Search for the production of four top quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV in the single lepton and opposite-sign dilepton channels with the ATLAS detector at the Large Hadron Collider

Dissertation

for the award of the degree
„Doctor of Philosophy“(Ph.D.)
of the Georg-August-Universität Göttingen

within the doctoral program „Physics“
of the Georg-August University School of Science (GAUSS)

submitted by

Paolo Sabatini

from Abbadia San Salvatore (Italy)

Göttingen, 2019

Thesis Committee

Prof. Dr. Arnulf Quadt
Prof. Dr. Stan Lai

Members of the Examination Board:

Reviewer:	Prof. Dr. Arnulf Quadt II. Physikalisches Institut, Georg-August-Universität Göttingen
Second Reviewer:	Prof. Dr. Stan Lai II. Physikalisches Institut, Georg-August-Universität Göttingen

Further members of the Examination Board:

Prof. Dr. Marcus Baum
Institute of Computer Science, Georg-August-Universität Göttingen

Prof. Dr. Ariane Frey
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Hans Hofsäss
II. Physikalisches Institut, Georg-August-Universität Göttingen

Prof. Dr. Steffen Schumann
Institut für Theoretische Physik, Georg-August-Universität Göttingen

Date of the oral examiniation: 10/02/2020

Reference: II.Physik-UniGö-Diss-2019/09

Search for the production of four top quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV in the single lepton and opposite-sign dilepton channels with the ATLAS detector at the Large Hadron Collider

Abstract

This thesis presents the search for the production of $t\bar{t}t\bar{t}$ from pp collision at $\sqrt{s} = 13$ TeV. The analysed dataset has been collected by the ATLAS detector during the LHC Run 2, corresponding to an integrated luminosity of 139 fb^{-1}. The analysis uses events containing either one lepton or two leptons with opposite-sign charges, at large jet and b-jet multiplicities. The output of a Boosted Decision Tree is used as a discriminating variable in a profile likelihood fit for the measurement of the $t\bar{t}t\bar{t}$ signal strength $\mu_{t\bar{t}t\bar{t}} = \sigma_{t\bar{t}t\bar{t}}/\sigma^{\mathrm{SM}}_{t\bar{t}t\bar{t}}$. The estimation of the dominant $t\bar{t}$+jets background is given by the data-driven ttTRFmethod. The measured signal strength in the single lepton and opposite-sign dilepton channel is $\mu^{\mathrm{1L+OS2L}}_{t\bar{t}t\bar{t}} = 0.9^{+0.5}_{-0.5}\ (\mathrm{stat.})^{+1.0}_{-0.9}\ (\mathrm{syst.}) = 0.9^{+1.1}_{-1.0}$, compatible with the Standard Model expectation and corresponding to a deviation of 0.9 (1.0) standard deviations from the background-only hypothesis. This result is combined with a complementary measurement, using events with two leptons with same-sign charges or three leptons in the final state. The combination provides a measurement of the $t\bar{t}t\bar{t}$ cross-section of 24 ± 6 fb, compatible with the Standard Model prediction within 1.8 standard deviations. The measured significance of the $t\bar{t}t\bar{t}$ process over the background of 4.4 standard deviations, providing evidence of the $t\bar{t}t\bar{t}$ production process.

Search for the production of four top quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV in the single lepton and opposite-sign dilepton channels with the ATLAS detector at the Large Hadron Collider

Zusammenfassung

Die folgende Arbeit präsentiert eine Messung des Prozesses $t\bar{t}t\bar{t}$ in pp-Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV. Der hierfür genutzte Datensatz wurde mit dem ATLAS-Detektor während des Run 2 des LHC genommen und entspricht einer integrierten Luminosität von 139 fb^{-1}. Die Analyse selektiert Ereignisse mit entweder einem Lepton oder zwei Leptonen mit unterschiedlichem Ladungsvorzeichen bei hohen Jet- und b-Jet-Multiplizitäten. Als diskriminierende Variable in einem Profile-Likelihood-Fit kommt das Ausgangssignal eines Boosted Decision Trees zum Einsatz, um die $t\bar{t}t\bar{t}$-Signalstärke $\mu_{t\bar{t}t\bar{t}} = \sigma_{t\bar{t}t\bar{t}}/\sigma^{\mathrm{SM}}_{t\bar{t}t\bar{t}}$ zu messen. Eine Abschätzung des dominanten Untergrund-Prozesses $t\bar{t}$+jets wurde mittels der datenbasierten $t\bar{t}$TRF-Methode durchgeführt. Die gemessene Signalstärke in den Ein-Lepton- und Zwei-Lepton-Kanälen (mit unterschiedlichem Ladungsvorzeichen) beträgt $\mu^{\mathrm{1L+OS2L}}_{t\bar{t}t\bar{t}} = 0.9^{+0.5}_{-0.5}\ (\mathrm{stat.})^{+1.0}_{-0.9}\ (\mathrm{syst.}) = 0.9^{+1.1}_{-1.0}$ und ist damit kompatibel mit dem Erwartungswert des Standardmodells; die Abweichung von der Untergrund-Hypothese beträgt 0.9 (1.0) Standardabweichungen. Dieses Ergebnis wird kombiniert mit einer komplementären Messung in Endzuständen mit zwei Leptonen gleichen Ladungsvorzeichens und mit drei Leptonen. Der gemessene Wirkungsquerschnitt im kombinierten Szenario beträgt 24 ± 6 fb und ist kompatibel mit dem erwarteten Wert des Standardmodells innerhalb der 1.8σ-Umgebung. Mit 4.4 Standardabweichungen gegenüber der Untergrund-Hypothese stellt diese Messung eine Evidenz für $t\bar{t}t\bar{t}$-Produktion dar.

Acknowledgments

I would like to thank Arnulf Quadt for giving me the opportunity to join the II. Physikalisches Institut. Without his supervision and support this thesis would not come into being. I am grateful to Thomas Peiffer and Elizaveta Shabalina, for the fruitful physics discussions and help throughout the PhD. The effort in reading and commenting this thesis despite your packed agenda has been amazing.
I would like to thank the whole analysis team for the innumerable chats about the analysis and problems. A special mention to Giancarlo and Michele, who introduced me to the $t\bar{t}$TRF world.
A sincere thank to all the institute colleagues. Perhaps "colleagues" is not the appropriate term to use here. "Friends" fits definitely better. The long chats, the many beers, the big laughs will always be with me.

A Silvia. Grazie per avermi sopportato e supportato in questi tre anni di treni persi, di EasyJet in ritardo e di (troppo poche) chiamate su Skype. È stato e sarà difficile, ma ce la faremo.
Ai miei amici. Nonostante sia raramente presente, voi siete sempre con me, ovunque io sia e qualsiasi cosa io faccia. Grazie.
Alla mia famiglia. Mi avete tenuto per mano, ci siamo guardati negli occhi e parlati al cuore. Non ho parole per dirvi grazie, casa è sempre con me.
A nonno. "Una cosa per volta e si fa tutto" mi dicevi sempre, avevi ragione. Questo lo abbiamo fatto insieme, e insieme andremo, da quí in avanti.

Contents

CHAPTER 1

Introduction

Mankind has always been fascinated by the study of the infinitesimally small structures, trying to explain the harmony of the surrounding world. The first answers came from religious belief: in the 6th century BC, the Indian *Jainism* religion formulated the first definition of indivisible and fundamental constituents that aggregate together into the non-living matter. One century later, Democritus independently developed the concept of *atoms* as indivisible components of the matter. Later in the 18th century, the philosopher Leibnitz in the manuscript *Monadology* described the Universe as made of infinite number of spiritual constituents (monads) with different degrees of consciousness.

The development of the scientific method and the enormous technical and theoretical achievements of the 17-18th centuries paved the way for the first experimental evidences of atomism. Firstly, Dalton, with his experiments on gases, proposed a model where all gases are composed of smaller similar components. Later, Thompson discovered the *electron* (1879) and Millikan measured its charge (1909). Meanwhile the knowledge of electromagnetism benefited Maxwell's exceptional works. The first half of the 20th century gave a huge boost to the formulation of the atomic structure. Rutherford's experiment inspired Bohr in modelling the atomic structure as composed of a positive heavy nucleus and electrons rotating in discrete orbits around it. Quantum mechanics and its relativistic formulation, quantum field theory (QFT), introduced the photons, particle spins and statistics, anti-particles and neutrinos.

However many questions were still unsolved in the 50's. The structure of the nuclei was not clear. The nature of the neutron decay was only described by Fermi's effective theory, including the elusive neutrino. The experimental setups were also limited: most of experiments based on chamber detecting particles from external sources (cosmic radi-

ation or nuclear reactors). Despite the simple setup, many new particles were discovered using this type of detector, such as *muons* (1936), *neutrinos* (1956) and *light mesons* (pions and kaons).

A more controlled production of particle was needed to investigate further in the *particle zoo*. Therefore the collider era started. Immediately, many unseen particles were discovered. Their classification in the *eightfold way* (1957) suggested those to be composite states of more fundamental particles, the partons, interacting with a new interaction, the strong force. Deep inelastic scattering experiments (1968) finally proved the existence of the partons, consisting of *quarks* and *gluons*. Meanwhile, Fermi's effective theory developed into the electroweak theory (1967), later confirmed by the discovery of the weak interacting bosons W and Z (1983). Eventually the so-called third generation of particles was completed with the discoveries of the *bottom* (1977) and *top* (1994) quarks, the *tau* lepton (1967) and corresponding neutrino (2000). The discovery of the *Higgs boson* in 2012 completes the set of the fundamental particles to be observed in the Standard Model, the current theory that describes interactions of the fundamental particles.

A brief summary of the Standard Model (SM) is given in Section 1.1. The physics of the top quark is described in Section 1.2. These two sections introduce the main topic of this PhD thesis: the measurement of the production of four top quarks ($t\bar{t}t\bar{t}$). The description of $t\bar{t}t\bar{t}$ production and its role in the modern particle physics is given in Section 1.3.

1.1. The Standard Model of fundamental interactions

The Standard Model (SM) is the current theoretical model of the interactions between the fundamental particles at the smallest scales of distance (or highest energy scales). It includes the *electromagnetic* force, acting on electrically charged particles, the *weak* interaction, involving particles with non-zero weak isospin and the *strong* interaction, which acts on particles with colour-charge. These are implemented in a *quantum field theory* (QFT) that describes the dynamics of the interacting particles.

The particles of the SM are summarised in Figure 1.1, separated in different coloured blocks depending on their role and their interactions. The violet and green blocks are the fermions of the SM: particles with half-integer spin whose quantum numbers are not predicted by the theory but experimentally measured and provided as input. The SM then predicts their interactions from the given quantum numbers. Green blocks are *leptons*, that do not interact via the strong force, and are arranged in three doublets (also called *generations* for historical reasons) of weak isospin. The up-isospin partners are the *neutrinos* (ν), massless and without electric charge. The down-isospin partners are the electron (e), muon (μ) and tau (τ): negatively charged and massive, with an increasing mass with the generation. The same structure is found in the violet blocks that contain the *quarks*. These are massive and electrically and weakly charged. In

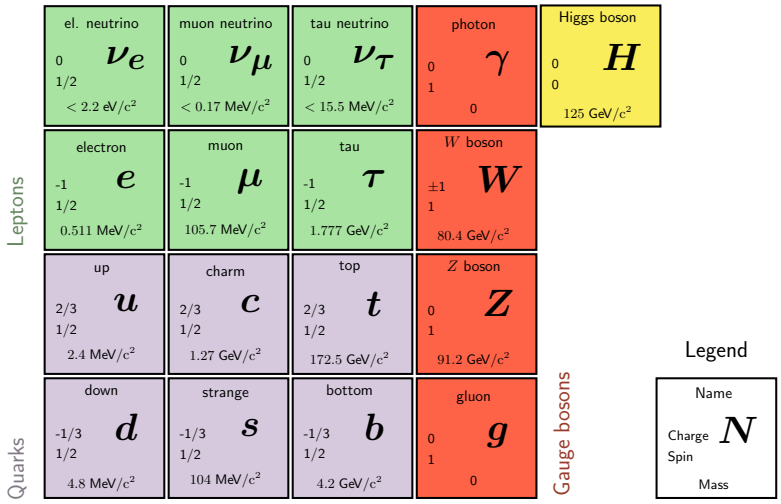


Figure 1.1.: Summary scheme of the fundamental particles of the Standard Model. Leptons are shown in green and quarks are shown in violet. Both are separated in three doublets of up-type and down-type quark or leptons. Gauge bosons are shown in red and the Higgs boson in yellow.

contrast to leptons, these particles interact via the strong force, and therefore they are coloured particles, featured by three possible colour states. To complete the picture an anti-particle (with opposite quantum numbers) has to be considered for each of the mentioned particles.

The remaining particles are the *gauge bosons*, with integer spin, each associated to an interaction of the SM. They are indeed called interaction *carrier* or *mediators*, since an interaction could be interpreted as a (real or virtual) exchange of a mediator. Starting from the top of the red column in Figure 1.1, there is the *photon* (γ): it is massless and it is the mediator for the electromagnetic interaction. It does not interact via the other forces. Gluons (g, at the bottom of the column in Figure 1.1) are also massless and they mediate the strong force. They are electrically and weakly not charged, but they have colour charge (eight colour variants). Therefore the strong interaction changes the colour of the involved particles and the name *quantum chromodynamics* (QCD) of the strong-interaction dynamics comes from this. The Z^0 and the $W^\pm$ bosons are the massive mediators of the weak force. They are not coloured and only the $W^\pm$ bosons are electrically charged. The exchange of these mediators permits the transitions of particles inside a weak isospin doublet. Quarks are a special case since the eigenstates for the weak interaction do not correspond to mass eigenstates (shown in Figure 1.1) and they are connected via the non-diagonal CKM matrix. Therefore the doublets shown in Figure 1.1 are not exact, and quarks can have off-doublet transitions via the weak interaction. All of the introduced particles are vector bosons; the only boson that is

instead a scalar is the *Higgs boson*, interacting with massive particles.

Here, a brief description of the theoretical construction of the SM is given. As mentioned above, the SM is a *quantum field theory* (QFT), that studies the dynamics of fields in space-time, $\varphi(x^\mu)$. The fields are interpreted as operators on the Hilbert space of the physical particle states: they correspond to the creation operator of a particle state in the x^μ coordinate. The generated states must have the correct quantum numbers, therefore different types of fields are needed to create different particles. Chiral fields are used to describe fermions, vector fields for the gauge bosons and scalar fields for particles with spin zero, such as the Higgs field. Once a field is associated to each particle of the SM, the *free theory* of the field can be written. This theory describes the evolution of the field without any sort of interaction and it is obtained by writing the most general Lagrangian fulfilling a few requirements:

- Lorentz (Poincaré) invariance of the terms of the Lagrangian.
- Second order in the fields and their derivative (φ,$\partial\varphi$).
- No quadratic terms in the fields (first consider a massless theory).

This Lagrangian describes the free evolution of the fields, and therefore of the particle states.

The interactions are included in the theory via the implementation of *gauge* symmetries. Given an invariance of the Lagrangian under a global transformation of the fields, the corresponding gauge transformation is obtained by making it depend on the coordinates x^μ, becoming a local transformation. The *covariant derivative* is introduced to make the theory invariant under the gauge transformation. This cures the distortion introduced by the implemented symmetry, providing a new formulation of the derivatives. Mathematically, it consists of replacing:

$$\partial_\mu\varphi(x^\nu) \to D_\mu\varphi(x^\nu) = \left[\ \partial_\mu + igT^aG^a_\mu(x^\nu)\ \right]\varphi(x^\nu) \tag{1.1}$$

where the additional term consists of the coupling constant g, measuring the strength, the generators of the gauge transformation group T^a, and newly needed fields G^a_μ. These fields, called *gauge vector bosons*, must follow fixed transformation rules under the gauge transformation to restore the invariance of the whole Lagrangian. Equation 1.1 shows that the multiplicity of the mediators is equal to the dimension of the gauge transformation group. *Quantum electrodynamics* (QED) is obtained by implementing the symmetry under local change the phase of the fields: the $U(1)_Q$ charge symmetry. The transformation group has dimension one and therefore one boson is required, the photon. The weak interaction is generated by requiring the invariance of the Lagrangian under rotation of the fields in the weak isospin doublets. The corresponding $SU(2)_L$ group needs three bosons: the $W^\pm$ and Z^0 bosons. Finally the strong interaction corresponds to the implementation of the symmetry under rotation of the colour states of the fields. The dimension of the corresponding group $SU(3)_c$ is eight, and as many different gluons indeed exist. By replacing these new derivative definitions into the kinematics terms of

the fields, the interactions between different particle fields are generated.

This lovely and clean theory, which works well, has however some caveats, among which all particles are assumed to be massless, both fermions and bosons. The mass terms for fermions, described by chiral fields, are explicitly non-invariant under $SU(2)_L$ transformation. The mass terms for $W^{\pm}$ and Z^0 bosons are instead not generated by the implementation of the corresponding gauge symmetry. The solution is given by the *gauge symmetry breaking* and the *Higgs mechanism*. The symmetry breaking occurs when the Lagrangian is invariant under a given transformation, but the physical states are not. In the SM, the massless Lagrangian, that is invariant under $SU(3)_c \times SU(2)_L \times U(1)_Y$ transformation group, breaks into another Lagrangian of the physical fields that is invariant under the sub-group $SU(3)_c \times U(1)_Q$. Note that the transformation group of the Lagrangian of the physical fields does not contain $SU(2)_L$ that is broken by the mass terms, and contains the colour $SU(3)_c$ and charge $U(1)_Q$ transformations, that are associated to real conserved quantities (colour and charge). Moreover the original $U(1)_Y$ symmetry is different from the final $U(1)_Q$ and involves the *hypercharge*, a combination of electric charge (Q) and weak isospin (I_3): $Q = I_3 + 1/2Y$.

