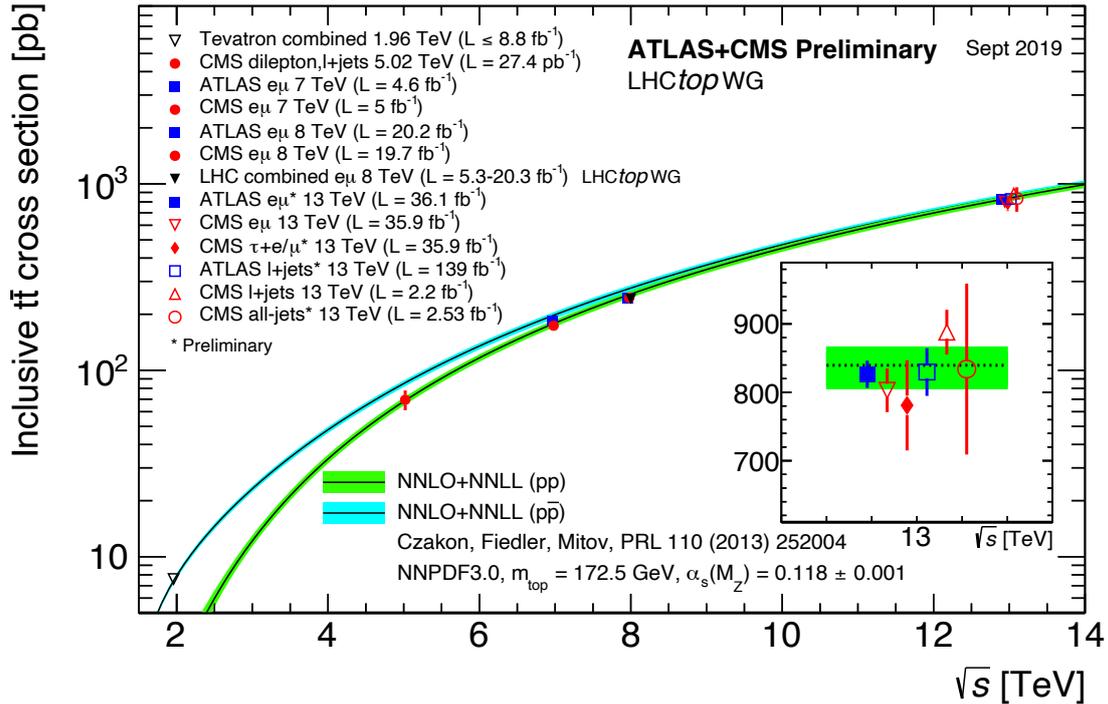
Figure 1.4: Summary of LHC and Tevatron measurements of the $t\bar{t}$ production cross section as a function of the centre of mass energy compared to theoretical calculations [17].

1.7.2 Top quark decay

As already discussed, the top quark does not form hadron states and decays into a b quark and a W boson. While the former hadronises producing a jet of particles, the latter can decay into a lepton and a neutrino ($\sim 1/3$ of the cases), with all the three lepton flavours being produced at equal rate, and in the remaining cases decays into a quark-antiquark pair, which hadronise producing jets. According to the number of leptons in the decay, the $t\bar{t}$ pair can be identified from three different signatures (see Figure 1.5):

- the *all hadronic* (or all jets) channel occurs when both the W bosons decay into a $q\bar{q}$ pair and corresponds to $\sim 45.7\%$ of the $t\bar{t}$ decays. There are six jets overall (at the LO), two from the b quarks and four from the light quarks coming from the W boson decay;

- the *single lepton* (or ℓ+jets) channel occurs when only one W boson decays leptonically. It is characterised by one isolated lepton, a neutrino and four jets. It corresponds to $\sim 43.8\%$ of the total decays;

- The *dilepton* channel occurs when both the W bosons decay into a lepton-neutrino pair. There are two leptons (with opposite charges), two neutrinos and two jets from b quarks. It makes the $\sim 10.5\%$ of the cases.
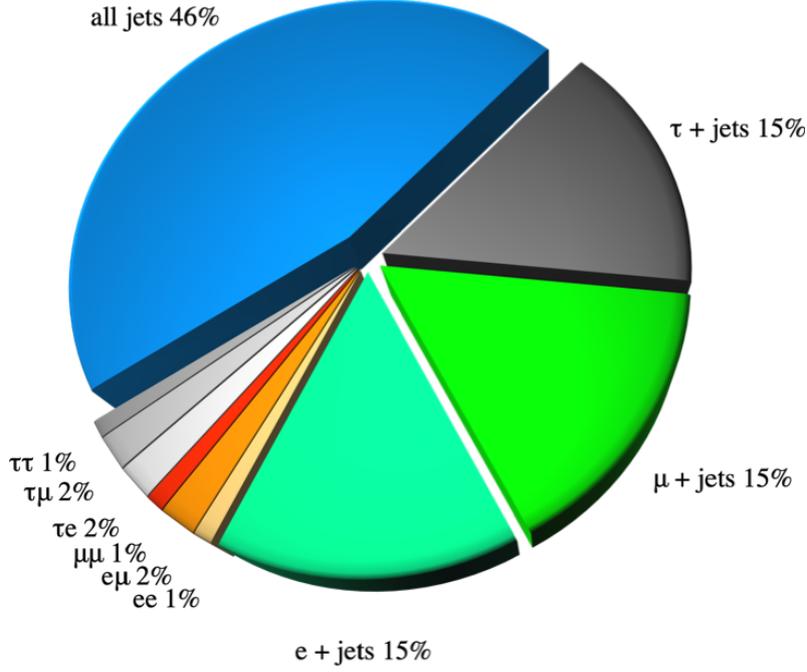
Figure 1.5: Top quark pair branching fractions.

1.7.3 Four tops production

Besides having very large coupling to the Higgs boson, the top quark is predicted to have large couplings to new particles and so it plays a special role in many hypothetical BSM models. Because of the large top quark mass, the production of $t\bar{t}t\bar{t}$ is one of the most difficult mechanisms accessible at the LHC. It is a rare process predicted by the SM and it has not been measured up to now. Many BSM scenarios lead to an enhancement of the $t\bar{t}t\bar{t}$ cross section [18]. In fact, the SM four top quark production cross section at NLO accuracy in QCD and EW couplings at $\sqrt{s} = 13$ TeV is predicted to be $\sigma_{SM}^{t\bar{t}t\bar{t}} = 11.97^{+18\%}_{-21\%}$ fb [19], and possible new phenomena BSM may enhance the cross section over the SM prediction. At LHC, at the current energy, the dominant mechanism for the production of $t\bar{t}t\bar{t}$ is gg fusion with a fraction of the total rate of 94%, followed by the $q\bar{q}$ annihilation with a 6% fraction: Figure 1.6 shows the representative tree-level Feynman diagrams for both mechanisms. As in the case of $t\bar{t}$, the final state topology of a $t\bar{t}t\bar{t}$ event is determined by the decays of each of the W bosons. In total, there are 35 possible final states[1] which can be grouped into five classes of channels (see Figure 1.7):

- the all hadronic (full hadronic or all jets) channel occurs when all the W bosons decay into a $q\bar{q}$ pair. This implies the presence of 12 jets, out of which four come from the b quarks. The branching ratio (BR) of this channel is the 31.1%;

- the single lepton (ℓ+jets or 1L) channel occurs when only one W boson decays leptonically. The final state is characterised by one charged lepton, 10 jets, out of which four come from the b quarks, and missing energy due to the neutrino (which can not be detected). It is the channel with the largest BR, 42.2%;

---

[1]Here leptons denote either electrons or muons, which can include those that have come from a tau decay.
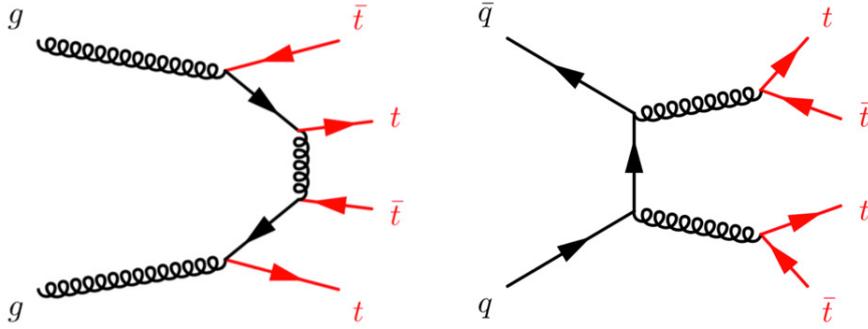
Figure 1.6: Leading order Feynman diagrams for the $t\bar{t}t\bar{t}$ production in case of gg fusion and $q\bar{q}$ annihilation.

- the dilepton opposite-sign (or 2LOS) channel occurs when two W bosons with different charges decay leptonically. The channel is characterised by two opposite-sign charged leptons, eight jets, out of which four come from the b quarks, and missing energy due to the neutrinos. It represents the 14.3% of the decays;

- the dilepton same-sign (or 2LSS) channel occurs when two W bosons with the same charges decay leptonically. The final state is composed by two same-sign charged leptons, eight jets, out of which four come from the b quarks, and missing energy due to the neutrinos. The BR is 7.2%;

- the multi-lepton (or ML) channel occurs when at least three W bosons decay leptonically. In this case there are four jets from the b quarks, missing energy from the neutrinos, and three leptons with two additional jets or four leptons without additional jets. The BR is 5.3% (4.9% for $\ell\ell\ell$ and 0.4% for $\ell\ell\ell\ell$).

The all hadronic and the 1L channels benefit form the largest BRs, but are penalised by massive backgrounds, like $t\bar{t}$ production associated with jets. Instead, the ML channel has a very clean signature, but a very small BR. In general, the searches for $t\bar{t}t\bar{t}$ production avoid the all hadronic channel because of the background and group the other final states into two bigger channels: single lepton with dilepton opposite-sign (1LOS) and dilepton same-sign with multilepton (SSML). Previous searches for $t\bar{t}t\bar{t}$ production using LHC data at $\sqrt{s} = 13$ TeV were performed by both the ATLAS ([20]) and CMS Collaborations ([21]), setting limits on the cross section, but not yet measuring the process.

In the single lepton channel the main and most important background source is the $t\bar{t}$ process associated with jets ($t\bar{t}$+jets). This background is irreducible and, moreover, the current predictions do not provide a good description of data in such extreme radiation regions. So a dedicated procedure is needed to cope this problem. Other important backgrounds are the single top quark production and the W/Z bosons production associated with jets. Smaller contributions come from the associated production of a gauge boson (W, Z or H) with $t\bar{t}$ and from diboson production. For these backgrounds, the simple Monte Carlo simulation is enough to have good predictions.

13
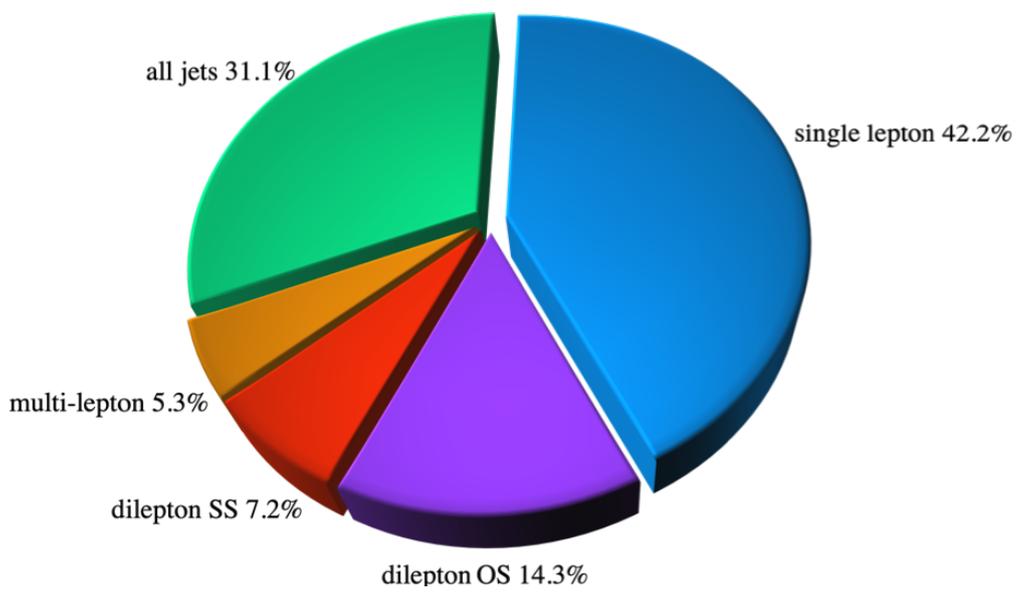
Figure 1.7: Branching ratio for the decays of $t\bar{t}t\bar{t}$.

Chapter 2

LHC and the ATLAS experiment

The search for experimental confirmations to the Standard Model brought to the development of new technologies and experimental setups. As precision tests were performed, providing more stringent constraints on the SM parameters, new energy regimes had to be explored. The Large Hadron Collider (LHC) was built in order to address this need. LHC is a proton-proton collider started up in 2008 as the latest addition to CERN's accelerator complex. The accelerator characteristics will be summarised in the first part of this chapter, while the second part is focussed on the description of the ATLAS experiment.

2.1   The Large Hadron Collider

The LHC [22] is the world's largest and most powerful particle accelerator. It was built between 1998 and 2008 near Geneva, between Switzerland and France. It has been installed in the underground tunnel which previously housed the Large Electron-Positron Collider (LEP), a 26.7 km ring at an average depth of 100 m. Being a hadronic collider, LHC can reach higher energies in the centre of mass frame with respect to LEP, since energy losses due to the synchrotron radiation are smaller. It can accelerate two proton beams, circulating in opposite directions, at an energy of 14 TeV in the centre of mass (currently 13 TeV); it can also accelerate heavy ion (lead) beams at a centre of mass energy of $\sqrt{s} = 5.5$ TeV. The beams circulate in two different pipes in a regime of ultra high vacuum (10^{-7} Pa). They are bent by NbTi superconducting magnets which produce a magnetic field above 8 T, cooled with superfluid helium to a temperature of -271.3 °C. Along the LHC ring there are four points of collision, where the following experiments are located (see Figure 2.1):

- ATLAS (A Toroidal LHC ApparatuS) [23];

- CMS (Compact Muon Solenoid) [24];

- ALICE (A Large Ion Collider Experiment) [25];

- LHCb (Large Hadron Collider beauty) [26].

Three further experiments are placed nearby the interaction points: LHCf (Large Hadron Collider forward) [27], TOTEM (TOTal Elastic and diffractive cross section Measurement) [28] and, from 2010, MoEDAL (MOnopole and Exotical Detector At
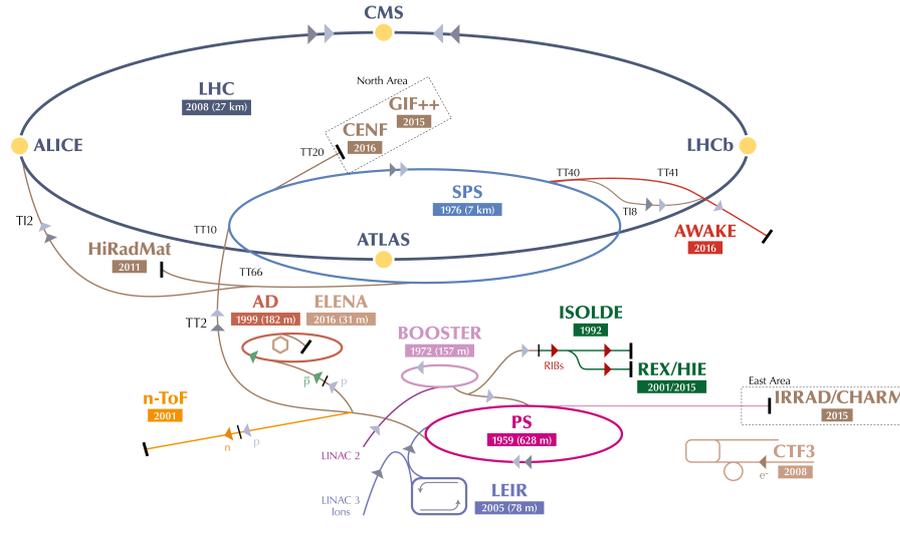
Figure 2.1: The CERN accelerators complex with the four main LHC experiments. Some lines are used to extract particles for non-LHC experiments [30].

the LHC) [29]. The principal aim of these experiments is the exploration of the high energy physics frontier:

- ATLAS and CMS have been designed to study as many as possible events produced in proton-proton collisions. They are both multifunctional detectors to mainly investigate the SM physics (discover the Higgs boson) and search for possible physics beyond the Standard Model (SUSY in particular);

- ALICE studies the strong interactions between heavy ions in a high energy density regime (when matter is in a quark-gluon plasma state). The physics programme also includes collisions with lighter ions and at lower energy and the collection of data from proton-proton collision to provide references for the heavy-ion programme;

- LHCb has been built to examine the heavy flavour physics, in particular to search for new sources of CP violation beyond the SM in hadrons with bottom and charm quarks;

- LHCf is close to the ATLAS detector and its goal is a better understanding of cosmic rays. It makes use of the particles produced in collisions to simulate cosmic rays in laboratory conditions;

- TOTEM is placed near CMS and its aim is to increase our knowledge on protons, measuring those that emerge from the collision region with such a small angle that the other experiments can not reveal them;

- MOEDAL is the latest experiment built (near LHCb) and it searches for magnetic monopoles.

16

2.1.1 Accelerator structure

As shown in Figure 2.1, the proton beam acceleration is performed in separated stages. The protons are produced by stripping electrons from a source of hydrogen gas, extracted with an energy of 91 keV and then injected in the LINAC2. This is a 36 m long linear accelerator which, using radio frequency (RF) quadrupoles and focusing quadrupole magnets, brings protons to the energy of 50 MeV. During the acceleration, protons are split in bunches interspersed by 25 ns (corresponding to a frequency of 40 MHz) using RF cavities. Each bunch contains about $1.1 \cdot 10^{11}$ particles. The beam is then brought to the Proton Synchrotron Booster (PSB), a 157 m long circular accelerator, in which the beam energy is increased up to 1.4 GeV. Then, the protons reach the Proton Synchrotron (PS), a 628 m ring, which raises the protons energy up to 25 GeV. Before entering in the LHC ring, the beam is further accelerated by the Super Proton Synchrotron (SPS, a 6.9 km circumference synchrotron) to the energy of 450 GeV. In the last stage of acceleration, inside the LHC, eight radio frequency cavities bring both the proton beams to the present energy of 6.5 TeV. Ions are instead produced in an ion-source, go through LINAC3 and the Low Energy Ion Ring (LEIR) before following the same path as the protons from the PS.

Inside LHC, particles are constrained into a circular orbit by 1232 superconducting magnetic dipoles (Figure 2.2), producing a 8.36 T magnetic field transverse to the beam direction. As stated before, the dipoles are cooled to a temperature of 1.9 K by a cryogenic system which exploits superfluid ^{3}He. The generated field is optimised in order to constrain both the beams, that are accelerated together in two different, but nearby, pipes. Near the collision points, where the experiments are located, the beams are further focused and collimated by a set of quadrupolar, sextupolar and octupolar magnets.
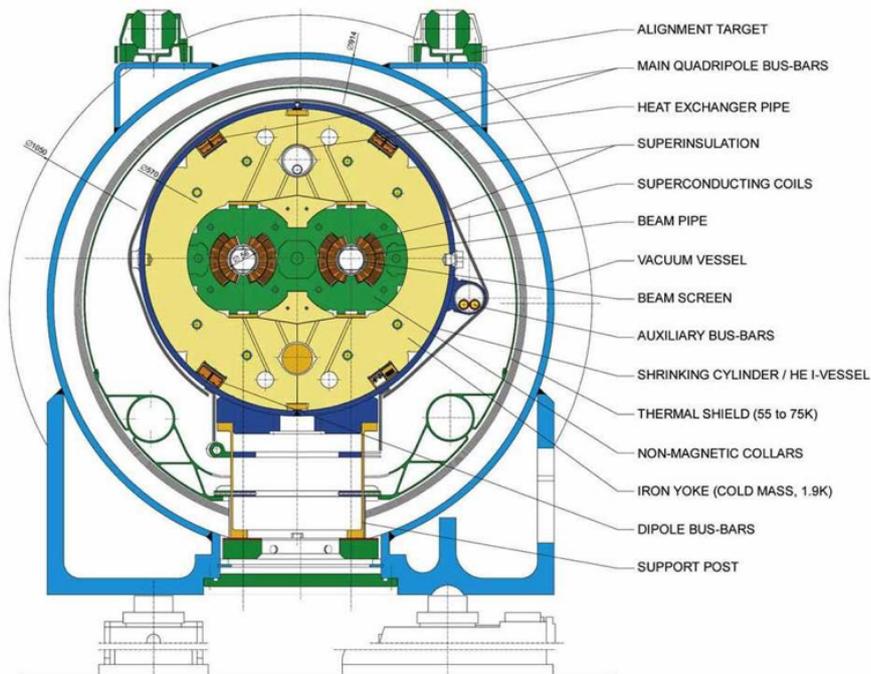
Figure 2.2: Section of one of the magnetic dipoles used to bend and stabilise the beam orbit [22].

For a given process, the number of events generated per second (N_e) at the LHC is simply:

$$N_e = L\sigma_e, \tag{2.1}$$

where L is the instantaneous luminosity (number of collisions per unit of time and transverse section of the beams) and σ_e is the cross section for the process. The luminosity depends only on beam parameters:

$$L = \frac{N_b^2 n_b f_r \gamma}{4\pi \epsilon_n \beta^*} F, \tag{2.2}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_r the revolution frequency, γ the relativistic factor, ϵ_n the normalised transverse beam emittance, which describes the dispersion of the protons in the phase space, β^* the beta function at the collision point, which describes the oscillation of the bunches around the ideal orbit, and F the geometric luminosity reduction factor due to the crossing angle at the interaction point. The peak of instantaneous luminosity delivered to ATLAS during proton-proton collisions was reached in 2018: $21.0 \cdot 10^{33}$ cm^{-2}s^{-1}, twice the designed luminosity.

The total number of interactions in a data acquisition is given by the integrate luminosity, which depends on the instantaneous luminosity and the acquisition time. The total integrated luminosity collected by ATLAS during the 13 TeV collisions period (2015-2018, the so-called *Run II*) was 139 fb^{-1} (see Figure 2.3).
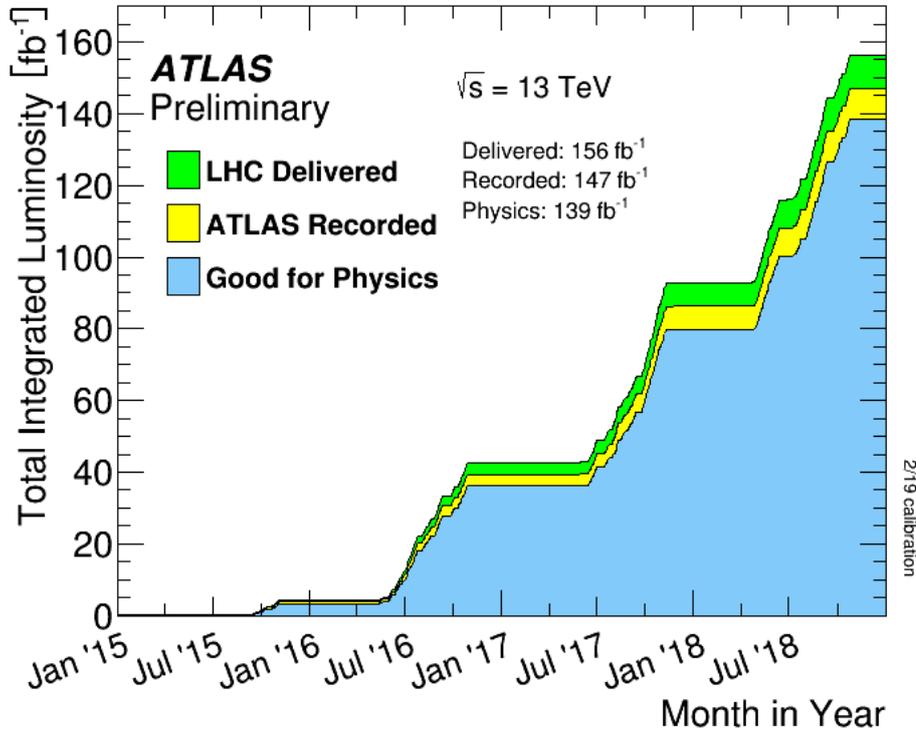
Figure 2.3: Integrated luminosity delivered to ATLAS (green), recorded (yellow), and certified to be good quality data (blue) during 13 TeV collisions in the period 2015-2018 [31].

2.1.2 Coordinate system

The coordinate system adopted by ATLAS, and by LHC in general, is a right-handed Cartesian system where the nominal interaction point is the origin of the system. The x-axis is defined as perpendicular to the beam in the origin and the positive direction points to the centre of LHC. The y-axis is also perpendicular to the beam and points upwards: the resulting $x - y$ plane is the so-called *transverse plane*. Finally, the z-axis is parallel to the beam direction.

The cylindrical structure of the ATLAS detector suggests to introduce some quantities useful to better define the event:

- $\phi = \cot\left(\frac{x}{y}\right)$ is the azimuthal angle measured around the z-axis;

- $\theta = \arctan\left(\frac{y}{z}\right)$ is the polar angle measured from the z-axis;

- $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$ is the rapidity, where E is the particle's energy and p_z is the component of its momentum projected along the beam axis;

- the rapidity can be approximated in case of highly relativistic particles ($m \ll p$) with the pseudorapidity $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$, a quantity that is still invariant under Lorentz boosts along the axis direction. The pseudorapidity is zero for $\theta = \frac{\pi}{2}$, grows up approaching the beam pipe, and tends to infinity for $\theta = 0$, as shown in Figure 2.4;

- $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is the distance in the $\eta - \phi$ space;

- $p_T = p\sin\theta$ is the transverse momentum of the particle;

- $E_T = E\sin\theta$ is the transverse energy of the particle.

In particular, p_T and E_T are more interesting variables than p and E, since for the colliding partons in the initial state it can be assumed that the transverse component is negligible compared to the longitudinal component and so $p_T^{init} = E_T^{init} = 0$, which brings a constrain on the total transverse momentum and energy of the final states.
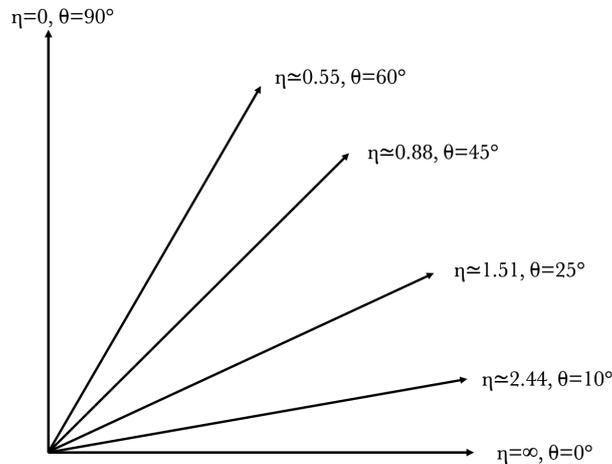
Figure 2.4: Variation of the pseudorapidity as a function of θ.
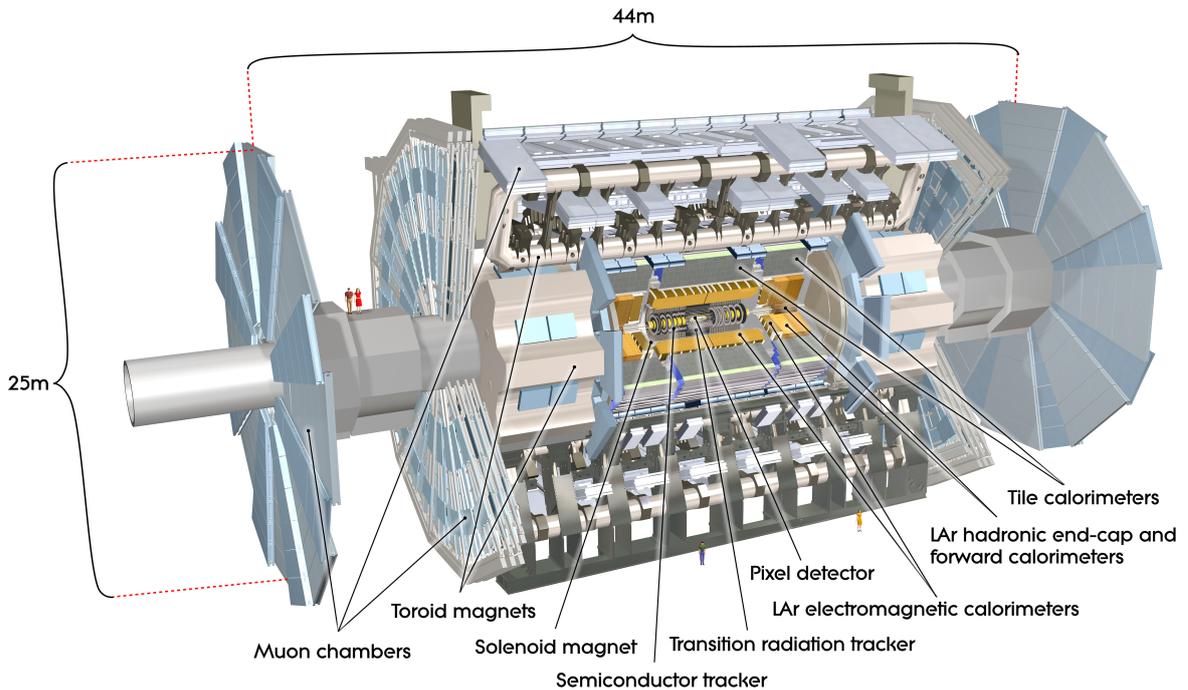
Figure 2.5: Digital view of the ATLAS detector with the indications of the various sub-detectors [23].

2.2 The ATLAS detector

The ambitious physics program of LHC required to have new and sophisticated detectors able to probe the proton-proton and ion-ion collisions. Two general purpose detector, ATLAS and CMS, were built to cope the request and have two different experiments investigating the Standard Model and Beyond physics. In order to reach the goals they had to satisfy some general requirements:

- fast and radiation-hard components and electronics;

- high detector granularity;

- large acceptance in pseudorapidity and angular coverage;

- good charged-particle momentum resolution and reconstruction efficiency;

- very good electromagnetic calorimeter for electron and photon identification and measurements;

- very good hadronic calorimeter for jets and missing transverse energy measurements;

- good muon identification and measurements;

- highly efficient trigger to reject background events and achieve an acceptable level of data stored.

The ATLAS detector is shown in Figure 2.5. With the dimensions of 25 m in height and 44 m in length and an overall weight of about 7000 tonnes, ATLAS is the biggest detector at LHC. It has a cylindrical structure, with a central part called barrel, closed by two end-caps: in this way the detector surrounds (almost) entirely the interaction point.

As many other particle detectors, the structure of ATLAS is arranged in different layers of subsystems, each of them with a specific target (see Figure 2.6):

- the Inner Detector (ID) is the closest system to the interaction point and provides the reconstruction of the trajectory of the charged particles. Being embedded in a 2 T solenoidal magnetic field, the ID measures also the momentum and the charge of those particles;

- the Electromagnetic Calorimeter (ECAL) surrounds the ID and its aim is the measure of the energy of photons and electrons;

- the Hadronic Calorimeter (HCAL) envelops the ECAL and it is designed to detect the energy of hadronic particles;

- the Muon Spectrometer (MS) is embedded in a toroidal magnetic field and it reconstructs the momentum, the charge and the trajectory of muons.

In addition to this layer structure, three smaller detectors are placed in the forward regions, in order to determine the luminosity delivered to ATLAS. A three-level Trigger and Data Acquisition (TDAQ) system is implemented in order to read out data from over 100 million electronic channels at a rate of approximately 40 MHz and save physics events with a rate of about 1 kHz. The general performance goals are summarised in Table 2.1. In the following sections, each detector will be described in details.
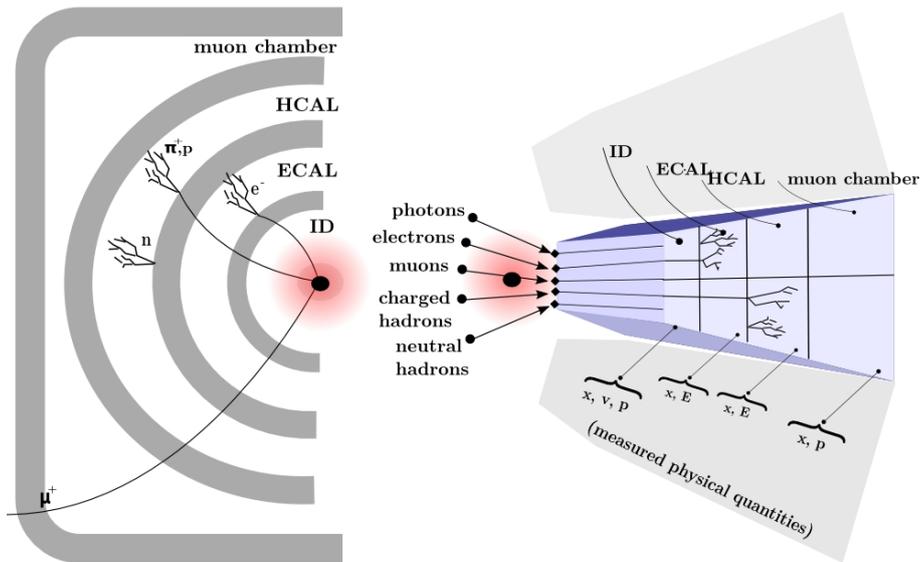
Figure 2.6: Schematic representation of a typical high energy physics detector. It is shown also how particles interact in the various subsystems.

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% \, p_T \oplus 1\%$	± 2.5	
EM calorimetry	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	± 3.2	± 2.5
Hadronic calorimeter(jets)			
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	± 3.2	± 3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 3.1\%$	$3.1 < |\eta| < 4.9$	$3.1 < |\eta| < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	± 2.7	± 2.4

Table 2.1: General performance goals of the ATLAS detector [23].

2.2.1 The Inner Detector

The ID is the innermost piece of ATLAS detector and it consists in three independent but complementary subsystems: with reference to Figure 2.7, there are two silicon detectors, the Pixel Detector and the SemiConductor Tracker (SCT), surrounded by a gaseous detector, the Transition Radiation Tracker (TRT). The overall dimensions are 6.2 m in length and 2.1 m in diameter, and it is embedded in a magnetic field of 2 T generated by the Central Solenoid (CS). The main task of the ID is the measurement of the tracks produced by the charged particles: thanks to the magnetic field and the high precision of track reconstruction, it provides the position, the transverse momentum and the charge of the particles and the positions of primary and secondary vertices.

The Pixel Detector is the closest subsystem to the interaction point, being built around the beryllium beam pipe. The basic element is the so-called *pixel*: a silicon sensor with a surface of 50×400 μm^2 and a thickness of 250 μm. A module contains 47232 pixels for a total of $\sim 80.4 \cdot 10^6$ pixels in 1744 modules, arranged in three barrel layers and two end-caps, each with three disk layers. The layers in the barrel are placed at radii 50.5 mm ($\sim 13.2 \cdot 10^6$ pixels), 88.5 mm ($\sim 22.8 \cdot 10^6$ pixels) and 122.5 mm ($\sim 31.2 \cdot 10^6$ pixels). Each disk in the end-caps contains $\sim 2.2 \cdot 10^6$ pixels and they are placed at $|z| = 495$ mm, 580 mm and 650 mm from the centre. In this way the Pixel Detector provides three measurement points for tracks in the $|\eta| < 2.5$ region with a spatial resolution of 10 μm $\times$ 115 μm. In 2014, before Run II, a fourth layer was installed at a radius of 25.7 mm in the barrel region: the Insertable B-Layer (IBL). It adds $\sim 6.02 \cdot 10^6$ pixels, improving the track reconstruction and the vertex measurement [33]. The supporting structure is made of thin carbon-fibre and it is integrated with the cooling system, in order to not perturb significantly the crossing particles, maintain a constant operating temperature of about -7 °C and be radiation-hard (~ 800 kGy are expected during the detector life time).

The SemiConductor Tracker exploits the same silicon technology, being made of silicon *micro-strips* with a pitch of 80 μm. The SCT is formed by 4088 modules, each with an intrinsic accuracy of 17 μm $\times$ 580 μm. The modules (2112) are distributed in four coaxial cylinders (the barrel) with a length of 1498 mm, placed at radii between 299 mm and 514 mm, while the remaining 1976 modules are arranged in two end-caps with nine disks each, placed between $|z| = 853.8$ mm and $|z| = 2720.2$ mm. The SCT provides eight strip measurements (four space points) for particles originating in the beam-interaction region. As the Pixel Detector, the SCT operates at a temperature of about -7 °C.
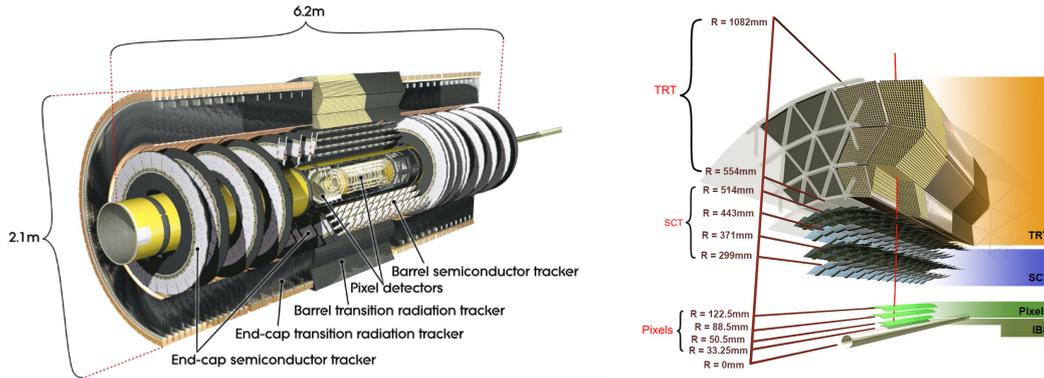
Figure 2.7: Schematic view of the ATLAS Inner Detector [23], [32].

The Transition Radiation Tracker is the ID outermost element. It includes gaseous drift tubes (*straws*) alternated with transition radiation material, covering the 563 mm $< |R| <$ 1066 mm region. With an average of 36 hits per tracks, the TRT provides a continuous tracking with an accuracy of 130 μm in the $R - \phi$ direction, for particles with $|\eta| < 2$. The basic elements of the TRT are the polyimide straw tubes of 4 mm diameter filled with a gas mixture (70% Xe, 27% CO_2, and 3% O_2) with $5 - 10$ mbar over-pressure. The anodes are 31 μm diameter tungsten (99.95%) wires plated with $0.5 - 0.7$ μm gold and they are directly connected to the front-end electronics. The free electrons created by the ionisation produced by a charged particle undergo avalanche multiplication near the wire. The TRT barrel contains up to 73 layers of straws interleaved with fibres, while 160 straw planes are alternated with foils in the end-caps, providing particle identification: transition radiation photons are emitted when a charged particle passes through the boundary region between materials with different dielectric constants. This effect is proportional to the relativistic gamma factor and allows to distinguish between particles. Unlike the other components of the ID, the TRT operates at room temperature, needing therefore a system of insulators and heaters.

The Inner Detector is fully immersed in a magnetic field generated by the Central Solenoid. The CS is designed to provide a 2 T axial field keeping the material thickness in front of the EM calorimeter as small as possible ($\sim$0.66 radiation lengths, X_0). It operates at a temperature of 4.5 K. Its length is 5.8 m and its inner and outer diameters are 2.46 m and 2.56 m. The single-layer coil is wound with a high-strength NbTi conductor inside an aluminium support cylinder. The flux is returned by the steel of the ATLAS hadronic calorimeter and its girder structure.

2.2.2 The calorimeter system

The ATLAS calorimeter system is made up of three different calorimeters (Figure 2.8): the Electromagnetic Calorimeter (ECAL or EM calorimeter), the Hadronic Calorimeter (HCAL) and the Forward Calorimeter (FCal). These calorimeters exploit different techniques and materials, covering the range $|\eta| < 4.9$. Their main task is to measure the particles energy and position. The calorimeter system has also to provide good containment for electromagnetic and hadronic showers, and to limit punch-through into the muon system. The EM calorimeter is designed for precision measurements of electrons and photons. Its total thickness is more than
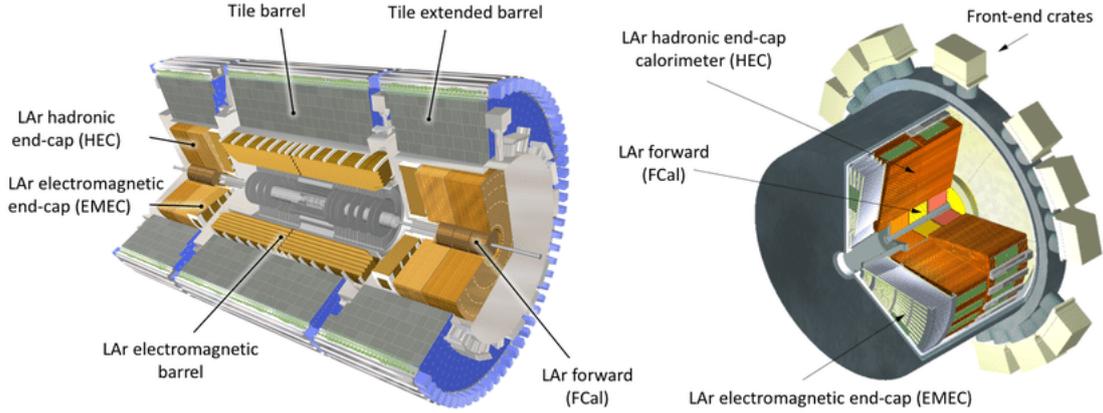
Figure 2.8: Schematic view of the ATLAS calorimeter system [23].

22 X_0 in the barrel and more than 24 X_0 in the end-caps: in this way the ECAL is able to contain electrons and photons up to 1 TeV and almost 2/3 of a typical hadronic shower. The rest of the calorimeter system is designed for the jet reconstruction and E_T^{miss} measurements. In fact, the thickness of the active calorimeter (ECAL + HCAL), ~9.7 interaction lengths (λ) in the barrel and ~10 λ in the end-caps, is enough to provide good resolution for high energy jets. The total thickness, considering an additional 1.3 λ from the outer support, is ~11 λ in the barrel and is sufficient to reduce punch-through below the irreducible level of prompt or decay muons and to ensure a good E_T^{miss} measurement. All the calorimeters use liquid argon (LAr) as active detector medium: incident particles shower in the absorber material, the LAr is ionised and electrons can drift in the gap where an electric field is applied, inducing a signal in the read-out electrodes. The choice of the material was done on the base of the intrinsic linear behaviour, the stability of response over time and the intrinsic radiation-hardness of the LAr.