The symmetry breaks thanks to the introduction of the Higgs field, consisting of two complex fields, arranged in a weak doublet $H(x^\mu) = (h_1, h_2)$, featured by a potential $V(H)$:

$$\mathcal{L}_{\text{Higgs}} = D^\mu \bar{H} D_\mu H - \overbrace{\left[-m_h^2 \bar{H}H + \frac{m_h^2}{2v^2}(\bar{H}H)^2 + \frac{m_h^2 v^2}{4}\right]}^{V(H)} \tag{1.2}$$

where m_h stands for the Higgs boson mass. The minimum of $V(H)$, the ground state, is not reached for $H = 0$, but occurs in the circular orbit in the (h_1, h_2) plane with radius v. A meaningful description of the field dynamics is obtained by expanding the potential around the minimum, by replacing

$$H(x^\mu) = (h_1(x^\mu), h_2(x^\mu)) \rightarrow (v, h(x^\mu)) \tag{1.3}$$

that is an arbitrary point on the orbit, breaking the original symmetry. By replacing Equation 1.3 in Equation 1.2, the massless original gauge fields $W^{\pm}$, W^0 and B_μ associated to the primitive $SU(2)_L \times U(1)_Y$ symmetry are combined into three massive $W^{\pm}$ and Z^0 bosons and the massless photon. Finally the mass terms for quarks and leptons are obtained by introducing Yukawa couplings with the Higgs fields (and their conjugate). The remnant from this mechanism is the scalar and neutral boson $h(x^\mu)$,

the Higgs boson. The resulting complete Standard Model Lagrangian is:

$$\begin{aligned}
\mathcal{L} = & -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{8}tr\left(\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu}\right) - \frac{1}{2}tr\left(\mathbf{G}_{\mu\nu}\mathbf{G}^{\mu\nu}\right) && \text{(U(1), SU(2) and SU(3) gauge terms)} \\
& + i\bar{l}_L\gamma^\mu D_\mu l_L + i\bar{l}_R\gamma^\mu D_\mu l_R + (h.r.) && \text{(Dynamical terms for leptons)} \\
& - \frac{\sqrt{2}}{v}\left[\bar{l}_L h M^l l_R + \bar{l}_R\bar{h}\bar{M}^l l_L\right] && \text{(Mass terms for down-type leptons)} \\
& + i\bar{q}_L\gamma^\mu D_\mu q_L + i\bar{q}_R\gamma^\mu D_\mu l_q + (h.r.) && \text{(Dynamical terms for quarks)} \\
& - \frac{\sqrt{2}}{v}\left[\bar{q}_L h M^d d_R + \bar{d}_R\bar{h}\bar{M}^d q_L\right] && \text{(Mass terms for down-type quarks)} \\
& - \frac{\sqrt{2}}{v}\left[\bar{q}_L \widetilde{h} M^u u_R + \bar{u}_R\widetilde{\bar{h}}\bar{M}^u q_L\right] && \text{(Mass terms for up-type quarks)} \\
& + D_\mu h D^\mu h - m_h^2\left[\bar{h}h - v^2/2\right]^2/2v^2 && \text{(Higgs dynamical and mass terms)}
\end{aligned} \tag{1.4}$$

The Lagrangian is then used to compute amplitudes between initial $\langle i|$ and final $|f\rangle$ states:

$$\mathcal{A}_{if} = \langle i|\exp\left(i\int \mathcal{L}(x^\mu)d^4x\right)|f\rangle \tag{1.5}$$

The Feynman rules allow a graphical interpretation of the non-vanishing terms of the amplitude. Thanks to the small values of the coupling constants[1], the amplitude is computed with a perturbative approach. A gauge invariant amplitude, thus physically meaningful, is obtained at each order of expansion in the number of loops (close paths) in the Feynman diagrams. This corresponds to the notations of Leading Order (LO), Next-to-Leading-Order (NLO) etc. Finally, once amplitudes are computed, the corresponding cross-section is extracted via Fermi's Golden Rule.

Amplitudes, and then cross-sections, are computed starting from fundamental particles, and not from composite states like protons. Therefore a bit of handling is needed to compute cross-sections in hadrons colliders. Protons are bound states of coloured particles following *parton density functions* (PDFs) that describe the probability of a given component of having a fraction x of the overall proton momentum. PDFs are experimentally measured [1–3] and used to compute any cross-section from pp collisions:

$$\sigma_{pp\to X}(\sqrt{s},\theta_X) = \sum_{i,j=q,\bar{q},g}\int dx_i\ dx_j f_i(x_i,\mu_F^2)f_j(x_j,\mu_F^2)\sigma_{ij\to X}(x_i,x_j,\alpha_s(\mu_R^2),\mu_R^2,\theta_X) \tag{1.6}$$

where $\sqrt{s}$ is the centre-of-mass energy and θ_X are the parameters of the final state X (e.g. masses). Every possible initial state ij for the transition is weighted by the corresponding PDF $f_{i/j}$ depending on the respective momentum fractions ($x_{i/j}$) and the *factorisation scale* μ_F. This scale gives the value that separates short-range interactions,

[1] In QCD, this condition is satisfied only for large transferred momentum interactions.

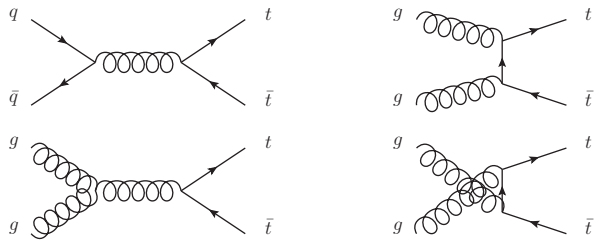


Figure 1.2.: Leading order (LO) QCD diagrams contributing to $t\bar{t}$ production at hadron colliders.

described with perturbative theory, and long-range interactions, modelled using PDFs (factorisation theorem [4]). The cross-section for a single configuration $\sigma_{ij\to X}$ depends also on the running QCD coupling constant (α_s) and on its subtraction point, the so-called *renormalisation scale* μ_R.

1.2. The top quark

The top quark (t) is the heaviest particle of the SM and the up-type quark of the third generation. Its very large mass (around 173 GeV) made its discovery very challenging, but places it in a very interesting position inside the SM.

The existence of the top quark was suggested by the discoveries in the 70's of the $b\bar{b}$ (Υ) [5] bound state and the τ lepton [6], opening the way to the third generation of particles. Later measurements on the b-quark charge and weak isospin [7, 8] indicated the bottom quark to be the down-type quark of a weak doublet, and therefore the top quark was predicted to be its partner. The discovery occurred in 1995 at the TeVatron $p\bar{p}$ collider, by the CDF [9] and DØ [10] experiments, announcing an excess over expected background of a signal compatible with the production of top quark pairs $t\bar{t}$.

The top quark in the SM is a fermion with spin 1/2 and electric charge 2/3. It is the up-type quark of a weak doublet with the b-quark and it is a strong triplet [11]. The decay mode $t \to W^+b$ is dominant over the decays in the other down-type quarks because the V_{tb} element of the CKM matrix is approximately unitary. Its decay width is around 1.5 GeV, giving a lifetime $\tau_t \approx 5 \times 10^{-25}$ s. By comparing the lifetime with the typical hadronisation time[2] $1/\Lambda_{\mathrm{QCD}} \approx 1/200$ MeV $\approx 3 \times 10^{-24}$ s, the top quark is expected to decay before hadronisation, allowing the measurement of its features via the decay products. An example is the top quark mass, typically measured as the invariant mass of its decay products[3], and equal to 172.9 ± 0.4 GeV [11]. The large top mass

[2]This estimation given by the inverse of the energy scale at which the strong interaction is no longer in perturbative regime.

[3]It is the definition of *pole mass*, but other mass definitions are possible. This ambiguity is due to

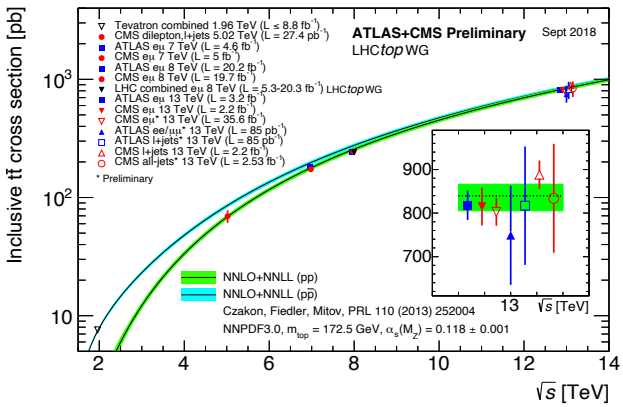


Figure 1.3.: Standard Model predictions compared with TeVatron and LHC measurements [12].

converts into a special role into the SM, since it is strongly related to other parameters of the theory, such as W and Higgs masses.

The main production mechanism for top quarks in hadron collisions is via $t\bar{t}$ pair production, shown in Figure 1.2. At the LHC, the dominant contribution to the cross-section comes from gg initiated production, differently from TeVatron $p\bar{p}$ collisions, where the $q\bar{q}$ initiated diagrams had a dominant contribution. The $t\bar{t}$ production cross-section has been measured extensively in different environments and at different $\sqrt{s}$ (Figure 1.3), resulting in an expected cross-section in the current LHC setup of around 800 fb.

A rarer, but still significant source of production of top quarks is *single top quark* production, whose diagrams are shown in Figure 1.4. The lower cross-section (around 350 fb overall) is due the mediation of a weak boson for the production a single top quark. Three distinct mechanisms for single top quark production are identified, and shown in Figure 1.4. The t-channel consists of an exchange of a space-like W boson and contributes with 70% of the total cross-section. The s-channel, with the propagation of a time-like W boson, has the lowest cross-section (5%). The Wt production requires a bottom quark in the initial state and contributes to the remaining 25% of the cross-section.

The top quark decay modes are classified in terms of the number of leptons in the final state in hadronic (around 2/3 of the cases) and semileptonic (around 1/3 of the cases) decays. This means that the $t\bar{t}$ is detected in *all-hadronic* final state in around 45% of the cases, in *single lepton* final state in 45% and dileptonic final state for the remaining 10% of decays.

Due to its large cross-section at LHC, top quark pair production is one of the dominat-

quarks confinement.

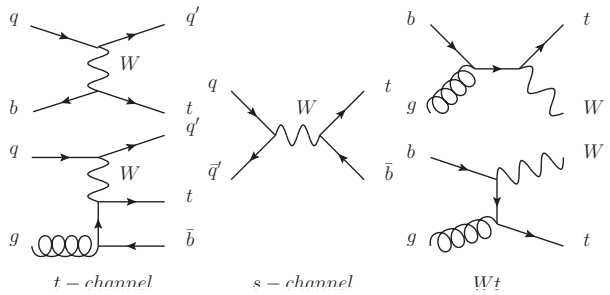


Figure 1.4.: Leading order (LO) QCD diagrams contributing to single top production at hadron colliders.

ing processes observed in the detectors, giving the LHC the name of a *top quark factory*. Therefore, many studies in the top quark sector have been performed in the past years. Cross-sections for the different production processes have been measured, also covering particular phase-spaces (boosted top regimes [13]) and differential in many kinematic variables [14–16]. The production in association with vector bosons has also been studied, from $t\bar{t}$+$W/Z/\gamma$ [17, 18] to the latest observation of the $t\bar{t}H$ production[4] [19, 20]. Those have provided sensitivity to the couplings of the top quark, confirming that the observed top quark behaves as the Standard Model predicts.

1.3. The four top quarks production process

The four top quark production ($t\bar{t}t\bar{t}$) can form some of the most spectacular signatures in the detector, featured by a large number of particles in the final state to be detected and reconstructed. Examples of diagrams contributing to the SM amplitude are shown in Figure 1.5: at the LHC, the main contribution comes from gg initiated diagrams (94%) and $q\bar{q}$ production contributes only for the remaining 6%. The cross-section is predicted by the Standard Model to be $\sigma_{t\bar{t}t\bar{t}} = 11.97^{+18\%}_{-21\%}$ fb [21] at NLO in QCD and electroweak coupling (EW)[5] from pp collision at $\sqrt{s} = 13$ TeV. The inclusion of the EW correction is motivated by the significant contribution due to the large tH Yukawa coupling ($\propto m_t^2$). The cross-section is far lower than for the other processes involving top quarks ($t\bar{t}$, $t\bar{t}$+X). However $t\bar{t}t\bar{t}$ can produce many different signatures in the detector because of the many top quarks in the final states, as shown in Figure 1.6.

As mentioned in Section 1.2, the top quark has an interesting role in the Standard Model because of its large coupling to the Higgs boson, due to the large top quark mass. Many Beyond-Standard-Model (BSM) theories introduce couplings that are proportional

[4] These processes of $t\bar{t}$ production associated with a other fundamental particles are noted as $t\bar{t}$+X in this thesis.

[5] Quoted uncertainties take into account scale and PDF uncertainties.

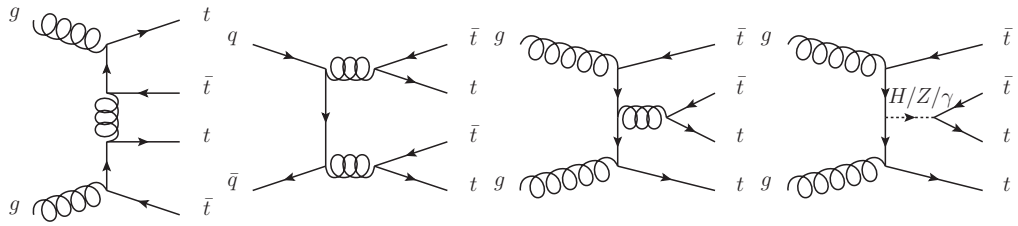


Figure 1.5.: Examples of Leading Order (LO) QCD diagrams contributing to $t\bar{t}t\bar{t}$ production at hadron colliders.

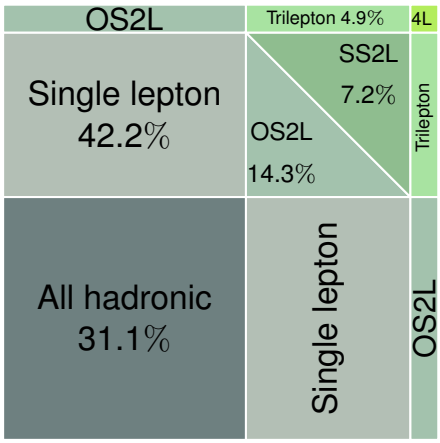


Figure 1.6.: The final states of $t\bar{t}t\bar{t}$ events are categorised in channels in terms of lepton multiplicity and their electric charges. Here, the branching ratios for the different channels are shown. The single lepton channel requires one lepton in the final state. The opposite-sign dilepton channel (OS2L) asks for two leptons with opposite-sign charges. Similarly, the same-sign dilepton channel (SS2L) is identified. The trilepton channel is obtained selecting three leptons in the finale state. The four-lepton channel (4L) requires all the top quarks to decay leptonically and corresponds to 0.4% of the events. Each channel is identified by a colour, and the corresponding value of branching ratio refers to the entire area associated to the channel. Here, a lepton is defined as muon and electron. Contributions from tau leptons are included in the different channels via its decay modes.

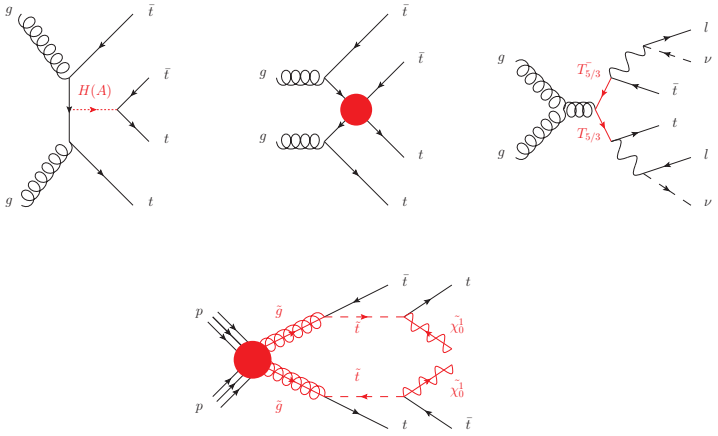


Figure 1.7.: Example of BSM diagrams that contribute to $t\bar{t}t\bar{t}$: presence of heavy resonances into $t\bar{t}$ (top left), contribution of contact interactions (top center), a diagram from vector-like quarks (VLQ) $T_{5/3}$ (top right) and SUSY gluino mediation (bottom). Here the VLQ contribution does not predict a real $t\bar{t}t\bar{t}$ enhancement, but populates a similar phase-space.

to the mass. Therefore the top sector is definitely promising for seeking new physics signals. The $t\bar{t}t\bar{t}$ production, having a large number of top quarks in the final state, is even more sensitive to new physics contributions. Moreover, the final state with multiple leptons of $t\bar{t}t\bar{t}$ is hardly obtained from other SM processes, while it is predicted to be populated by many BSM scenarios.

Essentially, BSM signals may result in the $t\bar{t}t\bar{t}$ measurement from two different mechanisms, as shown in Figure 1.7. The first mechanism is via actual enhancement of the $t\bar{t}t\bar{t}$ cross-section. This is commonly obtained by introducing new heavy particles decaying into a top quark pair. Examples are from SUSY with gluinos or sgluons, expected to be produced in pairs, each decaying to $t\bar{t}$ [22]. Another way is via the introduction of Kaluza-Klein gluons from extra-dimension theories [23, 24]. Also Higgs doublet models, predicting heavy Higgs bosons, contribute to enhance the $t\bar{t}t\bar{t}$ cross-section over the SM prediction [25, 26]. In case the top quark is not a fundamental particle, but rather composed of more fundamental fields, a $t\bar{t}t\bar{t}$ contact interaction term has to be introduced, increasing the expected total cross-section [27]. Secondly, many theories predict new contributions lying in the same phase-space as the $t\bar{t}t\bar{t}$ signal, that would result in an excess of the measured events. The assumption of wider symmetry groups breaking into the SM, leads to top-philic particles, providing excess of signal in regions with many top and bottom quarks [28]. The introduction of vector-like quarks (VLQ) decaying into tW provides a signal into $t\bar{t}WW$ phase space, overlapping with the interesting regions for the $t\bar{t}t\bar{t}$ search [29]. The BSM scenarios predict enhancements with a difference even of

orders of magnitudes with respect to the SM $t\bar{t}t\bar{t}$ cross-section. Therefore setting more restrictive limits on the $t\bar{t}t\bar{t}$ cross-section would constrain significantly the parameters of the theories and the terms of effective field theories [30].