The EM calorimeter is divided into barrel ($|\eta| < 1.475$) and end-caps ($1.375 < |\eta| < 3.2$), each one being inside its own cryostat, which keeps the argon temperature at 88 K. Due to the position of the CS in front of the ECAL, they share the same vacuum vessel in order to optimise the material distribution and increase the performance. The barrel calorimeter is composed by two identical half-barrels, separated by 4 mm at $z = 0$. The length of each half-barrel is 3.2 m, their inner and outer diameters are 2.8 m and 4 m respectively, and each half-barrel weighs 57 tonnes. The end-caps are, instead, two wheels, 63 cm thick and 27 tonnes weight, with internal and external radii of 330 mm and 2098 mm. They are divided into two coaxial wheels separated by 3 mm: one covers the region $1.375 < |\eta| < 2.5$, and the other the region $2.5 < |\eta| < 3.2$. The EM calorimeter is a lead-LAr detector with accordion-shaped kapton electrodes and lead absorber plates: these elements are laid out radially and folded so that the complete ϕ symmetry without azimuthal cracks is covered. Three layers of active material comes up in the $|\eta| < 2.5$ region, while two are in the remaining regions. In Figure 2.9 a sketch of a barrel module is shown: the first layer is finely segmented along η, the second layer collects the largest fraction of the energy of the electromagnetic shower, while the third layer collects only the tail of such shower and is therefore less segmented. Each half-barrel has been divided into 16 modules, each covering 22.5° in ϕ. The total thickness of a module depends from η, varying from 22 X_0 at $\eta = 0$ to 33 X_0 at $|\eta| = 1.3$. The
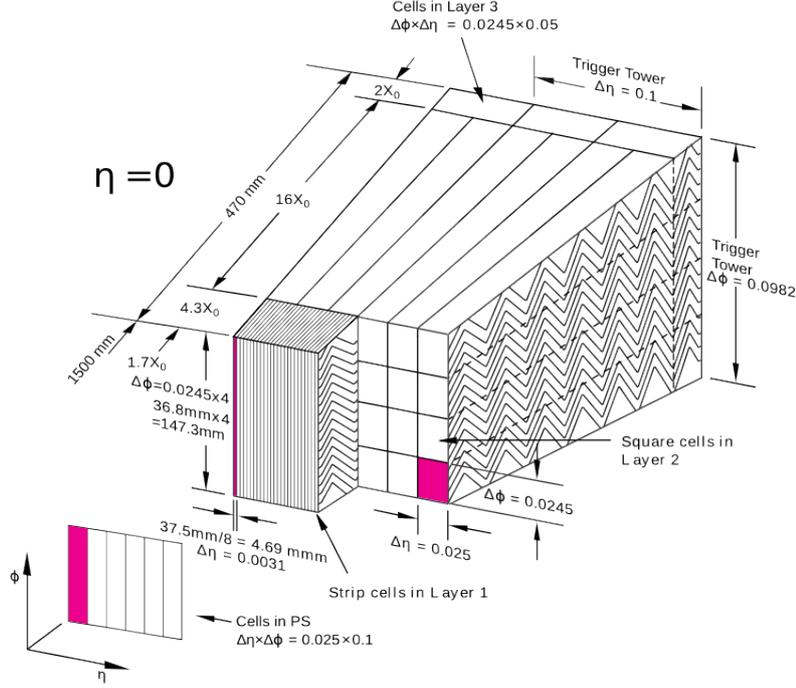
Figure 2.9: Sketch of an EM barrel module with the three different layers. The granularity of the cells is also shown [23].

supporting structure is made of stainless-steel outer rings. The end-cap wheels are divided into eight wedge-shaped modules. The total thickness of the end-caps varies from 24 X_0 to 38 X_0 for the outer wheel, and from 26 X_0 to 36 X_0 for the inner wheel. The EM calorimeters are complemented by presamplers, an instrumented argon layer, which provides a measurement of the energy lost in front of the electromagnetic calorimeters, due to the interaction with the ID or the solenoid. The presampler is 11 mm (5 mm) thick in the barrel (end-cap) and has no absorber layer. The transition region between the barrel and end-cap, $1.37 < |\eta| < 1.52$, has a poorer performance because of the higher amount of passive material in front. This region is often referred to as *crack region*.

The Hadronic Calorimeter can be divided into two parts that exploit different techniques: the Tile Calorimeter (TileCal) and the Hadronic End-cap Calorimeter (HEC). The TileCal is placed outside the ECAL and is made of a central part, the barrel, covering the region $|\eta| < 1.0$ and two extended barrels which cover the range $0.8 < |\eta| < 1.7$. It is a sampling calorimeter using steel as the absorber and scintillating tiles as the active material. The mechanical structure of the tile calorimeter is designed as a self-supporting, segmented structure comprising 64 modules. The Tile Calorimeter extends from an inner radius of 2.28 m to an outer radius of 4.25 m. The barrel is 5.8 m long, while each extended barrel is 2.6 m long. Radially, the TileCal is segmented in three layers: approximately 1.5, 4.1 and 1.8 λ thick for the barrel and approximately 1.5, 2.6, and 3.3 λ thick for the extended barrels. The total detector thickness is then $\sim 7.4\ \lambda$. Two sides of the scintillating tiles are read out by wavelength shifting fibres into two separate photomultiplier tubes. The gap region between the barrel and the extended barrel is instrumented with special modules, made of steel-scintillator sandwiches. Each HEC is divided into two wheels, placed behind ECAL end-cap in the same cryostat. It covers the range $1.5 < |\eta| < 3.2$.

25

Each wheel is built from 32 identical wedge-shaped modules and is divided into two segments in depth (so four layers per end-cap). The technology used is copper-LAr: copper plates are interleaved with 8.5 mm LAr gaps. The modules of the front wheels are made of 24 copper plates, each 25 mm thick, plus a 12.5 mm thick front plate. In the rear wheels, the modules are made of 16 copper plates, each 50 mm thick, plus a 25 mm thick front plate. The inner radius is 0.372 m for the first nine plates and 0.475 m for the others, while the outer radius is constantly 2.03 m. A stainless-steel structure supports the modules and the wheels.

The last piece of the calorimeter system is the Forward Calorimeter. It is placed in the end-cap cryostats, recessed by about 1.2 m with respect to the EM calorimeter front face. FCal covers the $3.1 < |\eta| < 4.9$ region and it is $\sim 10\ \lambda$ deep. In each end-cap there are three 45 cm thick modules: the first, made of copper, is optimised for electromagnetic measurements, while the other two, made of tungsten, measure predominantly the energy of hadronic interactions. The electrode structure consists of concentric rods and tubes parallel to the beam axis. The gap between the rod and the tube varies from 0.269 mm to 0.508 mm and it is filled by the LAr, which is used as sensitive medium.

2.2.3   The Muon Spectrometer

The MS is the outermost ATLAS sub-detector and it is made of different muon chambers, schematised in Figure 2.10. The main purpose of the MS is the triggering and measurement of muons, in particular of their track reconstruction and transverse momenta evaluation. In order to allow this operation, the MS is immersed in a toroidal magnetic field which bends the crossing muons: the tracks are measured at three points away from the interaction point. Therefore, knowing the sagitta s, it is possible to determine the transverse momentum: $p_T = L^2 qB/8s$, where L is the length of the trajectory in a constant magnetic field of intensity B and q is the electric charge of the muon. The magnetic field is provided by the large barrel toroid in the region $|\eta| < 1.4$ and by two smaller end-cap magnets for $1.6 < |\eta| < 2.7$. In the transition region ($1.4 < |\eta| < 1.6$) the deflection is due to a combination of the barrel and end-cap fields. Such magnet configuration provides a field which is mostly orthogonal to the muon trajectories. Each of the three toroids consists of eight coils assembled radially and symmetrically around the beam axis. The overall size of the barrel toroid system as installed is 25.3 m in length, with inner and outer diameters of 9.4 m and 20.1 m, respectively. The end-cap toroids are 5.0 m in length, with inner and outer diameters of 1.65 m and 10.7 m. The toroids work at a temperature of 4.6 K, needing thence a cooling system which uses liquid helium. The barrel toroid provides 1.5 to 5.5 T·m of bending power, while the end-cap toroids approximately 1 to 7.5 T·m.

In the barrel region, tracks are measured in chambers arranged in three cylindrical layers around the beam axis at radii of approximately 5 m, 7.5 m, and 10 m, while in the transition and end-cap regions, the chambers are installed in planes perpendicular to the beam, in four wheels located at distances of $|z| \sim 7.4$ m, 10.8 m, 14 m, and 21.5 m. The measurements of the tracks is entrusted principally to Monitored Drift Tubes (MDT), replaced by Cathode Strip Chambers (CSC) in the $2 < |\eta| < 2.7$ region. The trigger system, instead, covers the range $|\eta| < 2.4$ and exploits Resistive Plate Chambers (RPC) in the barrel and Thin Gap Cham-
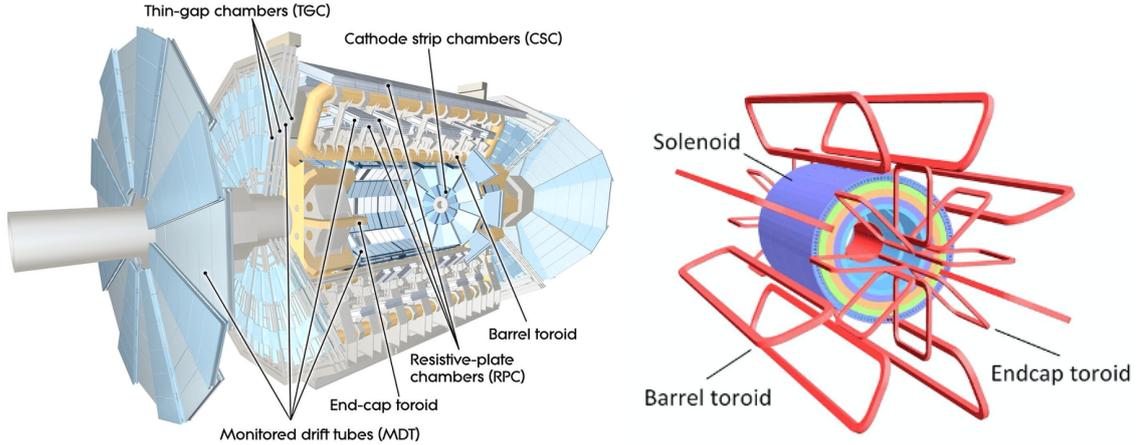
Figure 2.10: Schematic view of the ATLAS Muon Spectrometer and of the magnet system [23].

bers (TGC) in the end-caps. MDTs consist of three to eight layers of drift tubes with a diameter of 29.970 mm, operating with Ar/CO_2 gas at an absolute pressure of 3 bar, which achieve an average resolution of about 35 μm per chamber in the z direction. CSCs are used in the innermost tracking layer due to their higher rate capability and time resolution. They are multi-wire proportional chambers with cathode planes segmented into strips in orthogonal directions. The resolution of a chamber is 40 μm in the R direction and about 5 mm in the transverse plane. RPCs and TGCs provide bunch-crossing identification, well-defined transverse momentum thresholds and measure the muon coordinates in the direction orthogonal to that determined by the precision-tracking chambers. Both chamber types deliver signals with a spread of 15-25 ns, thus providing the ability to tag the beam-crossing. An RPC is a gaseous parallel electrode-plate (i.e. no wire) detector: two resistive plates are kept parallel to each other at a distance of 2 mm and filled with a gas mixture; the electric field between the plates allows the formation of an avalanche along the ionising tracks towards the anode. TGCs are multi-wire proportional chambers exploiting a gas mixture of CO_2 and n-C_5H_{12} for the ionisation process.

2.2.4 Forward detectors

In addition to the big structure described in the previous sections, three smaller systems cover the ATLAS forward region [34]. The first two are used in order to determine the luminosity delivered to the experiment, while the third one is employed to determine the centrality of heavy-ion collisions.

LUCID (LUminosity measurement using Cherenkov Integrating Detector) [35] detects inelastic proton-proton scattering in the forward direction. It consists of 16 polished aluminium tubes with a diameter of 15 mm which surround the beam pipe and point toward the interaction point. The 1.5 m long tubes are filled with C_4F_{10} at a constant pressure of $1.2 - 1.4$ bar. There are two detectors installed, one in each end-cap region of ATLAS, at a distance of approximately ± 17 m from the interaction point, and placed at a radial distance of approximately 10 cm from the beam line, covering the range $5.6 < |\eta| < 6.0$.

ALFA (Absolute Luminosity For ATLAS) [36] is located at ± 240 m and it con-

sists of scintillating fibre trackers located inside Roman pots which are designed to approach as close as 1 mm to the beam. It provides a luminosity measurement looking at elastic scattering at small angles (3 μrad).

The Zero-Degree Calorimeter (ZDC) [37] is located at ±140 m and it consists of layers of alternating quartz rods and tungsten plates which measure neutral particles at $|\eta| \geq 8.3$ in both proton-proton and heavy-ion collisions.

Moreover, the Beam Conditions Monitor (BCM) [38] monitors the stability of the particle beam. It consists of four 8×8 mm^2 diamond sensors arranged around the beam pipe in a cross pattern at $|z| = 1.84$ m, corresponding to $|\eta| = 4.2$.

2.2.5 The trigger system

ATLAS uses a two-levels trigger (see Figure 2.11): a hardware-based level (L1) is followed by a software-based High-Level Trigger (HLT). The HLT is composed by the Level 2 (L2) trigger and the Event Filter (EF), so the trigger system can be also described as a three-levels system.

The L1 performs the initial event selection based on reduced granularity signals sent from the calorimeters and muon detectors, searching for high energy objects. It reduces the event rate from the bunch-crossing rate of 40 MHz to 100 kHz. From the calorimeters, the L1 receives information about transverse energies and isolation of electrons, photons and jets. For muons, also the coincidence of hits among several layers of the MS is taken into account. The L1 decision must reach the front-end electronics within 2.5 μs after the bunch-crossing, providing Regions of Interest (RoI) to the L2: these are the regions of the detector where possible trigger objects within
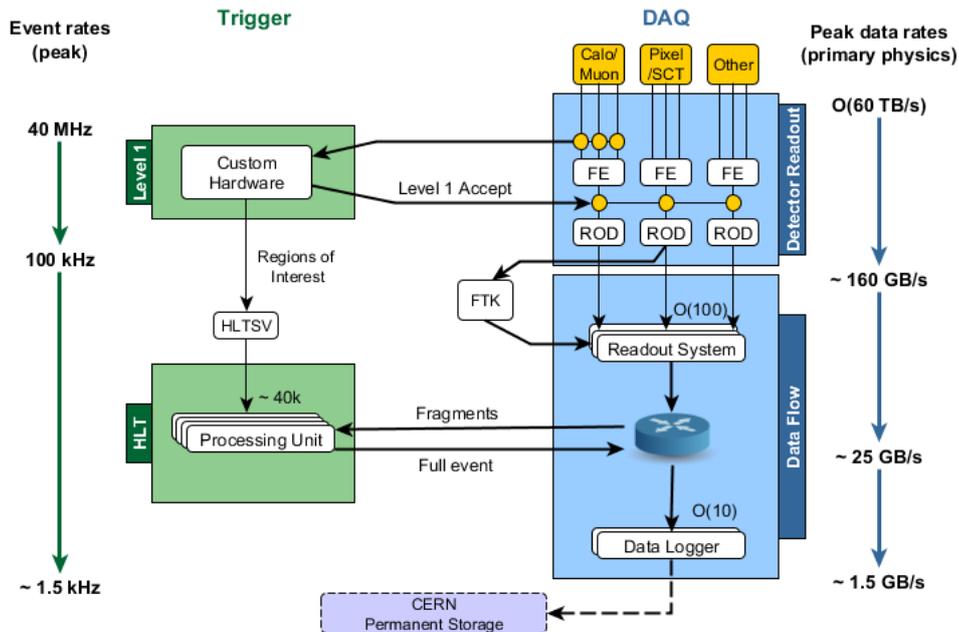
Figure 2.11: Functional diagram of the ATLAS Trigger and Data Acquisition system in Run II [39].

the event are identified. The L2 performs similar measurements as the L1, but with the full granularity and precision of calorimeters and muon chambers. The L2 uses RoI information on coordinates, energy, and type of signatures, taking a decision in less than 40 ms. The EF uses offline analysis procedures to fully reconstruct the event from all possible information coming from the sub-detectors and data calibrations. It makes a decision in less than four seconds. At the end of the HLT the average recording rate is ∼1 kHz. Trigger rates can be controlled by changing thresholds or applying different sets of selection cuts. After an event is accepted by the L1 trigger, the data from the pipe-lines are transferred off the detector to the Readout Drivers (ROD). Digitised signals are formatted as raw data prior to being transferred to the DAQ system, which receives and temporarily stores the data in local buffers. Those events selected by the L2 trigger are then transferred to the event filter. Events selected by the EF are moved to permanent storage at the CERN computer centre.

2.3   Performance of the ATLAS detector

The status of the ATLAS detector has been continuously monitored during all the Run II of LHC, ensuring the performance and the stability of the measurements. Table 2.2 summarises the number of electronic channels and the operational fraction of each of the ATLAS sub-detectors at the end of the 2018 data-taking (end of Run II).

Sub-detector	Number of channels	Approximate operational fraction
Pixels	$92 \cdot 10^6$	95.7%
SCT	$6.3 \cdot 10^6$	98.6%
TRT	$350 \cdot 10^3$	97.2%
ECAL	$170 \cdot 10^3$	100%
TileCal	$5.2 \cdot 10^3$	99.5%
HEC	$5.6 \cdot 10^3$	99.7%
FCal	$3.5 \cdot 10^3$	99.8%
L1 Calo Trigger	$7.2 \cdot 10^3$	99.9%
L1 RPC Trigger	$383 \cdot 10^3$	100%
L1 TGC Trigger	$320 \cdot 10^3$	99.9%
MDT	$357 \cdot 10^3$	99.7%
CSC	$31 \cdot 10^3$	93.0%
RPC	$383 \cdot 10^3$	93.3%
TGC	$320 \cdot 10^3$	98.9%

Table 2.2: Approximate values of the operational fraction of each of the ATLAS sub-detectors, together with the number of the electronic channels involved.

Chapter 3

Object reconstruction and event selection

The search for $t\bar{t}t\bar{t}$ production in the 1L channel presented in this thesis involves several physics objects, like electrons, muons, jets, b-jets and missing transverse momentum. The events reconstructed form these objects are then required to pass some selection cuts. Therefore in this chapter the event reconstruction and selection are described.

3.1 Object reconstruction

A charged particle crossing the ATLAS detector may leave some hits in the different layers of the Inner Detector which form a *track*, reconstructed using a χ^2-fit [40]. A set of five parameters is used to characterise such track: d_0 and z_0, which are the transverse and longitudinal impact parameters, ϕ and θ, which are the azimuthal and polar angles, and q/p_T, that is the charge over the transverse momentum. A two-steps procedure is employed to find the hits belonging to a track. First, an *inside-out* pattern recognition sequence, starting from *seed* space-points in the silicon detector, propagates outwards to include hits from the TRT. Then, an *outside-in* pattern recognition sequence (*back-tracking*) is applied, seeded in the TRT and extrapolated inwards adding silicon layer hits. The final track candidates are moreover required to fulfill a set of quality criteria.

The primary vertices are reconstructed by a dedicated algorithm in the HLT, as well as by an offline reconstruction algorithm. This way, each vertex found by reconstructing the tracks is associated to a vertex candidate and the position is then determined by a fit. The vertex that has the highest sum of the p_T of its tracks is considered as the primary vertex, while the others are assumed to be coming from *pile-up*[1] interactions or, in case they are far from the beam spot position, secondary vertices, which may come for example from the decay of a b-hadron: during the fragmentation process a b quark forms hadrons which can travel a few millimetres from the collision point before they decay creating secondary vertices.

Electrons are reconstructed from *clusters* (energy deposits) in the electromagnetic calorimeters associated to reconstructed tracks in the ID. The algorithms for electron reconstruction are optimised fo the discrimination between signal, isolated

---

[1]The pile-up is the occurrence of several inelastic collisions during the same bunch-crossing.

electrons and background electrons from photon conversion, object misidentification, and hadron decays. Different sets of identification criteria based on information from the calorimeter, the tracker and the matching between the tracker and the calorimeter, are used with different levels of background rejection and signal efficiency. Electrons passing all the identification requirements are called *tight* electrons, while *loose* and *medium* electrons are those which pass only some of the above mentioned requirements. The reconstruction and identification efficiencies are determined with the tag-and-probe method[2] from J/ψ and Z boson decay events.

Muons are reconstructed using information from the muon spectrometer, the ECAL system and the tracker. Depending from the information used, four different types of muons can be defined:

- stand-alone muons, for which the trajectory is reconstructed only from the MS and the track is extrapolated back to the point of closest approach to the beam line;

- segment-tagged muons, which are those with a track in the ID associated with at least one track in the MDT or CSC chambers;

- calorimeter-tagged muons which are those with a track in the ID and an energy deposit in the calorimeter compatible with a minimum ionising particle;

- combined muons, for which the reconstruction is done using both the ID and the MS. This is the muon definition which has been used in this thesis.

Also for the muons different sets of identification criteria can be used, leading to *tight*, *medium*, *loose*, *LowPt* and *HighPt* muons. In this case as well, the reconstruction and identification efficiencies are determined with the tag-and-probe method from J/ψ and Z boson decay events. For both muons and electrons, the momentum scale and resolution are studied, introducing correction factors to the simulation.

Due to the colour confinement, quarks and gluons produced in the collisions can not exist as free particles, so they undergo a hadronisation process that produces a collimated stream of charged and neutral hadrons, called jet. Jets are characterised by multiple energy deposits in the calorimeters and by the presence of several tracks in the ID. The reconstruction of jets is done using algorithms that cluster together adjacent particles. The criteria that such algorithms have to follow are two: they have to be collinear and infrared safe; this means that the jet configuration is independent of the emission of a soft or collinear particle. In ATLAS the jet reconstruction [41] starts from grouping together energy deposits in adjacent cells of the calorimeter, creating a *topocluster*. Topoclusters are formed by a growing-volume algorithm starting from a calorimeter cell with a highly significant seed signal. The cells near the seed are aggregate to the topocluster if their energy deposit is greater than a given threshold. This procedure is iteratively applied to further neighbours. The total energy of the topocluster is the sum of the energy deposits of the cells included, corrected with a local cluster weighting (LCW) calibration in order to take into account the effects of inactive material, non-compensation and out-of-cluster

---

[2]The tag-and-probe method exploits a resonance to find a *tag* lepton coming from that decay which passes a very tight set of cuts and a matching *probe* lepton which, instead, passes a looser set of cuts. The ratio between events with a positive match and the total number of events with a tag lepton is the efficiency of the looser set of cuts.

leakage. The topoclusters are then grouped: a distance between two clusters is calculated as $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where y_i and ϕ_i are the rapidity and the azimuth of the i-th cluster. If this distance is less than a certain value R, i and j are merged in a single jet. The procedure is repeated until all clusters are either part of a jet, or a jet itself. Practically the following distances are calculated [42]:

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

$$d_{iB} = k_{ti}^{2p}, \tag{3.2}$$

where k_{ti} is the transverse momentum of the i-th cluster. If $p = 1$ one speak of the k_t algorithm, if $p = 0$ of the Cambridge/Aachen algorithm, and if $p = -1$ of the anti-k_t jet-clustering algorithm. If $d_{ij} < d_{iB}$ i and j are combined in a single entity, otherwise i is considered as a single jet and it is removed from the list of entities. In this thesis jets are reconstructed using the anti-k_t algorithm with a radius requirement of $R = 0.4$.

Jets obtained with this procedure undergo a process of calibration and correction. Since the jet energy is measured by the ECAL, the Jet Energy Scale (JES) calibration consists of several consecutive stages derived from Monte Carlo (MC) simulations in order to correct the mismodelling due to non-EM particles. The jet is corrected to point back to the primary vertex, removing additional energy deposits coming from pile-up. An additional Jet Vertex Tagger (JVT) procedure is useful to suppress jets which originate from pile-up collisions, in particular when considering low-p_T jets: the scalar sum of the p_T of tracks matched to the jet and originating from the primary vertex is required to be at least 50% of the total p_T of the tracks matched to the jet. Further residual calibrations are derived using in-situ measurements and applied sequentially.

For the aim of top physics, a critical point is the capability to reconstruct jets coming from b quarks, discriminating b-jets from c- or light-jets. This type of discrimination is called *b-tagging* and is made with algorithms which provide the probability that a jet is originating from a b quark. The key ingredient for b-tagging is the relatively long lifetime of b-hadrons ($\sim 10^{-12}$ s) resulting in a significant mean flight length (even some millimeters in some cases). This leads to measurable secondary vertices and impact parameters of the decay products (see Figure 3.1).

The algorithm used in this thesis (MV2 [43]) is one of the most common and powerful in ATLAS. It is a multivariate (MV) algorithm, consisting in a boosted decision tree (BDT) that combines the output of the low-level taggers based on tracks and vertices information. The BDT algorithm is trained using the ROOT Toolkit for Multivariate Data Analysis (TMVA) [44] on a hybrid $t\bar{t} + Z'$ sample. For the sake of training, the c-jet fraction in the background sample is set to 7% (MV2c10), with the remainder composed of light-jets. This allows the charm rejection to be enhanced whilst preserving a high light-jet rejection. For this thesis, the chosen fixed cut on the MV2c10 output, called *working point* (WP), corresponds to 77% efficiency to tag a b-jet, with a light-jet rejection factor of 110 and a c-jet rejection factor of 5, as shown in Table 3.1.

Neutrinos do not interact at all with the detector, and so can be reconstructed only using the difference between the initial state and final state total momentum. In fact, since the initial transverse momentum of the partons is zero, it has to be the
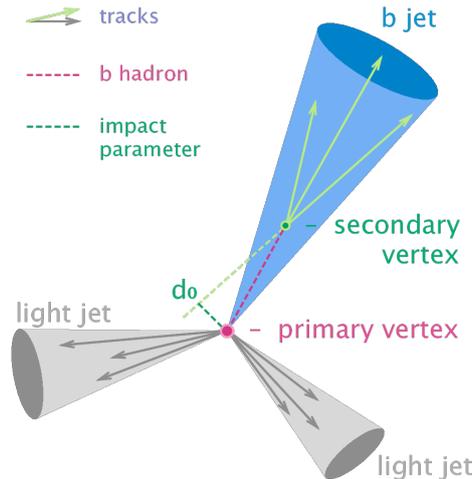
Figure 3.1: Schematic view of a b-hadron decay inside a jet coming from a secondary vertex.

b-jet efficiency	BDT output selection	c-jet rejection	light-jet rejection
60%	> 0.94	23	1200
70%	> 0.83	8.9	300
77%	> 0.64	4.9	110
85%	> 0.11	2.7	25

Table 3.1: Selection and c-jet and light-jet rejections corresponding to different b-jet tagging efficiency [45].

same for the final state. Candidate neutrinos are defined by the missing transverse energy $E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}$; the two components are calculated as:

$$E^{miss} = E_e^{miss} + E_\mu^{miss} + E_{jet}^{miss} + E_{soft}^{miss}, \tag{3.3}$$

where each term is the negative sum of all the object energies projected in the considered direction. E_{soft}^{miss} is the term taking into account the tracks not associated to any reconstructed object.

3.2 Event selection

The selection of $t\bar{t}t\bar{t}$ events is a procedure divided in steps: after a preselection based on data quality, trigger and physics objects, the selection for the 1L channel chategorises the events in "regions", depending on the number of jets and b-jets.

The quality of the events is checked to exclude those events with noise in the calorimeters. In addition, events are requested to pass single electron or single muon triggers, both for online and offline reconstructed objects.

The objects reconstructed as described in the previous section undergo a process of selection based on identification and isolation requirements, as recommended by the ATLAS Collaboration and summarised in Table 3.2:

- for electrons, tight criteria are used for identification and isolation. Moreover, electron candidates have to satisfy $p_T > 10$ GeV and $|\eta| < 2.47$, with the

	Electrons		Muons		Jets	b-jets
	loose	tight	loose	tight		
p_T [GeV]	> 10 or > 28		> 10 or > 28		> 25	> 25
$|\eta|$	< 1.37 or 1.52 − 2.47		< 2.5		< 2.5	< 2.5
ID quality	mediumLH	tightLH	medium		cleaning +JVT	MV2c10 77%
Isolation	none	FCTight	none	FixedCutTight-TrackOnly		
Track vertex: $|d_0/\sigma_{d_0}|$ $|z_0 \sin\theta|$ [mm]	< 5 < 0.5		< 3 < 0.5			

Table 3.2: Summary of object identification and definitions criteria.

exclusion of the region $1.37 < |\eta| < 1.52$ (known as "LAr crack region"), where electrons are discarded. Scale factors for identification and isolation are applied to the MC simulation in order to correct for the efficiency differences between data and simulation;

- for muons, medium quality criteria are used for the identification, while tight criteria are employed for the isolation. Moreover, muon candidates are requested to have $p_T > 10$ GeV and $|\eta| < 2.5$. As for electrons, scale factors are applied to the MC simulation;

- as already described, jets are reconstructed using the anti-k_t algorithm with a radius requirement of $R = 0.4$. The JVT procedure is applied to jets with $p_T < 60$ GeV and $|\eta| < 2.4$. Then, jet candidates have to satisfy $p_T > 25$ GeV and $|\eta| < 2.5$;

- for b-jet tagging, the MV2c10 algorithm at WP=77% is employed. In order to use the full b-tagging information, to each jet is associated also a *pseudo-continuous* b-tagging score, that defines if the jet passes the b-tagging working points of 85%, 77%, 70% and 60%, or if the jet does not pass any of the previous WPs.

The selected objects have also to pass an overlap removal procedure, in order to avoid double counting. For this procedure, loose criteria are employed:

- an electron candidate track overlapping with another electron or muon is removed;

- a calorimeter muon sharing a track with an electron is removed;

- if the ΔR between a jet and an electron is smaller than 0.2, the jet is removed. In cases of multiple jets fulfilling this criteria, only the closest jet (the one with the smaller ΔR) is removed;

- electrons are removed if their distance with a jet is $\Delta R < 0.4$;

- jets with less than three tracks and within $\Delta R < 0.2$ of a muon are removed;
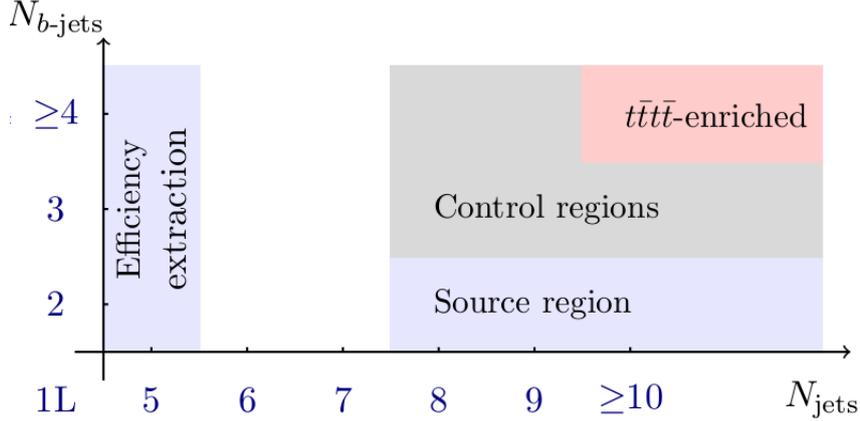
Figure 3.2: Schematic view of the regions definition.

- a muon is removed if the distance between a jet and the muon is $\Delta R < 0.4 + 10 \text{ GeV}/p_T^\mu$, with p_T^μ being the transverse momentum of the muon.

Jets that pass this procedure are calibrated and re-clustered using the anti-k_t algorithm with a radius requirement of $R = 1.0$, originating the so-called *RC-jets*. Such jets are "trimmed" by removing all small-R jets within a reclustered jet that have p_T below 5% of the p_T the RC-jet, suppressing further contributions from pile-up and soft radiation. The resulting RC-jets that satisfy $p_T > 200$ GeV and $|\eta| < 2.0$ are used to identify hadronically decaying top quark candidates, exploiting then the peculiar jet-substructure of the $t\bar{t}t\bar{t}$ process.

To all the events that pass the data quality and object selection, a further selection is applied. At least one vertex formed from at least two tracks with $p_T > 0.4$ GeV is required. The hard-scattering vertex is defined as the one with the highest summed track p_T^2. The transverse impact parameter divided by its estimated uncertainty $|d_0|/\sigma_{d_0}$ is required to be lower than five (three) for electron (muon) candidates. The longitudinal impact parameter must satisfy $|z_0 \sin\theta| < 0.5$ mm for both lepton flavours. Jets are further cleaned from those coming from non-collision background and from coherent noise or pathological cells in the calorimeter.

For the 1L channel, exactly one tight lepton satisfying $p_T > 28$ GeV and at least five jets among which at least two b-tagged are requested. Moreover, requirements are made on E_T^{miss} as well as on the transverse mass of the lepton[3] to suppress the background from multijet production: $E_T^{miss} > 20$ GeV and $E_T^{miss} + m_T^W > 60$ GeV. According to the number of jets and b-jets, the events can be split in different regions (see Figure 3.2): an event with n jets and m b-jets belongs to the $njmb$ region. When a region contains an exact number of jets it is labelled as *exclusive* (e), while if the region contains at least a number of jets it is labelled as *inclusive* (i). So, for example, $6je3be$ is the region with exactly six jets and exactly three b-jets, while $10ji4bi$ is the region with at least ten jets and at least four b-jets. For the purpose of the background evaluation method that will be described later, the regions can be classified as *extraction regions*, *source regions*, *control regions* and *signal regions* (or $t\bar{t}t\bar{t}$-enriched).

---

[3]$m_T^W = \sqrt{2p_T^l \cdot E_T^{miss} \cdot (1 - \cos\Delta\phi)}$, where p_T^l is the transverse momentum of the lepton and $\Delta\phi$ is the azimuthal angle separation between the lepton and the direction of the missing transverse momentum.

Chapter 4

Data and Monte Carlo simulation samples

The aim of this thesis is the search for $t\bar{t}t\bar{t}$ production in the 1L channel, confirming the evidence recently provided by the ATLAS experiment in the SSML channel [46]. The dataset used comes from the full Run II pp collisions period, collected by the ATLAS detector at $\sqrt{s} = 13$ TeV, from 2015 to 2018. It corresponds to an integrated luminosity of 139.0 ± 2.4 fb^{-1}. Monte Carlo techniques are employed to simulate and predict the event evolution from the collision to particles detection, both for the signal and the background processes, for which the MC simulation is assisted by a data-driven method.

4.1 Event simulation

The ATLAS simulation infrastructure [47] can be divided into four steps: event generation, detector simulation, digitisation and reconstruction.

The first, complex, part is the generation of events: the production of a set of particles which can be passed to the detector simulation. Event generation can be further divided into three substeps: hard-scattering, parton shower and hadronisation. In addition, this phase has to take into account the possibility of multiple interactions in the same event. More in details:

- the hard-scattering or matrix element, the big circle and lines in red in Figure 4.1, is the collision between two partons. It includes the parton-parton interaction at short distance, evaluated using perturbative QCD calculations, and the long distance interaction, modelled by the PDFs. As already discussed, such PDFs can not be directly calculated, but are provided from various collaborations fitting experimental data. For this thesis the PDF sets are provided by the NNPDF collaboration [49];

- the parton shower (PS), the blue lines in Figure 4.1, is the emission of coloured particles and photons by the scattered partons. The QCD theory allows three types of possible branchings, $q \rightarrow qg$, $g \rightarrow gg$ and $g \rightarrow q\bar{q}$. Initial state radiation (ISR) is the one emitted by the incoming partons before the scattering;

- the hadronisation process occurs after the PS emission, at energy scales comparable to the non perturbative limit of QCD, when partons are no more free
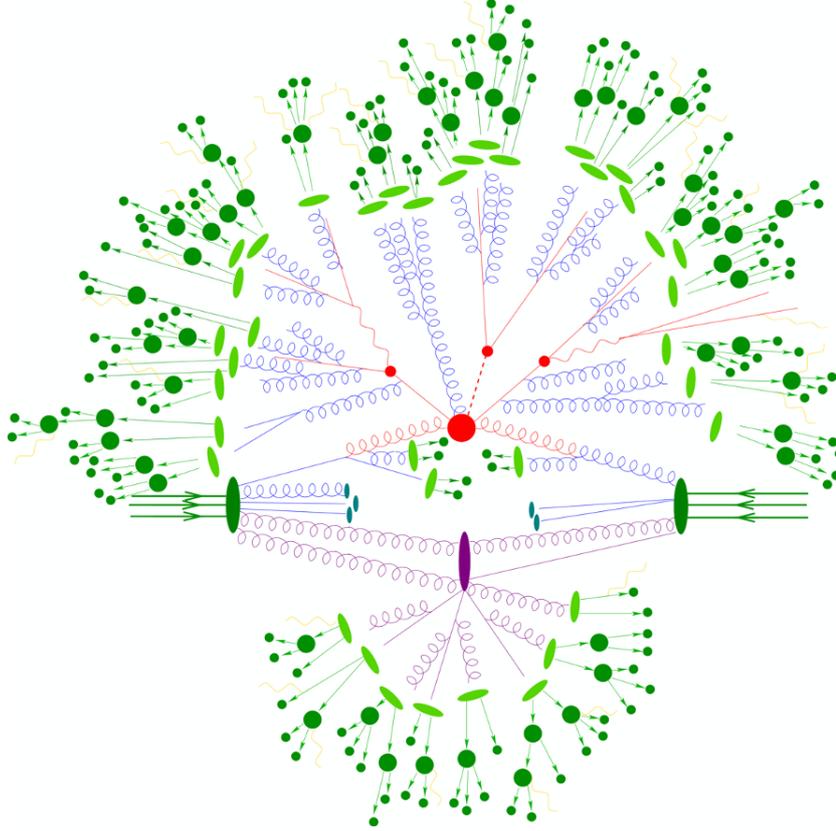
Figure 4.1: Schematic representation of an event as produced by a MC simulation [48].

particles and start to build hadrons (the light green ellipses in Figure 4.1), which then decay (dark green circles). Two different models can be employed to simulate the hadronisation, the Lund string fragmentation model [50] and the cluster fragmentation model [51];

- in addition to the "main" parton-parton collision, other interactions can occur originating a phenomenon known as multiple parton interactions (MPI) and the so-called underlying event (UE), the purple lines and ellipse in Figure 4.1. The energy of UEs is in general small and they are modelled using experimental data.

This analysis makes use of different event generators, depending from the process. For the matrix element computation, two generators are used:

- POWHEG-BOX [52] is a parton-level event generator computing matrix elements in perturbative QCD and using the POWHEG method [53] for interfacing NLO calculations with parton shower generators;

- MADGRAPH5_AMC@NLO [54] can simulate the matrix element for $2 \to n$ hard scattering processes without any approximation. The NLO calculation follows from the MC@NLO method [55] for the interface with parton shower generators.

For the parton shower and hadronisation steps, and for the modelling of the underlying events, the following generators are used:

- PYTHIA 8 [56] is a multi-purpose event generator, optimised for $2 \rightarrow 1$ and $2 \rightarrow 2$ hard scattering and it provides the PS, hadronisation and the decay of the unstable particles, giving back a list of stable particles. In this thesis it is not used for the matrix element computation;

- HERWIG 7 [57] is another multi-purpose event generator, but in this thesis it is used only for parton shower and hadronisation, being match with a different generator for the hard scattering;

- EVTGEN [58] is a generator specialised in the decays of bottom and charm hadrons, and it runs after the previous generators in order to have an accurate simulation for these processes;

- SHERPA [59] is a multi-purpose generator, specialised in multi-leg processes. It is employed, in particular, for the production of vector bosons in association with additional partons. In addition to the parton shower simulation, SHERPA can be interfaced with inbuilt libraries, like Comix [59], or external ones, like OpenLoops [60], for the matrix element calculation.

After the event generation, a detailed simulation of the ATLAS detector is needed, in order to take into account its geometry and response. The simulation is done with GEANT4 [61] that covers the majority of the electromagnetic, hadronic and optical processes, over a wide range of materials. The energy scale starts from 250 eV up to a few TeV. All aspects of the simulated process are included: the geometry of the system, the materials involved, the generation of primary particles of events, the tracking of particles through materials and external electromagnetic fields, the interactions between particles, the response of sensitive detector components, the generation of event data, the storage of events, the visualisation of the detector and particle trajectories, and the capture for subsequent analysis of simulation data. An alternative, less accurate but time saving, is provided by a fast detector simulation making use of parameterised showers in the calorimeters (FASTCALOSIM [62]).

In the digitisation step, the output of the detector simulation is converted to the same format of the output of the real detector. Moreover, correction are applied so that the simulated particles selection efficiencies, energy scales and energy resolutions match those determined from data control samples.

The last step is the reconstruction, and it is made for the MC simulation in the same way described in the previous chapter for real data.

4.2   Data and MC samples

As already mentioned the data used in this thesis correspond to the full Run II of LHC, collected by the ATLAS detector between 2015 and 2018 at $\sqrt{s} = 13$ TeV with stable beam conditions and with all detector systems operating normally. The pile-up, so the average number of additional collisions per bunch crossing, ranged between 14 and 38. The integrated luminosity is 139.0 ± 2.4 fb^{-1}, determined using the LUCID-2 detector [35]. Monte Carlo simulations are used to estimate the $t\bar{t}t\bar{t}$ signal acceptance and all the backgrounds.

The $t\bar{t}t\bar{t}$ simulation is performed by the MADGRAPH5_AMC@NLO v2.6.2 generator at NLO[1] accuracy in QCD with the NNPDF3.1NLO parton distribution function [63]. The renormalisation (μ_R) and factorisation (μ_F) scales are set to be $\frac{1}{4} \cdot \sum_i \sqrt{m_i^2 + p_{T,i}^2}$, where the sum runs over all the particles generated by the ME calculation, according to Reference [19]. Top quarks are decayed at LO using MADSPIN [64] to preserve all spin correlations. PYTHIA 8.230 is used for the hadronisation and parton shower, with the A14 set of tuned parameters [65] and the NNPDF2.3LO PDF set [66]. The decays of bottom and charm hadrons are simulated using the EVTGEN v1.6.0 program. Additional $t\bar{t}t\bar{t}$ samples are produced replacing HERWIG for parton shower or SHERPA for the full generation.

The main background, $t\bar{t}$, is simulated with the POWHEG-BOX v2 generator at NLO with the NNPDF3.0NLO PDF set and $h_{damp} = 1.5 \cdot m_{top}$ [67]. The h_{damp} parameter controls the p_T of the first additional emission beyond the LO Feynman diagram in the parton shower and therefore regulates the high-p_T emission against which the $t\bar{t}$ system recoils. The simulation is then interfaced with PYTHIA 8.230 using the A14 tune and the NNPDF2.3LO PDF set. The impact of the PS and hadronisation model is evaluated by comparing the nominal setup with a sample produced with the same ME generator, but interfaced with HERWIG 7.04 using the H7UE set of tuned parameters and the MMHT2014LO PDF set [68]. The uncertainty due to the ME generator choice is estimated with the comparison between the nominal setup with a sample generated with the same PYTHIA for PS but a different hard-scattering calculator: MADGRAPH5_AMC@NLO v2.6.0 with the NNPDF3.0NLO PDF. The mismodelling due to the choice of the value of h_{damp} is evaluated comparing the nominal setup with an alternative one produced with $h_{damp} = 3.0 \cdot m_{top}$.

The single top production is modelled with the POWHEG-BOX v2 generator at NLO with the NNPDF3.0NLO PDF set. The overlap between the $t\bar{t}$ and the tW final state is removed using the diagram removal technique [69]. The events are then interfaced with PYTHIA 8.230 using the A14 tune and the NNPDF2.3LO PDF set. The $t\bar{t}$ and single top simulated samples are normalised to the cross sections calculated at NNLO in QCD, including the resummation of NNLL soft-gluon terms.