The search of $t\bar{t}t\bar{t}$ has been conducted by ATLAS and CMS in the past years, using a partial dataset of about 36 fb^{-1} to look for signal from new physics. The ATLAS Collaboration has searched for the $t\bar{t}t\bar{t}$ signal in four different channels separately and combined [31, 32] to get the best measurement with the partial dataset, providing a measured (expected) limit at 95% confidence level (CL) of 48.7 fb (19.3 fb). The CMS Collaboration used same-sign dileptonic and multileptonic final states, obtaining a limit at 95% CL of 41.7 fb (20.8 fb) [33] and also the single-leptonic channel [34] with looser limits imposed on the cross-section. The current full Run 2 dataset aims for measuring the evidence of the SM signal, where evidence is defined as a deviation of three standard-deviations from the background-only hypothesis. A preliminary result with a dataset of 137 fb^{-1} has been released by the CMS Collaboration, using the same-sign dileptonic and multileptonic channels and measuring a cross-section of $12.6^{+5.8}_{-5.2}$ fb [35], with a significance of the $t\bar{t}t\bar{t}$ signal over the background-only expectation of 2.6 (2.7) standard deviations. The analysis presented in this thesis uses the dataset collected over the Run 2 by the ATLAS detector to achieve the first evidence of the SM $t\bar{t}t\bar{t}$ production by combining four different detection channels. Preliminary results of this analysis, using the channels featured by events with two same-sign leptons or three leptons, have been released [36] and are presented in this thesis. However, the main topic of this work is the analysis using the channels with a single lepton or two opposite-sign leptons and the combined measurement using all four detection channels.

CHAPTER 2

LHC and ATLAS

The Large Hadron Collider (LHC) [37] at the European Organization for Nuclear Research (CERN) in Switzerland, is the largest proton-proton collider in the world and it is the last out of a long history of hadronic accelerators. The first pp collider, the Intersecting Storage Rings (ISR) at CERN [38], opened the way for progressively more complex machines such as Sp$\overline{\text{p}}$S (CERN, 1981) [39] and TeVatron (Fermilab, 1987) [40]. Complementary physics was studied at different machines, like LEP (CERN, 1987) [41] and SLC (SLAC, 1991) [42], significantly boosting particle physics knowledge until the first decade of 2000s. After their decommission, the need for investigating physics at higher energy scales led to the construction of the LHC in the tunnel previously used by LEP accelerator. Since 2007, LHC is successfully running, providing pp collisions at unprecedented centre-of-mass energy to four detectors. An introduction to LHC accelerators complex is given in Section 2.1 with a focus on ATLAS detector in Section 2.2.

2.1. The Large Hadron Collider

The LHC is the final stage in the CERN's accelerator complex: it is a 27 km long acceleration ring, that boosts two proton beams up to the energy of 6.5 TeV intersecting in four crossing points, where the detectors are located.

Proton beams propagate in two vacuum pipes, accelerated and separated in bunches through 16 radio frequency cavities oscillating at 400 MHz. Their trajectory is bent by 1232 dipole magnets, in a quasi-circular path, while the beam shape is adjusted by 392 quadrupole magnets to prevent significant beam loss in the propagation and to focus them for the collision. Both cavities and magnets are kept at extremely low temperature

(−271.3 °C) in order to operate in the superconducting phase.

The CERN accelerator complex consists of a chain of accelerators, serving LHC and many beam facilities used by other experiments based at CERN, as shown in Figure 2.1. Protons are extracted from Helium gas and first accelerated up to 50 MeV in the LINAC2 linear accelerator. Before being injected into the LHC, the protons go through three smaller ring accelerators with increasing radii: first the Proton Synchrotron Booster (PSB, where they are accelerated up to 1.4 GeV), then the Proton Synchrotron (PS, up to 25 GeV) and finally the Super Proton Synchrotron (SPS, up to 450 GeV). The protons are then injected into the LHC, that takes approximately four minutes to get filled and needs other 20 minutes to ramp up the beam energy to the desired 6.5 TeV. During the acceleration stage, the beam is also arranged in 2808 bunches of 11.5×10^{11} protons spaced by an interval of 25 ns, resulting in a collision rate of 40 MHz.

LHC operations started officially on September 24th 2008. However, a magnet quenching incident delayed the start of the physics program (Run 1) to November 2009 and the beginning of stable data-taking at reduced beam energy of 3.5 TeV to March 2010. Run 1 ended in 2013 and the data collected led to the discovery of the Higgs boson (July 2012) [43, 44]. A long shut-down (LS1) came afterwards, with upgrade of the detectors and the accelerator to meet the design performances. The Run 2 data-taking period lasted from April 5th 2015 to December 2nd 2018, collecting an unprecedented amount of data from pp collision with a beam energy of 6.5 TeV. Currently, the second long shut down phase (LS2) is ongoing, aiming for detectors and accelerators upgrades for the Run 3, starting at the end of 2020 until the end of 2023. This Run is meant to be with stable beam energy of 7 TeV at an instantaneous luminosity of 20×10^{33} cm^{-2}s^{-1}, double the design value. The plan is then to move into the so-called High-Luminosity LHC (HL-LHC, starting in 2026) for which a significant detector and accelerator renewal is needed.

Proton collisions are detected by four experiments located along the tunnel: ALICE [45], ATLAS [46], CMS [47] and LHCb [48]. The ALICE detector investigates high-density environments of strongly interacting particles, the *quark-gluon plasma*. LHCb focuses instead on CP-violation measurements in the b-quark sector. ATLAS and CMS are two general purpose experiments, testing the fundamental features of the Standard Model, using different detector setups.

2.2. The ATLAS detector

The ATLAS (A Toroidal LHC ApparatuS) detector is a general purpose experiment at the LHC, located 100 m underground at P1 at CERN. With an astonishing dimension of 43 m × 25 m × 25 m, it is the largest detector at the LHC and one of the largest machines on Earth. Challenging performance requirements are needed to meet precise measurements over a wide range of physics analyses.

One of the first challenge for an experiment at the LHC is to face the high luminosity.

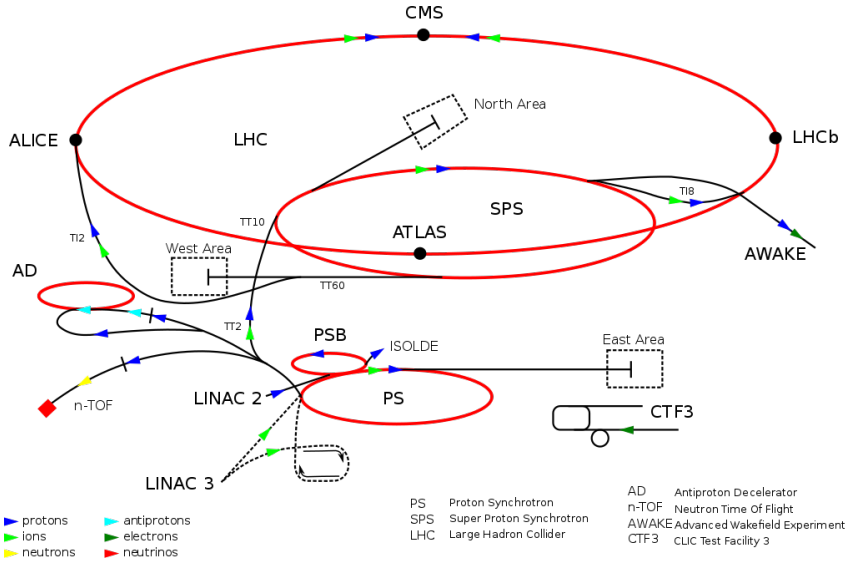


Figure 2.1.: Scheme of the CERN accelerators complex for the proton injection into LHC © CERN.

An ultra-fast read-out of data off the detector is needed to manage the high event rate in the detector. However, most of the events come from low momentum transfer pp elastic collisions, used for luminosity measurements (see Section 2.2.5), most not useful for physics searches. A very efficient trigger system is required to prevent most of those events to propagate through the data acquisition chain, reducing the event rate to a decent level to be recorded.

As in every hadronic collider, lots of particles are produced for each collision, and they need to be detected and reconstructed. High granularity detectors are used to manage the high particle flux and reduce the effect of overlapping events. An almost full solid angle coverage is important to ensure high event acceptance. The tracking system must ensure high efficiency, high momentum resolution and good vertexing performance, crucial for secondary vertex recognition. A precise electromagnetic calorimetry system facilitates a good electron and photon measurements. A full-coverage of the hadronic calorimeter is important for a good jet and missing transverse energy measurements. Finally, a muon detector is needed to precisely measure muons, important for triggering and for their clean signature.

The detector has been designed to meet the above requirements and first presented on December 15th, 1994, but formally approved almost three years later. Construction in the P1 cavern started in 1998 and the detector assembly finished on October 4th, 2008. Since the beginning of Run 1, the ATLAS detector has been operational and is expected to keep the current setup until the end of 2023.

To better understand the terminology used in this thesis, a description of the coordinate system used in ATLAS is given in Section 2.2.1. Then an overview over all ATLAS sub-systems is given in Sections 2.2.2 (tracking system), 2.2.3 (calorimetry system), 2.2.4 (muon system), 2.2.5 (forward detectors) and 2.2.6 (trigger and readout).

2.2.1. Coordinate system

In ATLAS the origin of the right-handed coordinate system is located at the nominal interaction point, with the z-axis lying on the beam line. The position along z distinguishes the detector into two sides: A-side ($z > 0$) and C-side ($z < 0$). The $x-$ and $y-$axes define the transverse plane to the beam direction, with the x-axis pointing towards the centre of LHC. To better suite the detector shape, a cylindrical coordinate system is also used based on the radial distance from the z-axis (r) and azimuthal angle (ϕ).

Since the rest frame of the parton-parton collisions might move along the z-axis, variables that are invariant under Lorentz boosts on z-axis are needed. Kinematic variables are usually projected onto the transverse plane, for instance a momentum p becomes transverse-momentum p_{T}. The longitudinal information is recovered using the *pseudo-rapidity* (η) defined from the polar angle θ from the z-axis as:

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \tag{2.1}$$

Therefore the angular distance of two objects is given by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

2.2.2. Tracking system

Tracking in ATLAS is provided by the Inner Detector (ID) [49] that is the closest system to the interaction point. It is composed of three different sub-detectors: the Pixel Detector, the Silicon microstrip detector (SCT) and Transition Radiation Tracker (TRT) as shown in Figure 2.2, covering up to $|\eta| < 2.5$. The whole detector is immersed into a 2 T magnetic field from the central solenoid.

The Pixel detector is located around the beam pipe and it is the detector with the highest granularity. It consists of four silicon pixel layers in the central region and two endcaps, with three disks on each side. The innermost layer has a reduced pixel dimension and an improved radiation hardness to face the extreme particle flux from the LHC. The typical pixel dimensions for this layer are 50×250 $\mu\mathrm{m}^2$ to be compared with the 50×400 $\mu\mathrm{m}^2$ of all the other components, resulting in a better z-position resolution (75 μm compared to 115 μm of the other layers). Resolution on r-ϕ position is constant over the whole detector (10 μm), given the same pixel pitches for all the layers. However, some effects such as radiation damage may lead to a non-homogeneous performance over the detector. Studies on the performance of this detector are also included in this thesis in Section 3.

The SCT has a similar structure, with eight layers of strips (corresponding to four space point measurements) and two endcaps. The strips in the barrel are half along the $z-$axis and half tilted by a stereo angle of 40 mrad. Strips in the endcaps are half in the radial direction and half inclined of a stereo angle as in the barrel. The typical length is 6.4 mm with 80 µm pitch giving an r-ϕ position resolution of 17 µm. The resolution on the z position is limited by the stereo angle to 580 µm.

The outer detector is the TRT which uses charged particle emission when changing medium of the propagation. It consists of 4 mm-diameter gas-filled tubes placed parallel to the beam axis and radially in the endcaps. The gas mixture (70% Xe, 27% CO_2 and 3% O_2) and operational conditions are chosen to maximise the conversion of the emitted transition radiation to electrons over the minimum ionising particle energy deposition. This detector provides only r-ϕ information with intrinsic resolution of 130 µm. The only information on z-axis is the side of the detector, since straws inside the tubes are segmented in two, approximately at $z = 0$.

The combination of these detectors provides a very robust measurement of charged particles r-ϕ and z positions. The TRT compensates the lower granularity with the longer measured track path and the larger number of measurements. It also improves the recognition of electrons when combined with calorimeter information, thanks to the large transition radiation of electrons in the TRT tubes gas. The resulting resolution on the reconstructed tracks p_{T} (in GeV) is [49]:

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = 0.04\% \times p_{\mathrm{T}} \oplus 2\% \tag{2.2}$$

At larger p_{T} the curvature of the track gets smaller, therefore the resolution on measured p_{T} is larger. At lower p_{T} the irreducible effect of multiple scattering becomes instead dominant.

2.2.3. Calorimetry system

The calorimetry system is located around the Inner Detector and covers a larger angular region, up to $|\eta| < 4.9$ to ensure large coverage for missing transverse energy measurement. As shown in Figure 2.3, it is divided into two main sub-detectors, the electromagnetic calorimeter (LAr) [50] and the hadronic calorimeter [51].

The LAr is a lead-liquid Argon sampling calorimeter: the lead initiates photon and electron shower from Bremsstrahlung and the produced particles ionise electrons in liquid Argon (active material), providing an electric signal proportional to the incident particle energy. It is divided into a central barrel ($|\eta| < 1.375$, three layers) and two endcaps ($|\eta| < 3.2$, two wheels). This sub-detector consists of more than 24 radiation lengths, important to fully contain high energy photons and electrons for accurate energy measurement. The granularity depends on the distance from the interaction point and η: for example the segmentation of the closest layer in the barrel is $\Delta\eta \times \Delta\phi = 0.025/8 \times 0.1$.

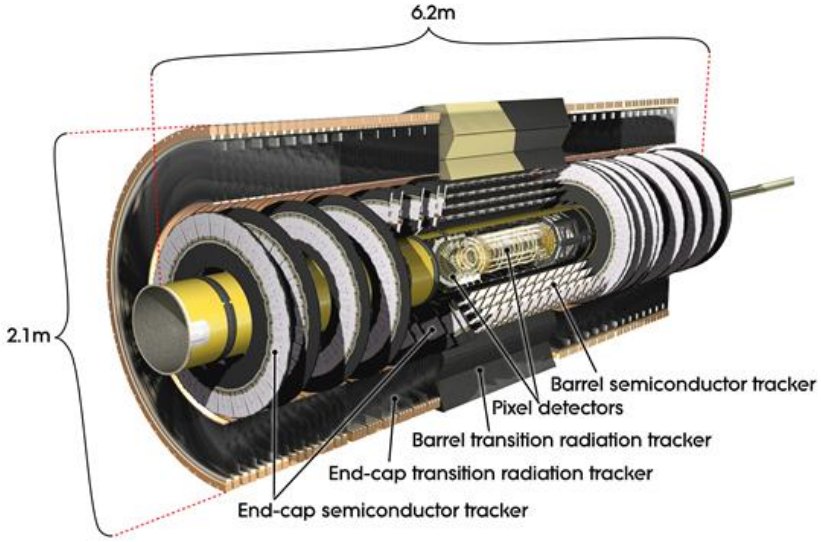


Figure 2.2.: Layout of the ATLAS Inner Detector ©CERN.

This setup provides an energy E resolution (in GeV) of:

$$\frac{\sigma_E^{\text{LAr}}}{E} = \frac{0.1}{\sqrt{E}} \oplus 0.01 \oplus \frac{0.3}{E} \tag{2.3}$$

At larger energy ($> 100\,\text{GeV}$) the calibration term (second term in Equation 2.3) becomes dominant, while for lower energy the sampling term is the main contributor to the total resolution (first term in Equation 2.3). At extremely low energy ($< 10\,\text{GeV}$) the constant resolution of 0.3 GeV given by noise level is the largest contribution (third term in Equation 2.3).

The Tile calorimeter is placed around the LAr and it is the central part of hadronic calorimeter system. It consists of a central barrel covering $|\eta| < 1.0$ with two appendices that reach $|\eta| < 2.5$ in coverage. It is an iron-scintillator tile sampling calorimeter with a depth corresponding to 9.7 radiation lengths. The working principle is similar to the LAr: the iron aims at initiating a hadronic shower from strong interacting particles that is measured in the scintillators, producing light that is transmitted via wavelength shifting fibres to off-detector photomultipliers. The segmentation changes over the detector, the with finer segmentation closer to the interaction point. Typical granularity ranges from $\Delta\eta{\times}\Delta\phi = 0.1{\times}0.1$ to $\Delta\eta{\times}\Delta\phi = 0.1{\times}0.2$. The achieved hadronic calorimeter energy E resolution is:

$$\frac{\sigma_E^{\text{TileCal}}}{E} = \frac{0.5}{\sqrt{E}} \oplus 0.03 \tag{2.4}$$

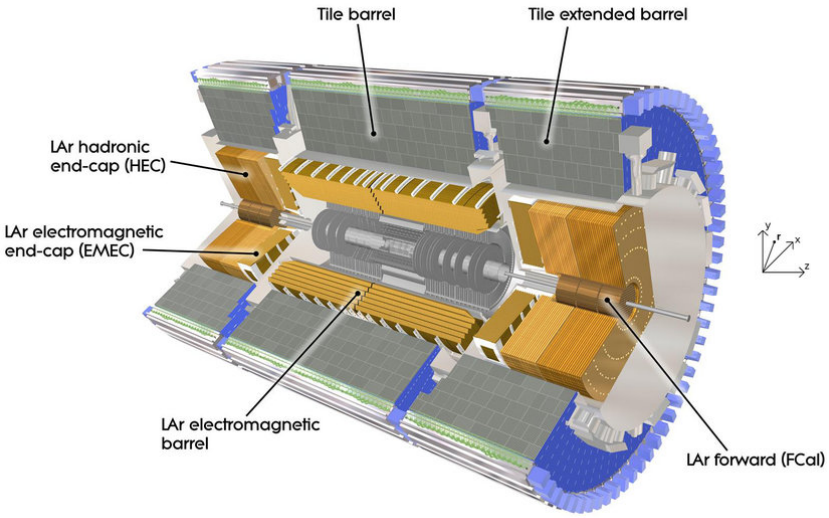


Figure 2.3.: Layout of the ATLAS Calorimetry system © CERN.

The large uncertainty from the first term is due to showers that are not fully-contained and to the response of the calorimeter to different components of the hadronic showers. Hadronic showers are indeed very complex objects to calibrate and measure, since they are made of many components with different signature in the detector and their composition significantly varies event by event.

The endcaps of the hadronic calorimeter are the Hadronic Endcaps (HEC) composed of two wheels with 32 identical wedge shaped modules, covering up to $|\eta| < 3.2$. The modules consist of copper plates separated by liquid Argon that is the active material. The HEC overlaps with the Forward Calorimeter (FCal), important to ensure maximum calorimetric coverage up to $|\eta| < 4.9$. The FCal is divided in three layers, the first is optimised for the detection of electromagnetic showers and uses copper as absorber, while the other two are optimised for hadronic shower measurements, using tungsten. All the layers use liquid Argon as active material.

2.2.4. Muon spectrometer

The muon spectrometer (MS) [52] is the outermost detector. It provides precise measurements of muon momenta and yields a fast response to muon signals for triggering. It is immersed in a toroidal magnetic field, provided by the central barrel toroid up to $|\eta| < 1.4$ and by two smaller endcap magnets in $1.6 < \eta < 2.7$ region. In the gap between these two regions, the superposition of the two magnetic fields is used for bending the particle trajectories. Because of the different magnets, the field intensity varies from 1.5

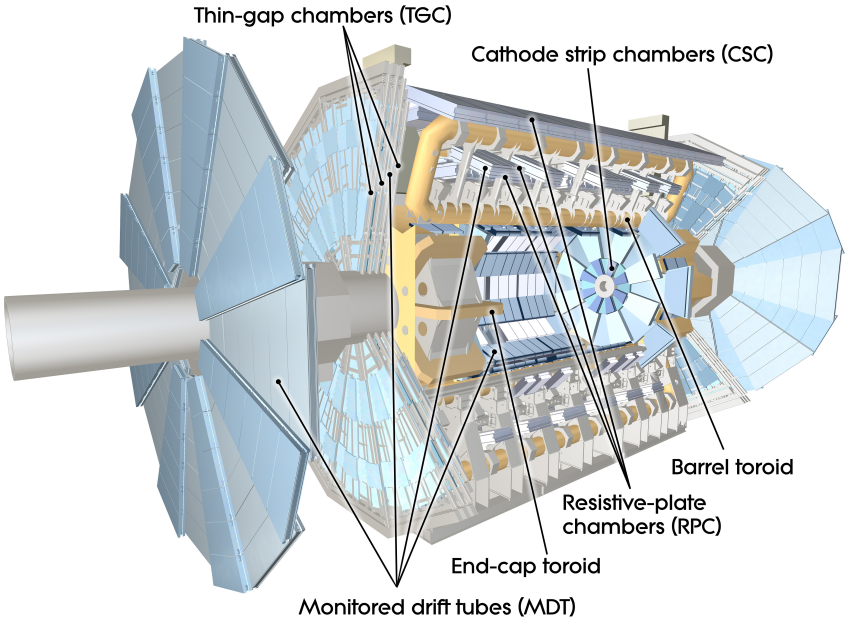


Figure 2.4.: Layout of the ATLAS Muon Detector © CERN.

to 5.5 T in the barrel, reaching 7.5 T in the endcaps.

The muon system is divided in different sub-systems, as shown in Figure 2.4. The Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs) are located in the central and peripheral regions respectively. The former consists of three to eight layers of drift tubes along the $z-$axis covering the region up to $|\eta| < 2.7$. The CSC is instead a multiwire proportional chamber with cathode segmentation into strips in orthogonal directions to measure both coordinates. The MDTs and CSCs are dedicated to precision measurements of particle momenta, with a resolution of $2-3\%$ for most of the momentum spectrum. The Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs) aim for giving a fast response for trigger purposes, and are located in the barrel and in the endcaps, respectively. The RPCs consist of three layers of resistive plates providing three independent measurements of η and ϕ in order to be used in coincidence to reject fake tracks in triggering. The TGC consists in multi-wire chambers similar to RPCs. It covers forward regions and it is also used to complete MDT information.

2.2.5. Forward detectors

A system of forward detectors is needed for independent ATLAS luminosity measurements, defined and described in Section 4.1. The LUCID (LUminosity measurement using Cherenkov Integrating Detector) detector consists of two tubes located at ± 17 m

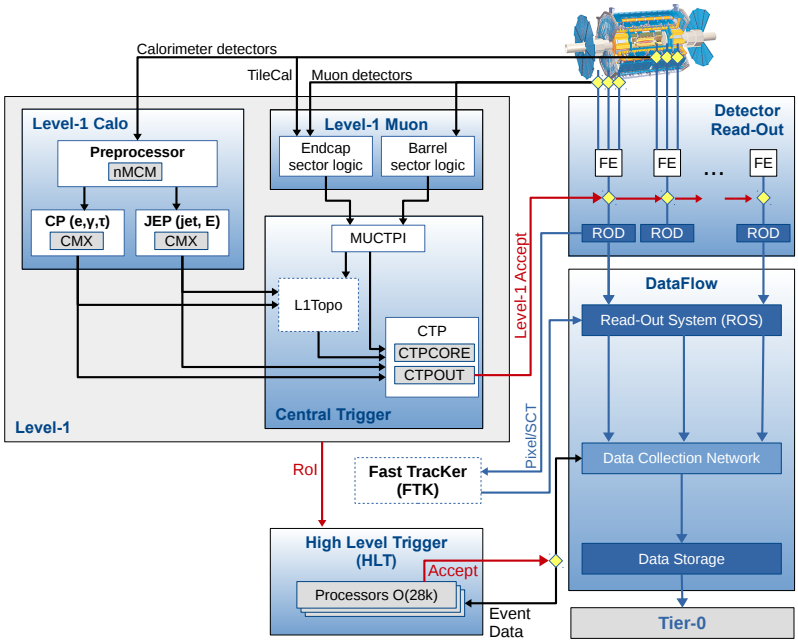


Figure 2.5.: Scheme of the ATLAS TDAQ system [53]. The Fast Tracker (FTK) system is not used for the results of this thesis.

and filled with C_4F_{10} gas measuring the Cherenkov light emitted by high energy protons. The ALFA (Absolute Luminosity For ATLAS) detector is made of a set of scintillators inside Roman pots located at ± 140 m from the interaction point and at 1 mm from the beam line. Both detectors measure the luminosity and monitor beam conditions by correlating them to signal amplitude or rate.

The ZDC (Zero Degree Calorimeter) is instead a calorimeter that measures forward photon and neutron emissions, used to estimate the centrality of heavy ion collisions.

2.2.6. Trigger and readout system

The Trigger and data acquisition system (TDAQ) [53] allows data recording and storing in ATLAS. The trigger filters the detected events based on physics criteria, since the whole detected dataset cannot be recorded due to disk space and bandwidth limitations. Therefore only potentially interesting events are selected and stored. The decision of keeping or not a recored event must be taken at a very high rate, leading to a multiple level decision structure. During Run2, ATLAS used two levels of trigger: a hardware-based first level trigger (L1) and a software-based higher level trigger (HLT). Data acquisition system represents the structure that brings on-detector recorded events

outside for further processing and eventually storage. Trigger and readout systems are strongly connected: the trigger system needs a smooth and fast data transmission off-detector for decision making, while the read-out depends on trigger system performance to not encounter busy bandwidth or issues in the data acquisition chain. A scheme of the ATLAS TDAQ system is shown in Figure 2.5.

The readout of data is detector-specific and classified into on-detector electronics and off-detector electronics. Events measured in a detector are read-out by on-detector electronics where they are temporarily stored into buffers and simultaneously sent to L1 trigger. Only muon system and calorimeters participate in the L1 decision: the muon detector via the trigger-dedicated sub-detectors and the calorimeter with signals from low-granularity cells. This information is elaborated by the L1 trigger to define Regions-of-Interest (RoIs) in ϕ and η. In case the RoI satisfies the conditions in the trigger menu, data are allowed to be transferred from the on-detector electronics to the off-detector readout systems (ROS). The access to only partial data from the L1 is motivated by the need of a quick decision, that has to be taken in 2.6 µs, reducing the event rate from 40 MHz of the bunch crossing to less than 100 kHz. The HLT inherits the information on RoIs from L1 and access to the readout data from all the systems. Its trigger decision reduces the event rate to around 1 kHz. The full information for events passing the HLT trigger is then stored in storage disks at CERN.

CHAPTER 3

Pixel detector: charge information and radiation damage

The ATLAS Pixel Detector [54] faces an unprecedented event rate, providing precise measurements of tracks from charged particles. To do so, a good position resolution is needed to well reconstruct tracks and vertices. Position reconstruction is based on measured charge from single firing pixels (hits) merged into clusters. A distorted charge information would therefore affect position resolution, and hence the detector performance. The study of the charge response of the detector is then important for the position measurement. Moreover, the charge collection is highly affected by radiation damage of the sensor. Therefore, the measurement of the charge provides information on the damage of the sensor bulk.

In this chapter, charge information and effects of radiation damage in the Pixel Detector are described: the description of the detector and an overview of performance during Run 2 is given in Section 3.1. A global study of charge information for the IBL is presented in Section 3.2. A closer look into radiation effects on detector performance is presented in Section 3.3. A brief plan for further studies is given in Section 3.4.

3.1. Structure and performance

The Pixel Detector is structured into four barrel layers in the central region and two endcaps composed of three disks on each side, as shown in Figure 3.1. This configuration allows to measure three points for tracks within $|\eta| < 2.5$.

For historical and technical reasons, the pixel barrels are divided into three plus one layers: the innermost Insertable B-Layer (IBL) [55,56] layer, and the outer three:

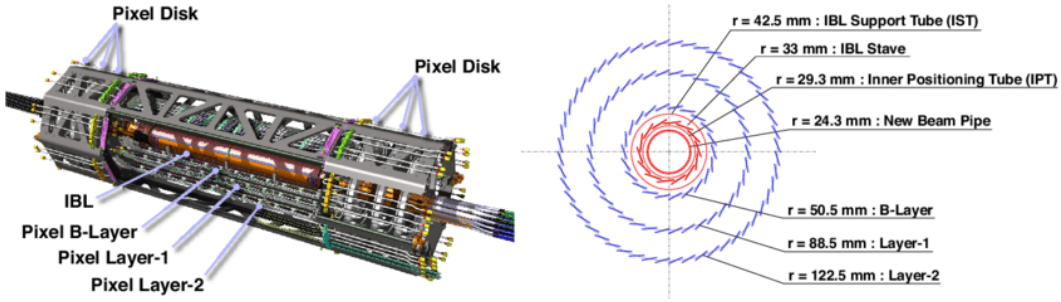


Figure 3.1.: Layout of the ATLAS Pixel detector. On the left the full 3D-scheme of the detector, on the right the transverse section with respect to the beam axis © CERN.

B-Layer, Layer1 and Layer2. The outer three layers and the endcaps were installed in 2007 and used for the Run 1 measurements. As shown in Figure 3.1, the B-Layer is the closest layer of the three installed for Run 1, facing the largest radiation fluence of $\approx 25 \times 10^{13}$ cm^{-2} per year. The absorbed radiation degraded the performance in charge collection of the B-Layer, consequently the position and vertex resolution. Therefore a new layer was inserted closer to the beam pipe. Thanks to the reduction of the beam pipe size, the IBL has been inserted during the LS1, at 33 mm from the interaction point. To better understand the whole detector, an overview of the layout, modules and read-out of all the layers are given in Sections 3.1.1 to 3.1.3. Finally a brief summary on the achieved performance in Run 2 is given in Section 3.1.4.

3.1.1. Layout

Referring to Figure 3.1, all the layers in the Pixel Detector have a common structure: barrels are located around the beam line and divided into staves, hosting a variable number of sensors. The IBL consists of 14 66 cm-long staves, hosting 20 modules each. The other layers are located at increasing distance from the interaction point, with an increasing number of staves (18, 42 and 56, respectively). Each stave hosts 13 modules, located over the stave length of 80 cm. All the staves are tilted by an angle from 11° to 16°, corresponding to the Lorentz angle of the charge drift in the sensor. This angle is different layer by layer because of the different electric and magnetic fields.

The endcaps are composed of three disks on each side and located at increasing distance from the interaction point, namely at 495, 580 and 650 mmalong the z-axis. The 72 modules are arranged in 8 sectors placed in radial direction from a distance of 12.6 cm to 18.7 cm.

All the mechanical support structures, such as staves and disks, are made of a carbon-fibre composite, used as cooling support for the hosted modules and electronics.

3.1.2. Sensors

Different sensors have been used in the IBL and the outer layers. The outer layers use 2×6 cm^2 sensors with a thickness of 250 µm. These sensors are n^+-in-n type, able to be depleted even after bulk type inversion due to radiation damage [57]. The sensor is then segmented into standard 50×400 µm^2 pixels and long 50×600 µm^2 pixels read out by 16 front-end chips. The long pixels cover the space between two neighbouring chips. Guard rings are used to dump the field at the edge of the modules.

The same sensor type is used in the central modules of IBL. However, few features have been changed. The overall size has been modified to approximately 2×4 mm^2, and the thickness to 200 µm. Since a different front-end chip has been used, the segmentation has also changed. In the IBL setup, just two front-ends read the whole module: the standard size of the segmentation has been reduced to 50×250 µm^2, the long 50×450 µm^2 pixels are located at the contact between the two front-end chips and longer 50×500 µm^2 pixels are used at the module edges. A reduced guard ring system is used to improve the fraction of active area.

The forward region of the IBL (4 modules on each side of the stave) is populated by 3D sensors. In this type of sensor, the anode and the cathode are not on the surface of the silicon sensor, but they are implanted as a column (with a diameter of 10 µm) into the sensor depth. This reduces the drift path of the charge, and therefore the sensitivity to radiation damage of the bulk. The size of the module is half of the planar ones and they are read by a single front-end, with the same segmentation as the planar. Due to the smaller depth to be depleted, a lower operational voltage is needed. The thickness of the sensor is 230 µm.

3.1.3. Readout

As mentioned in Section 3.1.2, the main difference between the IBL and the outer layer is the different readout chip, also called *front-end* chip in the following. Despite a similar logic, the two chips use different technologies, have different segmentation, and show different radiation hardness. The outer layers use FEI3 [58] chips, based on 230 nm CMOS technology, resistant to radiation up to 50 Mrad and segmented into 160×18 pixels. In contrast, the IBL uses instead FEI4 [55] chips based on 130 nm CMOS technology and are radiation-hard up to 250 Mrad with a segmentation into 336×80 pixels.

From a logic point of view, both are based on double-column logic. Groups of two columns at the time exchange input and output data with the on-chip processor that handles the communication with off-detector electronics, such as trigger, commands or readout. Each pixel measures the Time-over-Threshold (ToT) of the signal and the timestamp. This information is initially buffered waiting for the trigger signal to be read out. The FEI4 is also featured by a 4-pixel logic read-out: when a pixel fires, the signals from the four surrounding pixels are also saved and propagated to recover

Year	Fraction of operational modules		Efficiency of high-quality data-taking
	IBL	Others	
2015	99.5%	98%	93.5%
2016	99.5%	98%	98.9%
2017	99.3%	97.5%	100%
2018	99.3%	94.8%	99.8%

Table 3.1.: Summary table for Pixel Detector data-taking performance per year. The fraction of operational modules and runtime efficiency with good data quality conditions are shown [57, 59]. The label "Others" refers to the other layers: B-Layer, Layer1 and Layer2.

late signals due to *time-walk*[1]. Although both systems measure charge information, the granularity is different: FEI3 provides a 16-bit information, while FEI4 just 8-bit, in units of Bunch Crossing (BC, 25 ns). This aspect is further investigated in Section 3.2.

Both systems have the same optical-board for electronic to optical signal conversion for off-detector data transmission, and off-detector electronics.

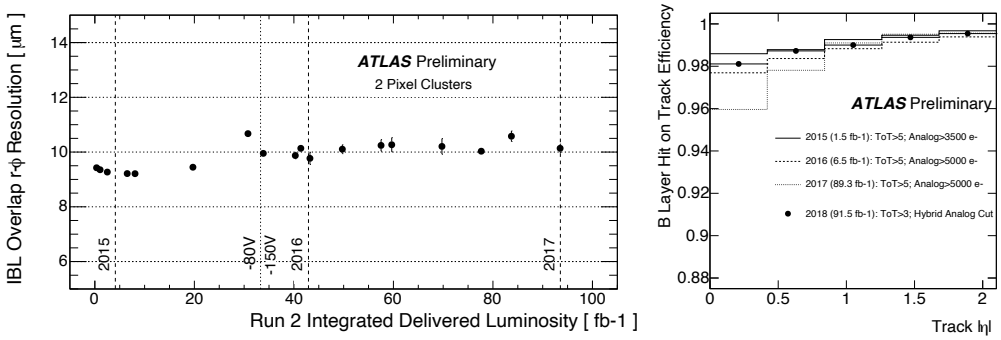
3.1.4. Performance in Run 2

During almost ten years of operation of the ATLAS detector, its performance has been exceptionally good in terms of stability and physics.

The detector has been running with high data-quality efficiency over the Run 2 and very low non-operational fraction of modules as shown in Table 3.1. The decrease of operational fraction of the detector is due to the increase of disabled modules over time, due to unrecoverable high noise or readout issues. For instance an important cause of disabled modules are failures of optical board links, preventing modules data transmission. However, the fraction of disabled modules reached only 4.47% in 2018, corresponding to only 98 out of 2192 modules of which only three are in the IBL. A very high data quality efficiency has been achieved over the years. This efficiency is a luminosity-weighted runtime efficiency of the detector in conditions of good quality of data.

Performance in terms of hit efficiency and position resolution met design expectations. A resolution on the position of 10 µm has been achieved and has shown to be stable over time (Figure 3.2a). A good position resolution in the innermost detector is crucial to

[1] Large and small signals have a different raising time, therefore two simultaneous signals with different amplitudes may end up in different time frames. The time difference between hits associated to two simultaneous signals with different amplitudes defines the *time-walk*.



(a) (b)

Figure 3.2.: Azimuthal r-ϕ position resolution versus integrated luminosity (a). It corresponds to the residual r-ϕ position of the cluster with respect to the reconstructed track crossing the overlap regions in the IBL. The hit efficiency for B-layer against η (b) is measured by associating a detected cluster to a reconstructed track. Inefficiencies come from outliers, that are clusters too far from the track, or missing clusters [60].

ensure secondary vertex reconstruction, and therefore for b- and τ-tagging.

The hit efficiency is also an important parameter to assess the tracking performance. It is defined as the efficiency of detecting a pixel cluster in association of a reconstructed track. An overall hit efficiency over 95% for all the layers has been measured over time. The B-Layer exhibits a significant drop in efficiency for the central region as shown in Figure 3.2b. It is due to the larger absorbed fluence with respect to the forward regions. To recover a homogeneous performance over η, different thresholds on signal amplitude have been set for modules along the stave.