The $t\bar{t}Z$ production is simulated with the MADGRAPH5_AMC@NLO v2.3.3 generator at NLO with the NNPDF3.0NLO PDF set, interfaced with PYTHIA 8.210 using the A14 tune and the NNPDF2.3LO PDF set. The $t\bar{t}W$ production is modelled using the SHERPA v2.2.1 generator with the NNPDF3.0NLO PDF set, interfaced with Comix and OpenLoops for the ME calculation. The $t\bar{t}H$ production is generated with POWHEG-BOX at NLO with the NNPDF3.0NLO PDF set and interfaced with PYTHIA 8.230 using the A14 tune and the NNPDF2.3LO PDF set. The $t\bar{t}WW$ is modelled using MADGRAPH5_AMC@NLO at LO interfaced with PYTHIA 8.230 using the A14 tune and the NNPDF2.3LO PDF set. The rare top quark backgrounds are normalised using their NLO theoretical cross sections.

Diboson production (VV) is generated with SHERPA v2.2.1 interfaced with the OpenLoops library and using NNPDF3.0NNLO PDF set and a dedicated set of tuned parton-shower parameters developed by the SHERPA authors. This sample is normalised to the theoretical cross sections calculated at NLO in QCD.

---

[1]As already mentioned in Section 1.7, LO stands for leading order, NLO for next-to-leading order, NNLO for next-to-next-to-leading order and NNLL for next-to-next-to-leading-logarithmic.

ME	PS	ME PDF PS PDF	PS Tune	σ (fb)
MADGRAPH5_AMC@NLO 2.6.2	PYTHIA 8.230	NNPDF3.1NLO NNPDF2.3LO	A14	$10.93^{+19.6\%}_{-23.2\%}$
MADGRAPH5_AMC@NLO 2.6.2	HERWIG 7.0.4	NNPDF3.1NLO MMHT2014LO	H7UE	$10.95^{+17.0\%}_{-22.4\%}$
SHERPA 2.2.8	SHERPA 2.2.8	NNPDF3.0NNLO NNPDF3.0NNLO	Author's tune	$11.15^{+3.3\%}_{-17.9\%}$

Table 4.1: A summary of $t\bar{t}t\bar{t}$ samples available. Cross sections quoted uncertainties are due to the QCD scale choice. The matching between the ME and the PS is MC@NLO in all the cases.

The production of Z+jets and W+jets is simulated with the SHERPA v2.2.2 generator interfaced with Comix and OpenLoops for the ME calculation. It uses the NNPDF3.0NNLO PDF set and the dedicated set of tuned parameters developed by the SHERPA authors. The samples are normalised to the NNLO cross sections.

In all Monte Carlo samples the top quark mass is set to 172.5 GeV and they are processed with GEANT4 in order to take into account the geometry and the response of the ATLAS detector. Except for the samples generated with SHERPA, the EVTGEN v.1.2.0 program is used for the decay of bottom and charm hadrons.

All the samples are divided in *campaigns* in order to reflect the different data conditions, triggers and pile-up between the years that compose the Run II: *mc16a* corresponds to the period 2015 + 2016, for an integrated luminosity of 36.2 fb^{-1}, *mc16d* is the 2017 period, for an integrated luminosity of 44.3 fb^{-1}, and *mc16e* corresponds to the 2018 period, for an integrated luminosity of 58.5 fb^{-1}.

4.3   Signal modelling

Besides the nominal MADGRAPH5_AMC@NLO + PYTHIA sample, two other samples have been simulated for the $t\bar{t}t\bar{t}$ process at NLO [70]. The settings are summarised in Table 4.1.

As already described, the MADGRAPH5_AMC@NLO samples are generated at NLO in QCD with a renormalisation and factorisation scales defined as $H_T/4$[2]. In the nominal sample, this generator is interfaced to the PYTHIA 8 parton shower using the A14 set of tuned parameters. The decays of bottom and charm hadrons are simulated using EVTGEN v1.6.0. The theoretical uncertainties due the QCD scale choice are estimated varying the renormalisation and factorisation scale independently by a factor of 0.5 and 2.0 with respect to the central value ($H_T/4$). An alternative NLO sample is also generated by showering the same events with HERWIG 7 instead of PYTHIA 8. Comparing these two samples provides an uncertainty related to the showering model and NLO matching (p_T-ordered versus angular-ordered) and the hadronisation model (Lund string versus clustering). A third alternative sample is generated at NLO accuracy in QCD with the SHERPA 2.2.8 generator using Comix and OpenLoops. No additional partons are considered in the ME calculations. The renormalisation and factorisation scales are set to $H_T/4$. The PDF set used for the

---
[2]H_T is the scalar sum of the jet transverse momenta.
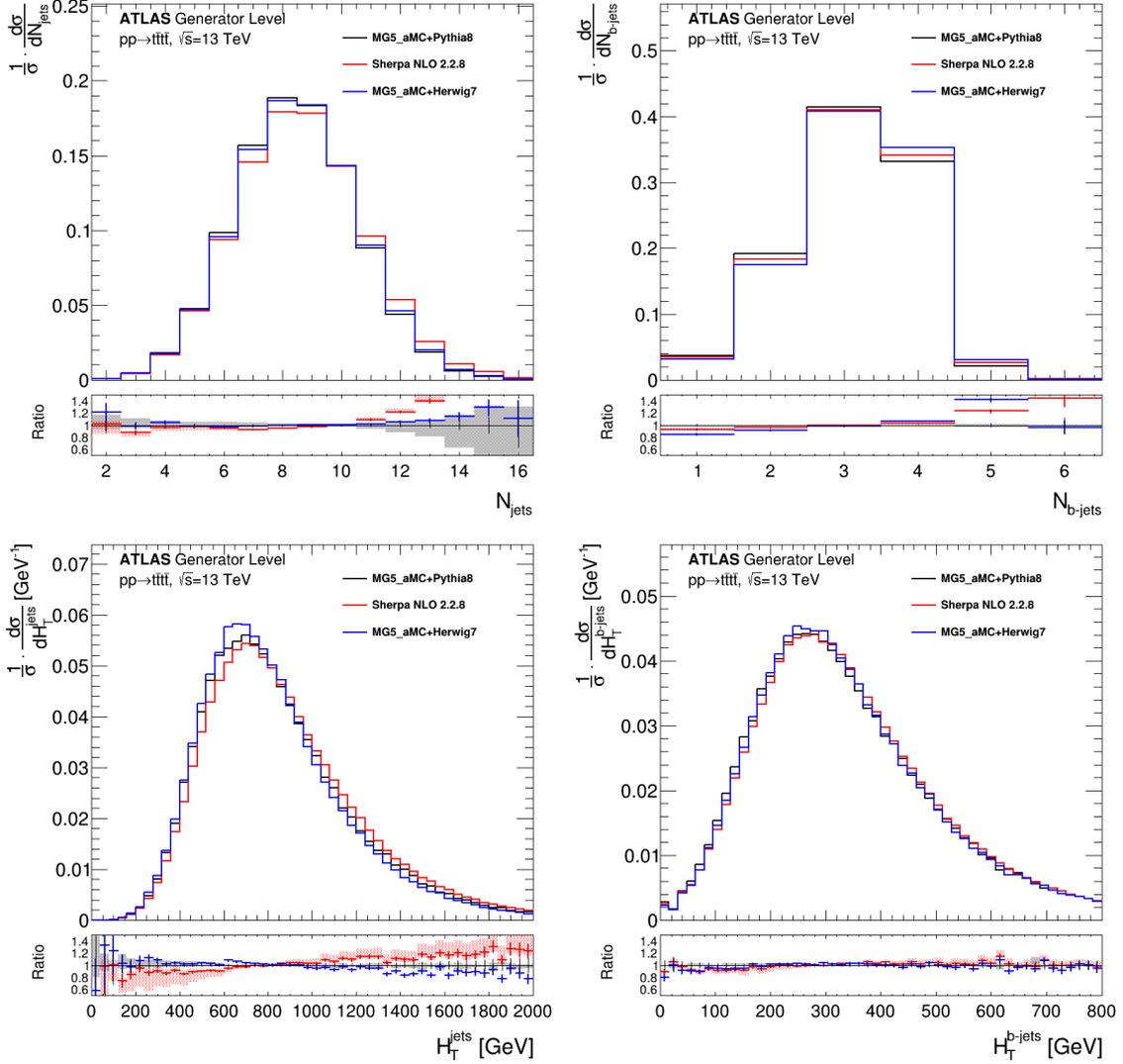
Figure 4.2: Distributions of the jet and b-jet multiplicities, the scalar sum of jets and b-jets p_T (H_T) from MADGRAPH5_AMC@NLO + PYTHIA, SHERPA and MADGRAPH5_AMC@NLO + HERWIG $t\bar{t}t\bar{t}$ generation at NLO. All distributions are normalised to one. The lower ratio is calculated with respect to the MAD-GRAPH5_AMC@NLO + PYTHIA NLO distribution. The uncertainties shown reflect the finite sample statistics (solid lines) and the effect of the different choices of factorisation and renormalisation scales (shaded bands) [70].

matrix element calculation is NNPDF3.0NNLO with a dedicated PS tuning developed by the SHERPA authors.

Figure 4.2 shows the comparison between the three signal samples, involving different generator choices, including the impact of varying the factorisation and renormalisation scales for MADGRAPH5_ AMC@NLO + PYTHIA and SHERPA samples. For all the samples the event selection cuts are applied to the objects as well as the overlap removal procedure already described in Section 3.2.

The two MADGRAPH5_AMC@NLO samples predict similar jet and b-jet multiplicity shapes, while the SHERPA sample has a higher jet multiplicity on average. The jet momenta are relatively different among all generators. The MADGRAPH5_AMC@NLO sample interfaced with HERWIG 7 predicts softer jets compared to PYTHIA 8, while SHERPA predicts harder jets compared to the MADGRAPH5_AMC@NLO samples. For the MADGRAPH5_AMC@NLO + PYTHIA sample, the shape uncertainty due to different choices of factorisation and renormalisation variations is relatively low in the jet p_T distributions, but starts to be relatively large in the jet multiplicity distributions. Instead, for the SHERPA sample, while the shape uncertainty associated to the scale variations is low for the jet multiplicities, it becomes larger for the jet p_T distributions, in particular for the H_T^{jets} distribution (order of 20%). The $t\bar{t}t\bar{t}$ cross section is currently known at complete-NLO (QCD and EW) accuracy using the MADGRAPH5_AMC@NLO framework and is calculated as $11.97^{+18\%}_{-21\%}$ fb, where the uncertainties are estimated by varying the renormalisation and factorisation scales [19].

4.4   Background modelling: the $TRF_{t\bar{t}}$ method

A purely Monte Carlo simulation-based method is not expected to model well the $t\bar{t}$+jets background in high jet and b-jet multiplicities regions, which are the signal regions for this thesis, where the exact perturbative order calculations needed to describe the relevant hard scattering features are still missing. Therefore, a data-driven, MC-assisted, approach is adopted to improve the prediction for this background. This estimate derives essentially from the method already used in the ATLAS collaboration for the $t\bar{t}H \to b\bar{b}$ analysis [71], and is referred to as Tag Rate Function (TRF) for multi-jet background. The method was used to estimate the background from multi-jet events for an all-hadronic selection with a large number of jets and three or four b-jets. Later, it was adopted and used for the estimate of the $t\bar{t}$+jets background in the previous ATLAS $t\bar{t}t\bar{t}$ 1LOS measurement [20], renamed as $t\bar{t}$ Tag Rate Function ($TRF_{t\bar{t}}$). The $TRF_{t\bar{t}}$ method measures the probability that a jet is b-tagged, with the simplifying assumption that the probability of tagging an additional jet is independent of the number of total jets. This assumption allows the b-tagging probabilities to be measured in lower jets multiplicity regions, where signal contamination is expected to be low, and then applied in higher jets multiplicity regions. Depending on the jet and b-tagged jet multiplicities the data are separated into efficiency extraction, source, control and signal regions (see Figure 4.3). Extracted efficiencies are applied to source regions to get yields in the higher b-jet multiplicity regions with same jet multiplicity. This is done by reweighting events in source regions by a factor, computed following TRF mathematics, which takes into account, given the measured efficiencies, of the probability for the event to populate higher b-tagged jet multiplicities.
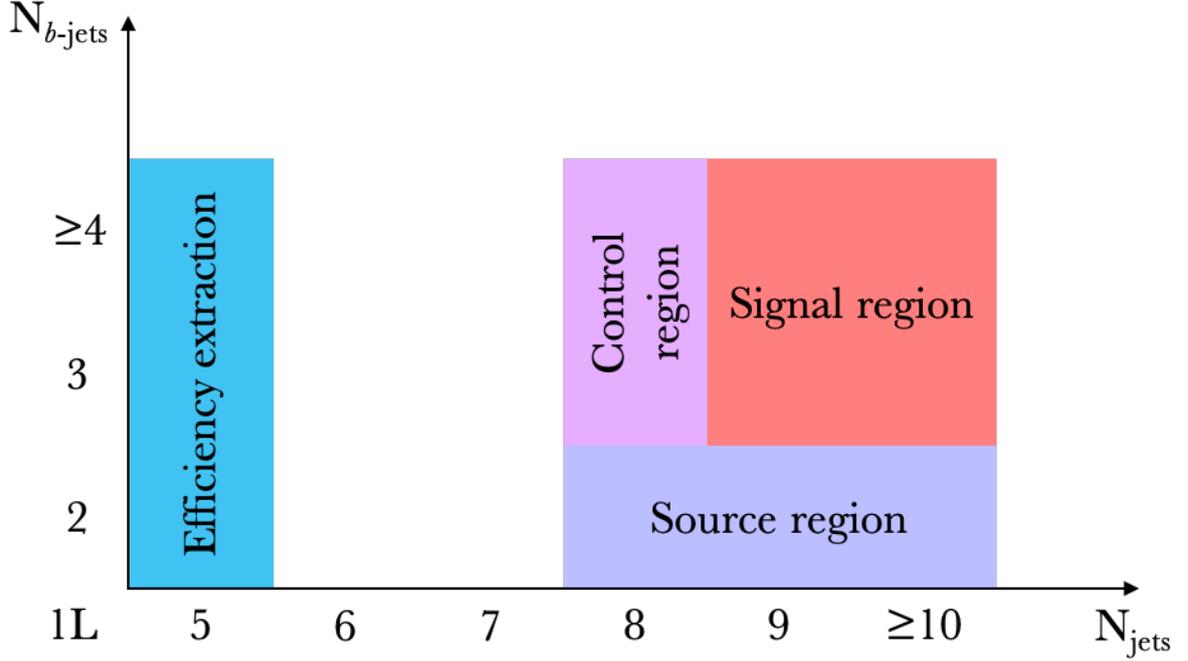
Figure 4.3: Schematic view of the regions definition.

The efficiencies are calculated as a function of jet variables, and propagated to the other variables straightforwardly. The $TRF_{t\bar{t}}$ prediction is calculated also for the $t\bar{t}$ Monte Carlo simulation, and the efficiencies are applied to the MC simulated events in source regions, deriving a pure MC correction factor: $MC_{corr} = t\bar{t}_{MC}/t\bar{t}_{MC}^{TRF_{t\bar{t}}}$. This ratio multiplies the $TRF_{t\bar{t}}$ prediction on data in the same region for the same predicted variable, acting as an effective reweighting correction. An illustrative example of the result of this procedure is displayed in Figure 4.4.

The efficiency extraction regions correspond to those with exactly five jets. The two jets with the highest b-tagging score are excluded in order to reduce the impact from the two b-jets from the $t\bar{t}$ decay. The efficiencies for the ≥ 4 b-jets regions are evaluated excluding also the third highest b-tagged jet. The efficiencies are parametrised in terms of jet p_T and $\Delta R_{jj}^{min} \times N_j$[3], looping over all the jets in data events, after the subtraction of the minor backgrounds contributions, estimated via the MC simulations. The b-tagging working point used for the estimation is MV2c10 77%. The efficiencies are calculated separately for the three campaigns.

Efficiencies for different working points applied to different jets are also extracted, used for the estimation of the pseudo-continuous b-tagging. For example, the notation 77% + 77% + 85% + 85% for the working point indicates that 77% working point is required to be passed for the leading two b-tagged jets (mostly coming from the $t\bar{t}$ decay), while 85% is required from the third b-jet in the event on. The efficiencies are used to promote events in data from source regions (those with exactly two b-jets) to higher b-jet multiplicity regions at the same jet multiplicity: the final $TRF_{t\bar{t}}$ weight estimates the probability of the event in 2b regions to be promoted to the 3b (or $\geq 4b$) regions by tagging one (or two) of the non-tagged jets. The $TRF_{t\bar{t}}$ prediction in the $\geq 4b$ regions sits on the $TRF_{t\bar{t}}$ prediction in the 3b regions.

---

[3]The variable ΔR_{jj}^{min} is the minimal distance in the $\eta - \phi$ plane for the given jet with respect to all the other jets in the event, weighted by the jet multiplicity.
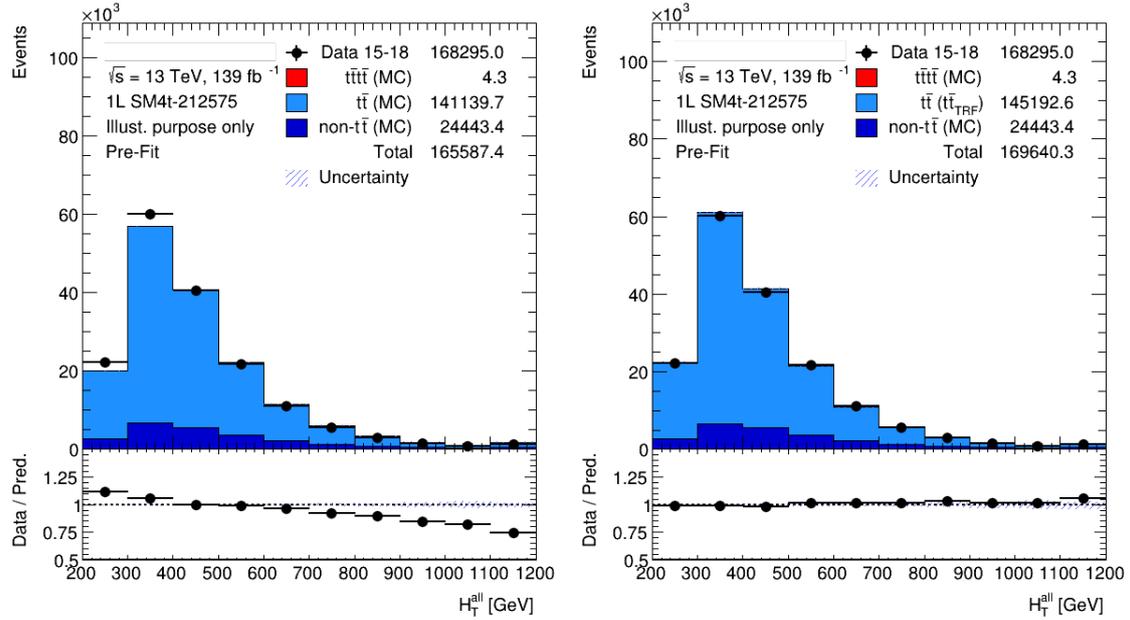
Figure 4.4: Illustrative example of the effect of the $TRF_{t\bar{t}}$ method in the region $5je3be$. On the left, the light-blue background is the simple MC prediction for $t\bar{t}$+jets. On the right, the same background is estimated using the $TRF_{t\bar{t}}$ method, resulting in a much better agreement with the data.

For every jet in the event, the probability of that jet to be b-tagged by different working points is calculated and a score is assigned based on these calculated probabilities. Given the new reassigned scores, the multiplicity of b-jet for any working point is recomputed. Therefore, the working point used for the $TRF_{t\bar{t}}$ estimation is independent from the one used for b-jet definition and region classification, allowing the prediction of any pseudo-continuous working point related variables.

The fully data-driven part of the method needs two assumptions: non-dependence of the efficiencies on jet multiplicity and absence of correlation between extra b-tagged jets. Any departure from such assumptions is corrected by the introduction of the MC correction factor. Using this correction factor it is also possible to exploit all the systematic uncertainties associated to the $t\bar{t}$ modelling to estimate the systematics on the $TRF_{t\bar{t}}$ prediction. The MC correction factors provide then better closure to the method.

In Table 4.2 the yields of the different processes separated in the signal regions are reported. For the $t\bar{t}$+jets background, the complete $TRF_{t\bar{t}}$ prediction is used.

	$9je3be$	$9je4bi$	$10ji3be$	$10ji4bi$
$t\bar{t}t\bar{t}$	23.0 ± 0.7	15.7 ± 0.8	24.7 ± 0.9	21.4 ± 1.2
$t\bar{t}$ ($TRF_{t\bar{t}}$)	5211.8 ± 1489.3	1434.3 ± 617.7	2503.9 ± 774.9	867.2 ± 459.2
Single top	303.5 ± 97.9	67.4 ± 23.6	106.1 ± 35.3	27.7 ± 10.5
$t\bar{t}$+X	137.6 ± 46.8	65.3 ± 22.4	66.0 ± 22.5	32.4 ± 11.1
V/VV+jets	143.7 ± 117.6	23.7 ± 19.5	73.3 ± 63.4	13.6 ± 12.0
Total	5819.6 ± 1497.9	1606.4 ± 618.9	2774.0 ± 778.0	962.3 ± 459.6
Data	5982	1426	2844	852

Table 4.2: Yields of the different processes in the four signal regions with the related uncertainties.

4.5 Uncertainties

Several sources of systematic uncertainty affect the measurement of the $t\bar{t}t\bar{t}$ cross section. Such uncertainties can be split in experimental and theoretical, and among the latter a special treatment is reserved to those related to the $t\bar{t}$+jets background prediction.

4.5.1 Experimental uncertainties

The uncertainty related to the luminosity of the full Run II is 1.7%. It is estimated using the LUCID-2 detector [35], following the same procedure already used for the Run I (2010-2012) [72]. The luminosity uncertainty (labelled as `Luminosity`) is applied to all the samples generated by the MC simulators. Moreover, to take into account the difference in pile-up distributions between data and MC simulations, a `Pile-up` uncertainty is applied to the MC samples. For leptons, some scale factors are applied in order to correct the mismatch between data and MC on reconstruction, identification, isolation and trigger performances. Such scale factors are estimated with the tag-and-probe method. More in details, the systematic uncertainties associated to electrons are `EL_SF_Trigger`, `EL_SF_Isol`, `EL_SF_Reco` and `EL_SF_ID`, in addition to those related to the energy scale (`EG_SCALE`) and energy resolution (`EG_RESOLUTION`). For muons, the uncertainties are split in statistical and systematic components: `MU_SF_Trigger_STAT`, `MU_SF_Trigger_SYST`, `MU_SF_Isol_STAT`, `MU_SF_Isol_SYST`, `MU_SF_ID_STAT`, `MU_SF_ID _SYST`, `MU_SF_TTVA_STAT` and `MU_SF_TTVA _SYST`, with the last two referred to track-to-vertex association (TTVA). Another group of uncertainties for muons is made of uncertainties related to inner detector track smearing (`MUON_ID`), muon spectrometer track smearing (`MUON_MS`), charge-independent scale momentum (`MUON_SCALE`), and charge-dependent scale momentum (`MU_SAGITTA_RESBIAS`). For jets, the JVT tool [73] is used to obtain the systematic uncertainty associated to the jet vertex tagging, varying up and down the JVT cut. This `JVT` uncertainty takes into account also the contamination from pile-up jets after pile-up suppression. For the uncertainties on the jet energy scale (JES), data from tests, simulations and collisions have been used [74]. This results in a set of 23 uncertainties, which are grouped as follows: 13 effective uncertainties labelled as `JET_EffectiveNP` (two detector-related, four modelling-related, three mixing both aspects, four statistical-related), two uncertainties related to η inter-

calibration labelled as `JET_EtaIntercalibration` (one modelling-related and one statistical-related), three uncertainties related to the flavour of the jet labelled as `JET_Flavor` and `JET_BJES`, four uncertainties related to pile-up subtraction labelled as `JET_Pileup`, and one uncertainty for non-closure for fast simulation labelled as `JET_RelativeNonClosure_AFII`. The jet energy resolution (JER) has been measured separately for data and MC using in situ techniques. The expected fractional jet p_T resolution is obtained using the JER tool as a function of the p_T and rapidity of the given jet. A systematic uncertainty is defined as the quadratic difference between the jet energy resolutions for data and MC simulation. A total of five independent uncertainties are used: four effective uncertainties labelled as `JET_JER_EffectiveNP` and one uncertainty labelled as `JET_JER_DataVsMC_AFII` accounting for the difference between data and simulation. For heavy flavour jets tagging, a total of 85 independent systematic variations are used. Each of them is obtained after the diagonalisation of the error matrix across every pairs of kinematic bins, which are used to derive heavy flavour efficiencies corrections. The obtained eigenvectors correspond to independent variations and include by construction the proper correlations across different kinematic regions. The b-jets identification efficiency is described by 45 parameters labelled as `FTAG_MV2c10_B[0-44]`, while the light-jets and c-jets tagging efficiencies are described by 20 parameters each, labelled as `FTAG_MV2c10_Light/C[0-19]`. Three independent systematic uncertainties are associated to the missing transverse energy. They are estimated using the ability of the simulation to model this observable in events without real E_T^{miss}, like $Z \to e^+e^-$. In that case, the presence of E_T^{miss} is due to resolution effects. This non-compensation is measured in data and compared to the simulations. The vectorial soft component is decomposed into a parallel and perpendicular component relatively to the object-based E_T^{miss}. The first uncertainty, labelled as `MET_SoftTrk_Scale`, is related to the scale of the parallel component, while the other two (labelled as `MET_SoftTrk_ResoPara` and `MET_SoftTrk_ResoPerp`) are related to the resolution of parallel and perpendicular components.

4.5.2 Non-$t\bar{t}$ backgrounds uncertainties

An uncertainty of 5% is applied to the cross section of single top production, which includes the t-channel, tW-channel and s-channel modes. For the $t\bar{t}$ production in association with a W, Z or H boson (denoted generally with $t\bar{t}+X$) an uncertainty of 15% on the cross sections of these processes is applied. A conservative uncertainty is associated to the cross sections of V+jets and minor processes: it includes a 5% from the normalisation of the cross section and an additional 24% normalisation uncertainty added in quadrature for each inclusive jet multiplicity bin. In total it amounts to $\sim$70%. Moreover, to all the non-$t\bar{t}$ background processes, an additional flat uncertainty of 30% is applied in order to cover the uncertainty on the choice of the generators.

4.5.3 $t\bar{t}$ background uncertainties

The uncertainties on $t\bar{t}$+jets process are evaluated comparing the nominal Powheg + Pythia simulation with three other MC simulations, covering the impact of the different modellings of the matrix elements, parton shower, hadronisation and radi-

ation. Each of the systematic uncertainties is chosen to be split by separating the effect on the different $t\bar{t}$+jets flavour contributions, namely $t\bar{t}+\geq 1b$ (if at least one jet is associated to a b-hadron not originated by the top quark decay), $t\bar{t}+\geq 1c$ (if at least one jet is associated to a c-hadron not coming from the W decay) and $t\bar{t}$+light (in all the other cases). On each of these contributions the fully $TRF_{t\bar{t}}$ method is applied, deriving the corresponding efficiencies and applying the proper MC correction factors. The uncertainty related to the choice of the matrix element generator is estimated by comparing the predictions of POWHEG and MADGRAPH5_AMC@NLO, and showering them with the same PYTHIA version. This procedure leads to three independent uncertainties (one for each flavour sub-process). For the uncertainty related to the parton shower and hadronisation modelling, the sample produced with POWHEG is matched with two different PS generators: PYTHIA and HERWIG. This adds three other independent uncertainties. Finally, three uncertainties come from the comparison of two samples generated with the same POWHEG + PYTHIA, but with different values for h_{damp} ($1.5 \cdot m_{top}$ and $3.0 \cdot m_{top}$), providing uncertainties on the radiation modelling. In addition, a flat uncertainty of 50% is applied to the normalisation of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ contributions separately to correct any mismodelling related to $t\bar{t}$ production in association with extra heavy flavour jets. This leads to two uncertainties.

Chapter 5

Multivariate Analysis

In the previous round of the $t\bar{t}t\bar{t}$ search [20], the main variable used for the discrimination between the $t\bar{t}t\bar{t}$ signal and the background was the scalar sum of the jet transverse momenta (H_T), since the signal events are characterised by higher values of H_T with respect to background (see Figure 5.1 for an example).

In this thesis, in order to improve the discrimination between the SM $t\bar{t}t\bar{t}$ signal and the main $t\bar{t}$+jets background, a multivariate analysis (MVA) is performed. In particular, a boosted decision tree (BDT) is trained using the TMVA framework [44] combining jets, b-tagging, leptons, missing energy and H_T information. The BDT is evaluated on all the samples described in the previous chapter and the output of this procedure (the *BDT score*) is used in the fit step of the analysis. Moreover, a Neural Network (NN) is trained with TMVA in order to have a comparison between two different methods.

5.1   MVA introduction

Every physics event is characterised by a multitude of quantities, like four-vectors of particles or energy deposits, which are commonly called *input variables* [75]. These variables are in general correlated and none of them is enough sensitive to provide alone a good separation between events that belong to the signal process and events that, instead, are background. For this reason it is necessary to treat the input variables in a fully multivariate way. The input variables describing an event can be seen as a vector of dimension d: $\vec{x} = (x_1, ..., x_d)$. The goal of the multivariate treatment is to construct a function $y = f(\vec{x})$ with properties that are useful for identifying patterns from the events. In practice, the target is to obtain an approximation of such function, $f(\vec{x}, \vec{w})$, where $\vec{w}$ is a set of parameters. The two main types of problems that are faced up with this approach are: the *classification* and the *regression*. In former type the target is to estimate a discrete value from the input variables, while in the latter one the goal is to estimate a continuous value.

The availability of vast amounts of data and the scientific progresses allowed the development of automated computer algorithms for learning from data, the so-called *machine learning*. It has origins in the pursuit of artificial intelligence, particularly in Frank Rosenblatt's creation of the *perceptron* around 1960 [76]. The primary goal of learning is to predict correctly the future data, and the main advantage of machine learning with respect to the conventional statistical approach is that an approximating function is inferred automatically from the given data without
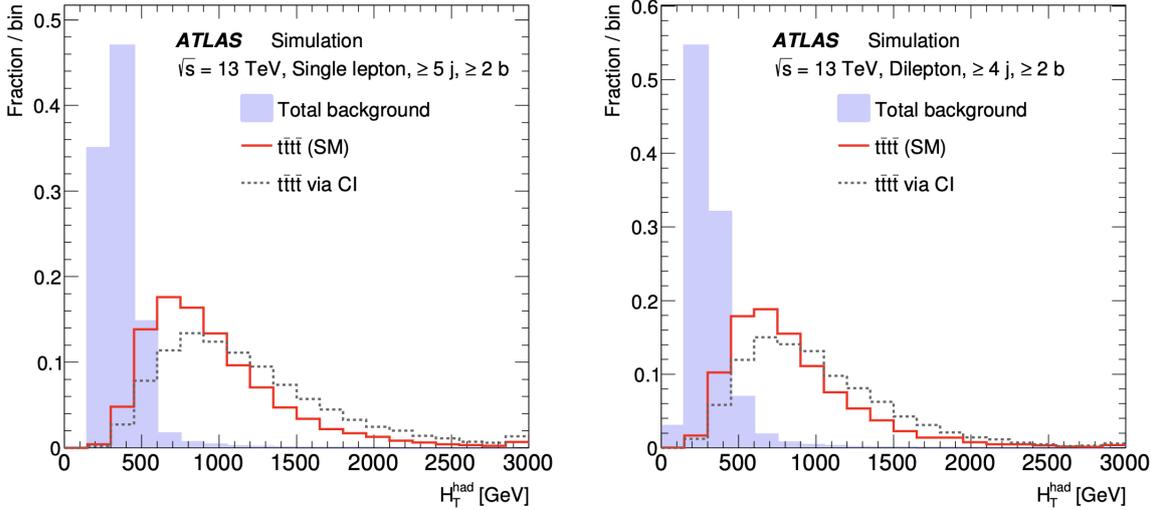
Figure 5.1: Examples of the H_T distributions in the first round of the analysis, after preselection, for the total predicted background with the $t\bar{t}$+jets background estimated via MC simulation (shaded histogram) and signals for the 1L ($5ji2bi$ region) and the 2LOS ($4ji2bi$ region) channels. The signals shown correspond to $t\bar{t}t\bar{t}$ production with SM kinematics (solid) and $t\bar{t}t\bar{t}$ production involving a four-fermion contact interaction (dashed). The distributions are normalised to unit area [20].

requiring a priori information about the function. In machine learning, the best approach to obtain $f(\vec{x}, \vec{w})$ is the so-called *supervised learning*, in which the training data set is formed by inputs and the corresponding outputs $\{\vec{x}, y\}$. This is because the real outputs y contain the information about the input-output relationship that the algorithm needs to learn. So, this *training* operation is necessary in order to find the best $\vec{w}$ parameters that model the input-output relationship. Since $f(\vec{x}, \vec{w})$ is only an approximation of y, the information loss has to be minimised during the training process. The information loss is quantified with a loss function $L(y, f)$. The machine learning algorithm minimises the average loss over the full training data set (the so-called *risk*) and not the loss event by event. The risk function $R(\vec{w})$ measures the cost of mistakes made in the predictions and finds the best $\vec{w}$ parameters. The empirical risk (an approximation of the true risk) is defined as the average loss over all (N) predictions:

$$R(\vec{w}) = \frac{1}{N} \sum_{i=1}^{N} L\left(y_i, f(\vec{x}_i, \vec{w})\right). \tag{5.1}$$

It is very common to use the mean squared error as risk function:

$$R(\vec{w}) = E(\vec{w}) = \frac{1}{N} \sum_{i=1}^{N} [y_i - f(\vec{x}_i, \vec{w})]^2. \tag{5.2}$$

The optimisation can also take into account any constraint $Q(\vec{w})$, simply adding it to the risk function to give the cost function:

$$C(\vec{w}) = R(\vec{w}) + \lambda Q(\vec{w}), \tag{5.3}$$

where λ is a parameter that determines the strength of the constraint imposed. The function $f(\vec{x}, \vec{w})$ obtained with this method converges to the function $f(\vec{x})$

that minimises the true risk. In practice, it is not always possible to find the absolute minimum, but only a local one. The constraint $Q(\vec{w})$ is used to have the control of the model complexity (*overfitting*) and is known as *regularisation*. The performance of the training is evaluated using a test data set, independent from the training set. Two other important approaches to learning are *unsupervised* and *reinforcement learning*. In the former approach, no targets are provided and the algorithm finds associations among the inputs. In the latter, correct outputs are rewarded and incorrect ones are penalised.

5.2   MVA methods

In the following, two MVA methods are going to be discussed in detail.

5.2.1   Neural networks

One of the most powerful and used MVA methods is the feed-forward neural network, also known as multilayer perceptron or artificial neural network. This method is vaguely inspired on the biological neural networks (see Figure 5.2, on the left). It consists of a collection of interconnected neurons arranged in layers (see Figure 5.2, on the right). Each neuron processes the input signal with an activation function, passing the result to the following layer. The first layer is the *input layer* and receives the input variables. It is followed by one or more *hidden layers*. The last layer outputs the final response of the network. Each interconnection is characterised by a weight, and each neuron may have a bias or a threshold. The weights and thresholds values are learned during the training phase. In general, the activation function is non-linear. Considering, for example, the neural network represented in Figure 5.2, the output is:

$$O(\vec{x}) = f(\vec{x}, \vec{w}) = g\left(\theta + \sum_j w_j h_j\right), \tag{5.4}$$

where g is the activation function and h_j are the outputs from the hidden layer:

$$h_j = g\left(\theta_j + \sum_i w_{ij} x_i\right). \tag{5.5}$$

There are multiple choices for the activation function, but the most used is the sigmoid:

$$g(x) = \frac{1}{1 + e^{-x}}. \tag{5.6}$$

This function is linear around $x = 0$, non-linear for higher values of x, and saturates for large values, mapping the input interval $(-\infty, +\infty)$ onto $[0, 1]$. Therefore, the linear model is a special case of this particular non-linear model. In general, the activation function of the output layer can be different from those of the hidden layers. The network weights and thresholds are determined by minimising the empirical risk function, commonly the mean square error.

   The choice of the neurons in the hidden layers depends on the density of data: too few neurons lead to underfitting, and too many lead to overfitting. To avoid the
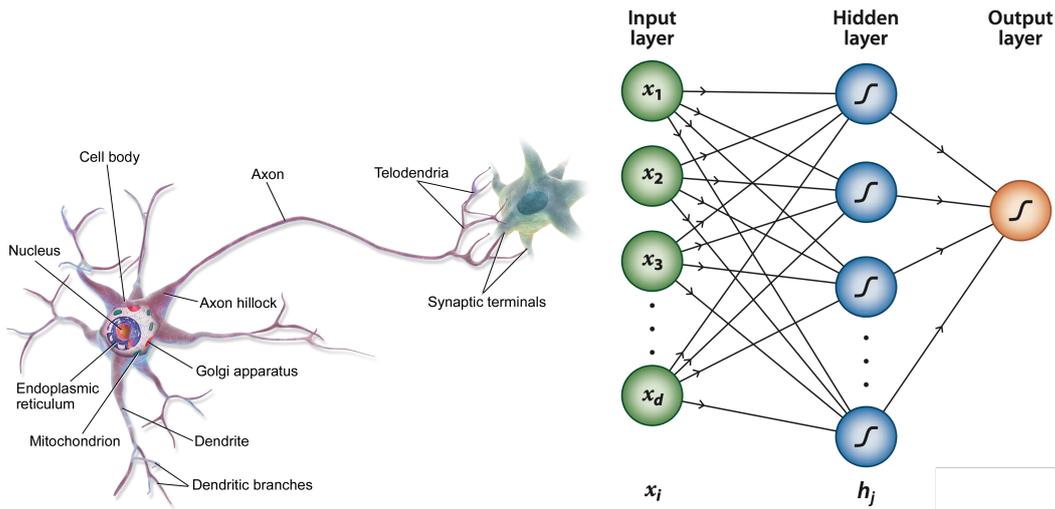
Figure 5.2: On the left, the anatomy of a real neuron. On the right, a schematic representation of a three-layer feed-forward neural network [75].

overtraining, there are many possibilities, like starting with large networks and then pruning connections or adding a penalty term to the risk function to control the complexity of the network. The most common algorithm for adjusting the weights that optimise the classification performance of a neural network is the so-called *back propagation*, where the desired output for every input event is known (so it is a supervised learning method). Back propagation is used by all neural networks in TMVA. Considering for simplicity a neural network with a single hidden layer with an activation function $g(x)$, and a linear activation function for the output layer, the output is:

$$O = \sum_{j=1}^{n_h} y_j^{(2)} w_{j1}^{(2)} = \sum_{j=1}^{n_h} g\left(\sum_{i=1}^{n_i} x_i w_{ij}^{(1)}\right) \cdot w_{j1}^{(2)}, \tag{5.7}$$

where n_i and n_h are the number of input and hidden neurons, $w_{ij}^{(1)}$ is the weight between input-layer neuron i and hidden-layer neuron j, and $w_{j1}^{(2)}$ is the weight between the hidden-layer neuron j and the output neuron. For each of the N training events, the neural network output O is computed and compared to the desired output $y \in [0,1]$ (in classification 1 for signal events and 0 for background events). The error function E is defined by:

$$E(\vec{x}_1, ..., \vec{x}_N | \vec{w}) = \sum_{a=1}^{N} E_a(\vec{x}_a | \vec{w}) = \sum_{a=1}^{N} \frac{1}{2}(O_a - y_a)^2, \tag{5.8}$$

where $\vec{w}$ is the vector of weights. The minimisation of the error function is done with the *gradient descent* method: starting from an initial set of weights $\vec{w}^{(\rho)}$, the weights are updated by moving a small distance in $\vec{w}$-space into the direction $-\vec{\nabla}_{\vec{w}} E$ where E decreases most rapidly,

$$\vec{w}^{(\rho+1)} = \vec{w}^{(\rho)} - \eta \vec{\nabla}_{\vec{w}} E, \tag{5.9}$$

where η is a positive parameter called *learning rate*. Since the target value is not known for the hidden nodes, the error has to be propagated from the output node

backwards to the hidden layer in order to perform the learning of the input-to-hidden weights, and the chain rule for differentiation is applied for the gradient descent. The weights connected with the output layer are updated by

$$\Delta w_{j1}^{(2)} = -\eta \sum_{a=1}^{N} \frac{\partial E_a}{\partial w_{j1}^{(2)}} = -\eta \sum_{a=1}^{N} (O_a - y_a) y_{j,a}^{(2)}, \tag{5.10}$$

and the weights connected with the hidden layers are updated by

$$\Delta w_{ij}^{(1)} = -\eta \sum_{a=1}^{N} \frac{\partial E_a}{\partial w_{ij}^{(1)}} = -\eta \sum_{a=1}^{N} (O_a - y_a) \frac{\partial g(x_{i,a}, w_{ij}^{(1)})}{\partial w_{ij}^{(1)}} \cdot w_{j1}^{(2)}. \tag{5.11}$$

5.2.2 Decision trees

Decision trees use sequential cuts to perform the classification of the events. At each step, the best cut is searched for and used to split the data, and this process is continued recursively on the resulting fractions until a given stop condition is fulfilled. A schematic representation of the process with a binary structure is shown in Figure 5.3. Decision trees start at the *root node*, with the full training data set containing signal and background events. At each iteration of the algorithm, and for each node, it finds the best cut (the cut that gives the largest reduction in impurity) and splits the data into two branch nodes. The stop to the splitting is done when no further reduction in impurity is possible (or when the number of events is too small to proceed further). A common choice for evaluating the impurity is the so-called *Gini index*:

$$G = (s + b)P(1 - P) = \frac{sb}{s + b}, \tag{5.12}$$

where $P = s/(s + b)$ is the signal purity, and s and b are the signal and background counts at any step respectively. The node that terminates the process is called *leaf*.

In principle, the splitting could continue until each leaf node contains only signal or only background events, which could suggest that perfect discrimination is achievable. However, such a decision tree would be strongly overtrained. To avoid overtraining a decision tree must be pruned. Pruning is the process of cutting back a tree from the bottom up after it has been built to its maximum size. Its purpose is to remove statistically insignificant nodes and thus reduce the overtraining of the tree.

Essentially, a decision tree creates M disjoint regions or a d-dimensional histogram with M bins of varying size, and a response value is assigned to each bin. As the training data set becomes arbitrarily large and as the bin sizes tend to zero, the predictions of the decision tree approach those of $f(\vec{x})$. The decision tree algorithm is applicable to discrimination of n classes and the concept can be extend from one tree to several trees which form a *forest*.

Decision trees are very popular because they are very simple to be implemented and understood. They have also advantages like insensitivity to irrelevant variables and invariance to one-to-one transformation of variables. On the other side, decision trees have some limitations like instability with respect to the training sample, suboptimal performance and poor global generalisation. Fortunately, such problems can be mitigated with ensemble learning techniques:
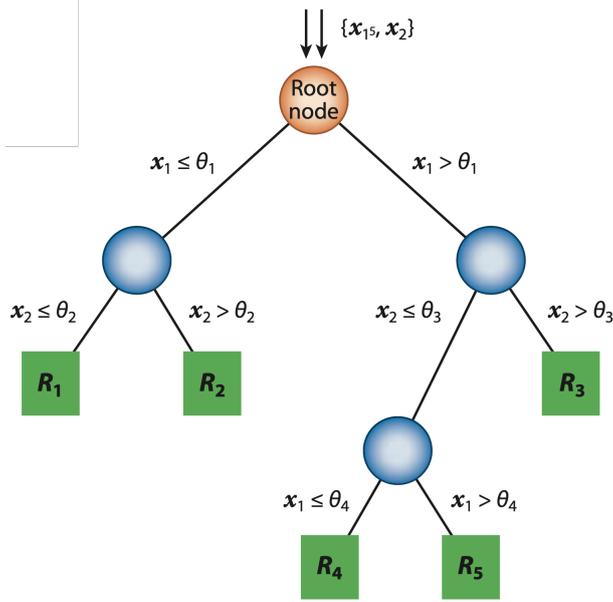
Figure 5.3: A schematic representation of a binary decision tree with two input variables x_1 and x_2 [75].