3.2. The charge information in the IBL

Despite the extraordinary performance of the detector during the Run 2, some issues have been observed, calling for a better understanding of the detector response. The puzzle of the charge information in IBL is one of those: the deposited charge in the sensor measured in data from pp collisions has been found to be lower of about 20% than the expectations. The distributions of the deposited charge were also wider in data than in the simulations. Studies for checking the impact of calibrations and operational temperature have been performed in different environments, but no specific cause has been found. A dedicated investigation on this issue is described in Section 3.2.1 [61]. The extraction of a correction for the predicted charge to restore the agreement with data is described in Section 3.2.2.

3.2.1. The cause of the charge defect

Before describing the performed measurement to understand the cause of the observed charge defect, a brief introduction on the charge measurement and calibration in IBL is given. A charged particle crossing the sensor ionises electrons in the sensor depth that are drifted towards the cathode by the applied electric field. The movement of this charge induces an electric signal[2] in the cathode that is propagated through the analog circuit of the attached front-end. The analog circuit consists of a charge amplifier that magnifies and shapes the input signal, connected to a discriminator, that requires the signal to be over a given threshold (usually around 3000 e). In case the signal is over the threshold, the time of the leading-edge of the signal is measured together with its Time-over-Threshold (ToT). Since the amplitude of the signal is proportional to the original ionised charge in the pixel and the ToT is a measure of the signal amplitude, the charge information is reconstructed from the ToT information. Proportionality is not satisfied because the ToT information is encoded in a 8-bit-long word, giving a saturation at large ToTs, as well as non-linearity close to the threshold.

The ToT-to-charge calibration consists of the extraction of the relation between the injected charge in the analog circuit, via an on-chip charge injection circuit, and the ToT of the measured signal. The injection of the charge is done by applying a given voltage onto an injection capacitance, so that the charge deposited on the plates is propagated through the front-end analog circuitry. A scan over different applied voltages is performed and the values of the ToT of the induced signals are measured. A distribution of the injected voltage for each ToT is reconstructed (Figure 3.3a) and its mean and standard deviation are associated to the value of the corresponding ToT. The conversion from voltage to charge is obtained by multiplying the voltage by the injection capacitance, obtaining the calibration curve in Figure 3.3b. This procedure is repeated for each pixel in the front-end matrix: the voltage is injected by a single pulser that is shared by all the pixels in the front-end, while two different injection capacitances per pixel are available. Both capacitances and their sum can be used for the injection.

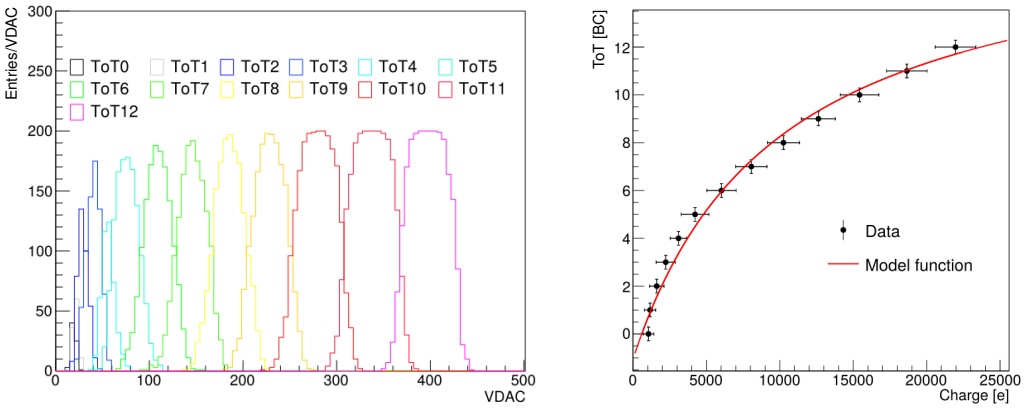
The function used to parametrise the ToT-to-charge calibration (Figure 3.3b) is:

$$\mathrm{ToT} = p_0 \frac{p_1 + Q}{p_2 + Q} \tag{3.1}$$

where p_0, p_1 and p_2 are the fitted parameters and Q the collected charge in the pixel. This formula has a physics interpretation by modelling the input signal with fixed raising and trailing times ($\tau_{l/t}$). Assuming a constant threshold voltage V_{th} and a charge offset Q_{off}:

$$\mathrm{ToT} = (\tau_t + \tau_l)\left(1 - \frac{V_{th}}{Q_{\mathrm{off}} + Q}\right) = a_0 + \frac{a_1}{a_2 + Q} \overset{\text{rewrite}}{=} p_0 \frac{p_1 + Q}{p_2 + Q} \tag{3.2}$$

[2] In this document, the electric charge is expressed in terms of the elementary charge e, approximately 1.6×10^{-19} C. For semplicity, the notation ke is also used, corresponding to 1000 e.



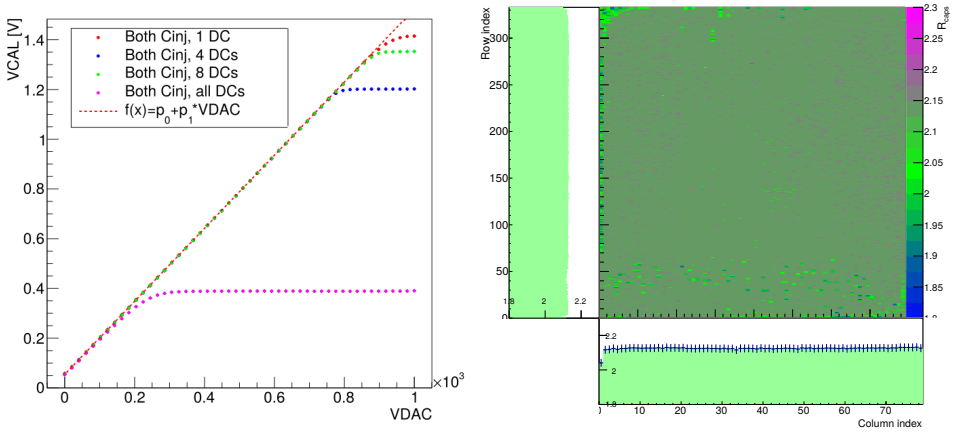
(a) (b)

Figure 3.3.: Distributions of the injection voltage for different ToT values (a). Each voltage is injected 300 times and the corresponding ToT is measured. The distributions correspond to the number of times a given voltage results in a signal with ToT=X BC (ToTX). The voltage is expressed in terms of digital units (VDAC) that are calibrated to real values of applied voltage by dedicated calibrations. By multiplying the applied voltage to the used injection capacitance for each ToT relation between injected charge and ToT is extracted (b).

The last form in Equation 3.1 is preferred because of stability against inversion. The calibration, then, relies on many parameters, such as the values of the injection capacitances and the calibration of the voltage pulser. They are indeed the first suspects for a miscalibration of the charge-to-ToT dependence.

Some measurements in laboratory have been performed to test the injection circuitry. A dedicated on-chip pin is available to directly measure the voltage pulser response. A typical behaviour is shown in Figure 3.4a. The calibration of the pulser is measured to be linear as expected, with a saturation that changes for different number of double columns (DC) of pixels that are injected at the same time. This measurement is however insensitive to possible voltage offsets, caused by residual charge on the capacitance plates, since only voltage differences are measured.

A test of the injection capacitances is more difficult to perform, since no direct measurements over the pixel matrix are possible. Therefore an indirect measurement has been performed. The calibration curve is expected to be independent from the value of the injection capacitance: by minimising the difference of the curves obtained via the two different capacitances, the ratio of the values of the two capacitances can be measured.

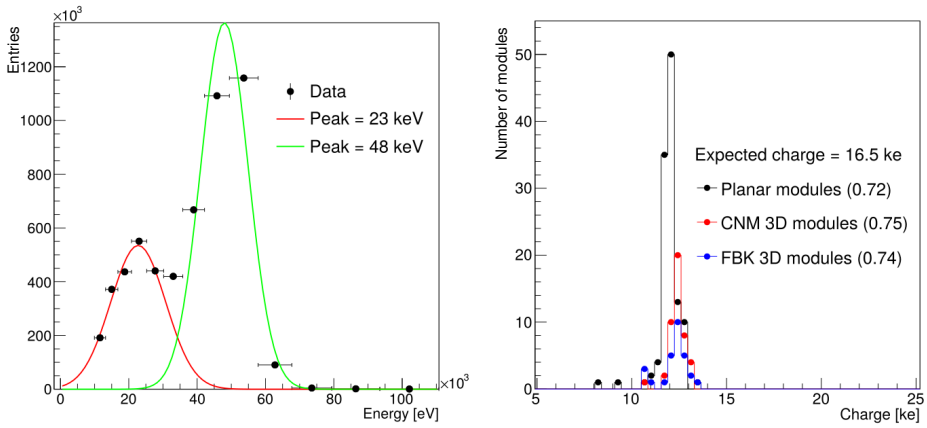


(a) (b)

Figure 3.4.: Calibration of the voltage pulser: VDAC stands for the input value of the pulser and VCAL corresponds the actual voltage in output (a). Different saturation is obtained for different number of double columns (DC) of pixels injected simultaneously. The voltage is applied on both the injection capacitances (C_{inj}) in parallel. A conversion factor around 1.5 mV/VDAC is measured. Measurement of injection capacitances ratio R_{caps} over the pixels in the front-end chip (b).

The capacitances ratio R_{caps} has been chosen as figure of merit to reduce the sensitivity to parasite capacitances. However, R_{caps} cannot measure the absolute values that are used in the calibration. As shown in Figure 3.4b, the R_{caps} is flat over the pixel matrix with a precision of 2%, with the mean value off from the expectation of 2 by 5%. The R_{caps} measurement provides a strong indication of the homogeneity along the module of the capacitances ratio, that also suggests a homogeneity of the values of the capacitances. Moreover, assuming the value of the smaller capacitance at the nominal value, the observed discrepancy of the ratio from the expectation corresponds to a change of 3% in the value of the sum of the capacitances, not enough to cover the observed charge defect.

Another possible cause for the discrepancy is radiation damage, since an irradiated module has lower charge collection efficiency (more details in Section 3.3). A check has been performed by analysing data from modules installed in the detector at the production stage. This dataset consists of validation and characterisation measurements of the modules, at zero absorbed fluence, containing calibrations of all the parameters and a final measurement of a source spectrum. The used source is ^{241}Am providing three photons at approximately 59, 17 and 13 keV. Here, only the most energetic photon is used in the full analysis, since the lower two are close to the signal threshold of 3ke,



(a) (b)

Figure 3.5.: Typical measured spectrum from ^{241}Am in production stage data (a). Summary of the photopeak energy for all the analysed modules is shown in (b). In parenthesis the fraction of average observed charge with respect to the expected 16.5ke for the different module types (planar and 3D from CNM and FBK factories).

corresponding to about 11 keV.

The data from the source acquisition have been analysed. Neighbouring and simultaneous hits have been merged into clusters, whose charge is converted into energy to reconstruct the spectrum of the source, shown in Figures 3.5a and 3.5b. The Figure 3.5a shows the typical observed energy spectrum from the source, with an observed highest photopeak at around 48 keV. The lower part of the spectrum is given by the superposition of the Compton spectrum of the most energetic photon and lower energy photopeaks that are cut down by threshold effects. Data from all the produced modules have been analysed: a summary of the observed highest photopeak charge with the respective fraction to the expectation is shown in Figure 3.5b.

These measurements rule out the hypothesis of module's radiation damage causing the observed charge defect. Moreover, this result gives a strong indication that the defect is not caused by actual poor charge collection in the sensor. Indeed the same fraction of collected charge is observed in planar and 3D sensors (for both production factories, CNM and FBK), that have a completely different charge collection mechanism and different sensitivity to bulk damage. Therefore the issue must be caused by the front-end chip that is instead the same.

As a cross-check, the same measurement has been performed in the laboratory, after a complete tuning and calibration procedure described above. The observed spectrum, shown in Figure 3.6, is consistent with the production data results.

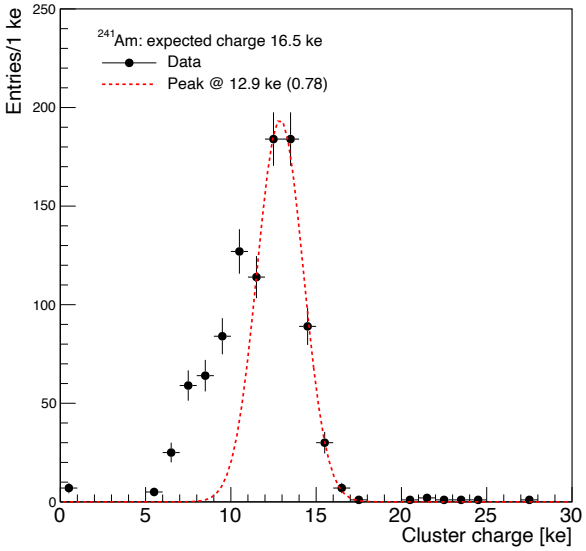


Figure 3.6.: Measurement of ^{241}Am spectrum in the laboratory. The different shape of the spectrum with respect to Figure 3.5a is due to higher digital cut on ToT.

To summarise the outcome of these studies, the cause of the observed charge discrepancy in data must be associated with the front-end charge injection circuitry. The main suspect would be an offset of the voltage threshold, changing along the module. This has been studied also by using a dedicated toy Monte Carlo simulation to reproduce the expected ToT distribution from the ^{241}Am source measurement [61]. The simulation predicts a much larger peak for emitted photons with lower energy, that could be cut down by the presence of an offset on the threshold value. This threshold offset corresponds to residual charge on the plates of the injection capacitances that cannot be directly measured with the current implementation of the chip. Complementary studies [62–64] give indications in this direction, finding out a degradation of the voltage supply with the distance from the pulser. This would explain the observation of this effect only in FEI4 and not in FEI3 despite the analog and charge injection circuitry is essentially the same: the FEI4 is much larger than FEI3, thus the voltage supplier might struggle more in providing the desired voltage at longer distances.

3.2.2. Charge correction for the ATLAS Pixel Simulation

The ATLAS Pixel Simulation does not predict correctly the charge spectrum of the measured tracks in data, and a 20% lower charge is observed. From the studies in Section 3.2.1, the discrepancy seems to be caused by undesired defects in the charge

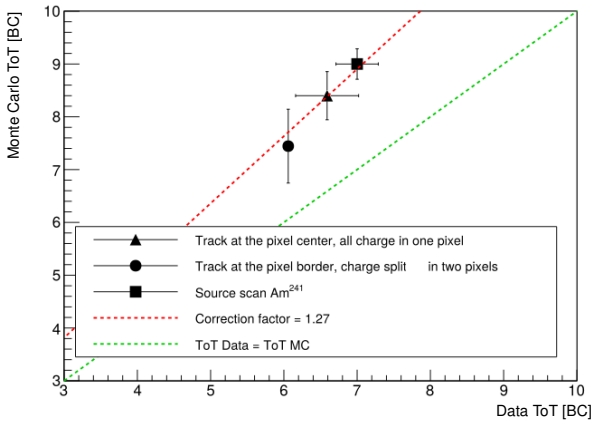


Figure 3.7.: Comparison between expected ToT (Monte Carlo ToT) and measured one (Data ToT) for different expected charge released in the sensor.

injection circuit of the front-end, ending up in an overestimation of the collected charge. Since the source of the effect is still not clear and the effect might be very difficult to parametrise and implement in the detector simulation, a data-driven correction is extracted to improve the simulation agreement with data.

A correction factor is extracted at ToT level, by comparing the expected and measured mean ToT value for the same expected charge released in the pixel. Two types of measurements have been used for this correction: collision data from Run 2 (2016) and in-laboratory ^{241}Am source measurement described in Section 3.2.1. In the first case, the expected ToT is provided by the official ATLAS MC simulation, while in the latter by the known calibration. Three points corresponding to three different deposited charges are considered:

- Track with perpendicular incidence in the centre of the pixel in both dimensions, inducing a single-pixel cluster: in this case the ionised charge from minimum ionising particles (MIPs) is expected to be released entirely inside a single pixel.
- Track with perpendicular incidence at the x-edge of the pixel and central in the other dimension inducing a 2-pixel cluster: the MIP ionisation charge is expected to be divided equally between the two pixels.
- ^{241}Am spectrum measurement (production data): the expected released charge is at the tuning point (ToT equal to 9 at 16 ke injected).

The outcome of these three measurements is shown in Figure 3.7: a correction factor of 1.27 ± 0.03 is extracted, with offset compatible with zero. This correction is then

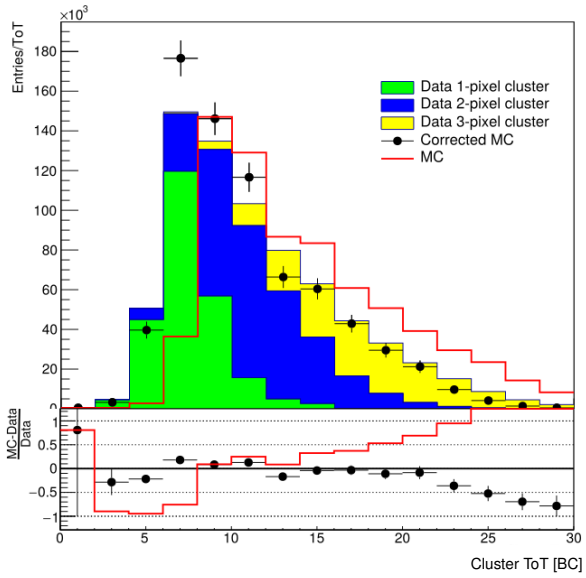


Figure 3.8.: Distributions of the ToT of the clusters in data, original ATLAS simulation and corrected ATLAS simulation.

applied to scale down the ToT values from the simulation to obtain a new data-corrected calibration to correct the simulated charge distribution, shown in Figure 3.8.

A clear improvement in the agreement between data and simulation is achieved. However, a further step to fully validate the correction is needed. The correction in Figure 3.7 is applied a-posteriori on simulated clusters with a digital cut on ToT applied. The ToT value used in the simulation is reconstructed from the simulated ionised charge using the non-corrected calibration. Therefore, the correction must be applied in the simulation at cluster level to be fully validated. The comparison of the peak position of the charge distributions in data, corrected and non-corrected simulations is shown in Table 3.2 [61].

This correction has been formulated in a tuning independent form and is now implemented in the ATLAS Pixel simulation. It provides percent accuracy on the peak of the charge distribution and also good modelling of its shape [61].