- boosting. Instead of seeking one high performance classifier, it is used an ensemble of classifiers, that collectively have a boosted performance. For an ensemble of N classifiers the prediction of the final one can be written as

$$f(\vec{x}, \vec{w}) = \sum_{n=1}^{N} \alpha_n y_n(\vec{x}, \vec{w}_n), \qquad (5.13)$$

where $\vec{w}_n$ are the parameters of the n-th classifier and α_n are the weighting coefficients. These coefficients depend from the particular algorithm used. The two most popular algorithms are *AdaBoost* (adaptive boost) and *GradientBoost*. When applied to decision tree it gives rise to the so-called boosted decision tree. As the boosting algorithms perform best on weak classifiers, pruning is not necessary;

- bagging. This is a simple average of the outputs of N predictors, where each is trained on a different randomly selected subset, drawn from a training sample of M events. In the previous equation for boosting, $\alpha_n = 1/N$ is the case of bagging;

- random forest. As already mentioned, when many decision trees are trained they origin a forest. If each one is trained on a randomly chosen subset of variables at each split, they provide a random forest. The output for each event is the average output of all trees. Further randomisation can be introduced through the use of bootstrap samples as in the case of bagging.

The gradient boosting and bagging are the ensemble learning techniques which are used in the BDT trained in this thesis, which is presented in the following section.

Option	Value	Description
NTrees	1000	Number of trees in the forest
MaxDepth	1	Max depth of the decision tree allowed
MinNodeSize	2.5%	Minimum percentage of training events required in a leaf node
nCuts	20	Number of grid points in variable range used in finding optimal cut in node splitting
BoostType	Grad	Boosting type for the trees in the forest
Shrinkage	0.1	Learning rate for GradBoost algorithm
UseBaggedBoost	True	Use only a random (bagged) subsample of all events for growing the trees in each iteration
BaggedSampleFraction	0.5	Relative size of bagged event sample to original size of the data sample
SeparationType	GiniIndex	Separation criterion for node splitting

Table 5.1: Summary of the TMVA options used in the training of the BDT.

5.3 BDT with TMVA

TMVA, acronym for Toolkit for Multivariate Analysis [44], provides a ROOT-integrated [77] environment for the processing, parallel evaluation and application of multivariate classification and multivariate regression techniques. All multivariate techniques implemented in TMVA are supervised learning. They make use of training events, to determine the mapping function that either describes a decision boundary (classification) or an approximation of the underlying functional behaviour defining the target value (regression). The mapping function can contain various degrees of approximations and may be a single global function, or a set of local models. The methods presented in the previous section are implemented in TMVA with a large selection of options. TMVA is specifically designed for the needs of high energy physics applications, and so it is a natural choice for the purpose of this thesis. The simplicity and the good performances of BDT are the keys of the choice of this method for the multivariate analysis. In particular, the MVA method uses a BDT with 1000 trees, the boosting type is the GradientBoost with a learning rate (shrinkage) of 0.1. A bagging procedure is also applied (using half of the original size of the sample for growing the trees in each iteration). The evaluation of the impurity makes use of the Gini index. Such options are summarised also in Table 5.1.

The training is done in four different regions separately: $9je3be$, $9je4bi$, $10ji3be$ and $10ji4bi$. The signal sample is provided by the MC simulation of the SM $t\bar{t}t\bar{t}$ process at NLO in QCD (MADGRAPH_AMC@NLO+PYTHIA), while the background is the $TRF_{t\bar{t}}$ prediction without the MC correction factor. The events with negative weights are not ignored, but fully exploited in the training. The events provided are split in two subsamples by odd or even event number: one is used for the real training, while the other is exploited for testing the overtraining. Then the two subsamples are swapped and the procedure is repeated. The BDT is evaluated in each sample (data and backgrounds) in order to associate a BDT score to each event and separate the SM $t\bar{t}t\bar{t}$ signal from the $t\bar{t}$+jets background, carefully taking the correct BDT according to the event number. For the training 16 variables are used as input:

- overall activity in the transverse plan. The $t\bar{t}t\bar{t}$ production leads to central collision with a more important transverse activity compared to $t\bar{t}$+jets. In this case the variables considered are the scalar sum of the p_T of all objects (H_T^{all}), the p_T of the most energetic jet, and the centrality of the event ($\sum p_T / \sum E$);

- b-jet information. The $t\bar{t}t\bar{t}$ production leads to more energetic b-jet with respect to $t\bar{t}$+jets, where additional b-jets come from radiations. The variables taken into account are the average invariant mass across all the triplets of b-tagged jets, the minimum invariant mass among all pairs of b-tagged jets, the minimum ΔR among all pairs of b-jets, the minimum ΔR among all pairs of b-jet and a lepton and the sum of the pseudo-continuous b-tagging score of the six leading jets ranked in MV2c10 score;

- jet information. The $t\bar{t}t\bar{t}$ production leads to more energetic jets with respect to $t\bar{t}$+jets for the same reason of above, and the radiation structure can be exploited using jet-to-jet correlations. For this reason the number of jets is used. Also the average ΔR across all pairs of jets and the invariant mass of the triplet of jets that has the minimum ΔR^1 are exploited;

- RC-jet information. Since $t\bar{t}t\bar{t}$ leads to more transverse objects, some of the produced top quarks could be boosted enough to be contained in RC-jets, while $t\bar{t}$+jets produces less of this boosted topology. The considered variables related to RC-jets are the number of such jets with a mass larger than 100 GeV, and the sum of the first (and second) k_t splitting scale d_{12} (d_{23}) of all RC-jets, where d_{ij} is defined as in Equation 3.1;

- E_T^{miss} and lepton-related information. The W boson from top quark decay should be more transverse in $t\bar{t}t\bar{t}$ than in $t\bar{t}$+jets, so the missing transverse energy (E_T^{miss}) and the W reconstructed transverse mass $m_T(\ell, E_T^{miss})$ are the chosen variables.

Figure 5.4 shows the separation of each variable in the $10ji4bi$ region before the splitting in even/odd events, i.e. the power a variable has to distinguish between signal and background, and it does not depend from the method used. Figure 5.5 shows the ranking of the input variables in the $10ji4bi$ region for the training done on even events. The ranking is done by TMVA according to the separation. From these plots it is clear that the most discriminating variables are the sum of the pseudo-continuous b-tagging score, those related to the RC-jets and those about transverse quantities.

Table 5.2 collects the number of signal and background events in each region in which the MVA is performed. Such events are split in two halves, one for the training and the other for testing. Signal and background events used for the training are further normalised in order to have the same amount of events in each of the two samples. The integral of the ROC curve is also reported. The ROC (Receiver Operating Characteristic) curve is a measure of how well a signal can be measured against a background. The integral of such curve measures in percentage how much signal is correctly obtained: the closer to 1, the better the modelling of the BDT output.

---

[1]The ΔR value of a triplet of jets is defined as $\Delta R_{ijk} = \sqrt{\Delta R_{ij}^2 + \Delta R_{ik}^2 + \Delta R_{jk}^2}$ where, i, j and k represent the indices of the three jets.

Region	Signal events	Background events	ROC (even)	ROC (odd)
9je3be	254420	229328	0.795	0.804
9je4bi	171140	229328	0.800	0.809
10ji3be	371736	109048	0.781	0.777
10ji4bi	320702	109048	0.778	0.781

Table 5.2: Summary of the number of events in each region before the splitting, with the ROC curve integral.

Figure 5.6 shows the overtraining test of the BDT in the $10ji4bi$ region for the training done on even events: a good agreement between points (trained samples) and curves (test samples) indicates a good training with low overtraining. In the right tail, the BDT shows some overtraining in the samples, given the different values of the trained points with respect to the test curves.

The plots shown in this chapter corresponding to the other training regions are collected in Appendix C.

For comparison purposes, a neural network is also trained with TMVA and the details are discussed in Appendix D.
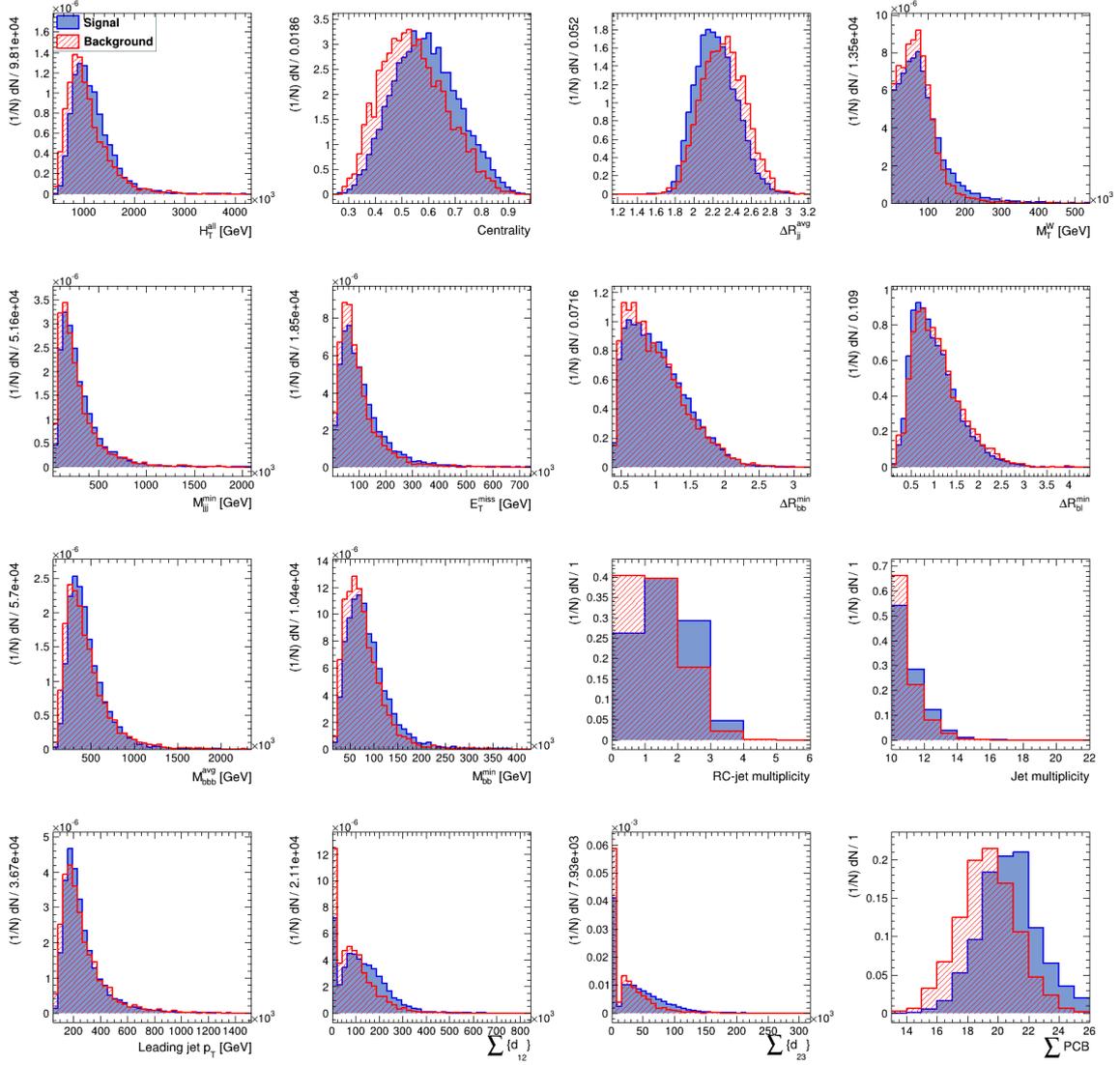
Figure 5.4: Separation of the variables used for the training in the $10ji4bi$ region.
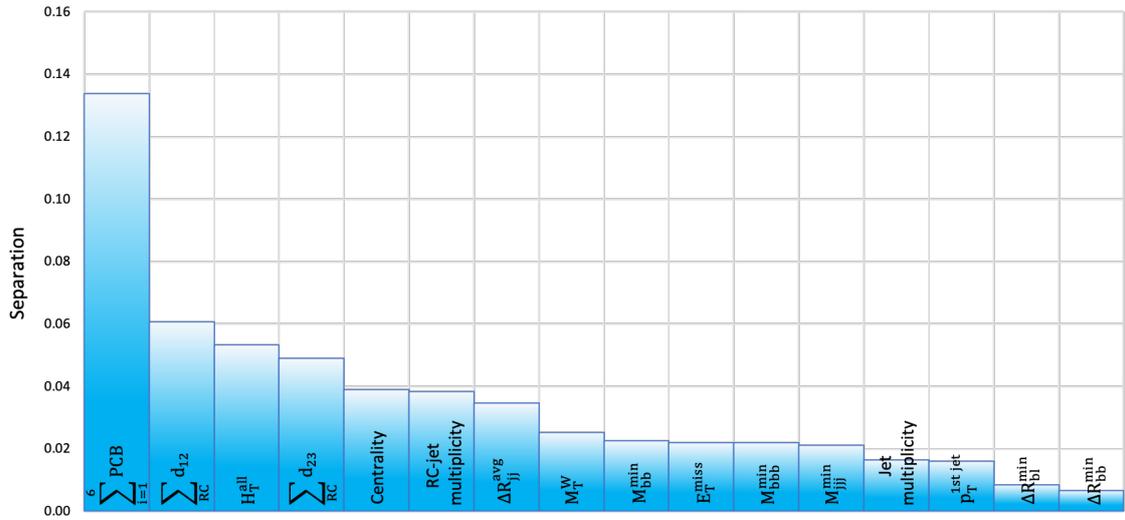
Figure 5.5: Ranking of input variables in terms of separation in the $10ji4bi$ region for the training done on even events according to TMVA.
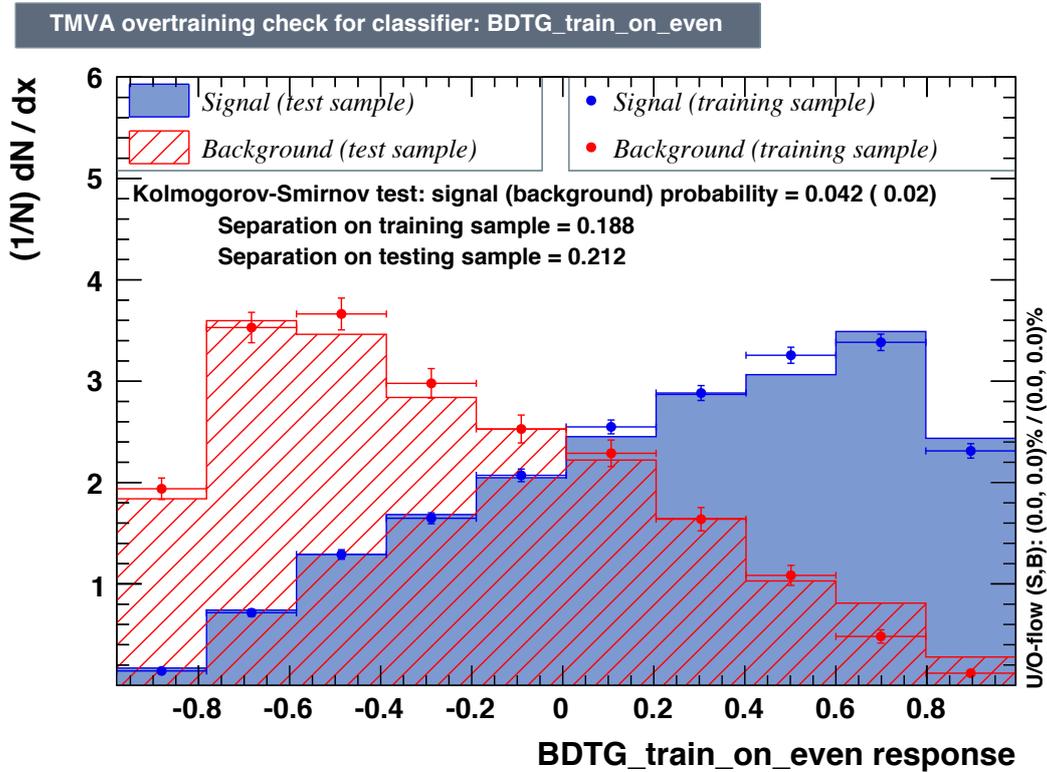
Figure 5.6: Training response for the BDT in the $10ji4bi$ region for the training done on even events.

Chapter 6

Data analysis

The aim of this thesis is the search for SM $t\bar{t}t\bar{t}$ production at LHC in the 1L channel, using data and the MC simulations described previously, and following the selection and classification procedure presented in Chapter 3. The main $t\bar{t}$+jets background has been estimated with the $TRF_{t\bar{t}}$ method as explained in Chapter 4 and the MVA procedure described in Chapter 5 has been used. For the purpose of the following analysis, two control regions and four signal regions are exploited: $8je3be$ and $8je4bi$ are the control regions, while $9je3be$, $9je4bi$, $10ji3be$ and $10ji4bi$ are the signal regions. In the control regions, given the very low number of expected signal events, no MVA has been trained, and the discriminating variable used is H_T, while in the signal regions the output of the MVA is used. Since the signal is expected to be very small, a "blind" analysis strategy is adopted: the optimisation and the systematic model must be established without fitting to data in signal regions, and fits to data in background dominated regions have to be performed in order to validate the fit model. The final result is obtained performing a fit only in the four signal regions just defined.

6.1 Profile likelihood fit

The statistical analysis is based on a binned likelihood, as described in the following [78]. For the sake of clarity, let us consider a fit done in only one region with a signal and a single background: the number of signal events is S, while the number of background events is B. The signal and background "shapes" are denoted as $f_S(x)$ and $f_B(x)$, being x the discriminating variable. These are probability density functions normalized so that $\int f(x)dx = 1$. The *parameter of interest* is the signal-strength μ, which is the ratio between the observed signal and the SM expectation: $\mu = 0$ corresponds to the background-only hypothesis, while $\mu = 1$ corresponds to the nominal signal-plus-background hypothesis. For a given data set $\{x_1, ..., x_n\}$, the probability of obtaining n events when $\mu S + B$ are expected is modelled by the Poisson distribution: $\mathcal{P}_\lambda(x) = \frac{\lambda^n}{n!}e^{-\lambda}$, with $\lambda = \mu S + B$. Indeed, the probability density depends also on $f_S(x)$ and $f_B(x)$. So then, the probability of μ given the data set (the *likelihood function*) is:

$$L(\mu) = \mathcal{P}(\{x_1, ..., x_n\}|\mu) = Pois(n|\mu S + B) \left[\prod_{i=1}^{n} \frac{\mu S f_S(x_i) + B f_B(x_i)}{\mu S + B} \right]. \quad (6.1)$$

The maximum of $L(\mu)$ provides an estimator for μ. The formula can be simplified taking the negative logarithm:

$$-\ln L(\mu) = (\mu S + B) + \ln n! - \sum_{i=1}^{n} \ln \left[\mu S f_S(x_i) + B f_B(x_i) \right]. \qquad (6.2)$$

In this analysis the distributions are binned in histograms. For the j-th bin of width Δ_j, $f_S(x_j)$ and $f_B(x_j)$ are replaced by $\frac{\nu_j^S}{S\Delta_j}$ and $\frac{\nu_j^B}{B\Delta_j}$, where ν_j^S and ν_j^B are the bin content for the signal and the background, respectively. In this case, the likelihood in terms of the individual bins is:

$$\mathcal{P}(n_b|\mu) = Pois(n_{tot}|\mu S + B) \left[\prod_j \frac{\mu \nu_j^S + \nu_j^B}{\mu S + B} \right] = \mathcal{N}_{comb} \prod_j Pois(n_b|\mu \nu_j^S + \nu_j^B), \quad (6.3)$$

where $\mathcal{N}_{comb}$ is a combinatorial factor taking into account the permutations in a binned sample and can be neglected since it is constant.

The analysis makes use of multiple regions (r) and samples (s). The uncertainties are included in the fit as nuisance parameters (NPs), and they can be constrained or unconstrained (ϕ_{rs}). The statistical uncertainties are constrained NPs (γ_{jrs}) and change the shape of the distribution (bin-by-bin). The systematic uncertainties are constrained NPs ($\boldsymbol{\theta}$) and can produce normalisation ($\eta_{rs}(\boldsymbol{\theta})$) and shape ($\sigma_{jrs}(\boldsymbol{\theta})$) variations of the samples. The luminosity uncertainty is considered as a parameter aside (λ), which affects equally all the samples. The expected number of events in a given bin in a region is then:

$$\nu_{jr}(\phi, \theta, \gamma) = \sum_s \lambda \gamma_{jrs} \phi_{rs}(\boldsymbol{\theta}) \eta_{rs}(\boldsymbol{\theta}) \sigma_{jrs}(\boldsymbol{\theta}), \qquad (6.4)$$

and the likelihood function becomes:

$$\mathcal{P}(n, \boldsymbol{a}|\phi, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \prod_r \prod_j Pois(n_{jr}|\nu_{jr}) Gauss(L_0|\lambda, \Delta_\lambda) \prod_{p \in \theta} f_p(a_p|\theta_p), \qquad (6.5)$$

where n is the total number of events, n_{jr} is the number of events in a bin in a region, $\boldsymbol{a}$ are the values for $\boldsymbol{\theta}$, L_0 is the nominal luminosity, and $f_p(a_p|\theta_p)$ is a constraint term describing an auxiliary measurement a_p that constrains the nuisance parameter θ_p and is usually assumed to be a normal Gaussian. The effect of the variation of the single NP on shape and normalisation of the distributions are given from independent measurements for variations of ± 1 standard deviation from zero. The effect on the distributions for intermediate values is obtained by linear interpolation. The likelihood is maximised to estimate the values of NPs and μ from data.

The statistical analysis is then based on a binned likelihood function $\mathcal{L}(\mu, \boldsymbol{\theta})$ constructed as a product of Poisson probability terms over all bins considered in the search, those in the four signal regions. The signal-strength parameter μ is the $t\bar{t}t\bar{t}$ cross section normalised to the SM prediction. The nuisance parameters are implemented in the likelihood function as Gaussian constraints and allow variations of the expectations for signal and background according to the corresponding systematic uncertainties, and their fitted values correspond to the deviations from the nominal expectations that globally provide the best fit to the data. This procedure allows a

reduction of the impact of systematic uncertainties on the search sensitivity. Individual sources of systematic uncertainty are considered to be uncorrelated. Correlations of a given systematic uncertainty are maintained across processes. The statistical uncertainty of the prediction, which incorporates the statistical uncertainty of the MC events, is included in the likelihood as a single bin-by-bin γ parameter. For optimisation reasons, nuisance parameters corresponding to systematic uncertainties with an effect smaller than 0.5% on a given sample are removed for that sample (this procedure is referred to as "pruning"). A smoothing algorithm is applied to the systematic variations. This is done in order to avoid bin-by-bin fluctuations resulting from statistical fluctuations. Therefore, the algorithm iteratively rebins the effect of the systematic variation, as long as the number of counted slope changes becomes less than a certain number (four for this analysis).

6.2   Studies on control regions

In order to validate the data-driven $t\bar{t}$+jets background estimation, the agreement between data and total background prediction is assessed in the different control regions before any fit to data. Figure 6.1 shows the effect of background modelling with the $TRF_{t\bar{t}}$ method. As already discussed previously, the pure MC simulation is not able to model well the main $t\bar{t}$+jets background at high jet and b-jet multiplicity. In fact, the disagreement between the data and the simulation is particularly evident: in the low H_T region the simulation underestimates data, while in the high H_T region it is the opposite. The $TRF_{t\bar{t}}$ method (with the MC correction factor) allows to cope a big part of the mismodelling: in the $8je3be$ region it can recover an excellent agreement, within the statistical uncertainty, while in the $8je4bi$ region some tension is left. Such remaining disagreement is indeed covered implementing the systematic uncertainties defined in Section 4.5. The result of this operation is shown in Figure 6.2: in the $8je3be$ region the total uncertainty is around 25% along all the H_T spectrum, while it varies between 25% and 35% in the $8je4bi$ region. The biggest contributions to the uncertainty come from the choice of the generators for the $t\bar{t}$+jets process: in particular, each flavour sub-process in parton shower and radiation modelling introduces an uncertainty of ∼10%. The bigger impact of the PS with respect to the ME is due to the fact that the matrix element calculation for the $t\bar{t}$ process produces only four jets (five at NLO). Thence, all the other jets are produced by the parton shower generator, which models their rates only at leading logarithmic accuracy. Figure 6.3 displays the comparison of the expected yields before the fit.
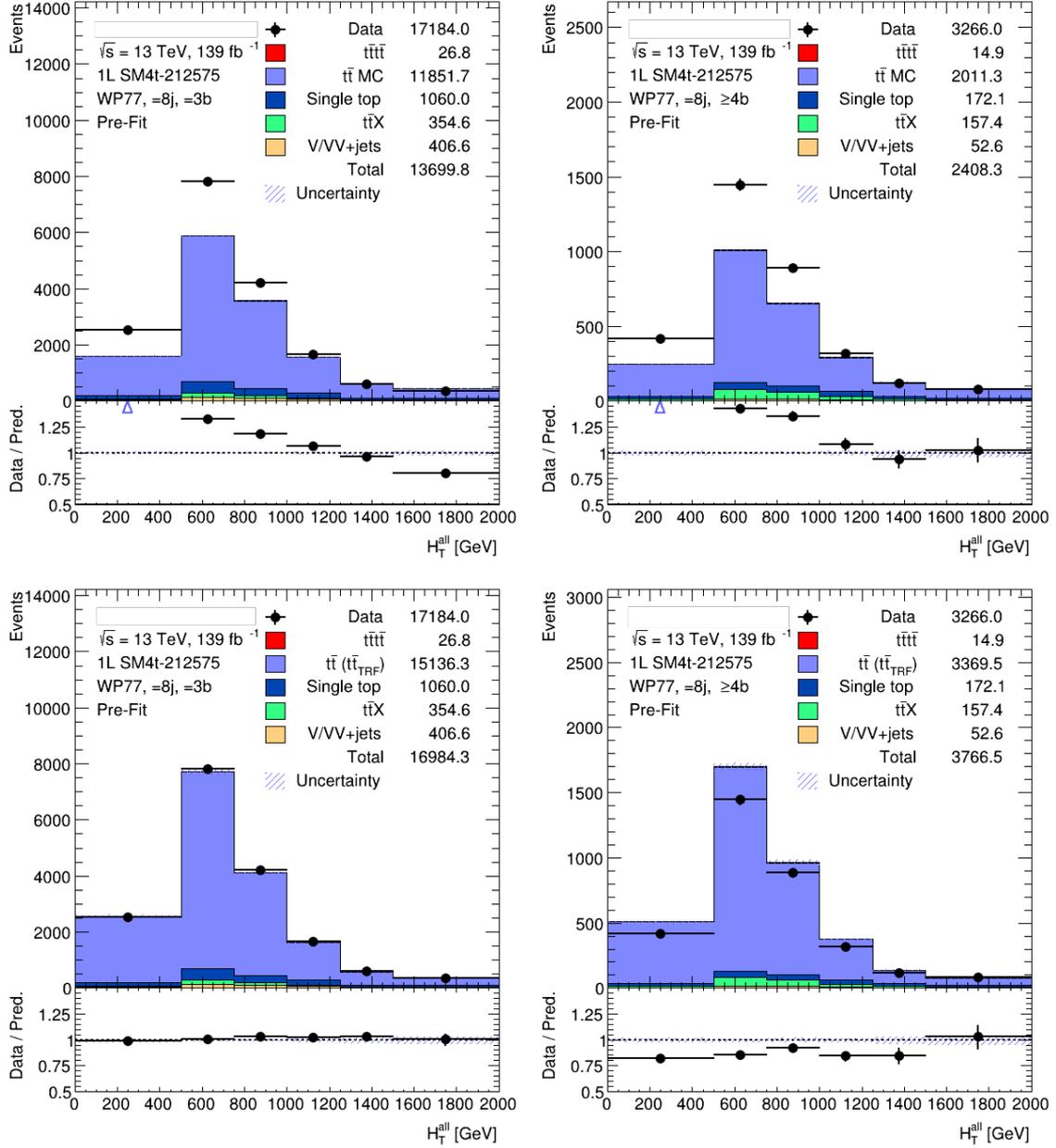
Figure 6.1: Comparison of the H_T distribution between the prediction from pure MC simulated $t\bar{t}$+jet events (upper plots) and from $TRF_{t\bar{t}}$ method (lower plots) applied in the control regions. The hashed area represents the statistical uncertainty on the background prediction. The last bin in all figures contains the overflow. The lower panels show the bin-by-bin ratios of data with respect to the predictions.
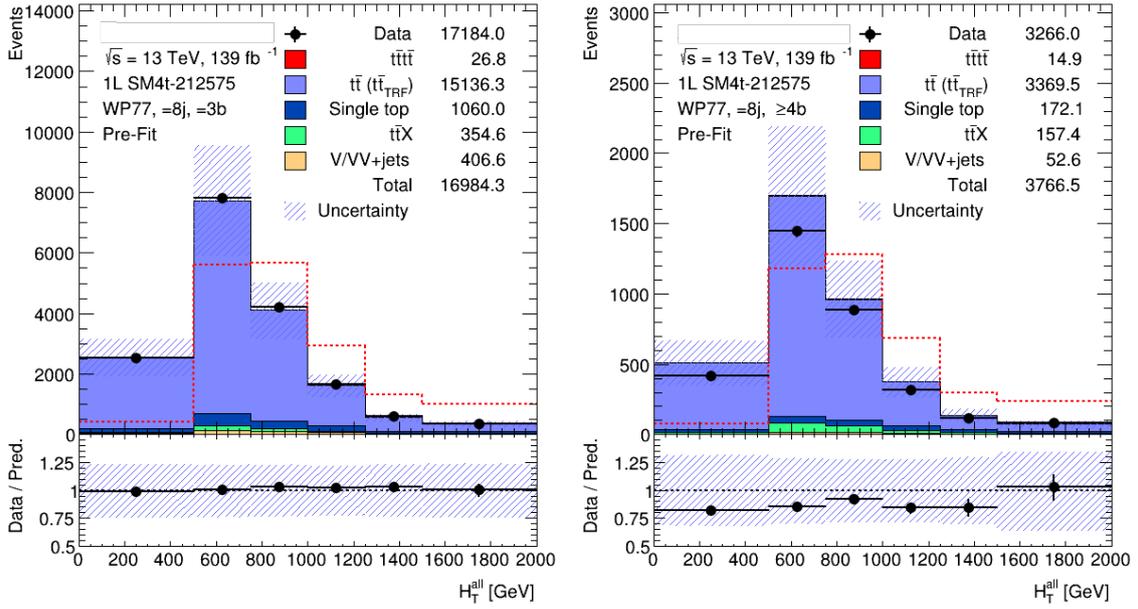
Figure 6.2: Comparison between data and prediction of the H_T distributions in the control regions prior to the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction. The last bin in all figures contains the overflow.
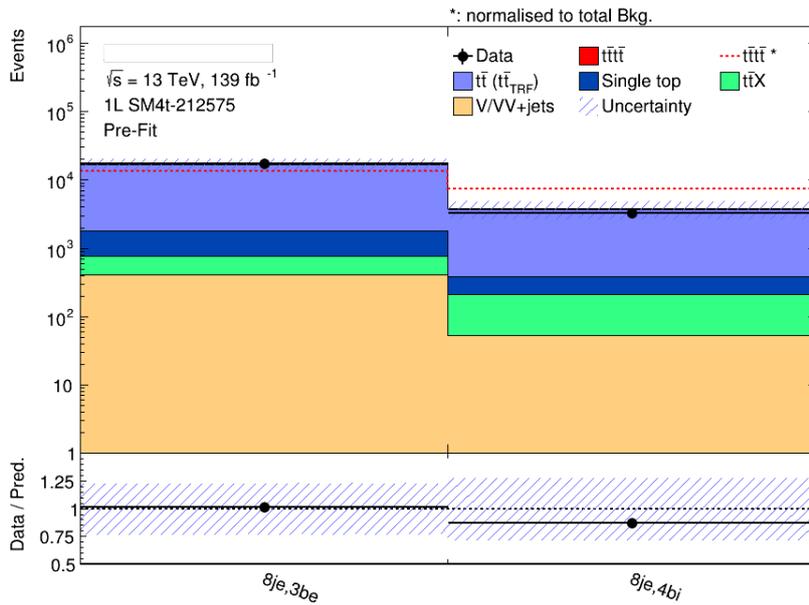
Figure 6.3: Comparison of prediction to data in control regions prior to the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction. The bottom panel displays the ratio of data to the total prediction.

6.2.1 Fit on control regions

Before looking the signal regions, the binned profile likelihood fit used for the search of four top quark production signal is performed in the control regions in order to inspect the expected constraints on the nuisance parameters in the fit to data. The fit is performed to the H_T distribution simultaneously in the control regions with the signal-plus-background hypothesis and the results are shown in Figures 6.4 and 6.5. Figure 6.6 displays the constraints on the non-pruned nuisance parameters from this fit. No strong constraints are observed: the most constrained NP is the $t\bar{t}+ \geq 1b$ shape uncertainty on the choice of the matrix element generator, which is constrained to 80%. Various pulls are observed on the nuisance parameters associated to the theoretical uncertainties on the $t\bar{t}$+jets background estimation. The nuisance parameter related to the choice of the parton shower generator on $t\bar{t}+ \geq 1b$ is pulled by a factor of $(58 \pm 87)\%$. The nuisance parameter related to the choice of the matrix element generator on $t\bar{t}$+light is pulled by a factor $(-59 \pm 84)\%$. The nuisance parameters related to the uncertainty on the $t\bar{t}+ \geq 1b$ and $t\bar{t}+ \geq 1c$ cross section are pulled by $(-18 \pm 97)\%$ and $(45 \pm 93)\%$, respectively. A part for the nuisance parameter related to the jet flavour composition, which is pulled by $(28 \pm 97)\%$, no other significant pull is observed on the nuisance parameters associated to detector and jet-tagging uncertainties. The nuisance parameters are also ranked according to their impact on the fitted value of the signal strength, when fixed to their pre- and post-fit errors. The resulting ranking plot, showing at the same time the impact on μ and the constraints of the 20 nuisance parameters with the highest impact on the fitted value of μ, is reported in Figure 6.7, confirming the importance of the uncertainties coming from the $t\bar{t}$+jet modelling. The fitted signal strength results to be $\mu = 4.68^{+19.30}_{-22.11}$, compatible with $\mu = 0$ (background-only hypothesis), as expected.
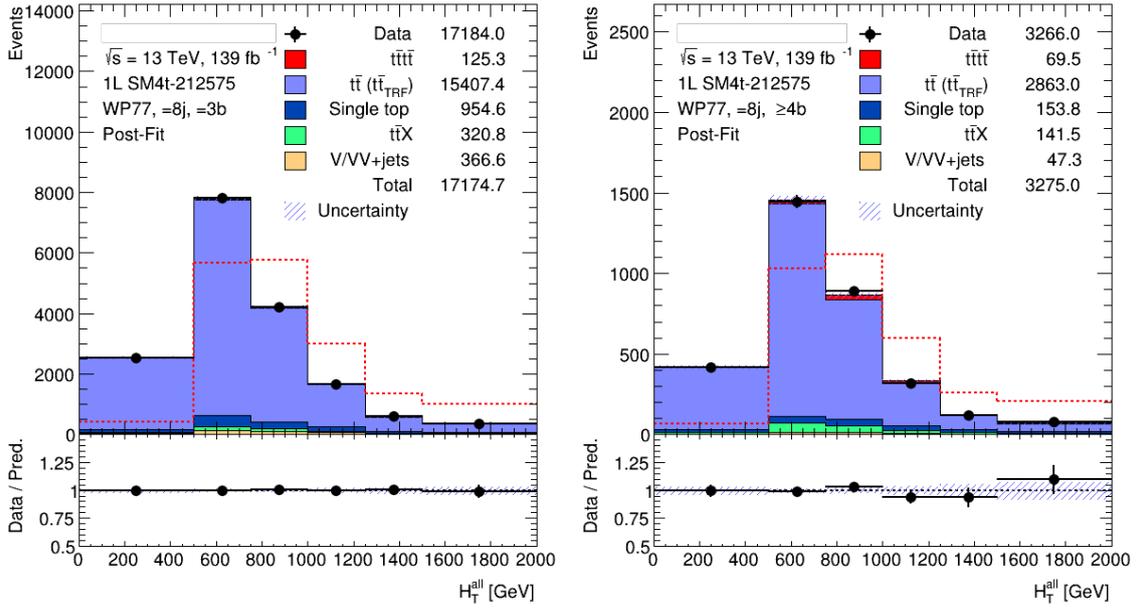
Figure 6.4: Comparison between data and prediction of the H_T distributions in the control regions after the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction. The last bin in all figures contains the overflow.
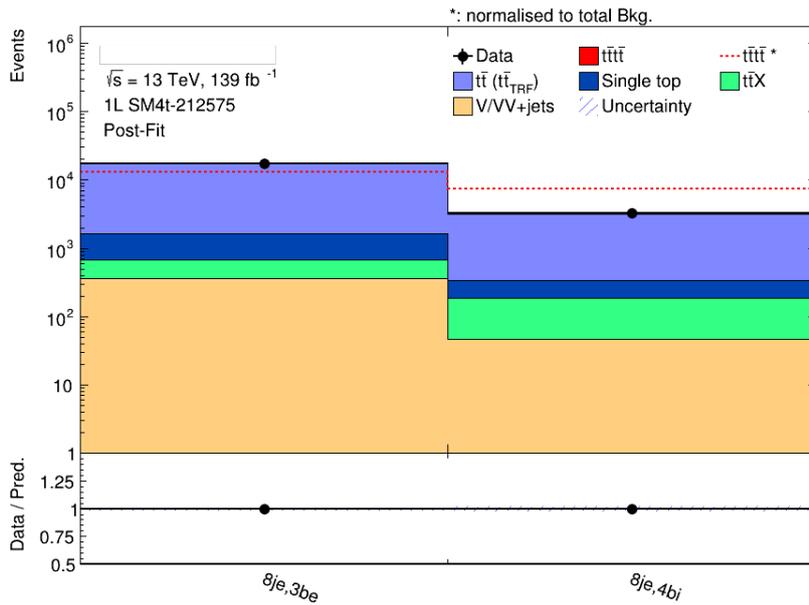
Figure 6.5: Comparison of prediction to data in control regions after the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction. The bottom panel displays the ratio of data to the total prediction.

$\bar{tt} + \geq 1c$ cross section
$\bar{tt} + \geq 1b$ cross section
$\bar{tt} +$ light Radiation Shape
$\bar{tt} +$ light Radiation Norm.
$\bar{tt} + \geq 1c$ Radiation Shape
$\bar{tt} + \geq 1c$ Radiation Norm.
$\bar{tt} + \geq 1b$ Radiation Shape
$\bar{tt} + \geq 1b$ Radiation Norm.
$\bar{tt} +$ light Parton Shower Shape
$\bar{tt} +$ light Parton Shower Norm.
$\bar{tt} + \geq 1c$ Parton Shower Shape
$\bar{tt} + \geq 1c$ Parton Shower Norm.
$\bar{tt} + \geq 1b$ Parton Shower Shape
$\bar{tt} + \geq 1b$ Parton Shower Norm.
$\bar{tt} +$ light Generator Shape
$\bar{tt} +$ light Generator Norm.
$\bar{tt} + \geq 1c$ Generator Shape
$\bar{tt} + \geq 1c$ Generator Norm.
$\bar{tt} + \geq 1b$ Generator Shape
$\bar{tt} + \geq 1b$ Generator Norm.
$\bar{tt}$ METSoftTrkResoPerp
$\bar{tt}$ METSoftTrkResoPara
Luminosity
$\bar{tt}$ JETPileupRhoTopology
$\bar{tt}$ JETPileupOffsetNPV
$\bar{tt}$ JETFlavorResponse
$\bar{tt}$ JETFlavorComposition
$\bar{tt}$ JETEtaIntercalibrationTotalStat
$\bar{tt}$ JETEtaIntercalibrationModelling
$\bar{tt}$ JETEffectiveNPStatistical2
$\bar{tt}$ JETEffectiveNPModelling1
$\bar{tt}$ JETEffectiveNPMixed1
$\bar{tt}$ JETBJESResponse
$non - \bar{tt}$ generator
V/VV+jets cross section
Pile-up
JVT
$\bar{tt} + X$ cross section
Single top cross section
EL SF ID
FTAG MV2c10 Light9
FTAG MV2c10 Light8
FTAG MV2c10 Light7
FTAG MV2c10 Light6
FTAG MV2c10 Light5
FTAG MV2c10 Light4
FTAG MV2c10 Light3
FTAG MV2c10 Light2
FTAG MV2c10 Light1
FTAG MV2c10 Light12
FTAG MV2c10 Light11
FTAG MV2c10 Light10
FTAG MV2c10 Light0
FTAG MV2c10 C6
FTAG MV2c10 C5
FTAG MV2c10 C4
FTAG MV2c10 C3
FTAG MV2c10 C1
FTAG MV2c10 C18
FTAG MV2c10 C15
FTAG MV2c10 C0
FTAG MV2c10 B9
FTAG MV2c10 B7
FTAG MV2c10 B6
FTAG MV2c10 B5
FTAG MV2c10 B4
FTAG MV2c10 B41
FTAG MV2c10 B3
FTAG MV2c10 B39
FTAG MV2c10 B38
FTAG MV2c10 B36
FTAG MV2c10 B35
FTAG MV2c10 B2
FTAG MV2c10 B25
FTAG MV2c10 B23
FTAG MV2c10 B22
FTAG MV2c10 B21
FTAG MV2c10 B1
FTAG MV2c10 B18
FTAG MV2c10 B17
FTAG MV2c10 B14
FTAG MV2c10 B0

-2 -1 0 1 2
$(\hat{\theta}-\theta_0)/\Delta\theta$

Figure 6.6: Fitted nuisance parameters under the signal-plus-background hypothesis in the control regions only.
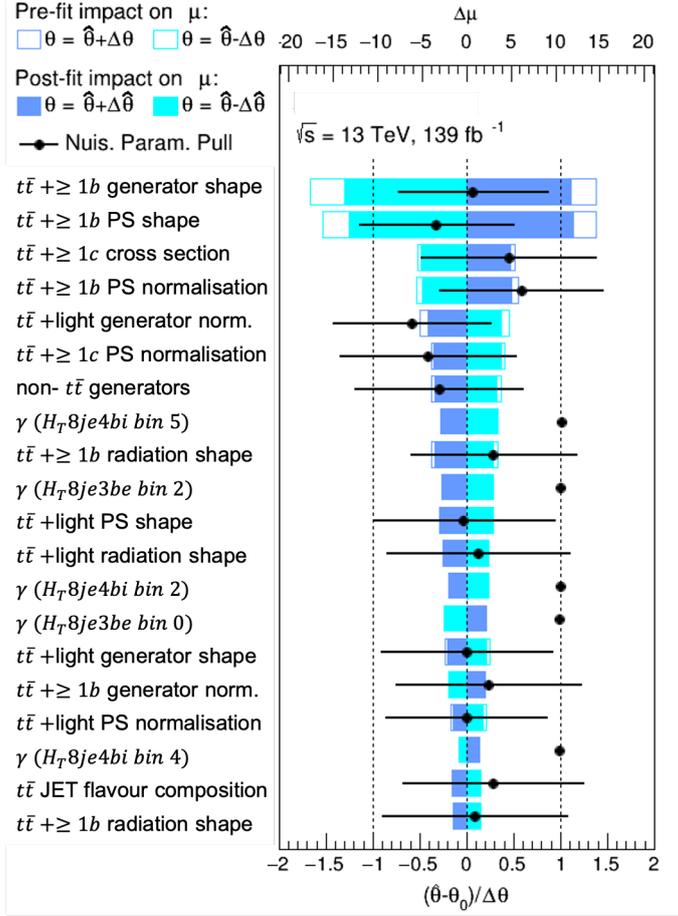
Figure 6.7: Ranking plot from the fit to data in control regions under the signal-plus-background hypothesis. The γ parameters refer to statistical uncertainty on the background model and are assumed to be Poisson-distributed NPs.