3.3. Studies on detector radiation damage

Most of the particles crossing the sensor behave as a minimum ionising particle (MIP), but sometimes non-ionising heavy particles or nuclei could heavily interact with the sensor material, damaging the bulk structure. The most common damage consists of

	ToT MPV	Charge MPV [ke]
Original MC	8.60 ± 0.01	13.19 ± 0.01
Corrected MC	7.26 ± 0.01	10.44 ± 0.01
Data	7.57 ± 0.01	10.65 ± 0.01

Table 3.2.: Comparison between data and simulation agreement on the Most Probable Values (MPVs) of the ToT and charge distribution before and after the correction is applied.

displaced silicon in the bulk structure that induces a leftover vacancy. New energetic levels inside the energy band gap are then created: if occupied, the effective doping of the bulk changes. Many effects are caused by the variation of the doping: the charge collection decreases because of the increase of the probability of charge trapping, the electric field changes as well as the voltage needed for the depletion of the sensor depth.

An understanding of the damage is crucial to improve detector performance at higher absorbed dose. Therefore, a dedicated effort for a Radiation Damage Simulation has started in ATLAS [65] to provide reliable simulation for the upcoming Run 3. In parallel, studies of detector performances over time are needed to tune and check the simulation. Many observables can indicate radiation damage:

- Charge collection efficiency: the detector tends to collect less deposited charge due to charge trapping.
- Lorentz angle evolution: changes in bulk structure and doping modifies the electric field and consequently the Lorentz angle.
- Leakage current: defects in the bulk lead to new energetic levels in the band gap. Therefore, a change of the leakage current is introduced.
- Depletion voltage: the change of the effective doping strongly affect the applied voltage needed to deplete the sensor. Also temperature conditions affect the depletion voltage evolution.

Here, the behaviour of two basic features is studied: the effect on charge collection efficiency is described in Section 3.3.1 and measurements of the depletion voltage are shown in Section 3.3.2.

3.3.1. Charge collection efficiency

As described in Section 3.2, the charge information is important for clustering and position resolution. A drop in charge collection efficiency causes a reduction of cluster size, degrading the position resolution and the hit efficiency. The missing charge is expected to come from deeper in the sensor depth, since a longer drift of the charge is

needed and the probability of the charge to be trapped becomes larger. To test this hypothesis, a measurement of charge collection efficiency as a function of the ionisation depth in the sensor has been done.

The measurement is based on reconstructed tracks at the beginning and at the end of Run 2 to assess the impact of radiation damage. All the layers are used to test the different fluences[3]. 3D sensors are not considered for this study, due to the different charge collection mechanism. The idea is to use the information of the track direction to predict the ionisation path inside each layer. From the path of the track in the layer, the path in each crossed pixel is reconstructed, allowing the computation of the average depth at which the charge is generated for each pixel. The measured and the expected (from the track length) charge release are finally compared against the average ionisation depth.

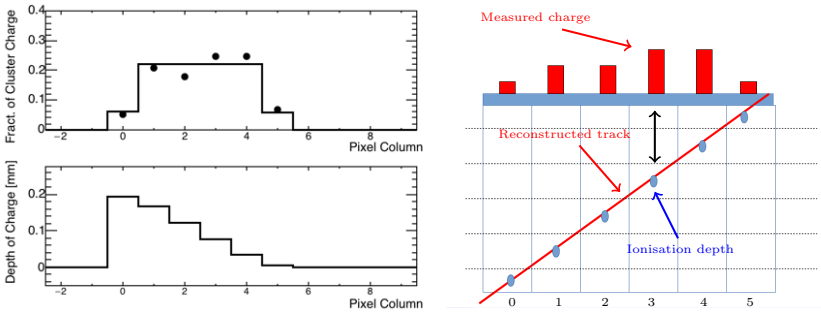
A selection on the tracks and clusters quality is performed: the measured cluster in the sensor is required to be contained in the expected one from track trajectory, and their centres to be compatible inside half a pixel pitch. Moreover only clusters with more than three pixels are used as best compromise between resolution on depth measurement and statistics. The resolution on depth close to the surfaces of the sensor strongly depends on the pixel multiplicity since larger clusters allow a better sampling of the sensor depth.

A final consideration on the charge estimation must be made. The expected and the measured charges are not directly compared, since charge large ionisations, corresponding to depositions in the long tail of the dE/dx Landau distribution, may affect the measurement. Therefore the fractions of the measured and expected charge in the pixels with respect to the cluster charge are used instead[4]. A typical cluster used in the analysis is shown in Figure 3.9a and the corresponding scheme in Figure 3.9b. In the example, the track propagates from the pixel on the "right" to the pixel on the "left": from the track direction (red line) the average ionisation depth for each pixel in the cluster is computed (blue dots) and the charge is measured (top red rectangles) for each pixel. In Figure 3.9a, the corresponding measured and expected charges in pixels are shown in the top pad, while the ionisation depths per pixel are shown below. As expected the computed values of the depths increase with the distance from the read-out (blue top rectangle in Figure 3.9a).

The results of this analysis are shown in Figure 3.10. A general disagreement in the outer bins is explained by charge sharing and threshold effects, therefore they are not considered for drawing the conclusions of this study. A clear trend is visible in IBL in 2018, giving lower collection efficiency for larger ionisation distances from the collection point. The cause is attributed to the radiation damage given the absence of such trend in data collected in 2015. A smaller discrepancy is also observed in the B-Layer, while no effect in the outer layers is visible. This is compatible with the expected absorbed dose by each layer.

[3] Corresponding to almost 10^{15} $n_{eq}cm^{-2}$ of absorbed dose for the innermost layer.

[4] The expected charge is given by the fraction of the ionisation path in the pixel with respect to the total path. It coincides with the charge fraction assuming a constant ionisation of the particle.



(a) (b)

Figure 3.9.: Typical measurement of the depth based on cluster and track information (a). The solid line corresponds to the expected charge fraction in the pixel, while the points correspond to the measurement. The lower pad in (a) shows the computed deposition depth for each pixel in the cluster. Here, the column indices are from 0 onwards for display purposes. In the explicative scheme (b), the track is indicated with a solid red line, the computed charge depths are shown as blue dots and measured charges for each pixel is sketched with red rectangles. The numbers refers to the column numbers in (a).

3.3.2. Depletion voltage evolution

The depletion voltage deeply affects the detector operations: since it controls the size of the depletion region in the sensor, it changes the profile of the electric field, the efficiency of charge collection and the magnitude of the leakage current. Generally, the higher applied voltage is, the more the sensor is depleted, but it must not exceed the maximum operational voltage of 600 V for safe running conditions. It strongly depends on the absorbed dose and on the temperature conditions. The measurement of the evolution of the depletion voltage as a function of these two parameters is therefore interesting to predict the future running conditions. In this section, the depletion voltage is measured as a function of the absorbed dose, comparing results from different test environments [66].

The depletion voltage is extracted by looking at the charge response of the sensors at different applied voltages. It is measured as the crossing-point between two regimes of the sensors:

- Under-depleted regime: the applied voltage (HV) is lower than the depletion voltage V_{dep}, thus the depleted depth is smaller than the sensor thickness. The collected charge is expected to grow with the applied voltage as $\sim \sqrt{\text{HV}}$.
- Over-depleted regime: the applied voltage is higher than the depletion voltage. The whole ionised charge is expected to be collected, reaching a plateau. How-

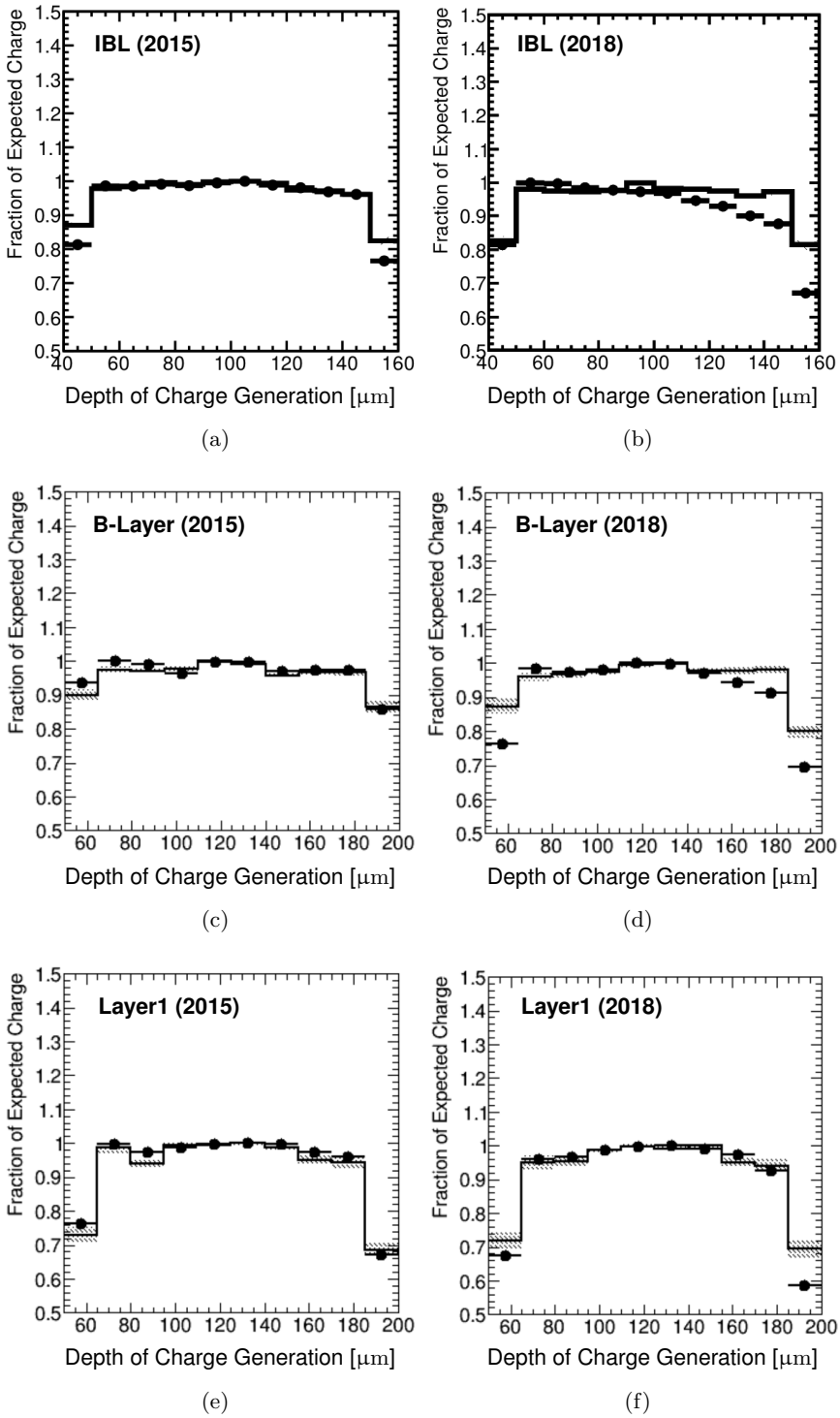


Figure 3.10.: Charge collection versus ionisation depth for IBL, B-Layer and L1 in 2015 and 2018. Data are shown with points, while MC prediction (not including radiation damage effects) is shown with the solid line.

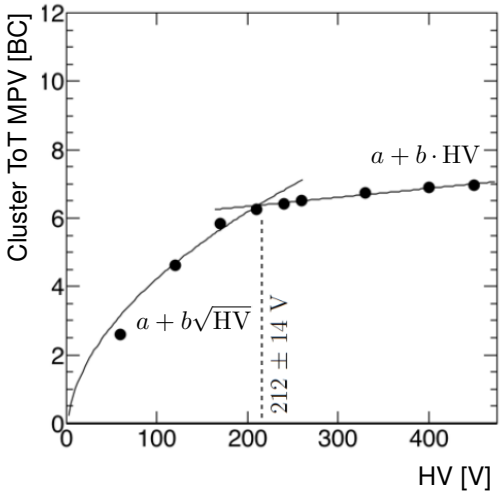


Figure 3.11.: Example for the measurement of the depletion voltage of a sensor.

ever, especially after radiation, defects in the bulk cause charge trapping. In this case, a larger applied voltage leads to a smaller drift time, reducing the trapping probability. A small linear increase of the collected charge with the voltage is then expected.

The depletion voltage is therefore determined to be at the intersection of the two curves representing the fit results of the two models at low and high applied voltages[5], as shown in Figure 3.11. Two independent measurements have been conducted using collision data in different run periods and dedicated beam tests. The analysis of data from collisions has been done by comparing the results for the measurement of the depletion voltage at different times during the Run 2, then at different absorbed doses. To obtain comparable results, two beam tests have been performed on the same sensor and front-end before and after 1×10^{15} neq/cm^2 absorbed irradiation dose. This value is compatible with the expected radiation dose absorbed by IBL at the end of Run 2. A summary of the two beam test setups is shown in Table 3.3.

Some technical issues made the data from the beam tests difficult to analyse. For example, measured ToT at low values were not acquired, cutting out the peak of the

[5]The charge is measured via ToT distribution, not converted to charge through the calibration. Results obtained by using ToT and charge distribution have shown to be compatible.

Fermilab	
Device	D31-18-01 with FE-I4
Absorbed fluence	0 neq/cm^2
Date	March 2018
Telescope	Four planes with MIMOSA sensors
Settings:	
Threshold	2500 e
Tuning	ToT 8 at 16 ke
Bias voltage	80, 100, 150, 250, 350 V

CERN H8	
Device	D31-18-01 with FE-I4
Absorbed fluence	1×10^{15} neq/cm^2
Date	September 2018
Telescope	Five planes with IBL sensors and front-ends
Settings:	
Threshold	2500 e, 2000 e
Tuning	ToT 8 at 16 ke, ToT 10 at 16 ke
Bias voltage	50, 80, 100, 150, 200, 250, 300, 350, 400, 500 V

Table 3.3.: Summary of settings and conditions of the two beam tests measurements.

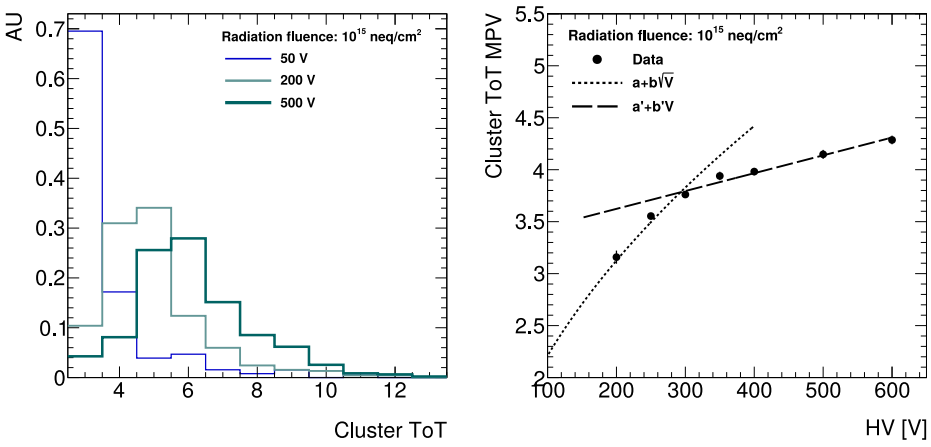
ToT distribution at low values of the applied voltage (Figure 3.12a). These data points have been excluded from the measurement, giving few points to test the under-depletion regime, as shown in Figure 3.12b.

A summary of the beam tests and collision data measurements is presented in Figures 3.13a and 3.13b. The conversion between fluence and corresponding luminosity is extracted from simulation [67]. A similar slope of 1.39 ± 0.10 V/fb^{-1} (Figure 3.13a) is observed for both measurements. In the same fashion, both measurements show similar linear trends in the over-depletion regime (Figure 3.13b).

3.4. Conclusions and plans

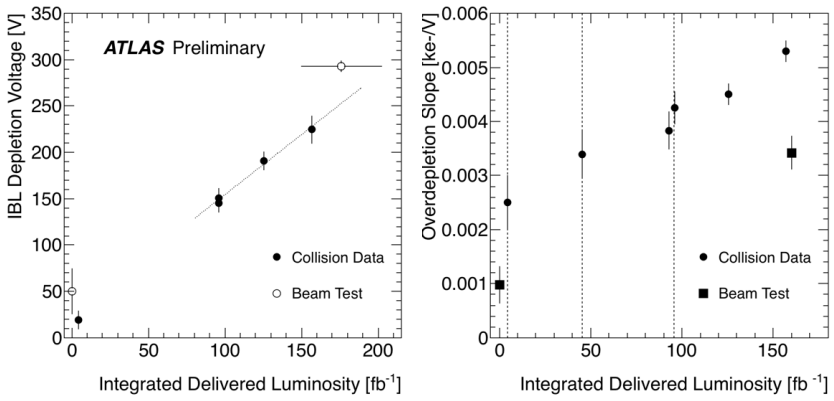
The described studies have a two main goals: to understand better the detector and to quantitatively evaluate the effects and quantities which modelling have to be improved to ensure better prediction of the detector performance in the future.

The puzzle of the charge defect for IBL has been attributed to an intrinsic issue of the front-ends. New studies started to directly address this problem. Measurements



(a) (b)

Figure 3.12.: Evolution of ToT distribution with the applied voltage (a). The corresponding curve for the depletion voltage extraction (b). Note that the points below 200 V, where the peak of the ToT distribution is not visible, are missing from the curve.



(a) Depletion voltage (b) Overdepletion slope

Figure 3.13.: Trend of depletion voltage observed in beam test and collision data (a) [60]. Comparison of the slope in the over-depletion regime (b). The large offset between the points corresponding to data from collisions and beam test is attributed to the different history of the sensor operations (such as temperature conditions), that strongly influences the depletion voltage. Therefore, only trends of the measurements in the two setups are compared.

of localised charge depositions from sources and tracks can be used to investigate a possible change of the threshold in different parts of the module. This might provide a possibility to tune the current simulation to correct for this effect. Also alternative calibration procedures are investigated. In the meantime, the proposed correction is used in the official ATLAS Pixel Detector simulation to correct the charge response and it has been refined and generalised to be used for different operational conditions.

Radiation damage is the biggest challenge for detector performance in Run 3, when the detector will be already damaged and will have to face even more challenging luminosity and fluence conditions. A dedicated effort to address the evolution of the detector features has been started to predict conditions and monitor detector degradation during the Run 2. The measurements of charge collection and depletion voltage can be used to tune the detector simulation. A successful example of this is the measurement of the leakage current against the longitudinal position over the IBL [57] that helped the simulation algorithms to provide more accurate predictions. The same will happen with the measurement described in this thesis. For example, the slope of the depletion voltage in the over-depleted regime can yield indications on the effective doping of the bulk since it is related to charge trapping probability. It can thus be used to monitor the conditions of the sensors and to further study the charge collection in a damaged sensor bulk.