6.2.2 BDT shape in control regions

In the control regions no BDT can be trained due to the low presence of signal. However, the MVA trained in the signal regions can be evaluated in the control regions to check the agreement between data and prediction in "signal-free" regions, i.e. without spoiling the blind analysis. Figure 6.8 shows the distribution of the MVA output in the control regions, before any fit. It is important to remark that such regions are not included in the final fit to data, which will be presented in the following sections. In particular, the BDT exploited for these plots is trained on even events in $10ji3be$ and $10ji4bi$ regions. As expected, the BDT classifies the events mostly as background (score < 0), with a rapidly decreasing slope. The agreement between data and prediction reflects what was already evident from the H_T distributions presented in Figure 6.2: in the $8je3be$ region there is an excellent agreement, while in the $8je4bi$ region there is some tension (order of 15%), which is in any case covered by the uncertainty band.
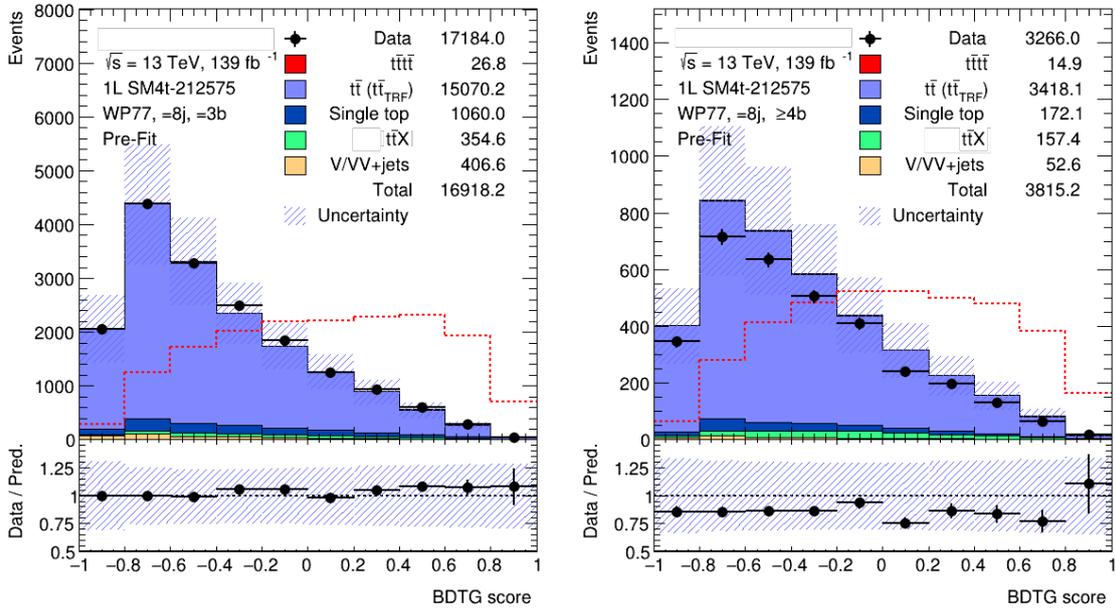
Figure 6.8: Comparison between data and prediction of the BDT distributions in the control regions prior to the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.

6.3 Studies on Asimov data set

In order to inspect the expected constraints on the nuisance parameters in the fit to data, before having the signal regions unblinded, a fit under the signal-plus-background hypothesis is performed on the Asimov data set, i.e. a pseudo-data set in which the observed values are set to the expected ones[1]. Only the signal regions are included in the fit: $9je3be$, $9je4bi$, $10ji3be$ and $10ji4bi$. Figures 6.9 and 6.10 show the distribution of the MVA output in the signal regions, before and after the fit, respectively. Before the fit, the total uncertainty varies from $\sim 25\%$ to $\sim 45\%$. As for the control regions, the biggest contributions to the uncertainty come from the choice of the generators for the $t\bar{t}$+jet background: depending from the region considered, some flavour sub-processes can introduce an uncertainty even greater than 20% (the biggest impact belongs to the $t\bar{t}$+light sub-process in the parton shower choice, which contributes with the 25.8% in the $10ji4bi$ region). The constraints on the nuisance parameters from the fit are shown in Figure 6.11. There is no strong constraint coming from the nuisance parameters associated to detector and jet-tagging uncertainties. Instead, some constraints are observed on the nuisance parameters associated to the theoretical uncertainties on the $t\bar{t}$+jet background estimation: the $t\bar{t}+ \geq 1b$ shape and normalisation uncertainties on the choice of the parton shower generator are constrained to 64% and 58%, the $t\bar{t}$+light shape uncertainty on the choice of the parton shower generator is constrained to 51%, and also the $t\bar{t}$+light shape uncertainty on the choice of the ME generator is constrained to 65%. Several other nuisance parameters related to these uncertainties are constrained to the range from 77% to 91%. The nuisance parameters are also ranked according to their impact on the fitted value of the signal strength, when fixed to their pre- and post-fit errors. The resulting ranking plot, showing at the same time the impact on μ and the constraints of the 20 nuisance parameters with the highest impact on the fitted value of μ, is reported in Figure 6.12: as already evidenced, the most important systematic uncertainties are coming from the choice of the PS generator for the $t\bar{t}$+jet background estimation. The fitted signal strength results to be $\mu = 1.00^{+1.92}_{-1.93}$. This result provides information about the uncertainty on the μ value from the fit to data, which is then expected to be ~ 1.9.

Before the unblinding, the same fit has been performed in different subsets of regions, in order to determine the sensitivity of the regions in terms of the uncertainty on the fitted μ. The results are reported in Table 6.1. Two different paths are followed. The first starts from the region with higher (expected) sensitivity, $10ji4bi$, adding the others one by one to see how much sensitivity is gained. The second, instead, starts from the less (expected) sensitive region, $9je3be$. This important step is needed to establish which regions are safe to be unblinded before the full unblinding, in order not to spoil the result of the official 1LOS channel ATLAS analysis. The criterion for this partial unblinding is based on the result of the previous round of the analysis [20], in which the fitted μ was $1.7^{+1.9}_{-1.7}$. From this measurement, the signal is smaller than $\mu = 3.6$ at 1σ. Assuming a fitted value of 3.6, it would be still compatible with 0 (the background-only hypothesis), within 1σ significance. The regions can be unblinded, starting from the less sensitive one, as long as the uncertainty on μ is greater than 3.6, still preserving the "blindness" of the full fit.

---

[1]From the short story "Franchise" by Isaac Asimov, where the election's results are determined by a single person, which results to have the average characteristics of the entire population.
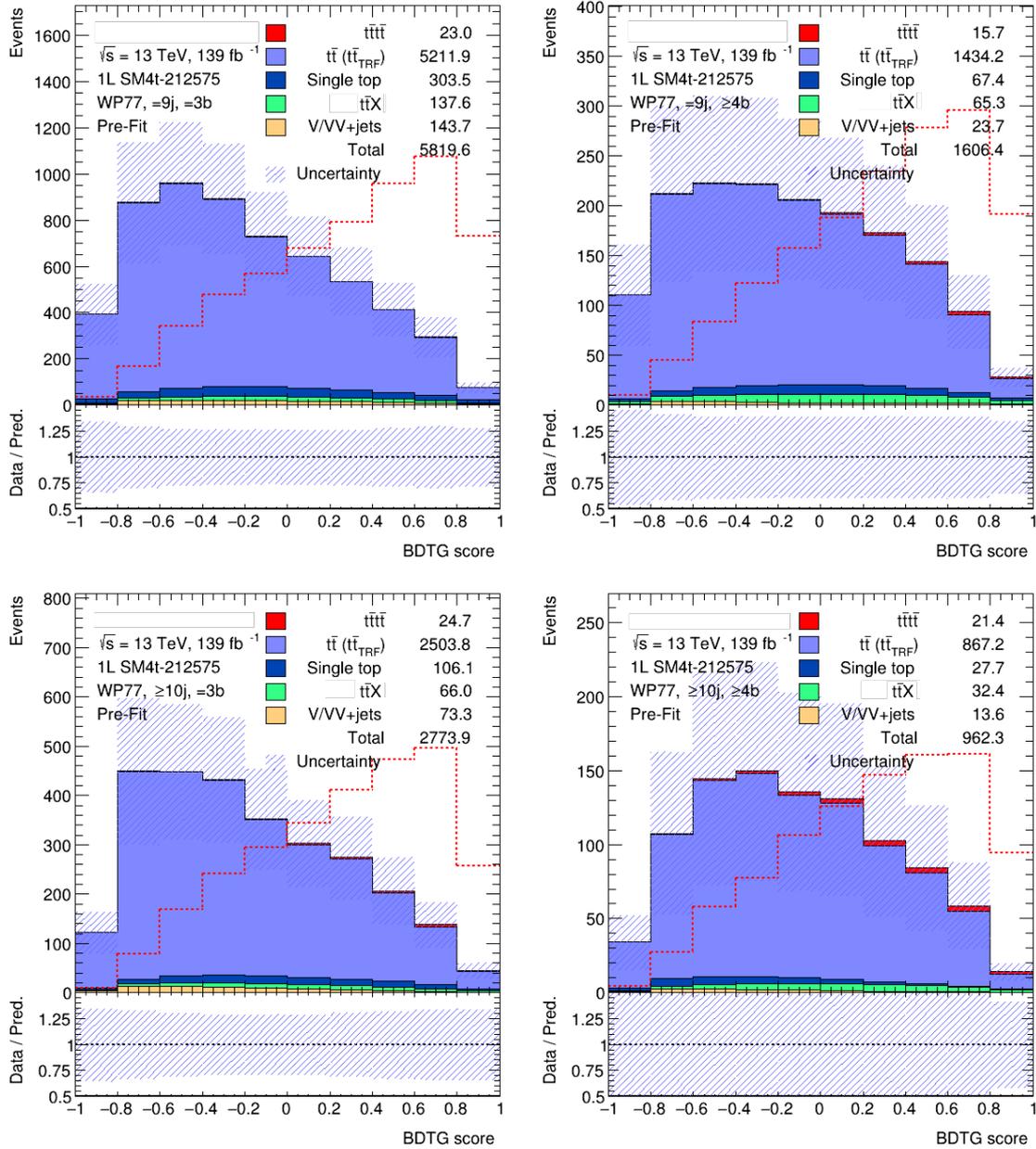
Figure 6.9: Comparison between Asimov data set and prediction of the BDT distributions in the signal regions prior to the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.
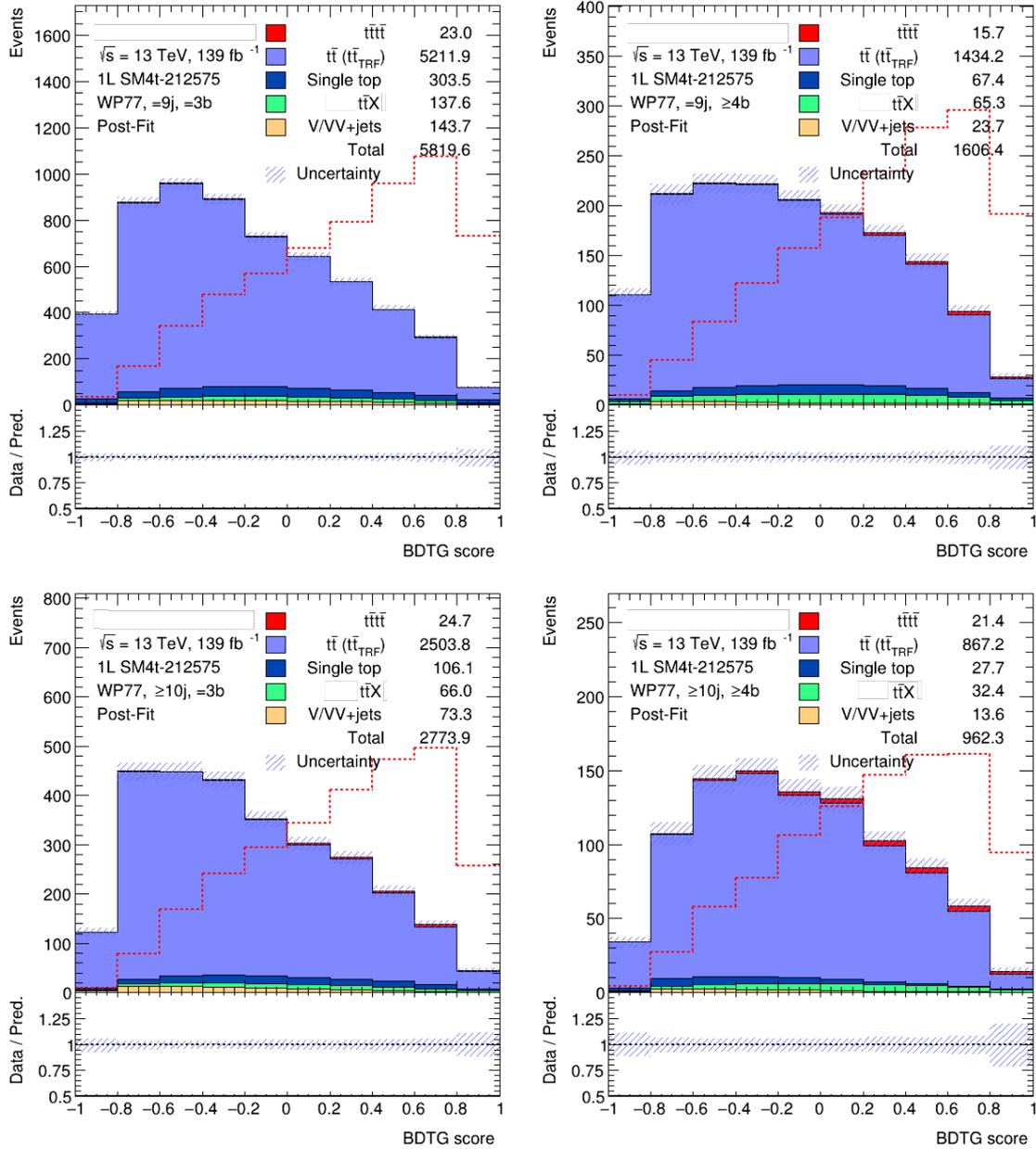
Figure 6.10: Comparison between Asimov data set and prediction of the BDT distributions in the signal regions after the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.
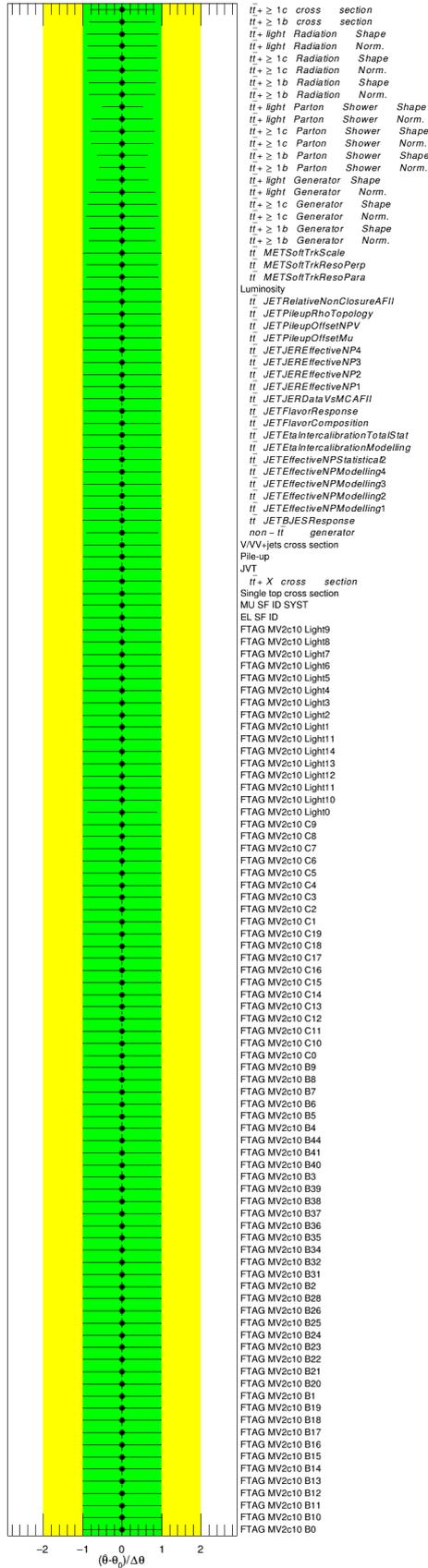
Figure 6.11: Fitted nuisance parameters on Asimov data set under the signal-plus-background hypothesis in the signal regions only.
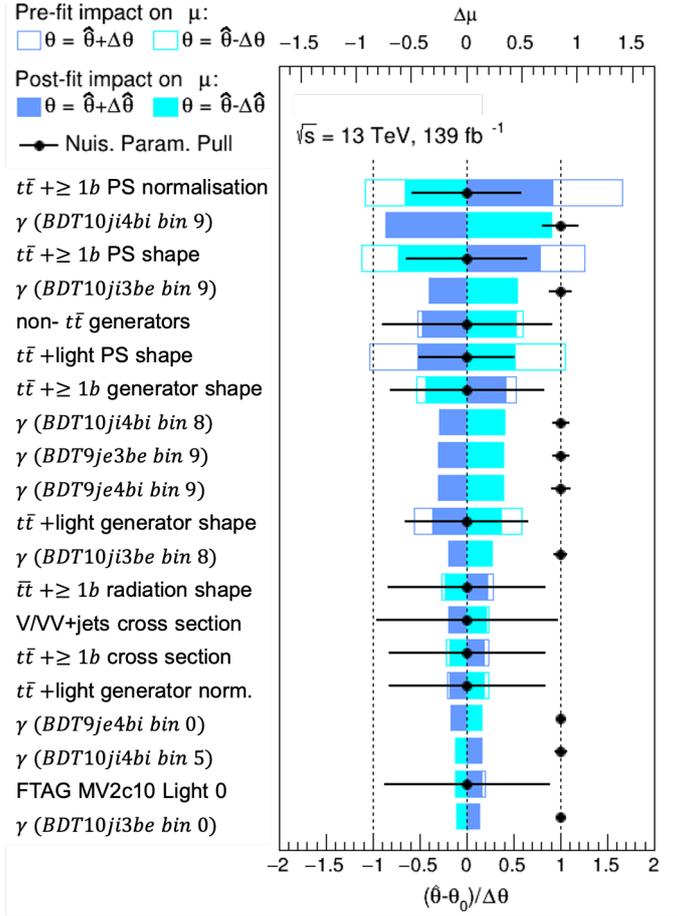
Figure 6.12: Ranking plot from the fit on Asimov data set under the signal-plus-background hypothesis. The γ parameters refer to statistical uncertainty on the background model and are assumed to be Poisson-distributed NPs.

Regions	Fitted μ
$10ji4bi$	$\mu = 1.00^{+2.48}_{-2.56}$
$10ji4bi+10ji3be$	$\mu = 1.00^{+2.30}_{-2.37}$
$10ji4bi+10ji3be+9je4bi$	$\mu = 1.00^{+2.00}_{-2.03}$
$10ji4bi+10ji3be+9je4bi+9je3be$	$\mu = 1.00^{+1.92}_{-1.93}$
$9je3be$	$\mu = 1.00^{+5.69}_{-5.69}$
$9je3be+10ji3be$	$\mu = 1.00^{+3.98}_{-4.20}$
$9je3be+9je4bi$	$\mu = 1.00^{+3.49}_{-3.51}$
$9je3be+9je4bi+10ji3be$	$\mu = 1.00^{+2.99}_{-3.03}$

Table 6.1: Results of the sensitivity tests performed in the different subsets of signal regions.

6.3.1 Studies on RC-jets splitting

An additional test on Asimov data set is done following the previous analysis recipe and splitting the signal regions according to the different multiplicity of the RC-jets with $p_T > 200$ GeV (J): $0J$, $1J$, and $\geq 2J$. The Asimov fit is then performed simultaneously in the 12 signal regions. Figures 6.13 and 6.14 show the distribution of the MVA output before and after the fit, respectively. The constraints on the nuisance parameters from the fit are shown in Figure 6.15. As in the case of the previous Asimov fit, without the splitting in J multiplicity, there is no strong constraint coming from the nuisance parameters associated to detector and jet tagging uncertainties, while some constraints are observed on the nuisance parameters associated to the theoretical uncertainties on the $t\bar{t}$+jet background estimation. The fitted signal strength results to be $\mu = 1.00^{+1.62}_{-1.62}$. Indeed, with this setup, some of the modelling systematics get smoothed so much that they lose completely any shape effect (see for example Figure 6.16, where the shape effect is dropped by the smoothing). Removing the smoothing procedure allows to cope this problem (see Figure 6.17), but the uncertainty on the value of μ increases to ~ 2.0. More sophisticated solutions to the problem could be a rebinning or a dedicated smoothing procedure, in order to exploit the benefit of the splitting in J multiplicity. However these possibilities are beyond the aim of this thesis, and thence the splitting in J multiplicity is chosen to be not used in the rest of the analysis.
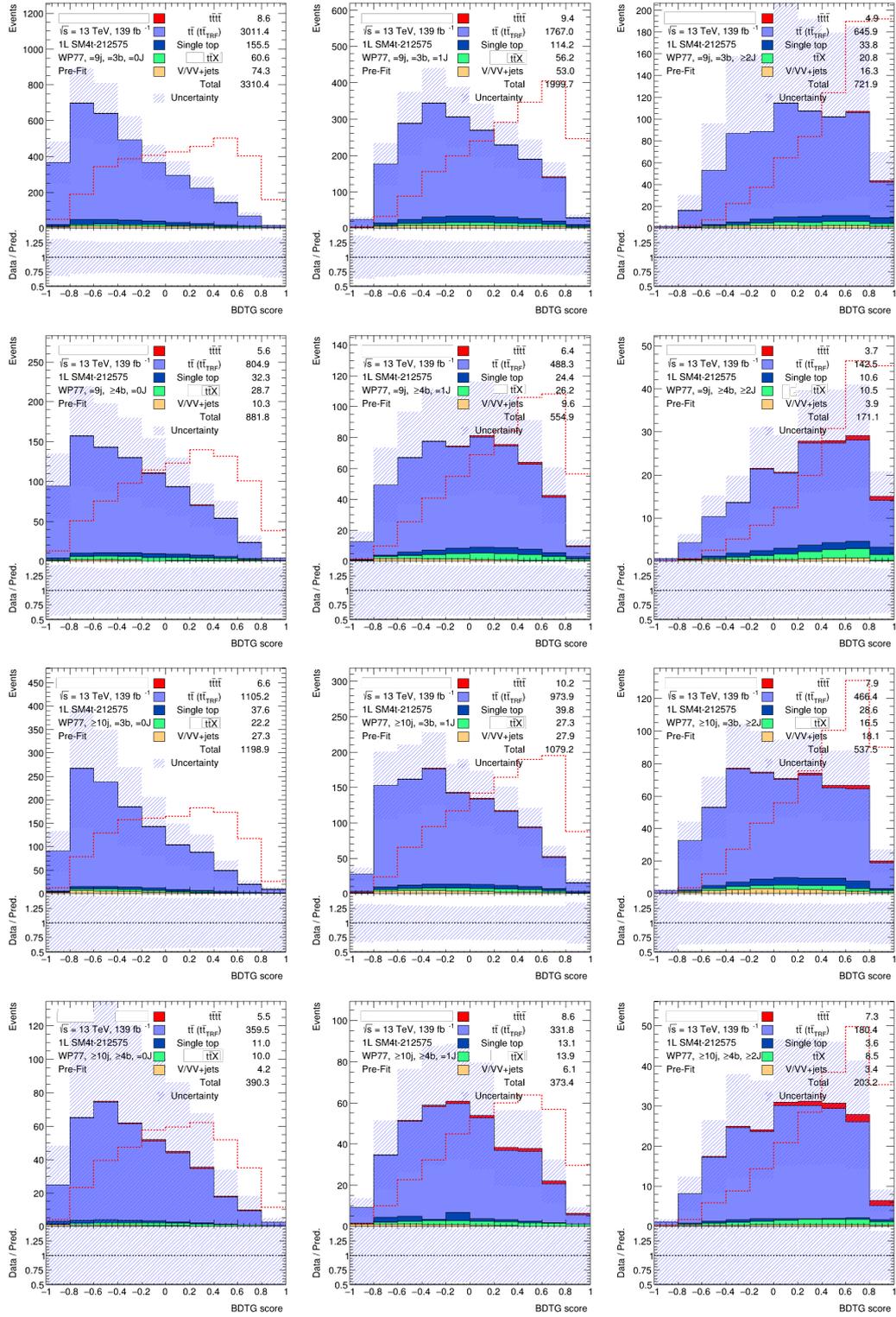
Figure 6.13: Comparison between Asimov data set and prediction of the BDT distributions in the signal regions split in multiplicity of J prior to the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.
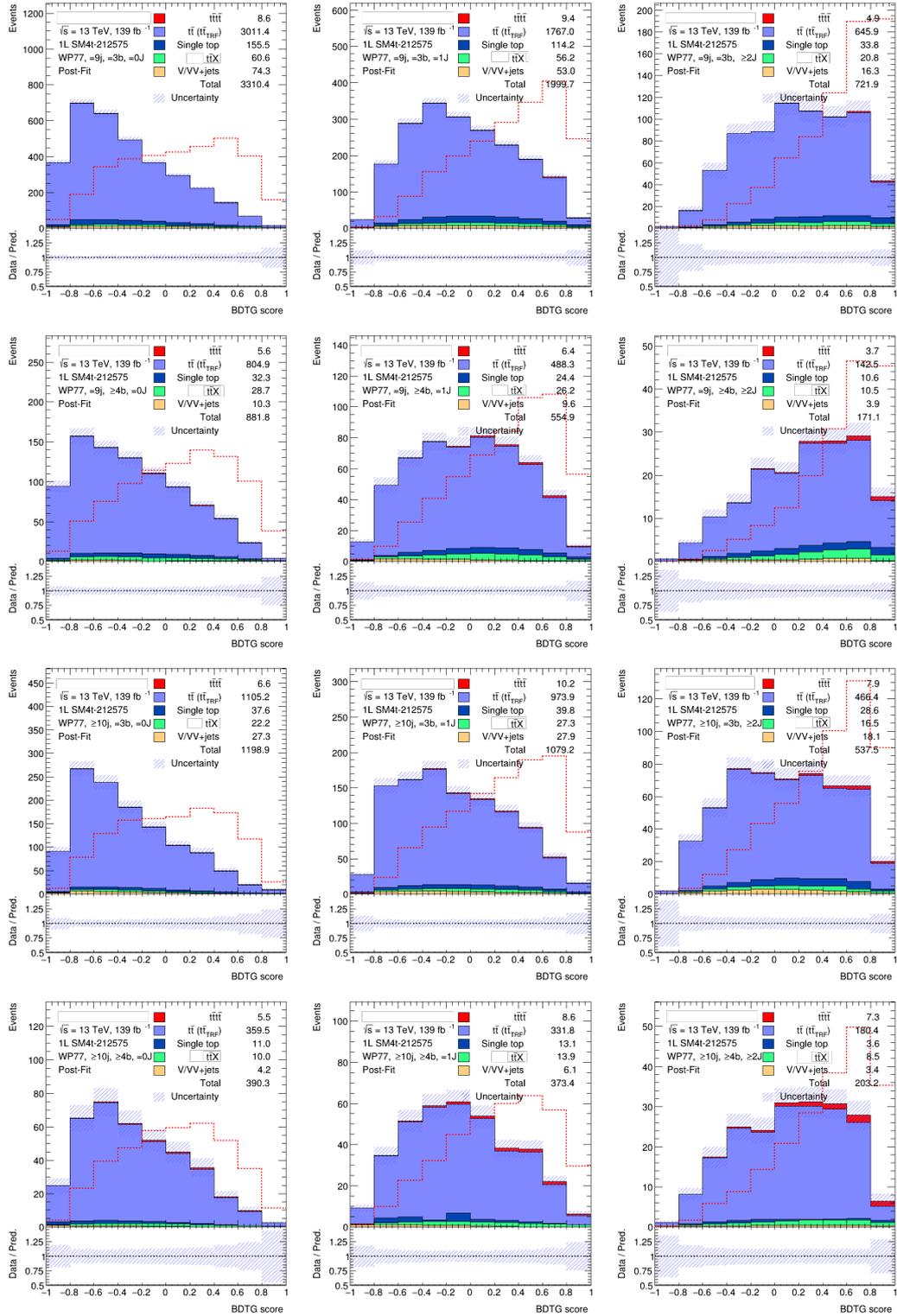
Figure 6.14: Comparison between Asimov data set and prediction of the BDT distributions in the signal regions split in multiplicity of J after the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.
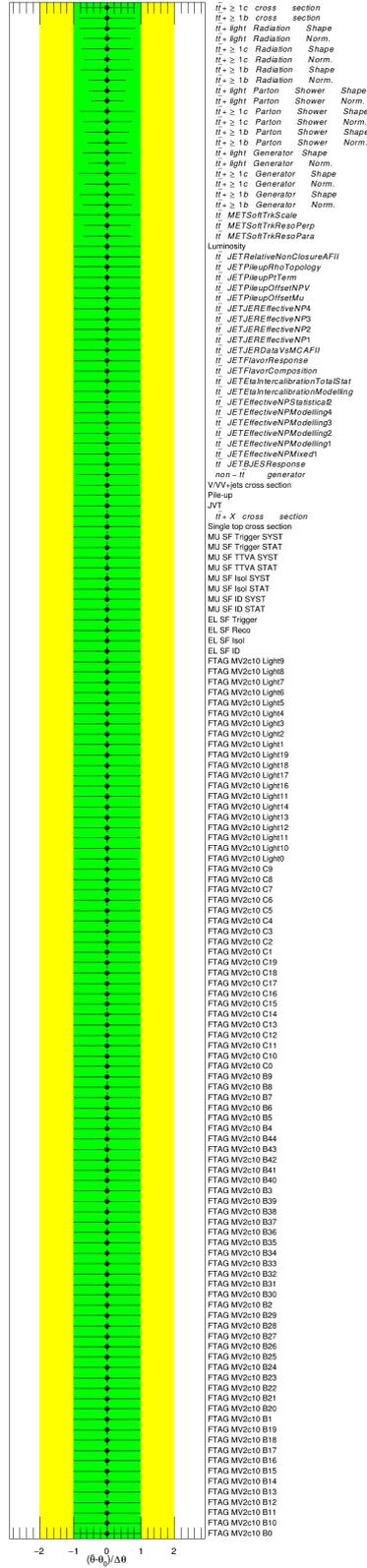
Figure 6.15: Fitted nuisance parameters on Asimov data set under the signal-plus-background hypothesis in the signal regions only.
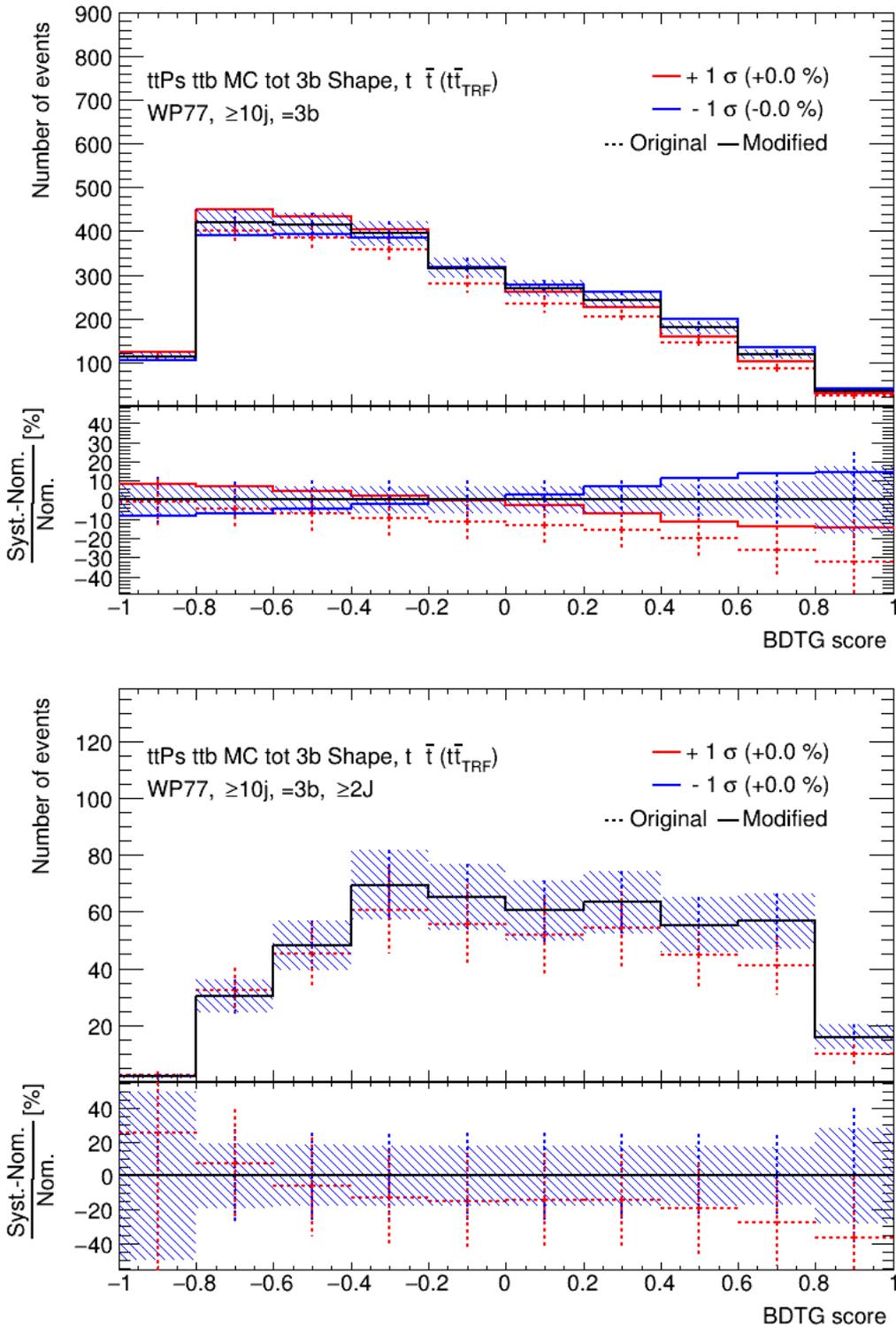
Figure 6.16: Pre-fit uncertainties associated to $t\bar{t}$ parton shower in $10ji3be$ region (on the top) and $10ji3be2Ji$ sub-region (on the bottom). Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.
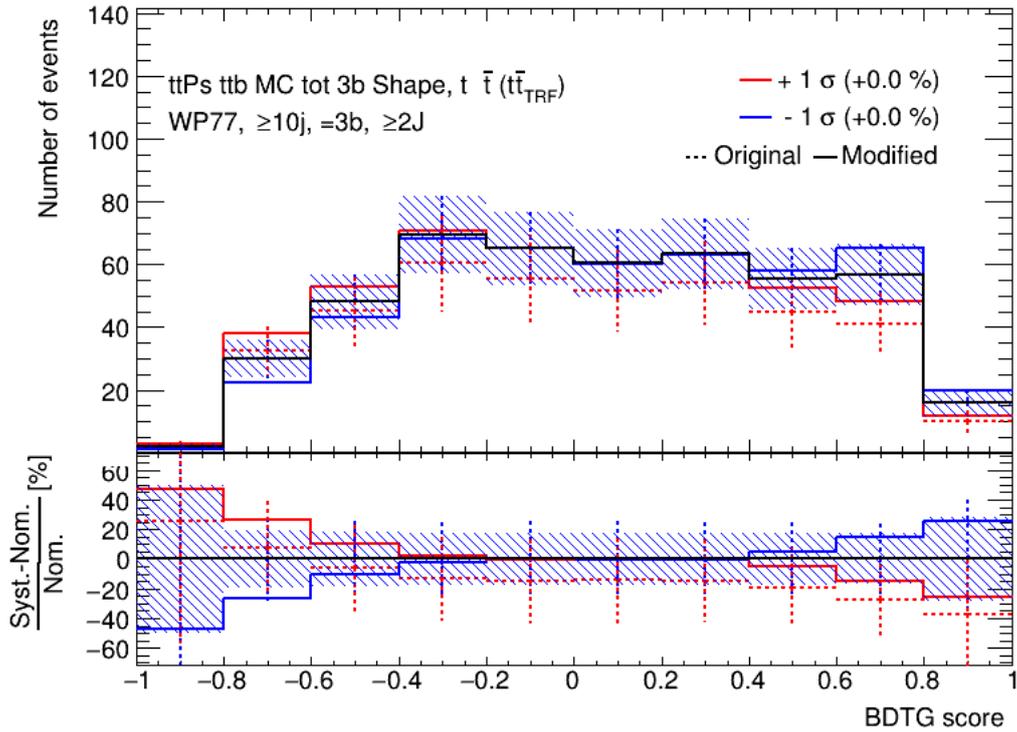
Figure 6.17: Pre-fit uncertainties associated to $t\bar{t}$ parton shower in the $10ji3be2Ji$ sub-region without the smoothing. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.

6.4 Fit to data

According to the blinding strategy illustrated in the previous pages, the signal regions are unblinded gradually starting from the less sensitive one, preforming a fit to data under the signal-plus-background hypothesis. Since the different fitted signal strengths result to be compatible with both the background-only hypothesis ($\mu = 0$) and signal-plus-background hypothesis within 1σ significance, a full unblinded fit is performed in the four signal regions simultaneously: $9je3be$, $9je4bi$, $10ji3be$ and $10ji4bi$. A first background-only hypothesis fit is performed to check the sensitivity: it is found to be in good agreement with the Asimov indication, being about 1.8. Then the signal-plus-background hypothesis fit is performed: Figure 6.18 displays the comparison of the expected yields before and after the fit. Figures 6.19 and 6.20 show the distribution of the MVA output in the signal regions, before and after the fit, respectively. As expected, the biggest contributions to the uncertainty come from the choice of the generators for the $t\bar{t}$+jet background. The constraints and pulls on the nuisance parameters from the fit are shown in Figure 6.21. While no strong constraint is observed on the NPs associated to detector and jet-tagging uncertainties, they show some pulls: the `JET_RelativeNonClosure_AFII` NP is pulled by a factor $(14 \pm 98)\%$, and the `Pile-up` NP is pulled by a factor $(22 \pm 99)\%$. The nuisance parameters associated to the theoretical uncertainties on the $t\bar{t}$+jet background estimation present both constraints and pulls, as expected from the studies done in the control regions and with the Asimov data set. The nuisance parameter related to the uncertainty on the $t\bar{t}+ \geq 1c$ cross section is pulled by $(55 \pm 80)\%$. The nuisance parameter related to the choice of the parton shower generator on $t\bar{t}$+light shape is pulled by a factor -36% and constrained to 51%. The nuisance parameter related to the choice of the parton shower generator on $t\bar{t}+ \geq 1c$ normalisation is pulled by a factor of $(-42 \pm 78)\%$. The $t\bar{t}+ \geq 1b$ shape and normalisation uncertainties on the choice of the parton shower generator are constrained to 67% and 54%, respectively, and the NP associated to the $t\bar{t}+ \geq 1b$ normalisation is also pulled of a factor 39%. The $t\bar{t}$+light shape uncertainty on the choice of the ME generator is constrained to 66%, while the corresponding normalisation uncertainty is pulled by a factor $(-41 \pm 85)\%$. As already discussed, it is not surprising that the PS provides the biggest pulls and constraints, given the fact that it has to add at least four jets modelled at leading logarithmic accuracy. The ranking of the 20 NPs with the highest impact on the fitted value of μ is reported in Figure 6.22, and it reflects the considerations made above. Figure 6.23 shows the correlation matrix of the most important nuisance parameters. The most significant correlations are observed between the different $t\bar{t}$ modelling uncertainties. The systematic uncertainties with the largest impact on signal strength are coming from the choice of the parton shower generator and the modelling of non-$t\bar{t}$ processes. In the current fit model, signal modelling uncertainties are not considered, due to the predominance of the $t\bar{t}$ systematics. A rough estimation is given in Reference [19], resulting in about 20%, as already observed in Section 4.3. Figures form 6.24 to 6.30 show the variation and the impact in the different signal regions of the most important systematics. The fitted signal strength results to be:

$$\mu = 3.47^{+0.72}_{-0.70}(\text{stat.})^{+1.75}_{-1.81}(\text{syst.}) = 3.47^{+1.89}_{-1.94}, \tag{6.6}$$

which implies an observed (expected) significance of the signal over the background-only hypothesis of 1.8σ (0.5σ). The corresponding measured cross section for the

$t\bar{t}t\bar{t}$ process is then $41.5^{+22.6}_{-23.2}$ fb, compatible with the SM value within 1.3σ. A final test is performed injecting a value for μ simular to the observed one, both in Asimov and real data configurations, in order to check the proper extraction of the signal by the fit procedure. The μ injected is 3.5 and the fitted signal strength results to be $\mu = 3.50^{+2.05}_{-2.05}$ and $\mu = 0.99^{+0.54}_{-0.55}$ for Asimov and real data, respectively, confirming the goodness of the fit.
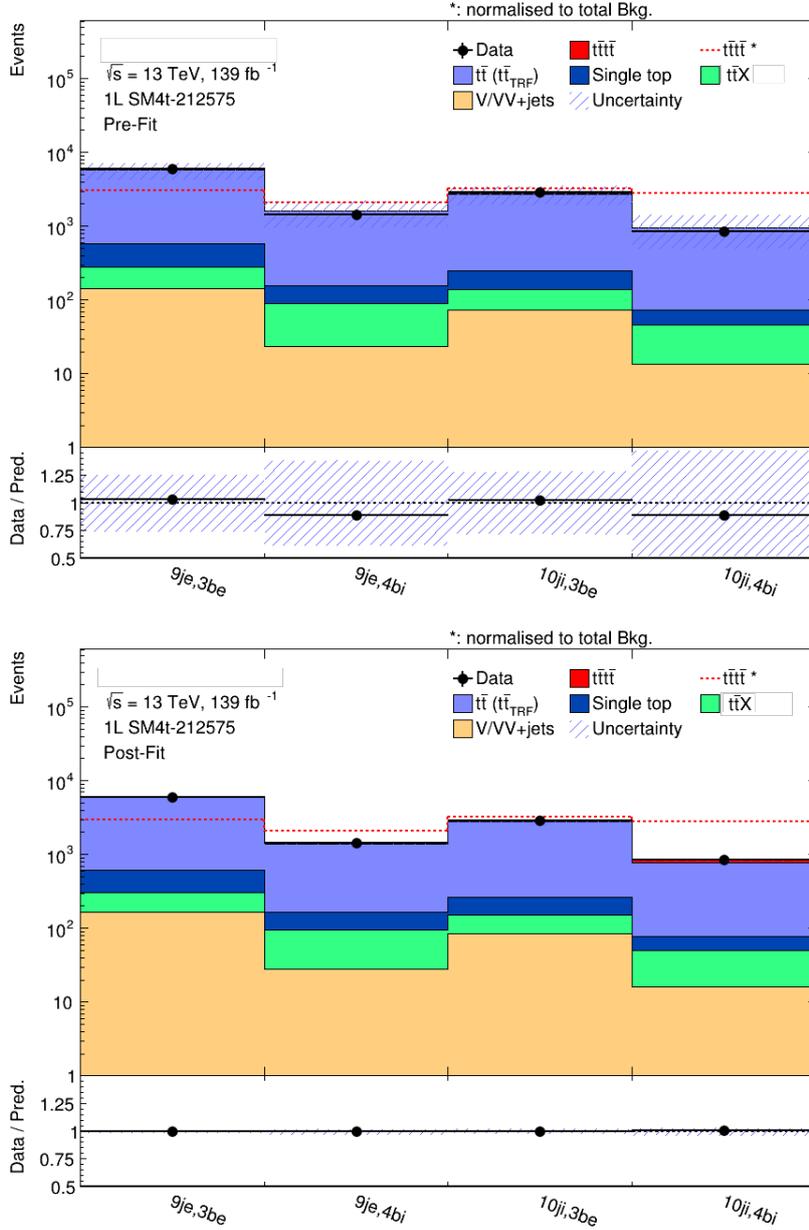
Figure 6.18: Comparison of prediction to data in the four signal regions prior and after the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction. The bottom panels display the ratio of data to the total prediction.
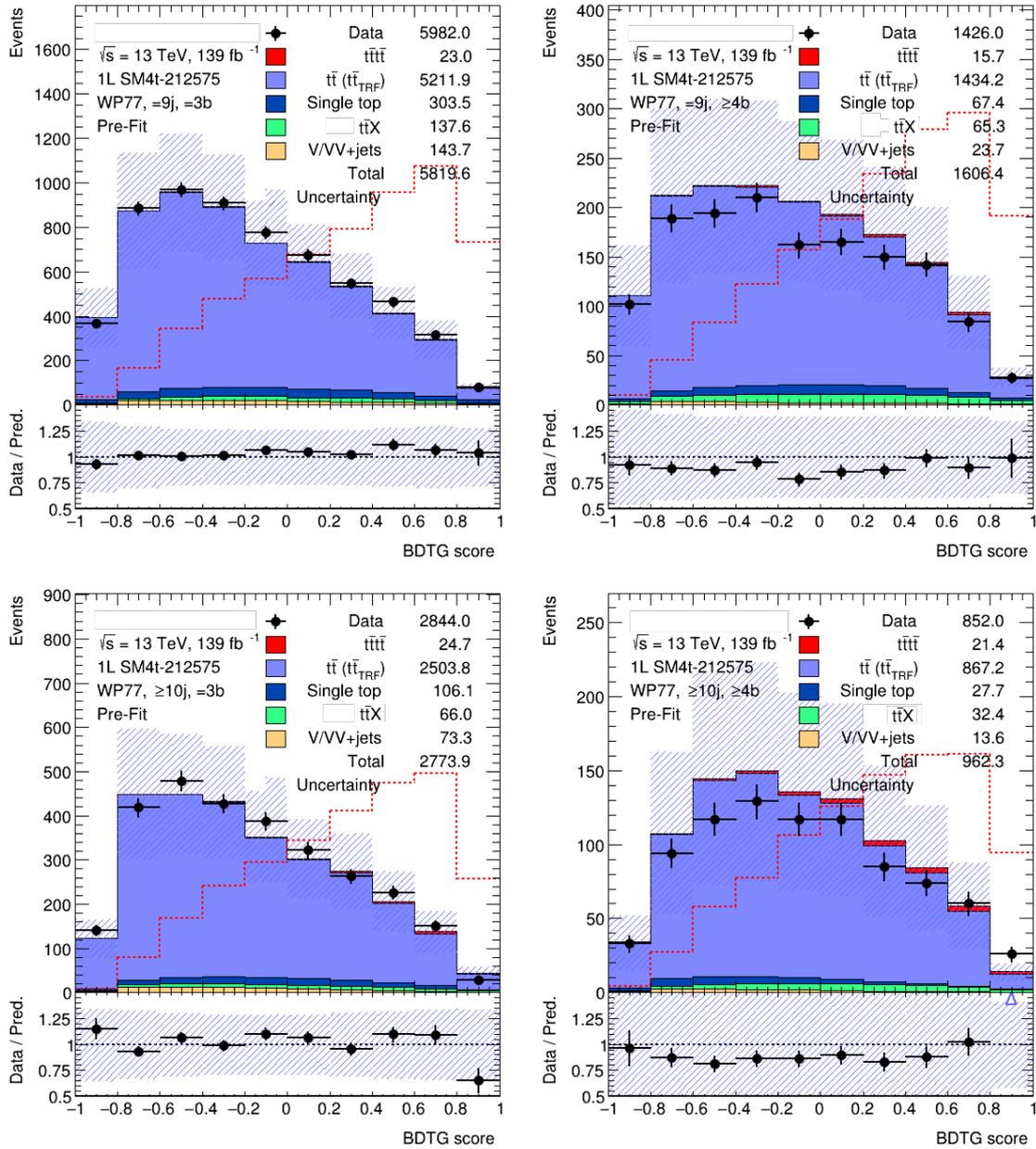
Figure 6.19: Comparison between data and prediction of the BDT distributions in the signal regions prior to the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.

Figure 6.20: Comparison between data and prediction of the BDT distributions in the signal regions after the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.

$t\bar{t} + \geq 1c$ cross section
$t\bar{t} + \geq 1b$ cross section
$t\bar{t} + light$ Radiation Shape
$t\bar{t} + light$ Radiation Norm.
$t\bar{t} + \geq 1c$ Radiation Shape
$t\bar{t} + \geq 1c$ Radiation Norm.
$t\bar{t} + \geq 1b$ Radiation Shape
$t\bar{t} + \geq 1b$ Radiation Norm.
$t\bar{t} + light$ Parton Shower Shape
$t\bar{t} + light$ Parton Shower Norm.
$t\bar{t} + \geq 1c$ Parton Shower Shape
$t\bar{t} + \geq 1c$ Parton Shower Norm.
$t\bar{t} + \geq 1b$ Parton Shower Shape
$t\bar{t} + \geq 1b$ Parton Shower Norm.
$t\bar{t} + light$ Generator Shape
$t\bar{t} + light$ Generator Norm.
$t\bar{t} + \geq 1c$ Generator Shape
$t\bar{t} + \geq 1c$ Generator Norm.
$t\bar{t} + \geq 1b$ Generator Shape
$t\bar{t} + \geq 1b$ Generator Norm.
$t\bar{t}$ METSoftTrkScale
$t\bar{t}$ METSoftTrkResoPerp
$t\bar{t}$ METSoftTrkResoPara
Luminosity
$t\bar{t}$ JETRelativeNonClosureAFII
$t\bar{t}$ JETPileupRhoTopology
$t\bar{t}$ JETPileupOffsetNPV
$t\bar{t}$ JETPileupOffsetMu
$t\bar{t}$ JETJEREffectiveNP4
$t\bar{t}$ JETJEREffectiveNP3
$t\bar{t}$ JETJEREffectiveNP2
$t\bar{t}$ JETJEREffectiveNP1
$t\bar{t}$ JETJERDataVsMCAFII
$t\bar{t}$ JETFlavorResponse
$t\bar{t}$ JETFlavorComposition
$t\bar{t}$ JETEtaIntercalibrationTotalStat
$t\bar{t}$ JETEtaIntercalibrationModelling
$t\bar{t}$ JETEffectiveNPStatistical2
$t\bar{t}$ JETEffectiveNPModelling4
$t\bar{t}$ JETEffectiveNPModelling3
$t\bar{t}$ JETEffectiveNPModelling2
$t\bar{t}$ JETEffectiveNPModelling1
$t\bar{t}$ JETBJESResponse
$non-t\bar{t}$ generator
V/VV+jets cross section
Pile-up
JVT
$t\bar{t} + X$ cross section
Single top cross section
MU SF ID SYST
EL SF ID
FTAG MV2c10 Light9
FTAG MV2c10 Light8
FTAG MV2c10 Light7
FTAG MV2c10 Light6
FTAG MV2c10 Light5
FTAG MV2c10 Light4
FTAG MV2c10 Light3
FTAG MV2c10 Light2
FTAG MV2c10 Light1
FTAG MV2c10 Light11
FTAG MV2c10 Light14
FTAG MV2c10 Light13
FTAG MV2c10 Light12
FTAG MV2c10 Light11
FTAG MV2c10 Light10
FTAG MV2c10 Light0
FTAG MV2c10 C9
FTAG MV2c10 C8
FTAG MV2c10 C7
FTAG MV2c10 C6
FTAG MV2c10 C5
FTAG MV2c10 C4
FTAG MV2c10 C3
FTAG MV2c10 C2
FTAG MV2c10 C1
FTAG MV2c10 C19
FTAG MV2c10 C18
FTAG MV2c10 C17
FTAG MV2c10 C16
FTAG MV2c10 C15
FTAG MV2c10 C14
FTAG MV2c10 C13
FTAG MV2c10 C12
FTAG MV2c10 C11
FTAG MV2c10 C10
FTAG MV2c10 C0
FTAG MV2c10 B9
FTAG MV2c10 B8
FTAG MV2c10 B7
FTAG MV2c10 B6
FTAG MV2c10 B5
FTAG MV2c10 B4
FTAG MV2c10 B44
FTAG MV2c10 B41
FTAG MV2c10 B40
FTAG MV2c10 B3
FTAG MV2c10 B39
FTAG MV2c10 B38
FTAG MV2c10 B37
FTAG MV2c10 B36
FTAG MV2c10 B35
FTAG MV2c10 B34
FTAG MV2c10 B32
FTAG MV2c10 B31
FTAG MV2c10 B2
FTAG MV2c10 B28
FTAG MV2c10 B26
FTAG MV2c10 B25
FTAG MV2c10 B24
FTAG MV2c10 B23
FTAG MV2c10 B22
FTAG MV2c10 B21
FTAG MV2c10 B20
FTAG MV2c10 B1
FTAG MV2c10 B19
FTAG MV2c10 B18
FTAG MV2c10 B17
FTAG MV2c10 B16
FTAG MV2c10 B15
FTAG MV2c10 B14
FTAG MV2c10 B13
FTAG MV2c10 B12
FTAG MV2c10 B11
FTAG MV2c10 B10
FTAG MV2c10 B0

$(\theta-\theta_0)/\Delta\theta$

Figure 6.21: Fitted nuisance parameters on data under the signal-plus-background hypothesis in the signal regions.

Figure 6.22: Ranking plot from the fit on data under the signal-plus-background hypothesis. The γ parameters refer to statistical uncertainty on the background model and are assumed to be Poisson-distributed NPs.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
FTAG MV2c10 Light0	100.0	8.4	2.5	-5.2	-5.2	-5.2	-4.1	-0.2	-7.0	-8.3	-11.8	-4.2	7.5	13.4	-5.0	-14.1	6.3	-1.1	6.7	-0.3	6.7	6.2	9.4	1.0	-1.5	12.7	7.1
$\mathrm{non}-t\bar{t}$ generator	8.4	100.0	2.1	0.2	1.2	0.9	4.7	-0.6	2.2	0.2	7.2	3.6	-20.6	-0.2	-4.0	-9.5	-3.5	10.1	-7.6	-2.0	-6.2	4.7	-5.3	0.9	-10.0	-2.6	-24.2
$t\bar{t}$ JETFlavorComposition	2.5	2.1	100.0	1.3	-3.9	-4.5	-5.8	-0.1	-2.8	1.6	1.6	-2.0	-18.5	1.9	0.5	4.2	-4.5	-9.2	2.8	-2.2	-1.9	-3.6	-0.2	1.4	6.9	-6.1	0.2
$t\bar{t}$ JETFlavorResponse	-5.2	0.2	1.3	100.0	0.2	0.1	-0.9	-7.3	0.9	-2.2	-0.4	-11.1	11.7	1.9	2.4	-4.1	-1.3	2.6	0.7	3.3	0.6	4.2	-0.4	2.4	-0.7	3.5	1.7
$t\bar{t}$ METSoftTrkResoPara	-5.2	1.2	-3.9	0.2	100.0	-21.4	-20.7	-0.1	-15.3	1.0	-10.2	0.6	-9.8	-3.9	-11.6	-0.8	-20.8	2.4	4.3	0.5	-4.2	1.0	4.2	0.4	-3.9	1.2	-3.3
$t\bar{t}$ METSoftTrkResoPerp	-5.2	0.9	-4.5	0.1	-21.4	100.0	-20.2	0.8	-15.2	1.9	-10.5	0.9	-8.5	-3.0	-11.8	-2.9	-20.5	0.9	4.0	-0.3	-4.1	2.4	4.2	2.2	-5.2	-1.3	-3.6
$t\bar{t}+ \geq 1b$ Generator Norm.	-4.1	4.7	-5.8	-0.9	-20.7	-20.2	100.0	-1.0	-9.8	-0.4	-0.5	1.2	-8.3	0.6	20.8	0.5	-22.0	2.1	26.0	-1.1	1.6	-0.5	2.4	-0.9	1.5	-7.0	-6.4
$t\bar{t}+ \geq 1b$ Generator Shape	-0.2	-0.6	-0.1	-7.3	-0.1	0.8	-1.0	100.0	-1.7	-20.5	-3.3	-25.3	7.5	-36.6	-2.4	0.1	-0.6	-22.7	0.1	18.0	0.8	4.7	1.6	-6.3	14.0	-5.7	19.1
$t\bar{t}+ \geq 1c$ Generator Norm.	-7.0	2.2	-2.8	0.9	-15.3	-15.2	-9.8	-1.7	100.0	0.0	-18.3	0.7	-15.7	-4.3	-20.3	2.3	-0.5	0.6	4.2	1.3	3.2	-0.4	13.1	-0.3	-3.5	6.4	-1.4
$t\bar{t}+ \geq 1c$ Generator Shape	-8.3	0.2	1.6	-2.2	1.0	1.9	-0.4	-20.5	0.0	100.0	1.3	4.3	-6.7	-4.4	0.7	-13.2	0.1	-52.2	0.5	8.3	-0.4	5.0	-0.3	-7.3	13.7	6.0	-9.1
$t\bar{t}+$ light Generator Norm.	-11.8	7.2	1.6	-0.4	-10.2	-10.5	-0.5	-3.3	-18.3	1.3	100.0	2.7	25.2	-11.3	-28.2	3.7	16.6	1.8	4.2	1.8	13.6	-0.2	23.3	0.4	-7.8	14.2	-7.1
$t\bar{t}+$ light Generator Shape	-4.2	3.6	-2.0	-11.1	0.6	0.9	1.2	-25.3	0.7	4.3	2.7	100.0	-10.3	18.5	-0.5	-3.6	-0.3	24.4	-1.6	1.9	-1.7	-21.8	-1.5	-13.8	-10.5	-1.8	-19.6
$t\bar{t}+ \geq 1b$ Parton Shower Norm.	7.5	-20.6	-18.5	11.7	-9.8	-8.5	-8.3	7.5	-15.7	-6.7	25.2	-10.3	100.0	29.4	-5.2	-1.8	24.6	-9.8	19.9	3.2	-7.1	-1.6	8.0	-3.2	3.7	-25.9	39.2
$t\bar{t}+ \geq 1b$ Parton Shower Shape	13.4	-0.2	1.9	1.9	-3.9	-3.0	0.6	-36.6	-4.3	-4.4	-11.3	18.5	29.4	100.0	-9.5	14.1	0.7	-13.2	-2.2	12.7	3.1	-3.6	5.2	-7.8	4.4	-40.2	36.6
$t\bar{t}+ \geq 1c$ Parton Shower Norm.	-5.0	-4.0	0.5	2.4	-11.6	-11.8	20.8	-2.4	-20.3	0.7	-28.2	-0.5	-5.2	-9.5	100.0	3.3	-5.4	-0.7	-41.5	3.3	-12.2	-0.1	8.2	0.6	-4.8	23.6	4.9
$t\bar{t}+ \geq 1c$ Parton Shower Shape	-14.1	-9.5	4.2	-4.1	-0.8	-2.9	0.5	0.1	2.3	-13.2	3.7	-3.6	-1.8	14.1	3.3	100.0	-1.3	-26.7	-0.6	-14.8	-1.3	8.3	-2.6	-0.3	-22.3	-1.7	-2.4
$t\bar{t}+$ light Parton Shower Norm.	6.3	-3.5	-4.5	-1.3	-20.8	-20.5	-22.0	-0.6	-0.5	0.1	16.6	-0.3	24.6	0.7	-5.4	-1.3	100.0	2.0	-23.7	-1.7	-32.1	0.1	-29.2	-0.3	8.7	0.4	-2.8
$t\bar{t}+$ light Parton Shower Shape	-1.1	10.1	-9.2	2.6	2.4	0.9	2.1	-22.7	0.6	-52.2	1.8	24.4	-9.8	-13.2	-0.7	-26.7	2.0	100.0	-1.0	-9.6	-0.6	-5.6	-0.4	1.5	15.9	4.1	-24.6
$t\bar{t}+ \geq 1b$ Radiation Norm.	6.7	-7.6	2.8	0.7	4.3	4.0	26.0	0.1	4.2	0.5	4.2	-1.6	19.9	-2.2	-41.5	-0.6	-23.7	-1.0	100.0	0.8	-24.4	0.5	-18.4	0.8	2.8	12.5	6.1
$t\bar{t}+ \geq 1b$ Radiation Shape	-0.3	-2.0	-2.2	3.3	0.5	-0.3	-1.1	18.0	1.3	8.3	1.8	1.9	3.2	12.7	3.3	-14.8	-1.7	-9.6	0.8	100.0	-0.3	-20.2	-1.4	-6.8	-12.8	7.3	11.4
$t\bar{t}+ \geq 1c$ Radiation Norm.	6.7	-6.2	-1.9	0.6	-4.2	-4.1	1.6	0.8	3.2	-0.4	13.6	-1.7	-7.1	3.1	-12.2	-1.3	-32.1	-0.6	-24.4	-0.3	100.0	0.2	-19.0	-0.1	5.2	0.6	4.7
$t\bar{t}+ \geq 1c$ Radiation Shape	6.2	4.7	-3.6	4.2	1.0	2.4	-0.5	4.7	-0.4	5.0	-0.2	-21.8	-1.6	-3.6	-0.1	8.3	0.1	-5.6	0.5	-20.2	0.2	100.0	0.4	-16.2	2.9	4.7	-4.9
$t\bar{t}+$ light Radiation Norm.	9.4	-5.3	-0.2	-0.4	4.2	4.2	2.4	1.6	13.1	-0.3	23.3	-1.5	8.0	5.2	8.2	-2.6	-29.2	-0.4	-18.4	-1.4	-19.0	0.4	100.0	0.0	6.8	-4.8	3.2
$t\bar{t}+$ light Radiation Shape	1.0	0.9	1.4	2.4	0.4	2.2	-0.9	-6.3	-0.3	-7.3	0.4	-13.8	-3.2	-7.8	0.6	-0.3	-0.3	1.5	0.8	-6.8	-0.1	-16.2	0.0	100.0	2.9	5.5	-4.2
$t\bar{t}+ \geq 1b$ cross section	-1.5	-10.0	6.9	-0.7	-3.9	-5.2	1.5	14.0	-3.5	13.7	-7.8	-10.5	3.7	4.4	-4.8	-22.3	8.7	15.9	2.8	-12.8	5.2	2.9	6.8	2.9	100.0	1.6	10.6
$t\bar{t}+ \geq 1c$ cross section	12.7	-2.6	-6.1	3.5	1.2	-1.3	-7.0	-5.7	6.4	6.0	14.2	-1.8	-25.9	-40.2	23.6	-1.7	0.4	4.1	12.5	7.3	0.6	4.7	-4.8	5.5	1.6	100.0	7.0
mu_signal	7.1	-24.2	0.2	1.7	-3.3	-3.6	-6.4	19.1	-1.4	-9.1	-7.1	-19.6	39.2	36.6	4.9	-2.4	-2.8	-24.6	6.1	11.4	4.7	-4.9	3.2	-4.2	10.6	7.0	100.0

Figure 6.23: Correlations between the nuisance parameters from the fit on data under the signal-plus-background hypothesis.
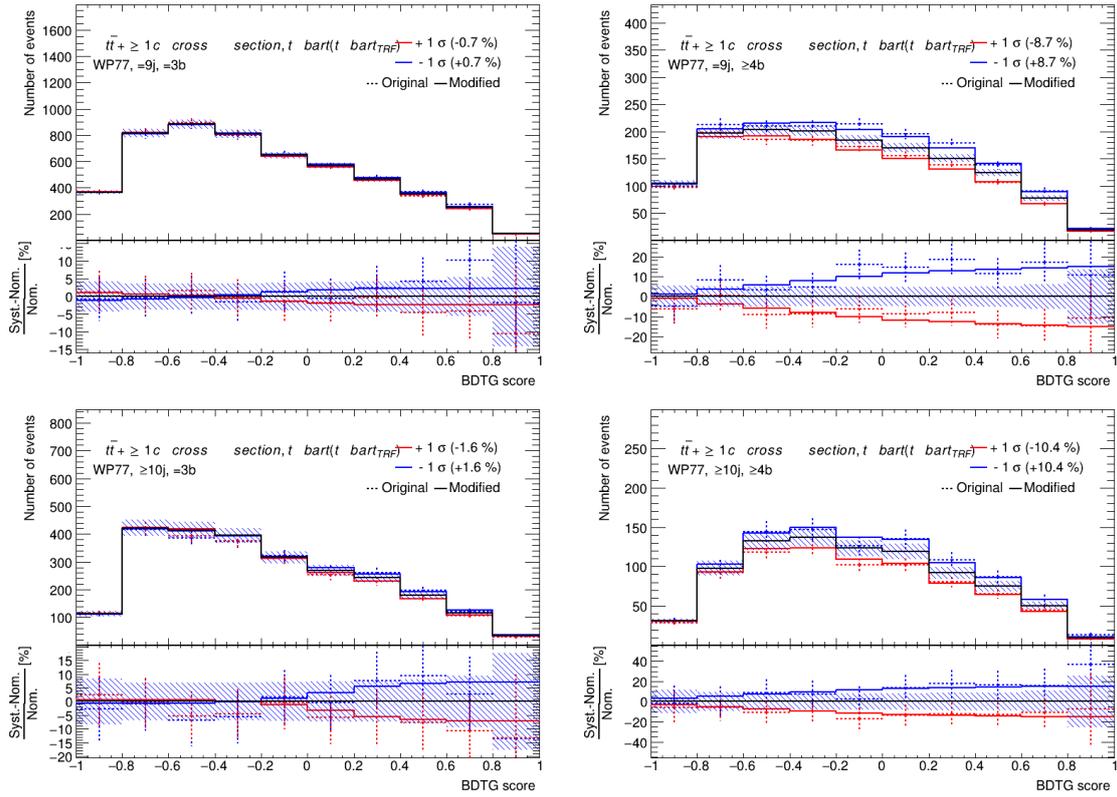
Figure 6.24: Pre-fit uncertainties associated to $t\bar{t}+ \geq 1c$ cross section in the signal regions. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.

90
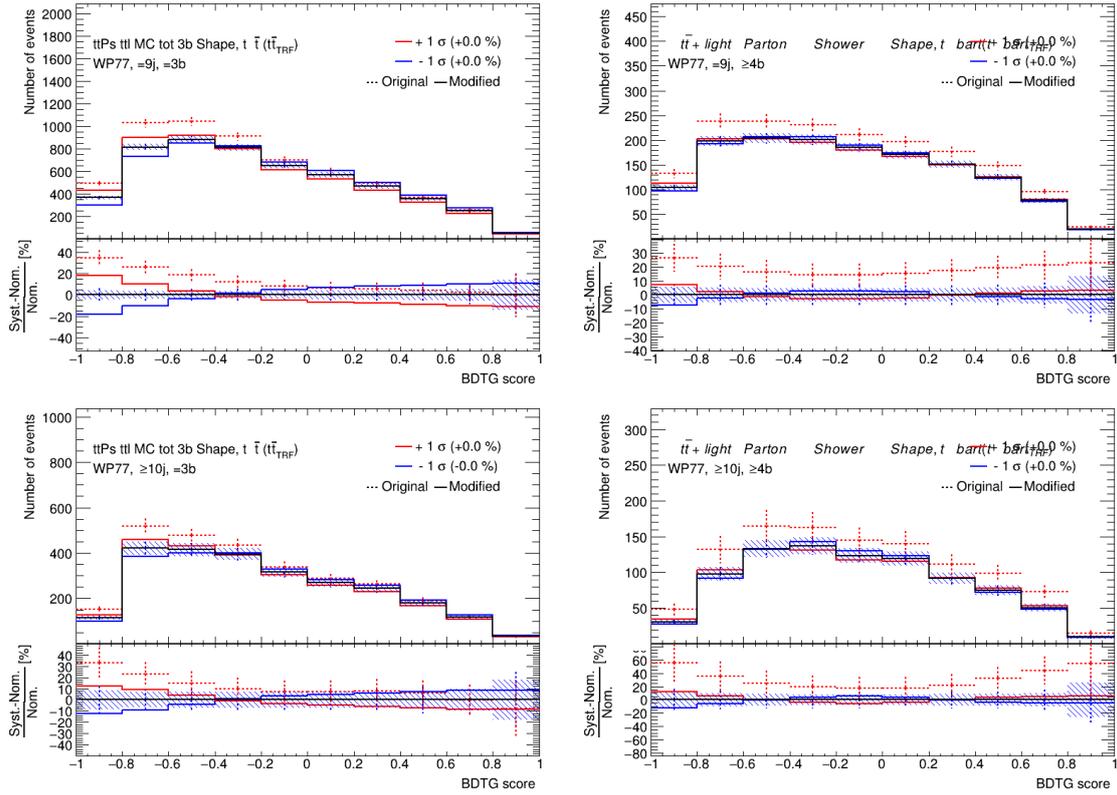
Figure 6.25: Pre-fit uncertainties associated to the parton shower generator on $t\bar{t}$+light shape in the signal regions. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.
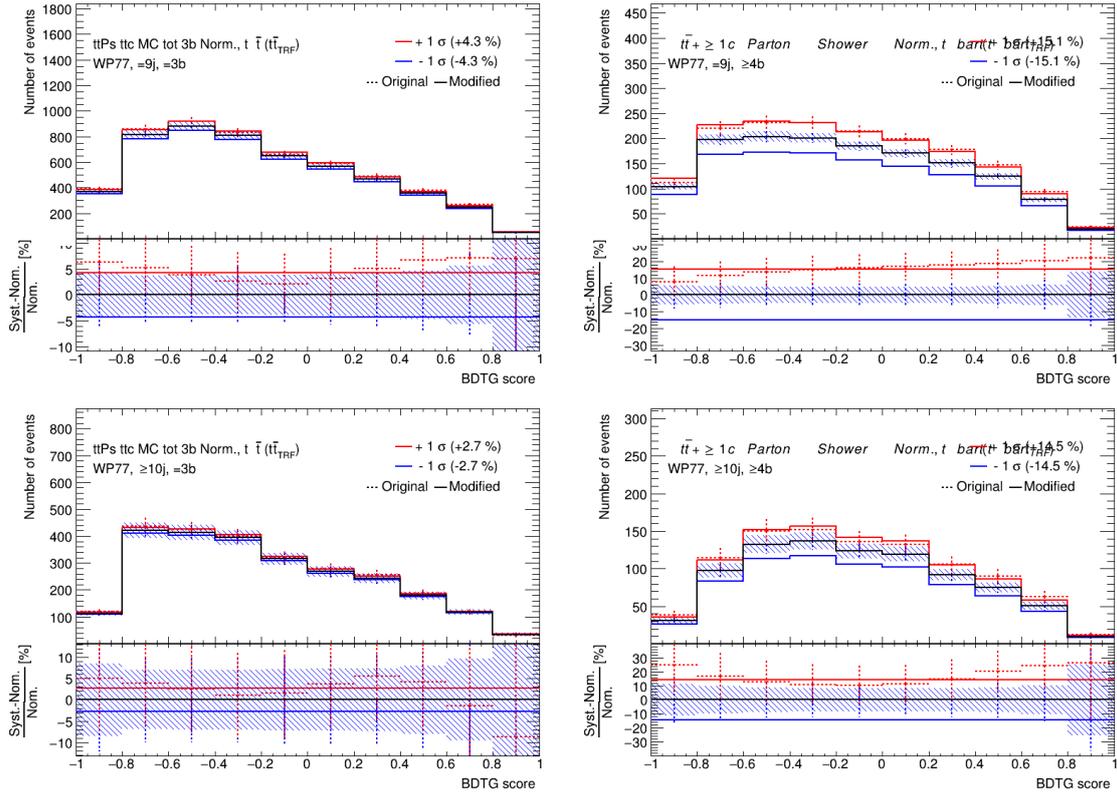
Figure 6.26: Pre-fit uncertainties associated to the parton shower generator on $t\bar{t}+ \geq 1c$ normalisation in the signal regions. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.
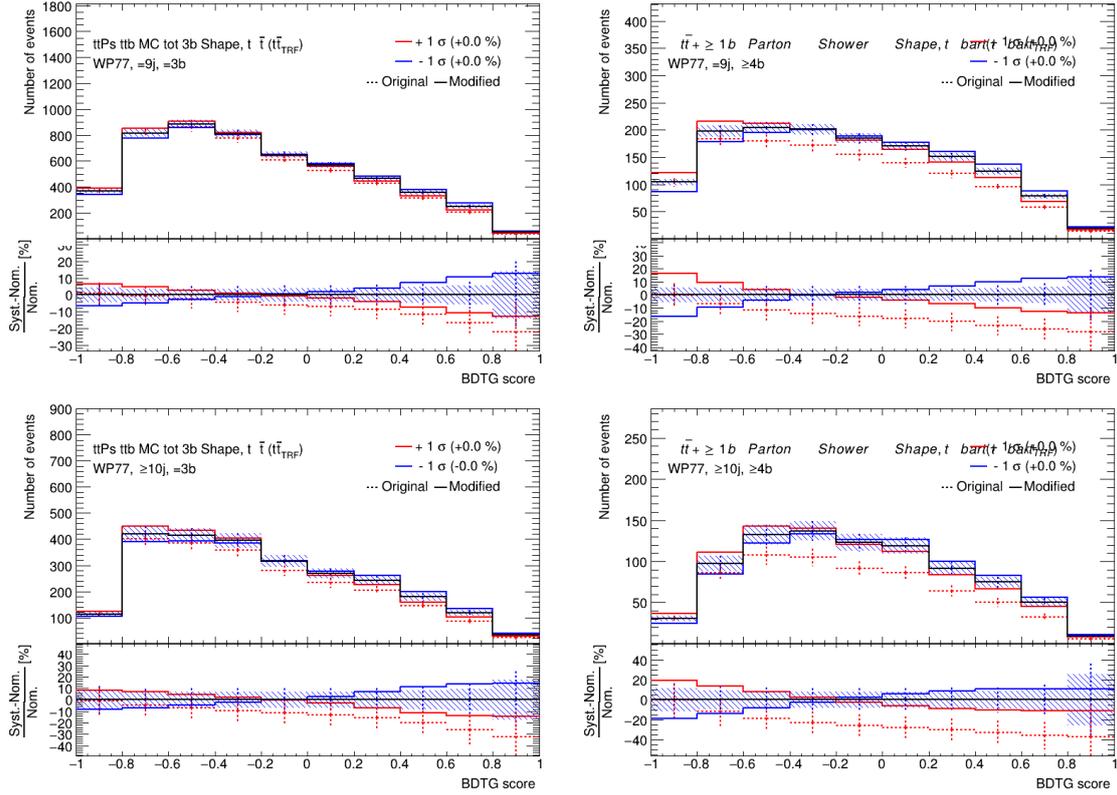
Figure 6.27: Pre-fit uncertainties associated to the parton shower generator on $t\bar{t}+ \geq 1b$ shape in the signal regions. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.

93
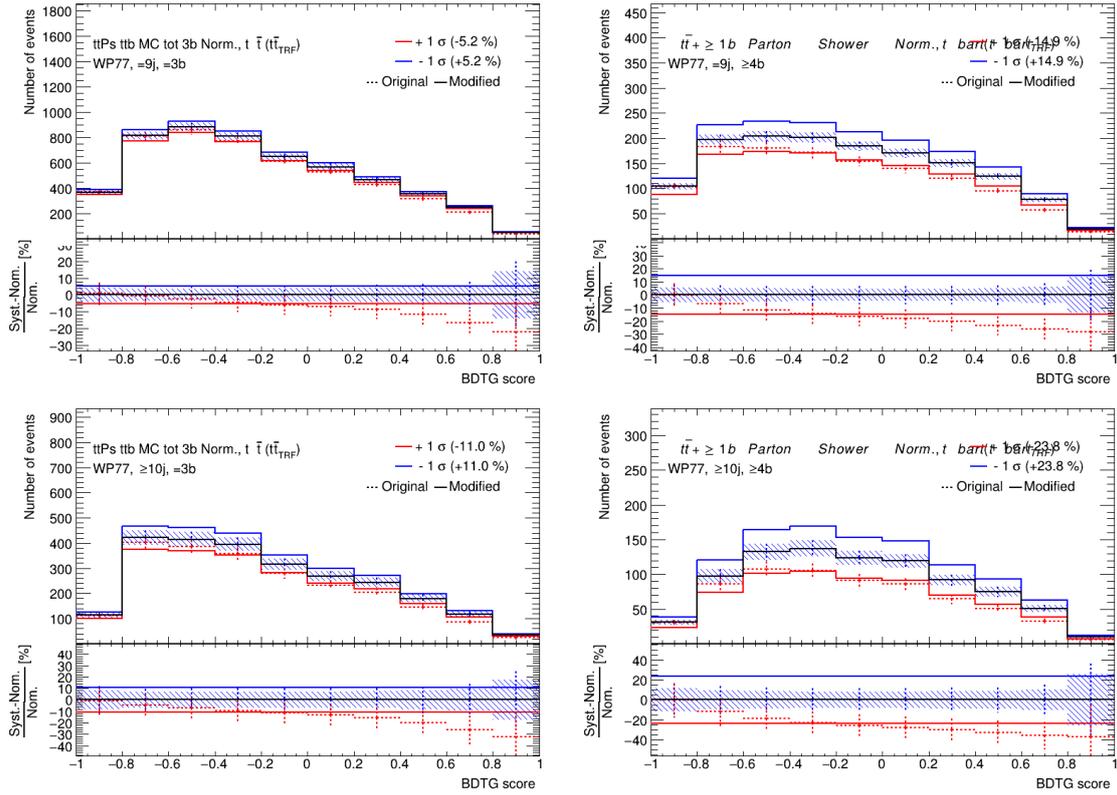
Figure 6.28: Pre-fit uncertainties associated to the parton shower generator on $t\bar{t}+ \geq$ $1b$ normalisation in the signal regions. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.
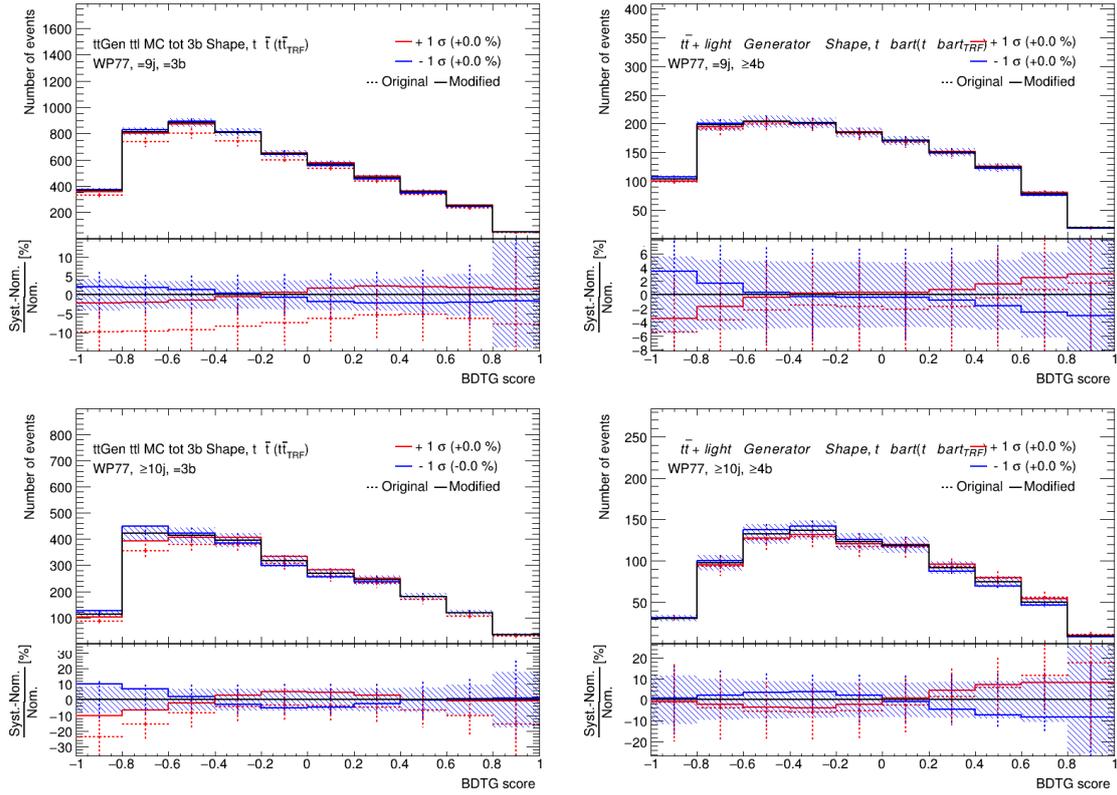
Figure 6.29: Pre-fit uncertainties associated to the matrix element generator on $t\bar{t}$+light shape in the signal regions. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.
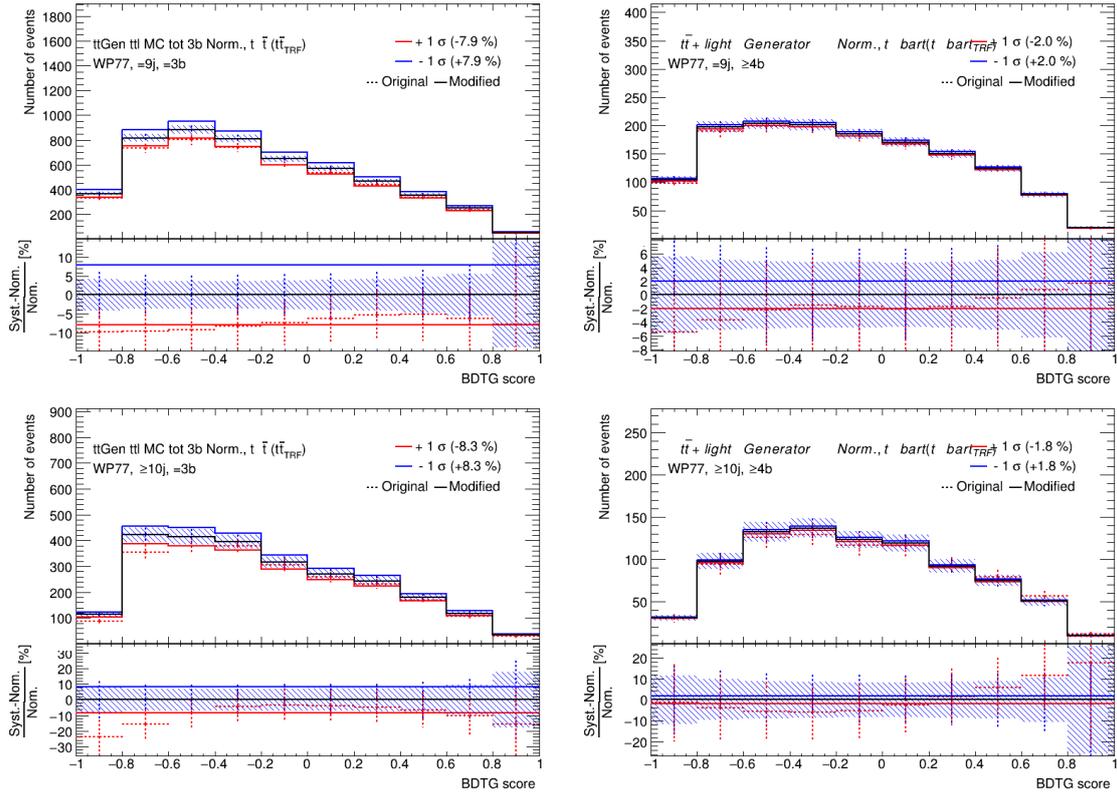
Figure 6.30: Pre-fit uncertainties associated to the matrix element generator on $t\bar{t}$+light normalisation in the signal regions. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistical uncertainty of the variation. The blue solid line is obtained by symmetrising the red one. Red and blue solid lines give the uncertainty on the prediction.

Conclusions

The production of $t\bar{t}t\bar{t}$ is one of the most rare processes accessible at LHC, given its small cross section, predicted to be around 12 fb at NLO in QCD. It is also one of the most interesting mechanisms, because many BSM phenomena may enhance the cross section of this process. This thesis exploits the data collected by the ATLAS detector at $\sqrt{s} = 13$ TeV, from 2015 to 2018, for an integrated luminosity of 139 fb^{-1}, performing the search in the single lepton channel. In the channel taken into account, the most important background source is the $t\bar{t}$+jets process. This process is not modelled well by pure Monte Carlo simulations in the regions with high jet and b-jet multiplicities. To address this problem, a data-driven, MC-assisted, approach is adopted, exploiting the $TRF_{t\bar{t}}$ method: assuming that the probability of b-tagging a jet is independent of the number of total jets, the b-tagging probabilities are measured in low jets multiplicity regions and applied in the higher jets multiplicity regions, where the signal is expected to be more prominent. In order to have a good discrimination between the SM $t\bar{t}t\bar{t}$ signal and the $t\bar{t}$+jet background, a multivariate approach is adopted: 16 variables are used to train a boosted decision tree with TMVA in the signal regions. A neural network is also trained in the most signal-like region in order to set up a comparison between different MVA methods. Given the small signal to background ratio, a "blind-analysis" approach is adopted: the analysis strategy, its optimisation and the systematic model have to be established on the Asimov data set in signal regions and real data in control regions, before looking at data in the most sensitive regions. An important step is the evaluation of the expected sensitivity, in order to be sure that the unblinding could be done without spoiling the official 1LOS ATLAS analysis. A binned profile likelihood fit is then performed in the signal regions on the BDT output distributions. The measured $t\bar{t}t\bar{t}$ signal strength $\mu = \sigma_{t\bar{t}t\bar{t}}/\sigma_{t\bar{t}t\bar{t}}^{SM}$ results to be $\mu = 3.5^{+0.7}_{-0.7}(\text{stat.})^{+1.8}_{-1.8}(\text{syst.}) = 3.5^{+1.9}_{-1.9}$. This value can be compared with the previous results obtained in the 1LOS channel using 36.1 fb^{-1}, $\mu = 1.7^{+1.9}_{-1.7}$ [20], and in the SSML channel using 139 fb^{-1}, $\mu = 2.0^{+0.8}_{-0.6}$ [46], and it is found to be compatible with both within 1σ. Moreover, the sensitivity is similar to the previous 1LOS result, despite the absence of the benefits from the combination with the dilepton opposite-sign channel. The sensitivity of the analysis is confirmed also by the NN result. At the time of the submission of this thesis no comparison can be done with the official ATLAS result for the 1LOS channel using 139 fb^{-1}, since it is not public. The outcome of the thesis implies an observed (expected) significance of the signal over the background-only hypothesis of 1.8σ (0.5σ). The corresponding measured cross section for the $t\bar{t}t\bar{t}$ process is 42^{+23}_{-23} fb, compatible with the SM value within 1.3σ. This small "excess" is compatible with the SSML results. The employment of a fully optimised neural network, or even better a deep machine learning method, would provide benefit to the result of the analysis, as well as the combination with the dilepton opposite-sign channel.