With more data, more radiation and higher luminosity, the detector will operate under untested conditions, therefore a continuous investigation of the performance and physics features is needed to properly tackle the Run 3 data-taking. These studies will represent a solid baseline for the next generation of tracker for ATLAS, the Inner Tracker (ITk), expected to be installed for the HL-LHC phase.

CHAPTER 4

Dataset, simulations and objects

This chapter presents all the elements used in a physics analysis. Prior to the description of the available dataset in Section 4.2, an introduction to the concept of *luminosity* is given in Section 4.1. Simulations of the physics processes as measured by the detector play a key role in physics at hadrons colliders. An overview of the several parts of these algorithms, from the computation of the quantum field theory amplitudes to the simulation of the detector response, is given in Section 4.3. Eventually simulated and data events are subject to the same algorithms for the reconstruction of the physics objects, described in Section 4.4.

4.1. Luminosity definition and measurement

A generic "counting experiment" consists of counting the number of detected events from the process of interest. The measured number $N_{\text{obs.}}$ can be predicted by the formula:

$$N_{\text{obs.}} = \mathscr{L}_{\text{int.}} \cdot \sigma_{\text{process}} \cdot \epsilon_{\text{detector}} \tag{4.1}$$

The cross-section σ_{process} can be predicted from theoretical calculations and is related to the probability of the process of interest to occur. The $\epsilon_{\text{detector}}$ stands for the efficiency of the detector to measure the process of interest. The integrated luminosity $\mathscr{L}_{\text{int.}}$ measures the total number of "performed measurements". In case of a rate measurement (dN_{obs}/dt), the total integrated luminosity is replaced by the instantaneous luminosity, $\mathscr{L}(t) = d\mathscr{L}_{\text{int.}}/dt$.

This equation can be used to predict the outcome from any counting experiment, such as the measurement of the number of times the number one is rolled from throwing a

single dice. The σ_{process} is replaced by the expected probability of rolling the number one (1/6). The $\epsilon_{\text{detector}}$ is equal to one, assuming that the experimentalist does not miss any roll. Finally, by replacing $\mathscr{L}_{\text{int.}}$ with the total number of rolls, the expected number of times the number one is rolled is predicted. In a collider experiment, the association of the correct values to each element is less trivial. The cross-section is predicted by the theory, while the efficiency of the detector is based on Monte Carlo simulation and/or dedicated measurements. The total integrated luminosity corresponds to the total number of pp collisions over the data-taking time. The instantaneous luminosity consists in the average number of pp collisions over a time unit.

Since all the processes resulting from pp collisions share the same luminosity, a well-known very-abundant process is usually chosen to measure it. Forward detectors (see Section 2.2.5) measure the luminosity from detecting the pp inelastic scattering process [68, 69] by using the following equation (derived from Equation 4.1):

$$\mathscr{L}(t) = \frac{d\mathscr{L}_{\text{int.}}}{dt} = n_b \frac{\langle\mu\rangle f_r}{\epsilon\sigma_{\text{inel.}}} \tag{4.2}$$

where n_b is the number of collided bunches per time unit, f_r the bunch revolution frequency, $\langle\mu\rangle$ is the average recorded number of pp interactions per bunch crossing (*pile-up*), $\sigma_{\text{inel.}}$ is the inelastic cross-section and ϵ is the efficiency of the detector used for the luminosity measurement.

Since the luminosity is proportional to the number of pp collisions provided by the LHC, another formulation, based on geometrical features of the beam, is possible. For this purpose, it is useful to use the luminosity per bunch crossing $\mathscr{L}_b$, corresponding to the total instantaneous luminosity per bunch-crossing time period (25 ns). It is expressed in terms of collider specifics as:

$$\mathscr{L}_b = \frac{f_r n_1 n_2}{2\pi\Sigma_x\Sigma_y} \tag{4.3}$$

where $n_{1/2}$ stand for the number of protons in the two beams and $\Sigma_{x/y}$ are their transverse dimensions. Dedicated *van der Meer* [68, 69] scans are performed to measure these geometric parameters of the beams to provide an independent estimation of the luminosity. These scans are vital to constrain the value of the visible cross-section ($\epsilon\sigma_{\text{inel.}}$) in Equation 4.2 to improve the luminosity measurements with the forward detectors [68, 69].

4.2. Dataset

The ATLAS Run 2 dataset is divided into data-taking years from 2015 to 2018. Each year is structured in *runs*, corresponding to shorter periods with stable data-taking conditions. Each run consists of *luminosity blocks* (LBs), around one-minute-long periods in which the average instantaneous luminosity is measured.

Year	GRL integrated luminosity
2015	3.2 fb^{-1}
2016	33 fb^{-1}
2017	44 fb^{-1}
2018	59 fb^{-1}
Total	139 fb^{-1}

Table 4.1.: "Good for Physics" integrated luminosity per data-taking year.

Quality requirements are applied at LB level on the recorded data, providing a list of accepted LBs to be used for measurements, the *Good Runs List* (GRL). The total integrated luminosity in Run 2 corresponds to around 139 fb^{-1}, divided into the different years as in Table 4.1. Every year of data-taking is featured by different running conditions in terms of instantaneous luminosity and pile-up as shown in Figures 4.1 and 4.2. The average pile-up doubled from 2015 to 2018 reaching a maximum of 80, more than three times the design value of 25. The instantaneous luminosity increased drastically from the beginning to the end of Run 2, having a stable profile in 2018 at 20×10^{33} $\text{cm}^{-2}\text{s}^{-1}$, double the design value.

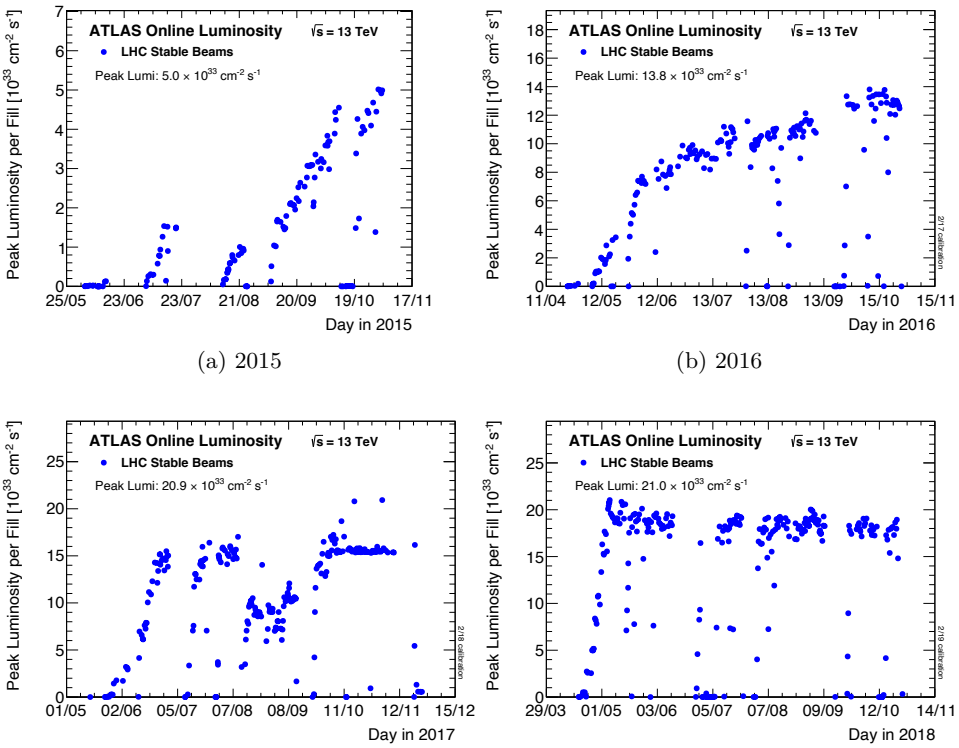
4.3. Monte Carlo simulations

Most of particle physics measurements make use of Monte Carlo (MC) simulations to understand and interpret the data from the detector. This is used in ATLAS to predict the observed distributions in the detector starting from an assumed physics model, for instance the Standard Model.

The simulation of the pp collisions is divided into several parts due to different models used to describe the different energy regimes in pp collisions, as shown in Figure 4.3. The first step consists of the matrix element calculation and event generation (red dot): the amplitude for a specific final state from parton-parton scattering is computed with Feynman rules of a given physics model. The involved initial state partons (curly red lines) come from the colliding protons (two green ovals aside) and their kinematics are described by parton density functions (PDFs). The matrix element is computed at a fixed order in SM perturbation theory providing MC events with fundamental particles in the final state (red lines). In this document, this stage is called *parton level* simulation.

Because of the increase of QCD radiation with lower energy, a step of *parton shower* is necessary, performed by a separate algorithm, interfaced with the matrix element generator. The final and initial state radiation of gluons (blue curly lines)[1] is also simulated in this step. Gluon radiation progressively reduces the momentum scale towards the

[1] Photon radiation is also included.



(a) 2015 (b) 2016

(c) 2017 (d) 2018

Figure 4.1.: Peak luminosity over time for each data-taking year [70].

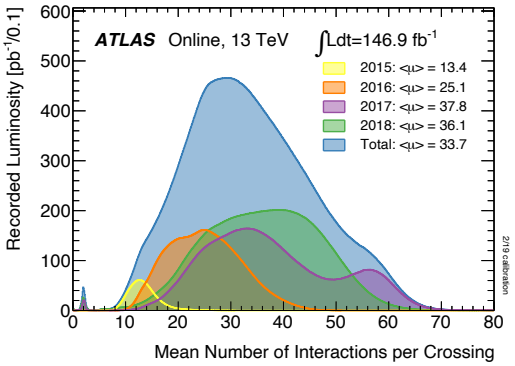


Figure 4.2.: Overall pile-up distribution for different data-taking years [70].

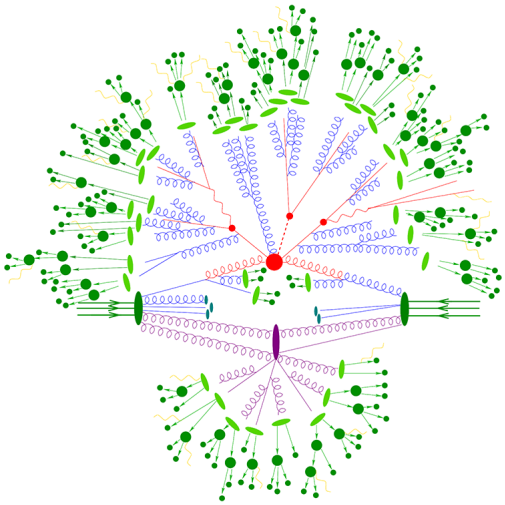


Figure 4.3.: Scheme for hadronic collision simulation in high energy physics [71]. The scheme refers to the generation of a $pp \to t\bar{t}H$ event.

point at which perturbation theory can no longer be used. Here, *hadronisation* takes place: effective models are used to arrange the generated set of partons into color-neutral particles (light green ovals). These unstable particles propagate and decay into stable or long-lived particles that are observed in the detector (dark green lines and dots). *Underlying events* (purple oval and lines) coming from the remaining components of the colliding protons are also simulated, going through the same steps as the hard interaction. In this document, the stage after hadronisation is called *particle level* simulation.

Eventually, the final set of particles is propagated through the detector and the detector response is simulated. Dedicated packages, such as GEANT4 [72], provide accurate predictions for interactions of the particles with matter. A lighter and faster AtlFast-II detector simulation [73, 74] is obtained by parametrising the hadronic and electromagnetic calorimeter response to incident particle showers. At the end of this stage, MC simulations and recorded data have exactly the same format to be processed by object reconstruction algorithms.

4.3.1. Relevant Monte Carlo samples

In the search for the production of four top quarks, MC samples of the signal and the relevant background processes are needed.

The $t\bar{t}t\bar{t}$ production process is modelled with aMC@NLO 2.6.0 [75] that uses the NNPDF3.1NLO PDF set [76] and provides a prediction at Next-to-Leading Order (NLO) in the strong coupling constant α_s (QCD). Dynamical factorisation and renormalisation

scales are set to $H_T/4$, where H_T stands for the sum of the transverse energy over all the particles in the final states at parton level, following the prescriptions in [21]. Top quark decays are simulated with MadSpin [77, 78] to take spin correlations into account. The generator is interfaced with Pythia 8.230 [79], using the A14 tuning [80] and the NNPDF23LO PDF set [76]. Heavy-flavour hadron decays are simulated with the EvtGen 1.6.0 [81] software. Using the same settings, a sample of $t\bar{t}t\bar{t}$ at LO in QCD has been produced to avoid the impact of negative weights during BDT training (more details in Section 5.3.2). An additional MC sample of $t\bar{t}t\bar{t}$ is simulated with the same matrix element generator and interfaced with Herwig 7.04 [82, 83] to assess the uncertainty on parton shower and hadronisation modelling. Herwig uses the H7UE set of tuned parameters [83] and the MMHT2014LO PDF set [84]. Samples are normalised to NLO in QCD+EW cross-section [21] to account for large electro-weak corrections from the large tH coupling.

The $t\bar{t}$ production is modelled by Powheg-Box v2 [85–88] at NLO QCD, using NNPDF3.0NLO PDF [76] with the h_{damp} parameter[2] set to $1.5 \cdot m_{\text{top}}$. Parton shower and hadronisation are performed with Pythia 8.230 [79] which uses the A14 set of tuned parameters [80], making use of the NNPDF23LO PDF set [76]. To assess the uncertainty on the choice of the value for the h_{damp} parameter, a sample with the same settings as above, but with $h_{\text{damp}} = 3m_{\text{top}}$ is simulated (see Chapter 6). The parton shower impact is evaluated by interfacing the matrix element generator with Herwig 7.04 [82, 83] with the H7UE set of tuning parameters set [80] and the MMHT2014LO PDF set [84]. A set of *filtered* samples has been produced to increase the simulation statistics in regions of interest. The *heavy-flavour-filtered* (HF-filtered) samples are classified according to the number of reconstructed heavy-flavour hadrons at particle level. The H_{T}-filtered samples cover different ranges in H_{T}. To evaluate the effect of a different matrix element generator, a simulation with aMC@NLO 2.6.0 [75] with the NNPDF3.0NLO PDF set [75] is used. This is interfaced with Pythia 8.230 [79] with the same settings as the nominal sample. All the samples are normalised to NNLO QCD cross-section, including resummation of soft gluon emissions at NNLL accuracy [89–92] calculated using the Top++2.0 [93] software.

The single-top tW process is generated with Powheg-Box v2 [85–88, 94] at NLO QCD in the five-flavour scheme with the NNPDF3.0NLO PDF set [76]. A diagram removal scheme [95] is used to remove the interference terms with $t\bar{t}$. The single-top t-channel is simulated with Powheg-Box v2 [85–88, 96] in the four-flavour scheme with NNPDF3.0NLOnf4 PDF set [76]. The single-top s-channel is modelled with PowhegBox-v2 [85–88, 97, 98] in the five-flavour scheme with NNPDF3.0NLO PDF set [76]. All the single-top samples are interfaced with Pythia 8.230 [79], using the A14 parameter set [80] and the NNPDF23LO PDF set [76]. Events are normalised to approximate NNLO cross-section [99, 100].

The production of a $t\bar{t}$ pair associated with a vector boson ($t\bar{t}V$) is modelled with

[2] The h_{damp} regulates the p_T of the first additional emission beyond leading-order Feynman diagram in the parton shower.

aMC@NLO 2.3.3 [75] at NLO QCD using the NNPDF3.0NLO PDF set [76]. The generator is interfaced with Pythia 8.210 [79] and the events are normalised to the NNLO QCD+EW cross-section [101]. The $t\bar{t}H$ process is simulated with Powheg-Box v2 [85–88, 94] at NLO QCD with the NNPDF3.0NLO PDF set [76]. Parton shower and hadronisation are modelled by Pythia 8.210 [79]. Rarer processes involving top quarks such as tZq and tWZ are simulated by aMC@NLO 2.3.3 [75] at NLO QCD using NNPDF3.0NLO PDF set [76]. The generator for tZq and tWZ are interfaced with Pythia 8.230 [79] and Pythia 8.212 [79], respectively. The A14 parameter set [80] and the NNPDF23LO PDF set [76] are used in parton shower and hadronisation modelling in the simulation of $t\bar{t}V$, tZq and tWZ processes.

Diboson and V+jets processes are modelled with Sherpa 2.2 [71] with the NNPDF23LO PDF set [76]. This algorithm includes matrix element generation, parton shower and hadronisation modelling. Tuning of parton shower parameters is provided by the authors. Diboson events are normalised to the NLO QCD theoretical cross-section [102]. V+jets samples are normalised to the NNLO cross-section [103].

4.4. Object reconstruction

As mentioned in Section 4.3, only long-lived or stable particles reach the detector such as $\mu^{\pm}$, $e^{\pm}$, $\pi^{\pm}$, p, γ, ν, K_L^0 and n. Each particle is identified via a specific signature in the detector, as shown in Figure 4.4. Electrons are associated with a large deposition in the electromagnetic calorimeter (EMC) with a corresponding ID track. Photons shower in the EMC, but no track is expected. Muons are detected in the MS, possibly in combination with a track in the ID and a MIP energy deposit in calorimeters. Hadrons and mesons generate large deposits in the hadronic calorimeter (HC). Finally, the missing transverse energy is evaluated by imposing an energy conservation requirement.

In the search for four-top-quark production, muons, electrons, missing transverse energy and jets (collimated cluster of charged particles) are used and described in the following sections (from Sections 4.4.1 to 4.4.3 and 4.4.5). A focus on the identification of jets induced by bottom quarks is given in Section 4.4.4. Finally, the description of the *overlap removal* procedure is given in Section 4.4.6.

4.4.1. Electrons

An electron crossing the ATLAS detector is expected to produce an EM shower in the EMC and leave a track inside the ID. Therefore, the information from these two detectors is used for its reconstruction [104]. Since photons are also expected to provide a large deposit in the EMC, electron and photon reconstruction is very similar and a discriminant is used for the identification.