Appendix A

NNLO QCD + NLO EW correction in $t\bar{t}$-enriched regions

Both ATLAS [79] and CMS [80] experiments had observed that the $t\bar{t}$ NLO predictions tend to overestimate the data, in particular for the top quark p_T distribution. An improved agreement between data and prediction can be obtained using NNLO calculations, and these improvements to the modelling are normally incorporated by reweighting the $t\bar{t}$ samples matching their top quark p_T distribution to that predicted at NNLO accuracy in QCD. However, precise predictions in top quark physics must take into account both the QCD and EW effects in order to correctly identify possible BSM effects. The correction of the mismodelling of top (or anti-top) p_T can be important for an accurate $t\bar{t}$+jets simulation, and therefore a correction in the modelling of the top quark p_T distribution (extended to all the currently available $t\bar{t}$ NLO generators) is described in the following.

The correction is based on the NNLO QCD + NLO EW $t\bar{t}$ calculation for the differential cross section at 13 TeV [81]. These theoretical calculations are performed with the top mass value of $m_t = 173.3$ GeV, using the central dynamical renormalisation and factorisation scales defined as $\frac{1}{2} \times \sqrt{m_t^2 + p_T^2}$ for the distribution of the top quark p_T, and NNPDF3.0QED as PDF set. Instead, the ATLAS nominal $t\bar{t}$ POWHEG + PYTHIA 8 sample makes use of $m_t = 172.5$ GeV, NNPDF3.0NLO PDF set, and $\sqrt{m_t^2 + p_T^2}$ as factor for the scales. Therefore the effects of the scale and PDF variations are also investigated. The comparison is done for the following $t\bar{t}$ NLO generators: POWHEG + PYTHIA 8 (the nominal one), POWHEG + HERWIG 7, MADGRAPH5_AMC@NLO + PYTHIA 8, POWHEG + PYTHIA 8 (with $h_{damp} = 3.0 \cdot m_{top}$) and SHERPA 2.2.1. The distributions are compared with the inclusive $t\bar{t}$ simulated sample (before any event selection) matching the normalised differential cross section predictions. The information of the top quarks is obtained at generator level (taking the last top/anti-top replicas in the MC event record).

Figure A.1 shows the comparison between the NNLO QCD + NLO EW calculations and the nominal NLO $t\bar{t}$ simulation sample for the top p_T. The difference between the top quark p_T predicted by the MC sample is clearly evident, confirming the behaviour observed in the various experimental measurements mentioned earlier. At high transverse momenta (greater than 1 TeV) the discrepancy becomes larger, reaching 50% above 2 TeV.

Figure A.2 shows the effects of the PDF and scale variations on the NNLO QCD + NLO EW calculation. The nominal MC prediction is compared to the two
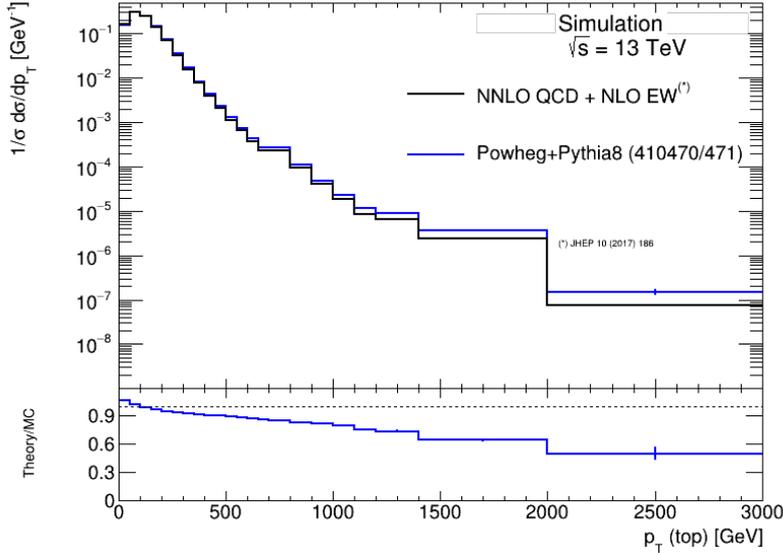
Figure A.1: Normalised $t\bar{t}$ differential cross section as a function of top quark p_T as predicted by the NNLO QCD + NLO EW calculations compared to the nominal NLO $t\bar{t}$ simulation sample (POWHEG + PYTHIA 8).

different theoretical predictions available, one calculated with NNPDF3.0QED PDF set and the other with PDF4LHC15LUXQED. The effect of the different PDF sets is very small and it is absorbed by the uncertainty due to the finite sample statistics. The impact of varying the renormalisation and factorisation scales is investigated using minimum and maximum scales variations for the theoretical calculations. Also in this case the difference is very small.

Not only the nominal POWHEG + PYTHIA NLO $t\bar{t}$ MC sample is expected to deviate from the NNLO QCD + NLO EW calculation, but also the other predictions available in ATLAS. Therefore, also for these the comparison is done, and it is shown in Figure A.3. While for POWHEG + HERWIG and POWHEG + PYTHIA (with $h_{damp} = 3.0 \cdot m_{top}$) the trend is similar to that of the nominal $t\bar{t}$ MC, MAD-GRAPH5_AMC@NLO + PYTHIA and SHERPA show a closer agreement with the theoretical calculation, but they suffer of a more limited statistics, reflected in the bigger statistical uncertainty and in the needed of merging the bins in the tail of the distribution.

The correction is applied reweighting the $t\bar{t}$ samples in order to match the top quark p_T distribution to that predicted by NNLO QCD + NLO EW calculation. The correction is tested on partonic variables, like the rapidity of the top quark or the rapidity of the $t\bar{t}$ system (see Figure A.4). The effect of the procedure is a clear improvement in the agreement with the NNLO QCD + NLO EW prediction, in particular in the tails of the distributions. In order to further testing the correction, it is applied to the $t\bar{t}t\bar{t}$ analysis. Figure A.5 shows the effect on the leading-jet p_T in the $5je2be$ region of the 1L channel. A small but non-negligible effect is observed, but, unfortunately, it is covered by the uncertainties. Since in the signal regions the uncertainties are even bigger, this correction is not included in the current $t\bar{t}t\bar{t}$ analysis. However, it is successfully employed in other ATLAS analyses [82], and it is in continuous development due to the new MC samples and theoretical calculations that are developed.
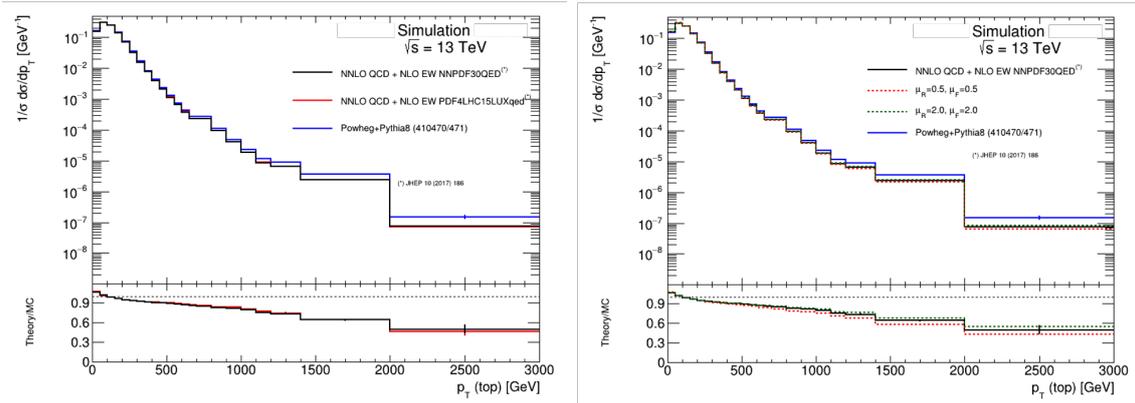
Figure A.2: Normalised $t\bar{t}$ differential cross section as a function of top quark p_T in case of PDF variations (on the left) and scale variations (on the right) of the NNLO QCD + NLO EW prediction compared to the nominal NLO $t\bar{t}$ simulation sample (POWHEG + PYTHIA 8).
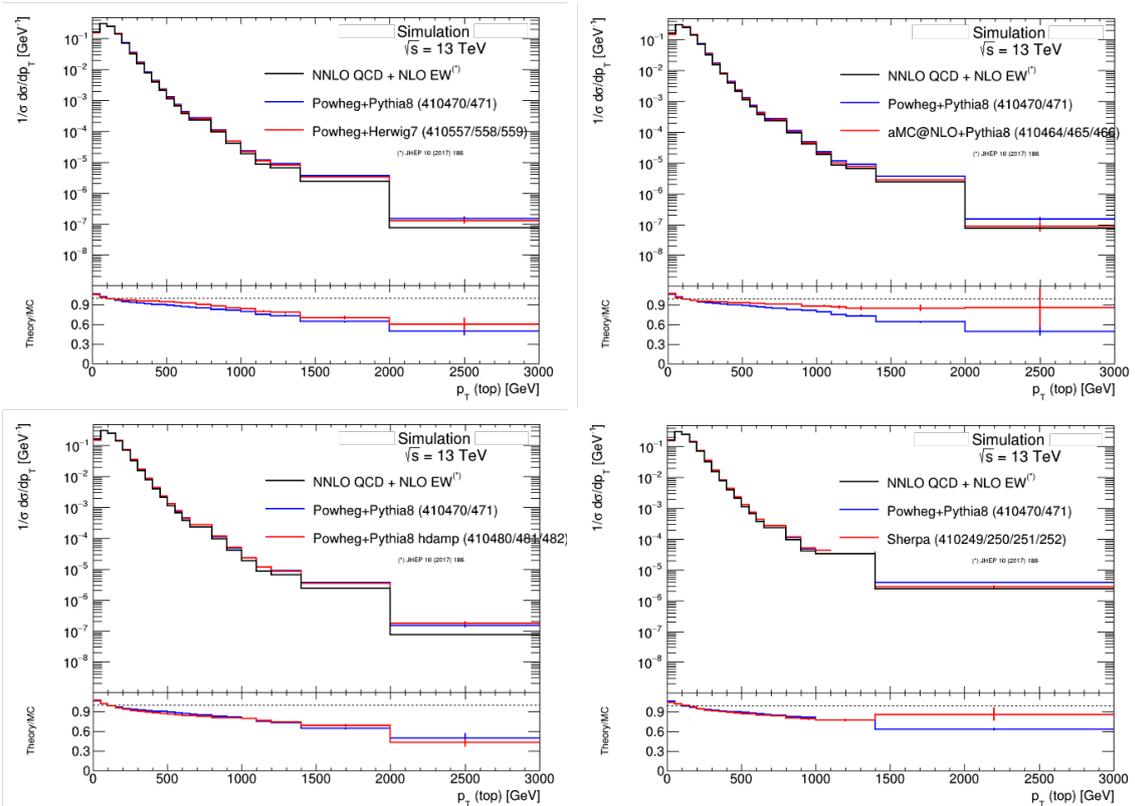
Figure A.3: Normalised $t\bar{t}$ differential cross section as a function of top quark p_T for different NLO $t\bar{t}$ simulation samples, compared to the nominal MC sample and the NNLO QCD + NLO EW prediction.
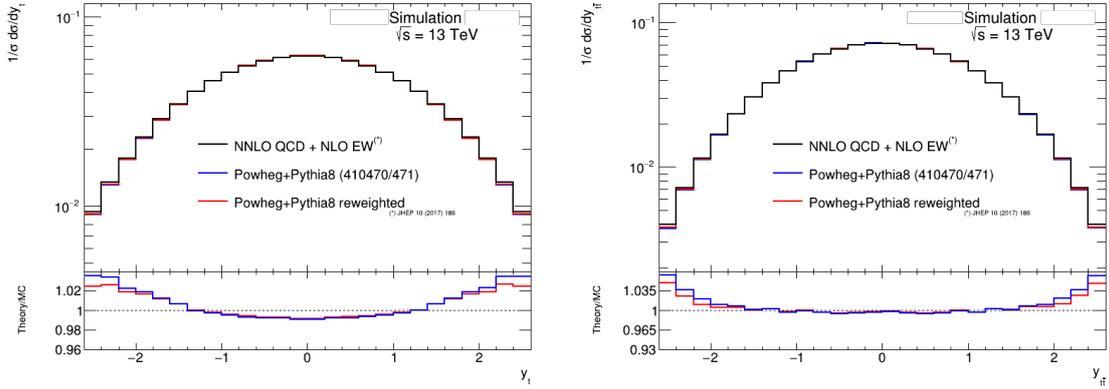
Figure A.4: Normalised $t\bar{t}$ differential cross section as a function of top quark rapidity (on the left) and rapidity of the $t\bar{t}$ system (on the right) before and after the correction of top quark p_T to the NNLO QCD + NLO EW prediction.
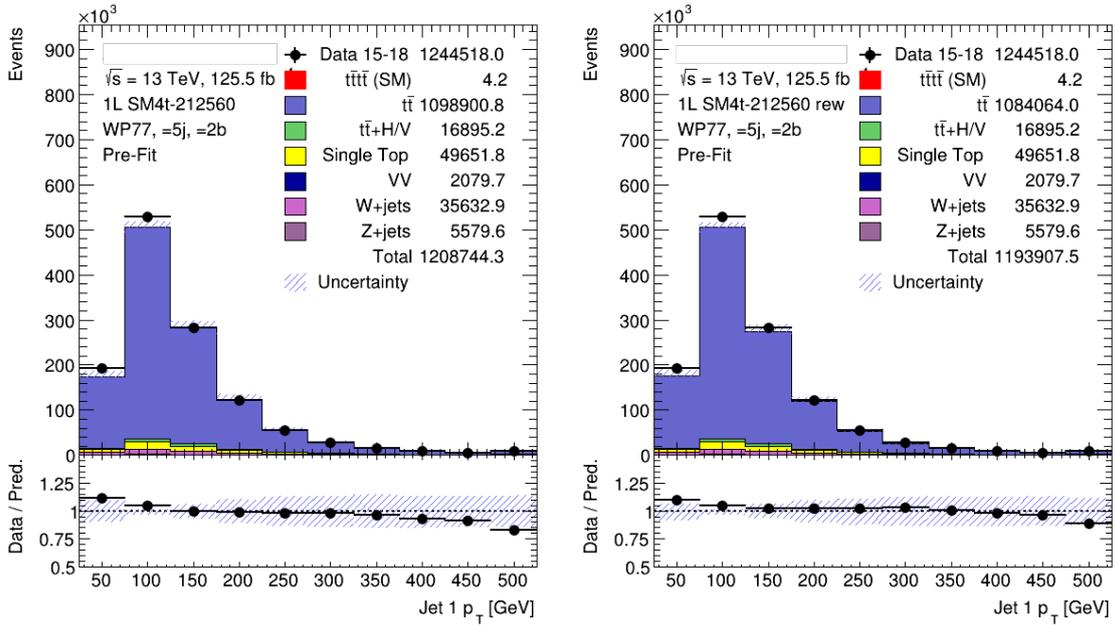
Figure A.5: Comparison between data and prediction of leading-jet p_T before any fit to data. The hashed area represents the combined statistical and systematic (on $t\bar{t}$ modelling) uncertainties of the prediction. Distributions are shown before (left) and after (right) the correction procedure. The comparison is done in the $5je2be$ region of the 1L channel of the $t\bar{t}t\bar{t}$ analysis.

Appendix B

$TRF_{t\bar{t}}$ efficiencies

In the following, the plots of the efficiencies extracted for the $TRF_{t\bar{t}}$ method are reported. Figures B.1 and B.2 show the efficiencies as a function of jet p_T for $3b$ and $\geq 4b$ regions, respectively. Figures B.3 and B.4 show, instead, the efficiencies as a function of $\Delta R_{jj}^{min} \times N_j$ for $3b$ and $\geq 4b$. All the efficiencies are extracted in the single lepton channel, in $5j$ region, and using the 77% WP for all the b-jets. They are calculated separately for the different campaigns and for the different flavour sub-processes.
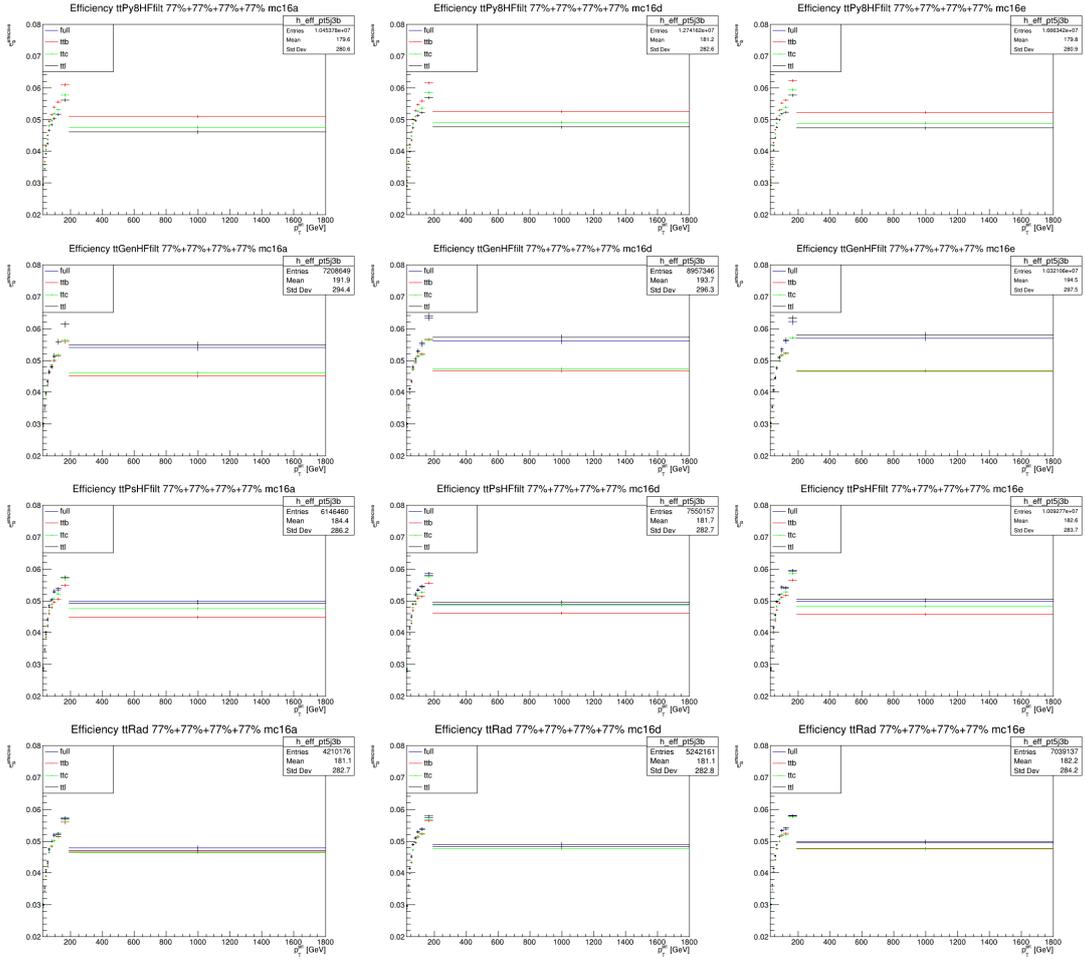
Figure B.1: Efficiencies for $TRF_{t\bar{t}}$ method for 3b regions as a function of jet p_T.
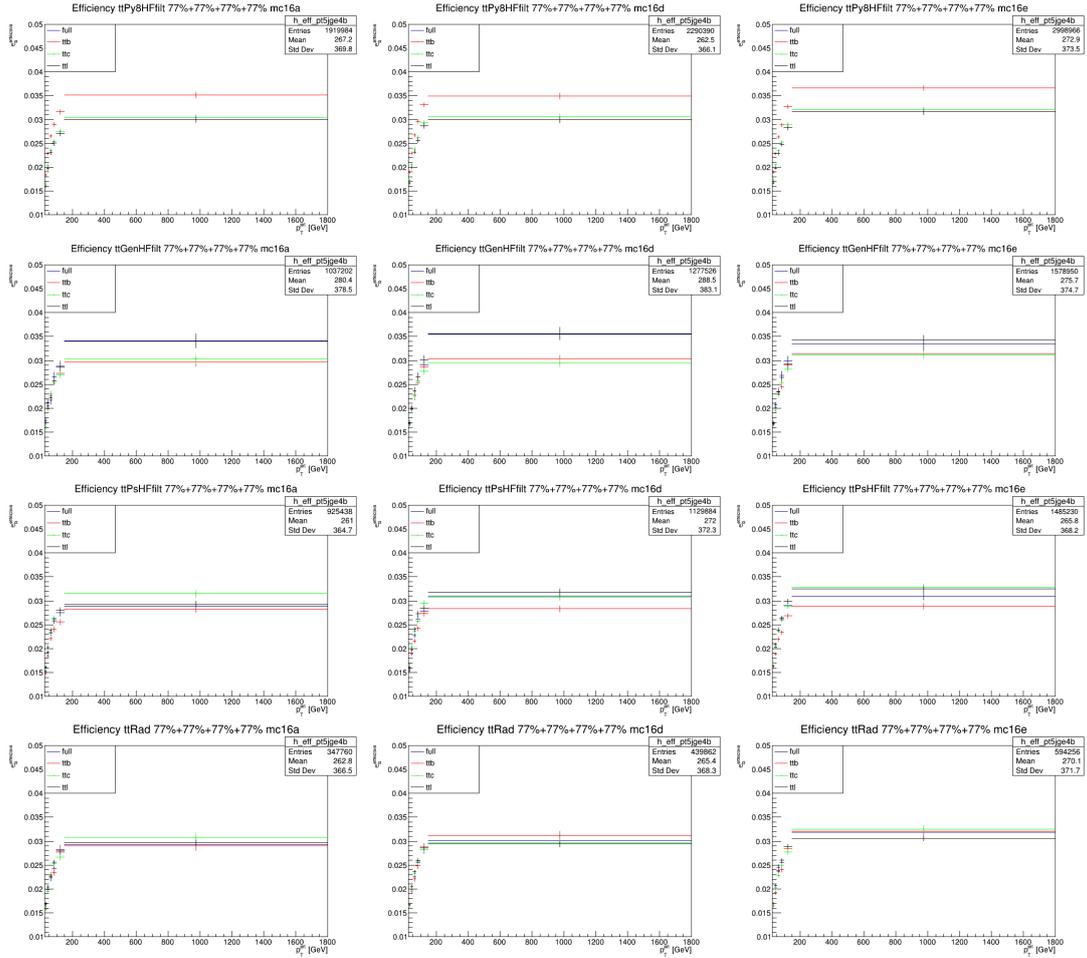
Figure B.2: Efficiencies for $TRF_{t\bar{t}}$ method for $\geq 4b$ regions as a function of jet p_T.
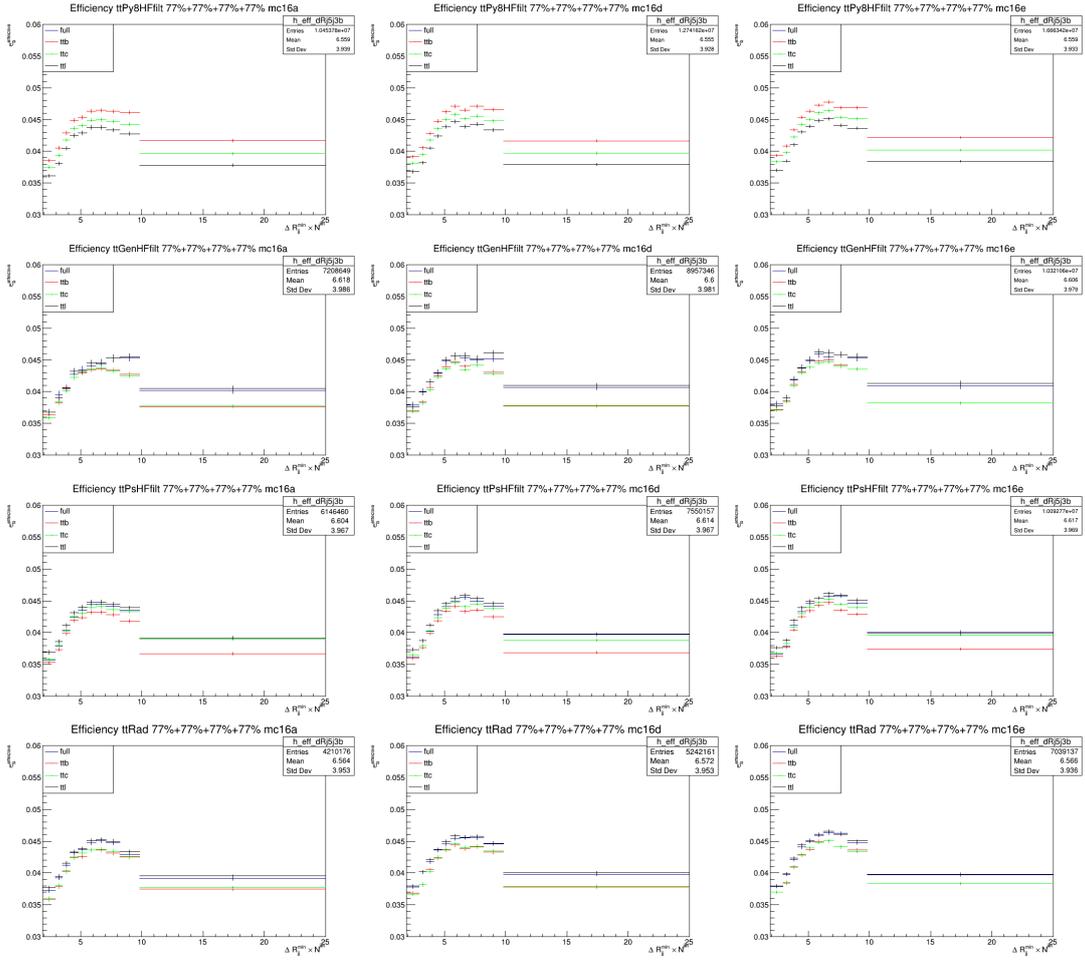
Figure B.3: Efficiencies for $TRF_{t\bar{t}}$ method for $3b$ regions as a function of $\Delta R_{jj}^{min} \times N_j$.
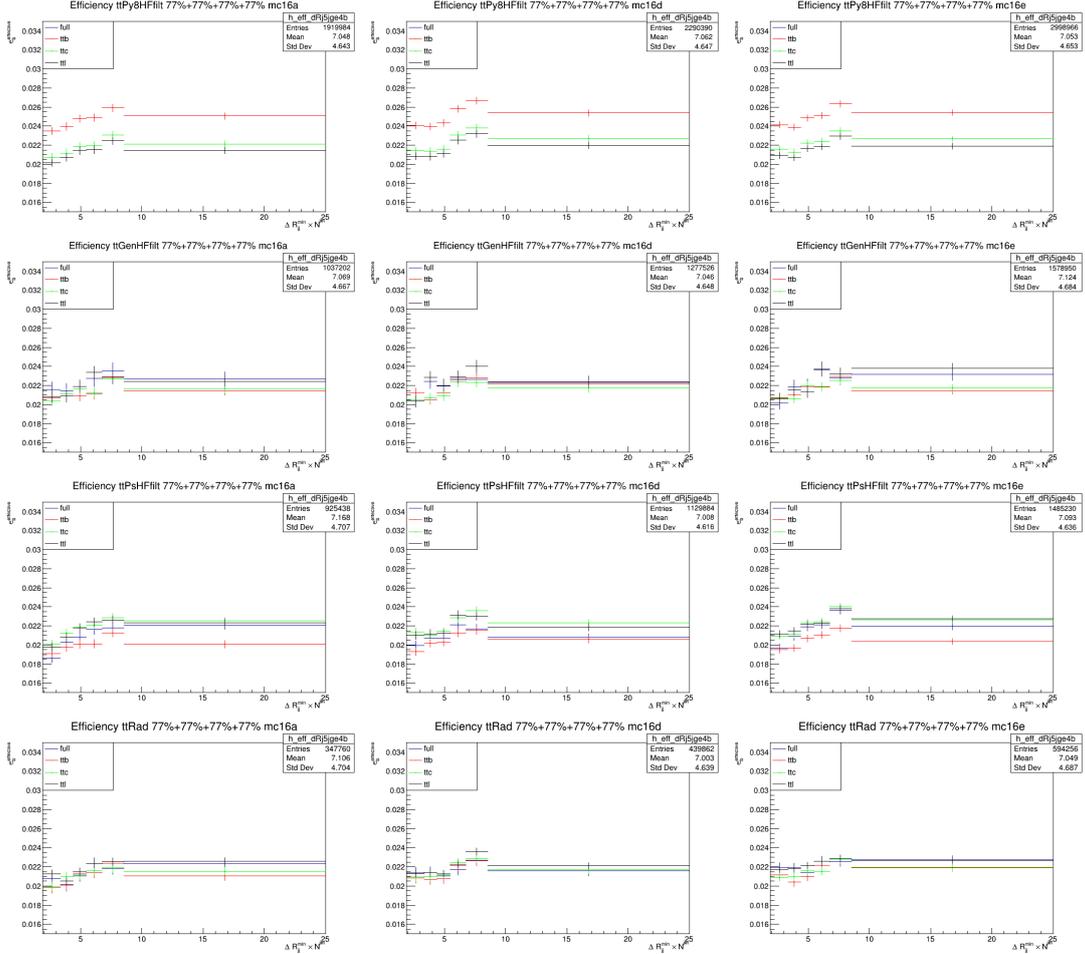
Figure B.4: Efficiencies for $TRF_{t\bar{t}}$ method for $\geq 4b$ regions as a function of $\Delta R_{jj}^{min} \times N_j$.

Appendix C

MVA ranking plots and overtraining tests

In the following, the plots already showed in Section 5.3 for the $10ji4bi$ region are reported for the other regions in which the training of the BDT is done. For each region, the ranking of the variables used in terms of separation and the overtraining test are displayed. For the regions with exactly nine jets, the jet multiplicity variable is automatically ignored during the training, since it is not able to provide any separation.
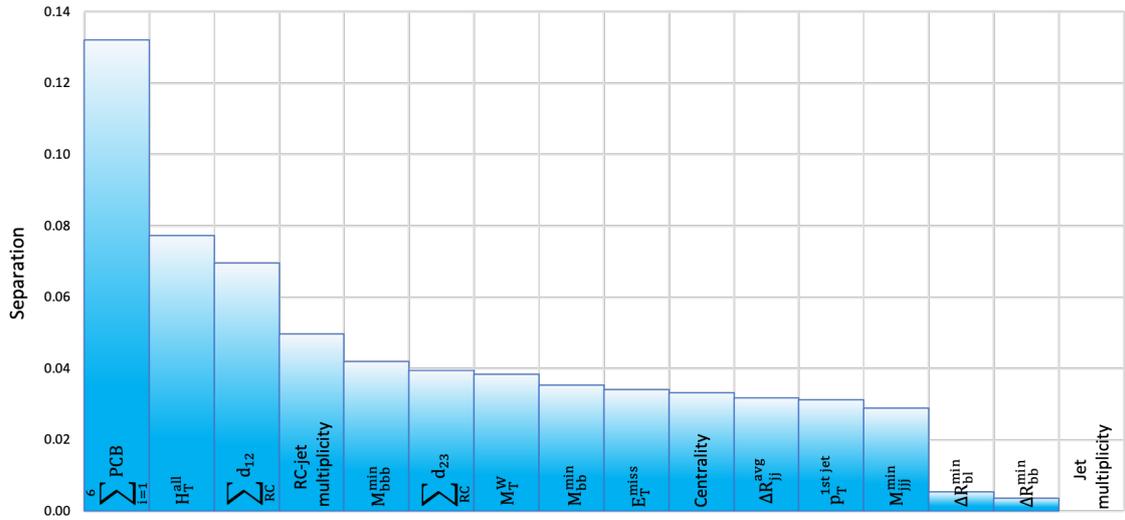
Figure C.1: Ranking of input variables in terms of separation in the $9je3be$ region for the training done on even events according to TMVA.
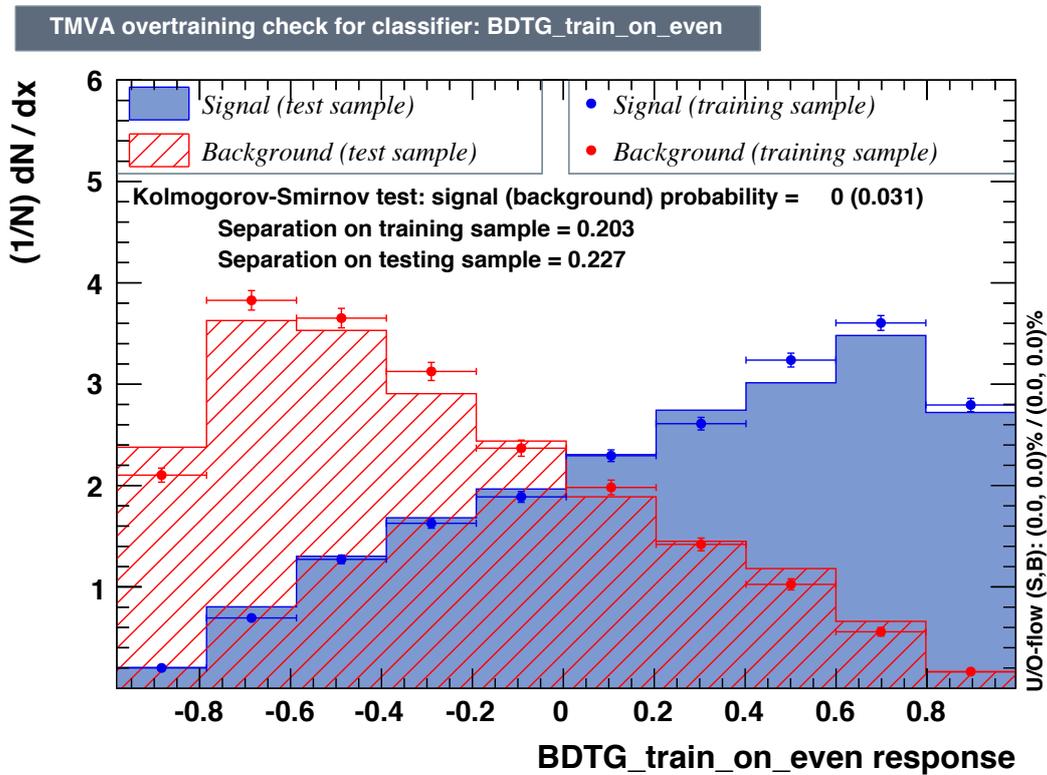
Figure C.2: Training response for the BDT in the $9je3be$ region for the training done on even events.
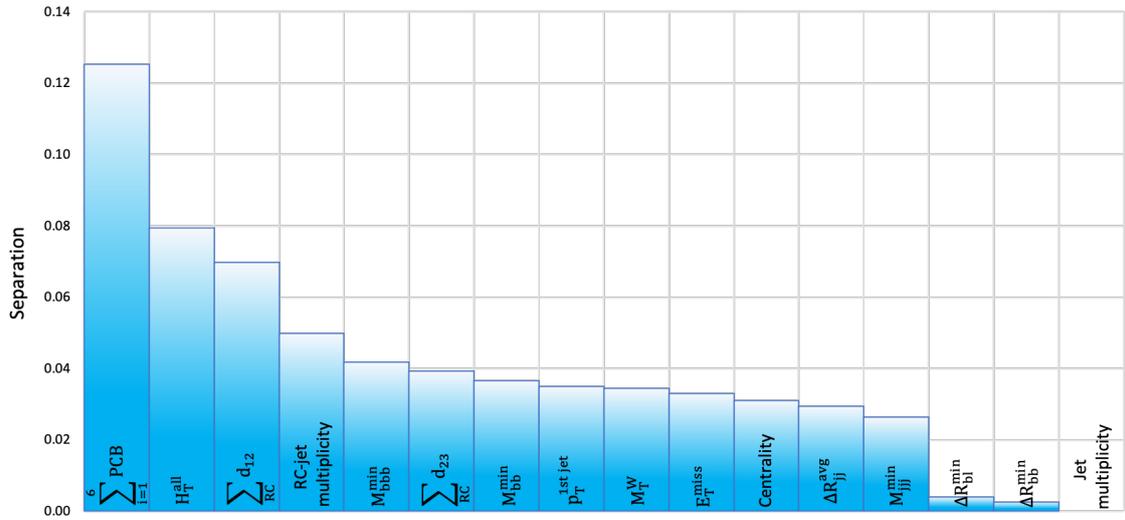
Figure C.3: Ranking of input variables in terms of separation in the $9je3be$ region for the training done on odd events according to TMVA.
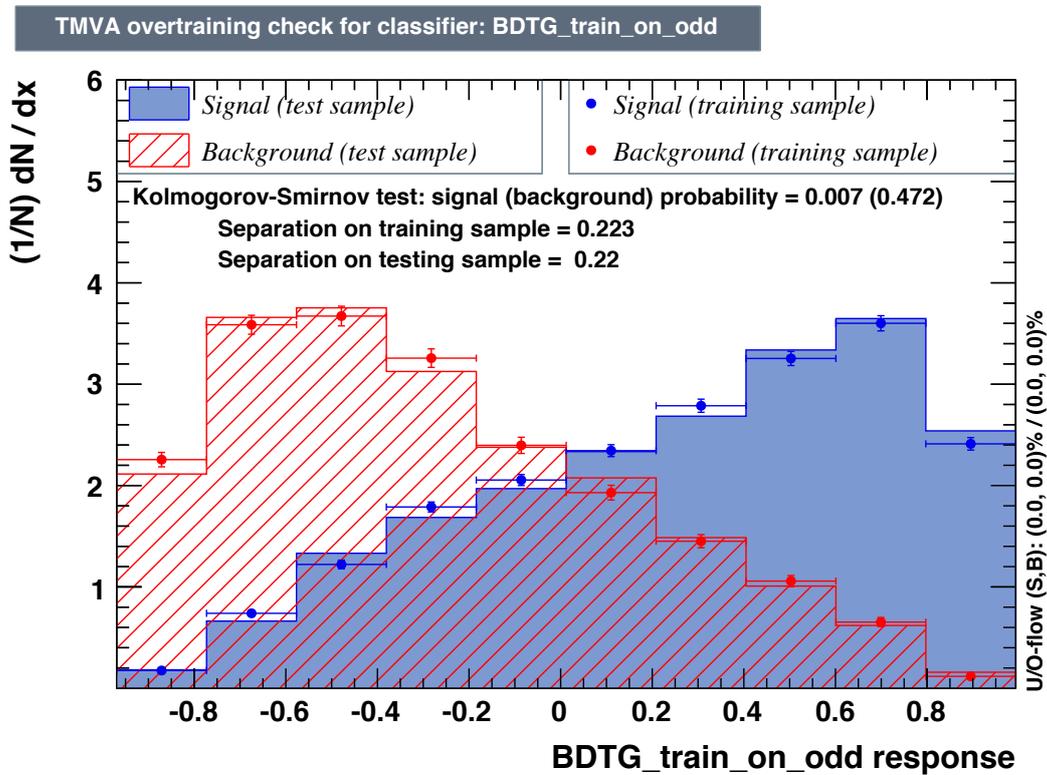
Figure C.4: Training response for the BDT in the $9je3be$ region for the training done on odd events.
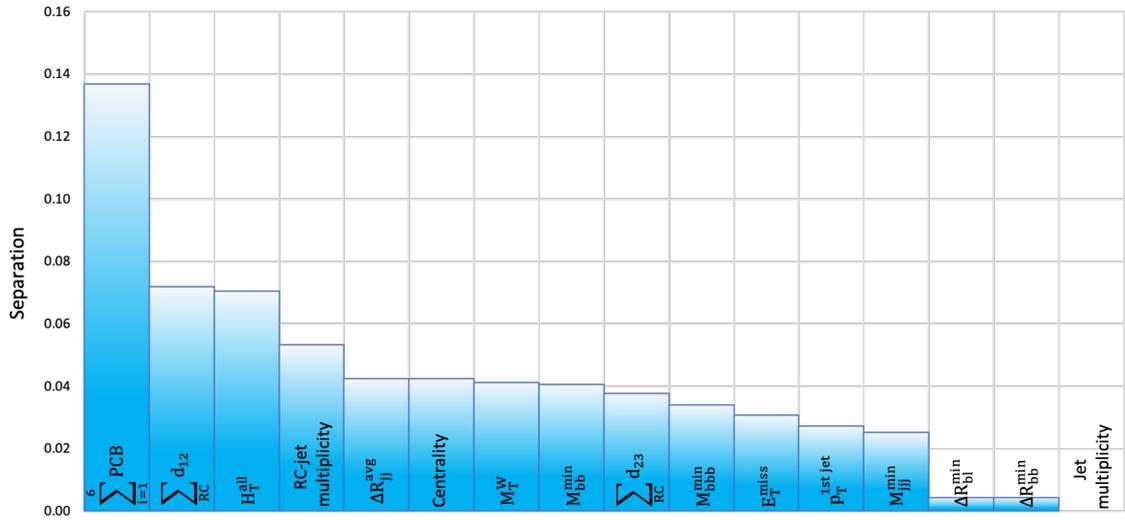
Figure C.5: Ranking of input variables in terms of separation in the $9je4bi$ region for the training done on even events according to TMVA.
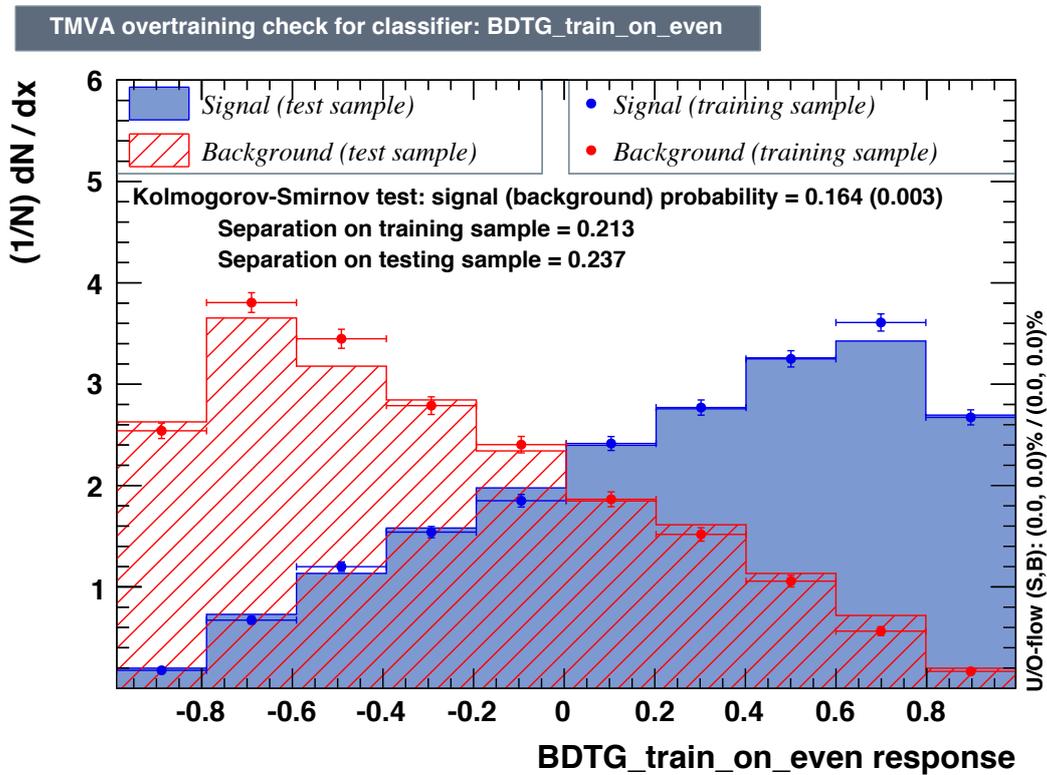
Figure C.6: Training response for the BDT in the $9je4bi$ region for the training done on even events.
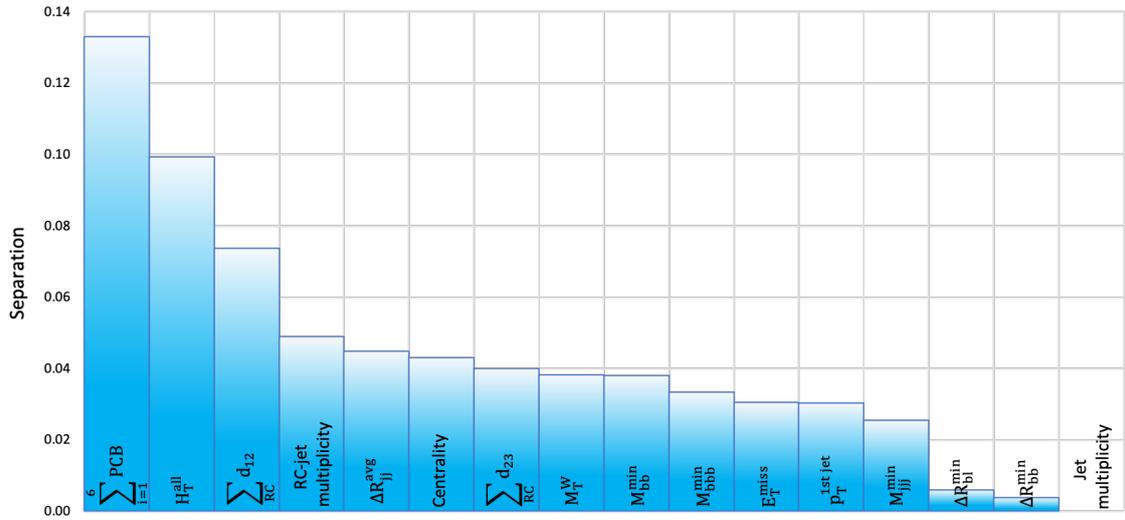
Figure C.7: Ranking of input variables in terms of separation in the $9je4bi$ region for the training done on odd events according to TMVA.
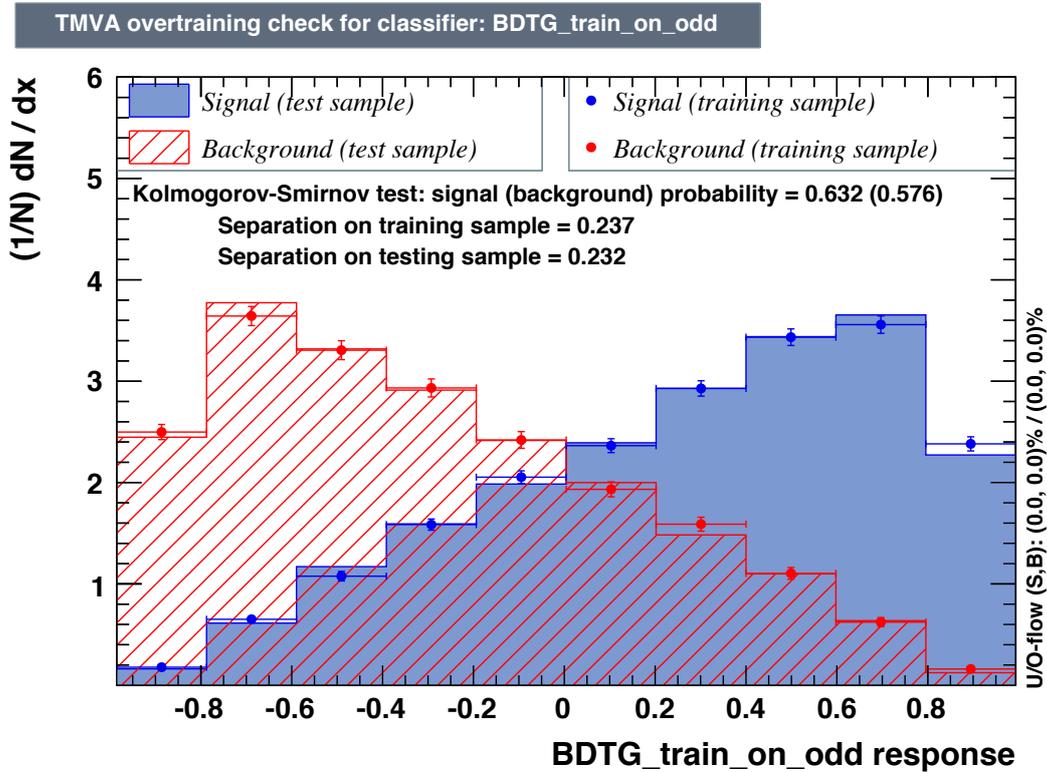
Figure C.8: Training response for the BDT in the $9je4bi$ region for the training done on odd events.
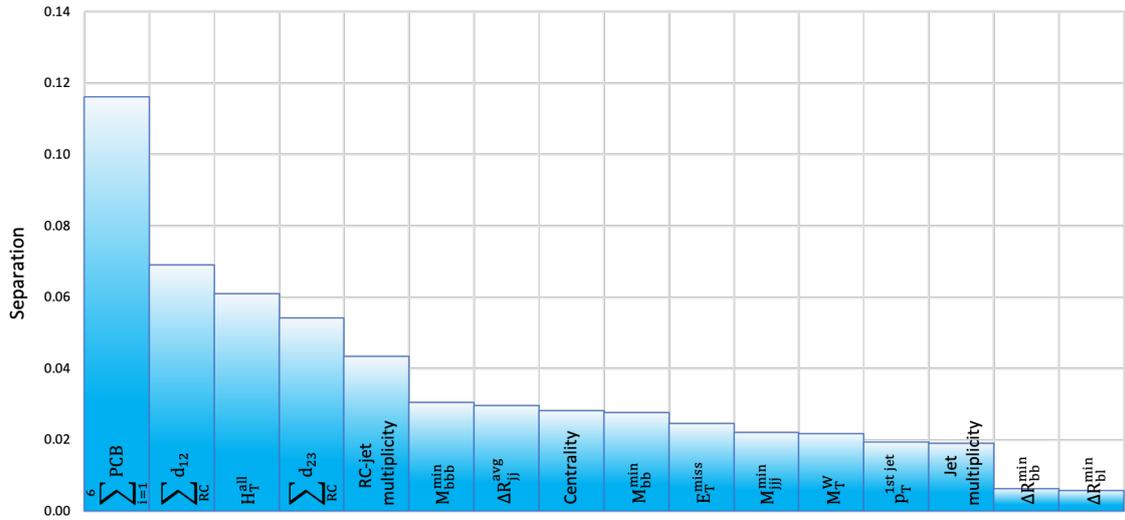
Figure C.9: Ranking of input variables in terms of separation in the $10ji3be$ region for the training done on even events according to TMVA.
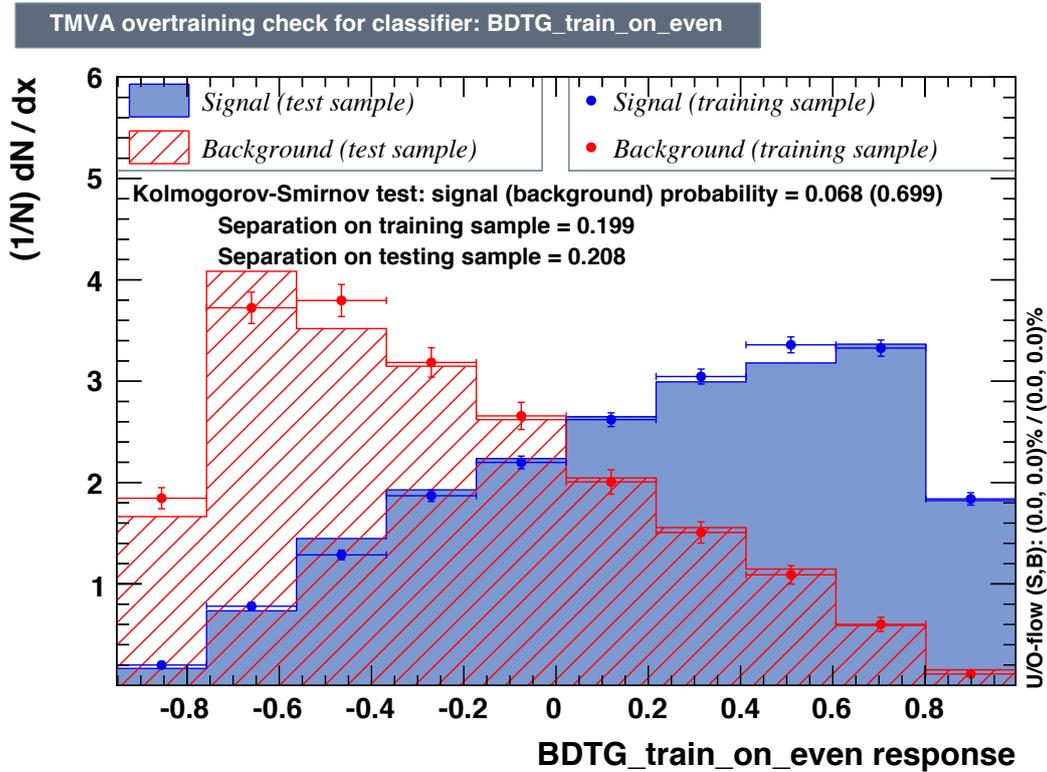
Figure C.10: Training response for the BDT in the $10ji3be$ region for the training done on even events.
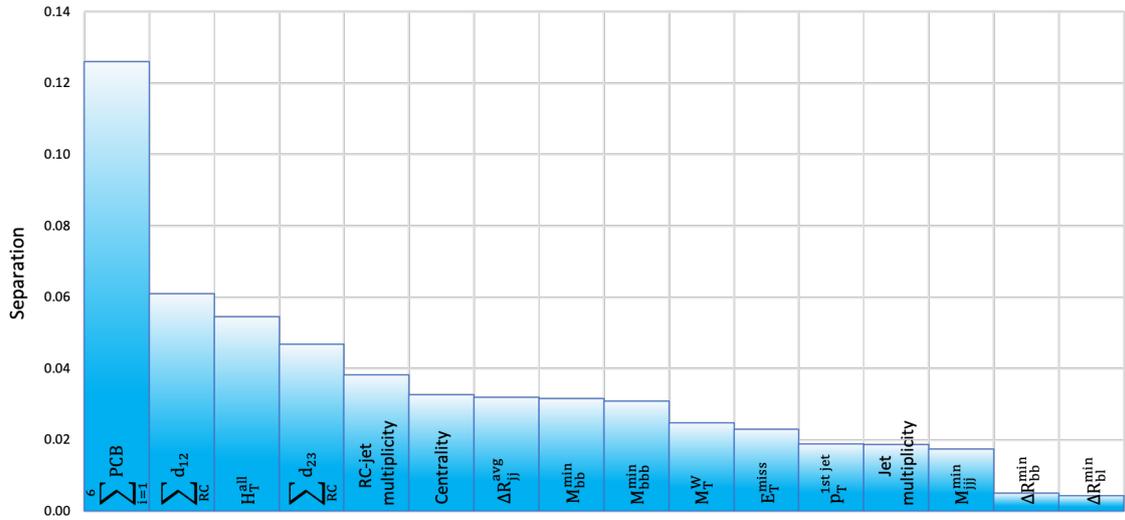
Figure C.11: Ranking of input variables in terms of separation in the $10ji3be$ region for the training done on odd events according to TMVA.
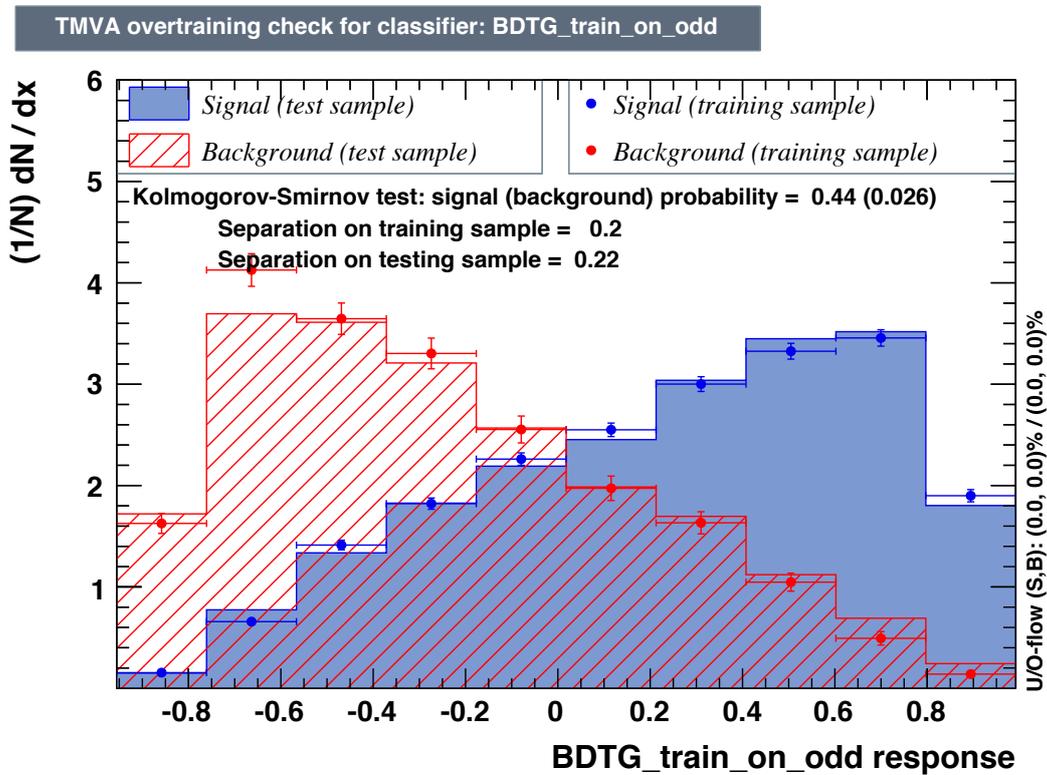
Figure C.12: Training response for the BDT in the $10ji3be$ region for the training done on odd events.
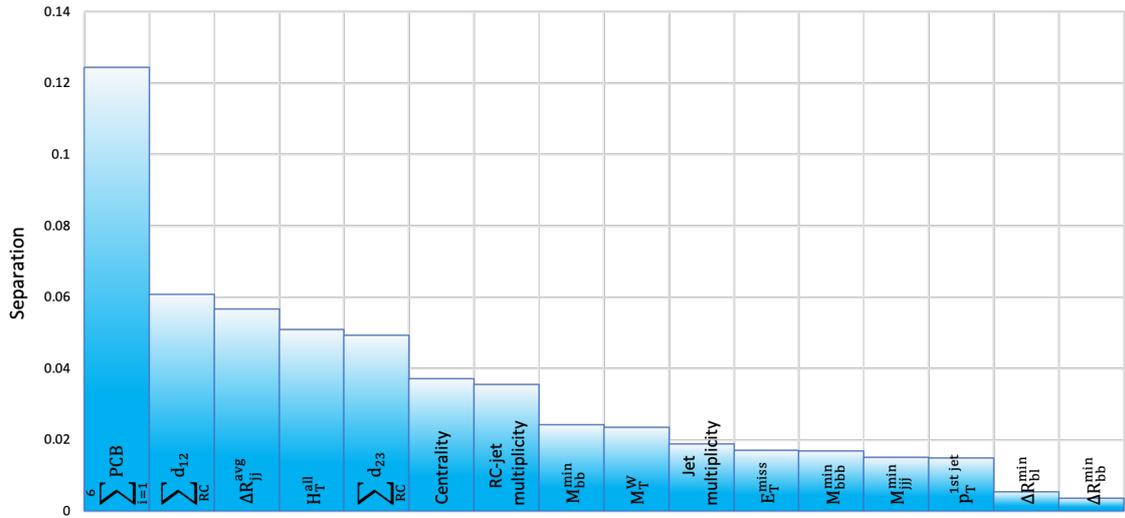
Figure C.13: Ranking of input variables in terms of separation in the $10ji4bi$ region for the training done on odd events according to TMVA.
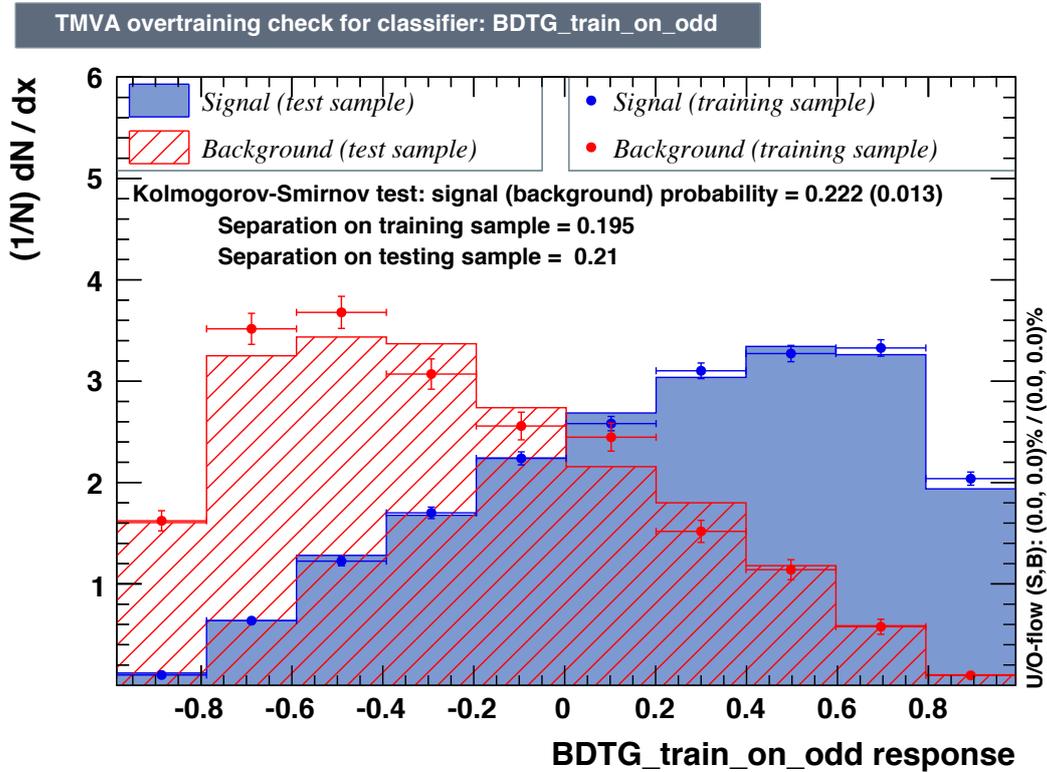
Figure C.14: Training response for the BDT in the $10ji4bi$ region for the training done on odd events.

Appendix D

Neural network test

D.1 Neural network with TMVA

For comparison purposes, a neural network is trained with TMVA. As for the BDT, the signal sample is provided by the MC simulation of the SM $t\bar{t}t\bar{t}$ process at NLO in QCD (MADGRAPH_AMC@NLO+PYTHIA), while the background is the $TRF_{t\bar{t}}$ prediction without the MC correction factor. The events with negative weights are not ignored, but fully exploited in the training. The events provided for the training are split randomly in two halves with the same amount of events: one is used for the training, and one for testing the overtraining. The input variables are the same of the BDT training. In this case the training is done only in the most signal-like region, $10ji4bi$. The neural network consists of three layers. The first layer is made of 16 neurons which accept the input variables. The second (hidden) layer is made of 300 neurons which are activated by the following function:

$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x \text{ otherwise} \end{cases}. \tag{D.1}$$

The last layer classifies the events as signal or background. The neurons are activated by the softmax function:

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}}. \tag{D.2}$$

The number of training cycles is 10. The integral of the ROC curve is 0.775, comparable with the BDT result. The overtraining test is displayed in Figure D.1: the signal sample shows a small overtraining between 0.6 and 0.8, while the background evidences the same effect between 0.1 and 0.3. This method does not perform better than the BDT, because it is not fully optimised since it would have needed much more input variables, instead of the only 16 variables used. But it is still useful to have another method to compare the BDT results with.
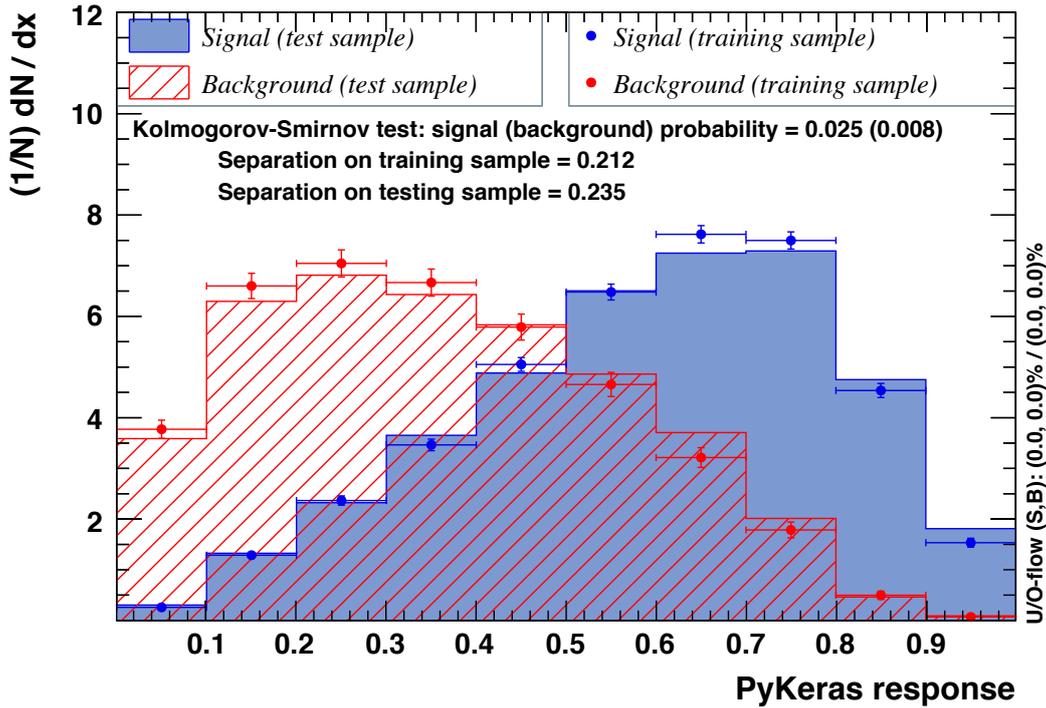
Figure D.1: Training response for the neural network in the $10ji4bi$ region.

D.2 Neural network results

For the comparison, the BDT distribution has been replaced by the neural network output only in the $10ji4bi$, while in the other signal regions the BDT output has been employed. Then a fit to the Asimov data set was performed. Figure D.2 shows the distribution of the MVA outputs after the fit. The constraints on the nuisance parameters from the fit are shown in Figure D.3. They look very similar to those already presented in Figure 6.11: the detector and the jet-tagging NPs do not present strong constraints, while some constraints are observed on the NPs associated to the theoretical uncertainties on the $t\bar{t}$+jet background estimation. In particular, the $t\bar{t}+ \geq 1b$ and $t\bar{t}$+light shape uncertainties on the choice of the parton shower generator are constrained to 51% and 59%, respectively. The fitted signal strength results to be $\mu = 1.00^{+2.19}_{-2.26}$. This result is a bit worse than the fit on the full BDT output, in terms of sensitivity on μ, but it is not unexpected, given the non-optimal configuration of the NN. In any case, it confirms the order of magnitude of the expected sensitivity of the analysis and the importance of some theoretical uncertainties with respect to others.

Figure D.2: Comparison between data and prediction of the BDT/NN distributions in the signal regions after the fit to data. The $t\bar{t}$+jet background is estimated through the data-driven method, with the MC-correction applied. The hashed area represents the combination of the statistical and systematic uncertainties on the background prediction.

tt̄ + ≥ 1c   cross   section
tt̄ + ≥ 1b   cross   section
tt̄ + light   Radiation   Shape
tt̄ + light   Radiation   Norm.
tt̄ + ≥ 1c   Radiation   Shape
tt̄ + ≥ 1c   Radiation   Norm.
tt̄ + ≥ 1b   Radiation   Shape
tt̄ + ≥ 1b   Radiation   Norm.
tt̄ + light   Parton   Shower   Shape
tt̄ + light   Parton   Shower   Norm.
tt̄ + ≥ 1c   Parton   Shower   Shape
tt̄ + ≥ 1c   Parton   Shower   Norm.
tt̄ + ≥ 1b   Parton   Shower   Shape
tt̄ + ≥ 1b   Parton   Shower   Norm.
tt̄ + light   Generator   Shape
tt̄ + light   Generator   Norm.
tt̄ + ≥ 1c   Generator   Shape
tt̄ + ≥ 1c   Generator   Norm.
tt̄ + ≥ 1b   Generator   Shape
tt̄ + ≥ 1b   Generator   Norm.
tt̄   METSoftTrkScale
tt̄   METSoftTrkResoPerp
tt̄   METSoftTrkResoPara
Luminosity
tt̄   JETRelativeNonClosureAFII
tt̄   JETPileupRhoTopology
tt̄   JETPileupOffsetNPV
tt̄   JETPileupOffsetMu
tt̄   JETJEREffectiveNP4
tt̄   JETJEREffectiveNP3
tt̄   JETJEREffectiveNP2
tt̄   JETJEREffectiveNP1
tt̄   JETJERDataVsMCAFII
tt̄   JETFlavorResponse
tt̄   JETFlavorComposition
tt̄   JETEtaIntercalibrationTotalStat
tt̄   JETEtaIntercalibrationModelling
tt̄   JETEffectiveNPStatistica l2
tt̄   JETEffectiveNPModelling4
tt̄   JETEffectiveNPModelling3
tt̄   JETEffectiveNPModelling2
tt̄   JETEffectiveNPModelling1
tt̄   JETBJESResponse
non – tt̄   generator
V/VV+jets cross section
Pile-up
JVT
tt̄ + X   cross   section
Single top cross section
MU SF Trigger SYST
MU SF ID SYST
EL SF ID
FTAG MV2c10 Light9
FTAG MV2c10 Light8
FTAG MV2c10 Light7
FTAG MV2c10 Light6
FTAG MV2c10 Light5
FTAG MV2c10 Light4
FTAG MV2c10 Light3
FTAG MV2c10 Light2
FTAG MV2c10 Light1
FTAG MV2c10 Light11
FTAG MV2c10 Light14
FTAG MV2c10 Light13
FTAG MV2c10 Light12
FTAG MV2c10 Light11
FTAG MV2c10 Light10
FTAG MV2c10 Light0
FTAG MV2c10 C9
FTAG MV2c10 C8
FTAG MV2c10 C7
FTAG MV2c10 C6
FTAG MV2c10 C5
FTAG MV2c10 C4
FTAG MV2c10 C3
FTAG MV2c10 C2
FTAG MV2c10 C1
FTAG MV2c10 C19
FTAG MV2c10 C18
FTAG MV2c10 C16
FTAG MV2c10 C15
FTAG MV2c10 C14
FTAG MV2c10 C13
FTAG MV2c10 C12
FTAG MV2c10 C11
FTAG MV2c10 C10
FTAG MV2c10 C0
FTAG MV2c10 B9
FTAG MV2c10 B8
FTAG MV2c10 B7
FTAG MV2c10 B6
FTAG MV2c10 B5
FTAG MV2c10 B4
FTAG MV2c10 B44
FTAG MV2c10 B43
FTAG MV2c10 B42
FTAG MV2c10 B41
FTAG MV2c10 B3
FTAG MV2c10 B39
FTAG MV2c10 B38
FTAG MV2c10 B37
FTAG MV2c10 B36
FTAG MV2c10 B35
FTAG MV2c10 B34
FTAG MV2c10 B32
FTAG MV2c10 B31
FTAG MV2c10 B2
FTAG MV2c10 B26
FTAG MV2c10 B25
FTAG MV2c10 B24
FTAG MV2c10 B23
FTAG MV2c10 B22
FTAG MV2c10 B21
FTAG MV2c10 B20
FTAG MV2c10 B1
FTAG MV2c10 B19
FTAG MV2c10 B18
FTAG MV2c10 B17
FTAG MV2c10 B16
FTAG MV2c10 B15
FTAG MV2c10 B14
FTAG MV2c10 B13
FTAG MV2c10 B12
FTAG MV2c10 B11
FTAG MV2c10 B10
FTAG MV2c10 B0

$(\theta-\theta_0)/\Delta\theta$

Figure D.3: Fitted nuisance parameters on data under the signal-plus-background hypothesis in the signal regions.

Bibliography

[1] S. L. Glashow, *"Partial symmetries of weak interactions"*, Nucl. Phys. **22**, 579 (1961).

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

[3] A. Salam, *"Elementary Particle Physics: Relativistic Groups and Analyticity"*, Eighth Nobel Symposium, 367 (1968).

[4] G. Arnison et al., *"Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c^2 at the CERN SPS Collider"*, Phys. Lett. **B126**, 398 (1983).

[5] ATLAS collaboration, *"Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC"*, Phys. Lett. **B716** (2012), `arXiv:1207.7214 [hep-ex]`.

[6] CMS Collaboration, *"Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC"*, Phys. Lett. **B716** (2012), `arXiv:1207.7235 [hep-ex]`.

[7] P.A. Zyla et al. (Particle Data Group), *"Review of Particle Physics"*, Prog. Theor. Exp. Phys. 2020 **083C01** (2020).

[8] P. W. Higgs, *"Broken symmetries, massless particles and gauge fields"*, Phys. Rev. Lett. **12** (1964);
F. Englert and R. Brout, *"Broken Symmetry and the Mass of Gauge Vector Mesons"*, Phys. Rev. Lett. **13** (1964);
G. Guralnik, C. Hagen, and T. Kibble, *"Global Conservation Laws and Massless Particles"*, Phys. Rev. Lett. **13** (1964);
P. W. Higgs, *"Broken Symmetries and the Masses of Gauge Bosons"*, Phys. Rev. Lett. **13** (1964).

[9] H. E. Logan, *"TASI 2013 lectures on Higgs physics within and beyond the Standard Model"*, `arXiv:1406.1786 [hep-ph]`.

[10] N. Cabibbo, *"Unitary Symmetry and Leptonic Decays"*, Phys. Rev. Lett. **10**, 531 (1963);
M. Kobayashi and T. Maskawa, *"CP Violation in the Renormalizable Theory of Weak Interaction"*, Prog. Theor. Phys. **49**, 652 (1973).

[11] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, *"Advantages of the Color Octet Gluon Picture"*, Phys. Lett. **B47**, 365 (1973).

[12] S. Weinberg, *"Nonabelian Gauge Theories of the Strong Interactions"*, Phys. Rev. Lett. **31**, 494 (1973).

[13] D. J. Gross and F. Wilczek, *"Asymptotically Free Gauge Theories"*, Phys. Rev. **D8** (1973).

[14] G. Altarelli and G. Parisi, *"Asymptotic Freedom in Parton Language"*, Phys. Lett. **B126** (1977).

[15] CDF Collaboration, *"Observation of top quark production in $p\bar{p}$ collisions with the Collider Detector at Fermilab"*, Phys. Rev. Lett. **74** (1995).

[16] D0 Collaboration, *"Observation of the top quark"*, Phys. Rev. Lett. **74** (1995).

[17] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots.

[18] B. S. Acharya et al., *"Identifying Multi-Top Events from Gluino Decay at the LHC"*, `arXiv:0901.3367 [hep-ph]`;
A. Deandrea and N. Deutschmann, *"Multi-tops at the LHC"*, JHEP **1408** (2014) 134, `arXiv:1405.6119 [hep-ph]`;
E. Alvarez et al., *"Four tops for LHC"*, Nucl. Phys. **B915** (2017) 19.

[19] R. Frederix, D. Pagani and M. Zaro, *"Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions"*, JHEP **02** (2018) 031, `arXiv:1711.02116 [hep-ph]`.

[20] ATLAS Collaboration, *"Search for four-top-quark production in final states with one charged lepton and multiple jets using 3.2 fb^{-1} of proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector at the LHC"*, ATLAS-CONF-2016-020, `https://cds.cern.ch/record/2144537`;
ATLAS Collaboration, *"Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector"*, JHEP **12** (2018) 039, `arXiv:1807.11883 [hep-ex]`;
ATLAS Collaboration, *"Search for four-top-quark production in the single-lepton and opposite-sign dilepton final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector"*, Phys. Rev. **D99**, 052009 (2019), `arXiv:1811.02305 [hep-ex]`.

[21] CMS Collaboration, *"Search for standard model production of four top quarks with same-sign and multilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV"*, Eur. Phys. J. **C78** (2018) 140, `arXiv:1710.10614 [hep-ex]`;
CMS Collaboration, *"Search for standard model production of four top quarks in final states with same-sign and multiple leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV"*, Eur. Phys. J. **C80** (2020) 75, `arXiv:1908.06463 [hep-ex]`;
CMS Collaboration, *" Search for standard model production of four top quarks with the single-lepton and opposite-sign dilepton final states in proton-proton collisions at $\sqrt{s}$ = 13 TeV"*, JHEP **11** (2019) 082, `arXiv:1906.02805 [hep-ex]`.

[22] L. Evans and P. Bryant, *"LHC Machine"*, JINST **3**, S08001 (2008).

[23] ATLAS Collaboration, *"The ATLAS Experiment at the CERN Large Hadron Collider"*, JINST **3**, S08003 (2008).

[24] CMS Collaboration, *"The CMS experiment at the CERN LHC"*, JINST **3**, S08004 (2008).

[25] ALICE Collaboration, *"The ALICE experiment at the CERN LHC"*, JINST **3**, S08002 (2008).

[26] LHCb Collaboration, *"The LHCb Detector at the LHC"*, JINST **3**, S08005 (2008).

[27] LHCf Collaboration, *"The LHCf detector at the CERN Large Hadron Collider"*, JINST **3**, S08006 (2008).

[28] TOTEM Collaboration, *"The TOTEM Experiment at the CERN Large Hadron Collider"*, JINST **3**, S08007 (2008).

[29] MoEDAL Collaboration, *"Technical Design Report of the MoEDAL Experiment"*, CERN-LHCC-2009-006, MoEDAL-TDR-001.

[30] https://cds.cern.ch/record/2197559.

[31] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[32] ATLAS Collaboration, *"Alignment of the ATLAS Inner Detector in Run-2"*, CERN-EP-2020-108, `arXiv:2007.07624 [hep-ex]`.

[33] ATLAS Collaboration, *"ATLAS Insertable B-Layer Technical Design Report"*, Tech. Rep. CERN-LHCC-2010-013, ATLAS-TDR-19.

[34] ATLAS Collaboration, *"Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC"*, Eur. Phys. J. **C76** (2016) 653, `arXiv:1608.03953 [hep-ex]`.

[35] G. Avoni et al., *"The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS"*, JINST **13** (2018) P07017.

[36] S. Abdel Khalek et al., *"The ALFA Roman Pot Detectors of ATLAS"*, JINST **11** (2016) P11013, `arXiv:1609.00249`.

[37] S. White, *"The ATLAS zero degree calorimeter"*, Nucl. Instrum. Meth. **A617** (2010) 126.

[38] ATLAS Collaboration, *"Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC"*, Eur. Phys. J. **C73** (2013) 2518, `arXiv:1302.4393`.

[39] https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ.

[40] ATLAS Collaboration, *"Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC"*, Eur. Phys. J. **C77** (2017) 332, `arXiv:1611.10235 [physics.ins-det]`.

[41] ATLAS Collaboration, *"Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1"*, Eur. Phys. J. **C77** (2017) 490, `arXiv:1603.02934 [hep-ex]`.

[42] M. Cacciari, G. P. Salam, and G. Soyez, *"The Anti-k_t jet clustering algorithm"*, JHEP **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[43] ATLAS Collaboration, *"Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run"*, ATL-PHYS-PUB-2017-013 (2017), `https://cds.cern.ch/record/2273281`.

[44] A. Höcker et al., *"TMVA-toolkit for multivariate data analysis"*, (2007), `arXiv:physics/0703039 [physics.data-an]`.

[45] ATLAS Collaboration, *"ATLAS b-jet identification performance and efficiency measurement with tt events in pp collisions at 13 TeV"*, Eur. Phys. J. **C79** (2019) 970, `arXiv:1907.05120 [hep-ex]`.

[46] ATLAS Collaboration, *"Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector"*, Eur. Phys. J. C 80 (2020) 1085, `arXiv:2007.14858 [hep-ex]`.

[47] ATLAS Collaboration, *"The ATLAS Simulation Infrastructure"*, Eur. Phys. J. **C70** (2010) 823, `arXiv:1005.4568 [physics.ins-det]`.

[48] E. Bothmann et al., *"Event Generation with Sherpa 2.2"*, SciPost Phys. **7** (2019) 034, `arXiv:1905.09127 [hep-ph]`.

[49] NNPDF Collaboration, R.D. Ball et al., *"Parton distributions for the LHC run II"*, JHEP **04** (2015) 040, `arXiv:1410.8849 [hep-ph]`.

[50] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *"Parton fragmentation and string dynamics"*, Physics Reports **97** (1983) 31.

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

[52] 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]`.

[53] P. Nason, *"A new method for combining NLO QCD with shower Monte Carlo algorithms"*, JHEP **11** (2004) 040, `arXiv:hep-ph/0409146`.

[54] J. Alwall et al., *"The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations"*, JHEP **07** (2014) 079, `arXiv:1405.0301 [hep-ph]`.

[55] S. Frixione and B. R. Webber, *"Matching NLO QCD computations and parton shower simulations"*, JHEP **06** (2002) 029.

[56] T. Sjöstrand et al., *"An introduction to PYTHIA 8.2"*, Comput. Phys. Commun. **191** (2015) 159, `arXiv:1410.3012 [hep-ph]`.

[57] J. Bellm et al., *"Herwig 7.0/Herwig++ 3.0 release note"*, Eur. Phys. J. **C76** (2016) 196, `arXiv:1512.01178 [hep-ph]`.

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

[59] E. Bothmann et al., *"Event Generation with Sherpa 2.2"*, SciPost Phys. **7** (2019) 034, `arXiv:1905.09127 [hep-ph]`.

[60] F. Cascioli, P. Maierhofer and S. Pozzorini, *"Scattering Amplitudes with Open Loops"*, Phys. Rev. Lett. **108** (2012) 111601, `arXiv:1111.5206 [hep-ph]`.

[61] GEANT4 Collaboration, *"GEANT4: A Simulation toolkit"*, Nucl. Instrum. Meth. **A506** (2003) 250.

[62] ATLAS Collaboration, *"The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim"*, ATL-PHYS-PUB-2010-013, 2010, `https://cds.cern.ch/record/1300517`.

[63] NNPDF Collaboration, *"Parton distributions from high-precision collider data"*, `arXiv:1706.00428 [hep-ph]`.

[64] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *"Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations"*, JHEP **1303** (2013) 015, `arXiv:1212.3460 [hep-ph]`.

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

[66] NNPDF Collaboration, *"Parton distributions with LHC data"*, `arXiv:1207.1303 [hep-ph]`.

[67] ATLAS Collaboration, *"Studies on top-quark Monte Carlo modelling for Top2016"*, ATL-PHYS-PUB-2016-020, 2016, `https://cds.cern.ch/record/2216168`.

[68] L. Harland-Lang, A. Martin, P. Motylinski and R. Thorne, *"Parton distributions in the LHC era: MMHT 2014 PDFs"*, Eur. Phys. J. **C75** (2015) 204, `arXiv:1412.3989 [hep-ph]`.

[69] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber and C. D. White, *"Single-top hadroproduction in association with a W boson"*, JHEP **07** (2008) 029, `arXiv:0805.3067 [hep-ph]`.

[70] ATLAS Collaboration, *"Modelling of rare top quark processes at $\sqrt{s} = 13$ TeV in ATLAS"*, ATL-PHYS-PUB-2020-024.

[71] L. Serkin, *"Search for the Standard Model Higgs boson produced in association with a pair of top quarks and decaying into a $b\bar{b}$-pair in the single lepton channel at $\sqrt{s} = 8$ TeV with the ATLAS experiment at the LHC"*, CERN-THESIS-2016-089, II.Physik-UniGö-Diss-2016/02.

[72] ATLAS Collaboration, *"Luminosity Determination in pp Collisions at $\sqrt{s} = 8$ TeV using the ATLAS Detector at the LHC"*, Eur. Phys. J. C **76** (2016) 653, `arXiv:1608.03953 [hep-ex]`.

[73] ATLAS Collaboration, *"Tagging and suppression of pileup jets with the ATLAS detector"*, ATLAS-CONF-2014-018, 2014, `https://cds.cern.ch/record/1700870`.

[74] ATLAS Collaboration, *"Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV"*, ATL-PHYS-PUB-2015-015, 2015, `https://cds.cern.ch/record/2037613`.

[75] P. C. Bhat, *"Multivariate Analysis Methods in Particle Physics"*, Ann. Rev. Nucl. Part. Sci. **61** (2011) 281.

[76] F. Rosenblatt, *"The perceptron: A probabilistic model for information storage and organization in the brain"*, Psychological Review **65(6)** (1958) 386, `https://doi.org/10.1037/h0042519`.

[77] R. Brun and F. Rademakers, *"ROOT - An Object Oriented Data Analysis Framework"*, Nucl. Instum. Meth. in Phys. Res. **A389** (1997) 81.

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

[79] ATLAS Collaboration, *"Measurements of top-quark pair differential cross-sections in the eµ channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector"*, Eur. Phys. J. **C77** (2017) 292, `arXiv:1612.05220 [hep-ex]`.

[80] CMS Collaboration, *"Measurements of differential cross sections for $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons"*, tech. rep. CMS-PAS-TOP-17-014, CERN, 2018, `https://cds.cern.ch/record/2621975`.

[81] M. Czakon et al., *"Top-pair production at the LHC through NNLO QCD and NLO EW"*, JHEP **10** (2017) 186, `arXiv:1705.04105 [hep-ph]`.

[82] ATLAS Collaboration, *"Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s}=13$ TeV with the ATLAS experiment"*, Phys. Lett. **B810** (2020) 135797, `arXiv:2006.13076 [hep-ex]`.