Electron candidates are built starting from proto-clusters in the EMC. The proto-

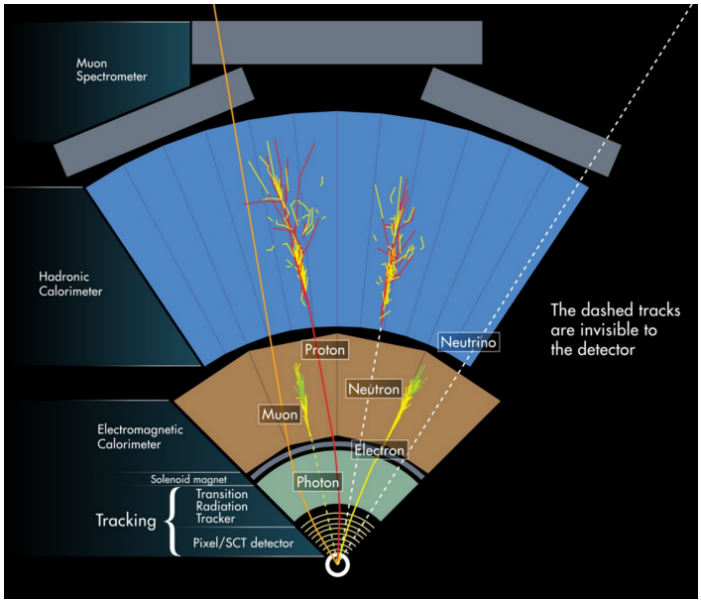


Figure 4.4.: Examples for the particle identification in the ATAS detector.

clusters are seeded by EMC cells measuring a large signal-to-noise ratio that are then merged with the neighbouring cells with a significant energy deposition. Only proto-clusters with an energy above 400 MeV are matched with a track reconstructed in the ID coming from the interaction point [105–107]. Track-matched proto-clusters within a window of 5×3 cells are merged into a *super-cluster*. Special super-cluster creation procedures address the cases of photon radiation from electrons and photon conversion. A super-cluster and the associated track identify an electron candidate.

The electron candidates are required to fulfil identification and isolation criteria [104]. Identification criteria aim to suppress the contribution from hadrons and photons faking the electron signature and electrons from photon conversion. The decision is taken with a likelihood discriminant based on the properties of the primary track, the shape of the shower in the calorimeter and the compatibility of the track with the cluster. In this analysis, the *Tight* working point for electron identification is used. It requires the tightest cut on the likelihood value and the electron to have $E/p < 10$ and a primary track with $p_{\mathrm{T}} > 2$ GeV. Isolation criteria are applied to reject electrons coming from heavy-flavour hadron decays and further suppress the contribution of misidentified hadrons. The criteria are based on calorimetric and track isolation variables: $E_{\mathrm{T}}^{\mathrm{cone20}}$ and $p_{\mathrm{T}}^{\mathrm{varcone20}}$, respectively[3]. The working point for electron isolation in this analysis is

[3] The $E_{\mathrm{T}}^{\mathrm{cone20}}$ variable corresponds to the sum of the transverse energy from ECM cells in $\Delta R < 0.2$ from the track. The $p_{\mathrm{T}}^{\mathrm{varcone20}}$ variable is the sum of p_{T} of tracks lying in a $\Delta R < 0.2$ angular cone around the electron candidate track. The cone varies depending on track p_{T}, with the maximum of $\Delta R < 0.2$ for low p_{T}.

FCTight, requiring $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}} < 0.06$, $p_{\mathrm{T}}^{\mathrm{varcone20}}/p_{\mathrm{T}} < 0.06$ and a loose matching of the track with the reconstructed vertex to reduce the pile-up dependency of the criteria.

The used *Tight* identification has a rather high average efficiency of 80% providing up to 4.5 times more background rejection power than the looser operational points. The *FCTight* isolation provides $> 95\%$ average efficiency and is robust against pile-up. The efficiencies of isolation and identification criteria are derived from *tag-and-probe* analyses [104, 108]. By comparing the efficiencies measured in data and MC on $Z \to e^+e^-$ and $J/\psi \to e^+e^-$ events, dedicated scale-factors are derived and applied to simulation. A similar procedure is used to correct MC electron energy scale and resolution to agree with observed data [104].

4.4.2. Muons

The ATLAS Muon Spectrometer (MS) plays a key role in the identification of muons coming from the interaction point [109]. The muon reconstruction starts from MS track candidates built from hits in the spectrometer segments. Information from reconstructed tracks in the MS is combined with the measurements of the other sub-detectors to identify different types of muon candidates [109]. In this analysis, only *combined muons* (CB) are used: they are featured by a matching track in the MS and ID sub-detectors.

Identification and isolation criteria are then applied to muon candidates [109]. Identification algorithms aim to reject muons from kaon and pion decays and select prompt muons from the hard scattering with high efficiency and a robust momentum measurement. Muons coming from long-lived particle decays have a kink in the reconstructed track, therefore track quality assessment is used to reject them. Specifically, the compatibility of the track charge and momentum measurements in MS and ID and the χ^2 of the track fit are used to identify three identification working points. In this analysis, the *Medium* operating point is chosen: this asks for a robust track reconstruction in the MS that loosely matches the ID track. This working point has been shown to minimise uncertainties from muon reconstruction and calibration [109]. Isolation criteria are applied to further suppress the contribution from non-prompt muons that are expected to be produced in association with other particles. The track of a muon candidate is asked to fulfil $p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.06$, with $p_{\mathrm{T}}^{\mathrm{varcone30}}$ defined as in Section 4.4.1 using a maximum cone angle of $\Delta R < 0.3$. No requirements on calorimeter isolation are applied. As in the electron isolation criteria, the muon candidate is required to loosely match the reconstructed vertex.

This configuration for the muon reconstruction, identification and isolation ensure more than 98% efficiency, with a strong rejection of muons coming from hadrons ($> 80\%$). Efficiencies are measured in a *tag-and-probe* analysis. In the same way efficiency scale-factors, momentum scale and resolution are extracted from data to improve the MC agreement with data and reduce the associated systematics [109].

4.4.3. Jets

Colour-charged particles created in pp collisions combine with each other to create colour-neutral states due to strong interaction confinement. A *jet* is defined as a collimated spray of particles propagating in a cone around the initial colour-charged particle. The reconstruction of jets is therefore crucial to identify strongly interacting constituents coming out of the hard scattering. The jet signature in the detector is a collimated ensemble of tracks in the ID, with an associated large depositions in the calorimeter system.

The jet reconstruction [110] is based on topological energy clusters in the calorimeter, following a similar procedure as for proto-clusters in the electron reconstruction (see Section 4.4.1). Clusters are then merged together into jets via the *anti-k_t* algorithm [111], using the FastJet 3.2.2 software [112]. The anti-k_t algorithm is based on clustering elements based on distances as defined in Equation 4.4.

$$\begin{aligned} d_{ij} &= \min\left(k_{ti}^{-2}, k_{tj}^{-2}\right)\frac{\Delta_{ij}^2}{R^2} \\ d_{iB} &= k_{ti}^{-2} \end{aligned} \tag{4.4}$$

In Equation 4.4 the distance between the i-th and j-th elements d_{ij} depends on their transverse momenta $k_{ti/j}$, their geometrical distance Δ_{ij} and the dumping radius R. For each couple of elements the algorithm computes d_{ij} and d_{iB} and only takes the minimum. In case d_{ij} is the minimum, the two elements are merged and the algorithm is iterated over the resulting set of elements. The algorithm stops merging when d_{iB} is lower than any possible d_{ij}. The algorithm tends to merge the soft components with the hard ones in a cone of radius R, while it keeps distant components separated. This procedure results in circular-shaped jets that are experimentally very convenient to remove the overlap with neighbouring objects. In this analysis, $R = 0.4$ is used.

Since jets are very complex objects, many calibrations and corrections are applied [110]. Firstly, pile-up corrections are applied to remove the contribution from other interactions, coming from the same or nearby bunch crossings. After that, a MC-based correction is applied to calibrate the measured jet four-momentum to the particle-level jet four-momentum from simulated dijet events. The jet kinematics are also calibrated against tracking and calorimeter information to reduce the dependence on the flavour of the initial parton, since different flavours yield different properties of the jet in the detector.

Eventually, a set of *in-situ* calibrations is performed to calibrate the MC prediction to observed data [110]. These calibrations aim to reduce the discrepancies between jets in simulation and data that come from a wide spectrum of effects: from detector imperfections to physics effects in the hadronic shower. The jet energy scale (JES) is calibrated to correct the simulated jet energy to data as a function of the jet kinematics. The first calibration step takes into account the inhomogeneity of the detector response over η, restoring the same response for central ($|\eta| < 0.8$) and forward ($0.8 < |\eta| < 4.5$)

regions. Afterwards, other calibrations are based on Z+jets, γ+jets and multi-jet events and they calibrate the recoil jet against the well-calibrated central Z, γ or jet. The use of different objects for the calibration allows to span over different ranges of jet kinematics. The calibration of the jet energy resolution (JER) is performed with similar analyses by using jets recoiling against a well-calibrated object (like Z boson), or by balancing the jet p_{T} in dijet events.

To further reduce the effect of pile-up, a *jet vertex tagger* (JVT) [113] has been developed for the Run 2. This tool provide a likelihood discriminant based on the compatibility of the jet tracks with the primary vertex, providing an estimation of the fraction of the tracks in the jet coming from the hard scattering. A specific calibration is derived in tag-and-probe analysis using Z+jets events [113] by looking at the angular separation of the jet components from the Z boson.

4.4.4. Bottom quark induced jets

Since the top quark decays almost 100% of the times into a W boson and a b-quark, the identification of b-quark induced jets (b-jets) is particularly important in top quark physics for signal-to-background discrimination. The b-jet identification algorithms use unique properties of b-quarks to tag a jet. First of all, b-quarks produce jets containing B hadrons, with a mean lifetime up to 1.5 ps [11], enough to travel a visible distance before the decay. Due to their large mass, they tend to produce decay products higher p_{T}. The decay products are usually not stable inducing decay chain that gives a larger particle multiplicity. These features are exploited by the b-jet tagging (or briefly b-tagging) algorithms.

The algorithm used in this analysis is *MV2c10* [114, 115]. It is a *Boosted Decision Tree* (BDT) discriminant developed in the ROOT Toolkit for Multi Variate Analysis (TMVA) [116]. The input variables to the algorithm come from three low-level algorithms: the IP2/IP3D algorithm, [114] looking at the jet impact parameter; the SV1 tagger [117], identifying the presence of a displaced secondary vertex and the JetFitter algorithm [118], that uses the structure of the jet to reconstruct the decay chains of b- and c-hadrons. The output of the tagger provides a score within $[0, 1]$, peaking at one for true b-jets and at zero for light quark and gluon jets. Four operating points are identified by four different cuts on the score distribution, corresponding to four different efficiency points for true b-jets. The four cuts provide five bins in the distribution, used to estimate how much a jet is b-jet-like, the so-called *pseudo-continuous* (PC) b-tagging. The working points are listed in Table 4.2.

Jets passing each operating point have been calibrated with tag-and-probe techniques on $t\bar{t}$ events. Each value of the pseudo-continuous b-tagging is also calibrated following the same procedure. Dedicated efficiency scale factors have been extracted [115] to cover the discrepancy between data and simulation.

Name	b-jets efficiency	Cut value	PC value
MV2c10 60%	60%	≥ 0.92	5
MV2c10 70%	70%	≥ 0.79	(< 0.92) 4
MV2c10 77%	77%	≥ 0.58	(< 0.79) 3
MV2c10 85%	85%	≥ 0.05	(< 0.58) 2
			(< 0.05) 1

Table 4.2.: List of the available operating points for MV2c10, with the corresponding efficiency on true b-jets, cut on the discriminant distribution and pseudo-continuous (PC) value. The PC value for each working point means that the jet is tagged by that working point but not by the tighter one. For instance, a jet with PC equal to four, is tagged by MV2c10 70% but not by MV2c10 60%, meaning a value of the discriminant within $[0.79, 0.92]$.

4.4.5. Missing transverse energy

A hard scattering in pp collisions involves partons inside the proton structure. The fraction of proton momentum of each parton follows the PDF (see Section 1.1) and the value of total initial state momentum in the direction of the beam is unknown. However the transverse momentum of the initial state is expected to be zero. Conservation of momentum in the transverse plane is used to evaluate the invisible component from the hard scattering, the missing tranverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$. In the SM, only neutrinos are sources of missing transverse energy, but many BSM theories predict low-interacting or sterile particles that may result in an excess of measured $E_{\mathrm{T}}^{\mathrm{miss}}$ [119].

The calculation and calibration of $E_{\mathrm{T}}^{\mathrm{miss}}$ needs, first, the reconstruction of all the other physics objects in the event, the *hard objects*, and the identification of all the tracks in the event to evaluate the *soft object* component. Hard objects consist of electrons, muons, hadronically-decaying taus (τ_{had}) [120], photons [104] and jets originating from the same interaction point. Since calorimeter-based objects, such as electrons, photons and jets, may overlap, a dedicated overlap-removal procedure is performed to avoid double-counting of the energy. The soft component consists of well-identified tracks reconstructed in the ID coming from the interaction point, but not associated to any object. Those may arise from underlying events associated with the hard interaction or diffuse particles from pile-up interactions.

The vector $\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} = (E_x^{\mathrm{miss}}, E_y^{\mathrm{miss}})$ is defined from the transverse momenta $\boldsymbol{p} = (p_x, p_y)$ of all the reconstructed objects, as in Equation 4.5.

$$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\text{electrons}} \boldsymbol{p}_{\mathrm{T}}^{e} - \sum_{\text{muons}} \boldsymbol{p}_{\mathrm{T}}^{\mu} - \sum_{\substack{\text{hadronic}\\ \text{taus}}} \boldsymbol{p}_{\mathrm{T}}^{\tau_{\mathrm{had}}} - \sum_{\text{jets}} \boldsymbol{p}_{\mathrm{T}}^{\mathrm{jet}} - \sum_{\substack{\text{soft}\\ \text{components}}} \boldsymbol{p}_{\mathrm{T}}^{\mathrm{soft}} \tag{4.5}$$

From the definition in Equation 4.5, the scalar missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ and

its corresponding azimuthal angle ϕ^{miss} are computed:

$$E_{\text{T}}^{\text{miss}} = |\boldsymbol{E}_{\text{T}}^{\text{miss}}| \tag{4.6}$$

$$\phi^{\text{miss}} = \tan^{-1}\left(\frac{E_y^{\text{miss}}}{E_x^{\text{miss}}}\right) \tag{4.7}$$

Dedicated calibrations aim to extract $E_{\text{T}}^{\text{miss}}$ scale and resolution [119]. $Z \to \mu^+\mu^-$ events are used to check the reconstruction in the absence of sources of the missing energy, while $W \to e\nu$ or $W \to \mu\nu$ events provide a test of the reconstruction performance with a well-defined neutrino in the final state. Finally, $t\bar{t}$ samples are used to validate the reconstruction at higher jet multiplicity.

4.4.6. Overlap removal

The procedure of overlap removal is applied to prevent a given signature in the detector to be reconstructed as multiple physics objects. For example a large calorimeter deposition from an electron is also reconstructed as a jet signature. This procedure is performed on non-isolated and loosely-identified leptons and consists of the following steps[4]:

- Electron with a track within $\Delta R < 0.2$ of another electron is removed.
- An electron is removed when sharing an ID track with a muon.
- Jets within $\Delta R < 0.2$ from an electron are removed. In case multiple jets are within the cone, only the closest one is removed.
- Electrons within $\Delta R < 0.4$ from a jet are removed.
- Jets within $\Delta R < 0.2$ from a muon and made of less than three tracks or with a p_{T} compatible with a muon are removed.
- Muons within $\Delta R < \min\left(0.4, 0.04 + 10\ \text{GeV}/p_{\text{T}}\right)$ from jets are removed.

The sliding ΔR window for the muon-to-jet rejection aims to suppress low p_{T} muons from decay chains.

[4] Rejected objects are not considered in the later steps.

CHAPTER 5

Analysis strategy for the search for the $pp \to t\bar{t}t\bar{t}$ process

The $t\bar{t}t\bar{t}$ production process is experimentally a challenging process to measure since it is rare and lies in a complicated phase space, not only for a precise measurement but even for observation. Therefore, the measurement of the $t\bar{t}t\bar{t}$ relies on a solid understanding of the background and on separating the signal signature from the dominant background processes.

A peculiarity of the $t\bar{t}t\bar{t}$ production process is the larger number of particles in the final state with respect to background processes such as $t\bar{t}$ and $t\bar{t}+X$. In particular, the jet and b-jet multiplicity, for a fixed number of leptons in the final state, is expected to be larger. The b-tagging information is indeed expected to play a crucial role in the signal-to-background discrimination. The high threshold for the production of $t\bar{t}t\bar{t}$ ($\sqrt{\hat{s}} \gtrsim 4m_{\text{top}}$) requires similar momentum from both the colliding partons, giving more central events. Therefore all the distributions for transverse kinematic variables (such as jet, lepton p_{T} etc.) are harder than in the background, despite the produced top quarks are expected to be (on average) less energetic with respect to top quarks in $t\bar{t}(+X)$ events.

As introduced in Section 1.3, many decay modes of the $t\bar{t}t\bar{t}$ final state are possible and are classified by the number of leptons in the final state. This classification is used to identify the analysis channels for the $t\bar{t}t\bar{t}$ measurement, since a different number of leptons in the final state leads to different background compositions. In this analysis, the four-lepton and all hadronic channels are not considered: the four lepton channel because of its very low branching ratio ($\approx 1\%$) and the all hadronic channel because of the very challenging environment, especially for triggering and background modelling. Based on background composition, the analysis is divided into two separate main channels:

the combination (1LOS) of the single lepton (1L) and opposite-sign dilepton (OS2L) channels and the combination (SSML) of the same-sign dilepton and trilepton channel combined. The main challenge for the 1LOS channel is the modelling of $t\bar{t}$ process associated with large number of jets, especially originating from heavy-flavour quarks. The SSML channel instead has to properly deal with instrumental background in a busy environment and it requires modelling the $t\bar{t}V$ production associated with heavy-flavour jets. These two channels are then combined to improve the sensitivity to $t\bar{t}t\bar{t}$ signal.

Some concepts are introduced prior to the sections describing the analysis strategy. An introduction to Multi Variate Analysis (MVA) techniques with a focus to Boosted Decision Trees (BDTs) is given in Section 5.1. The statistical approach of *profile likelihood fit* is described in Section 5.2. The analysis strategy for the 1LOS channel, that is the main topic of this thesis, is described in Section 5.3. An overview of the analysis strategy in the SSML is also given Section 5.4.

5.1. Multivariate analysis

Particle physics represents an ideal playground for the application for multivariate analysis (MVA) techniques. In case of the search for new processes, like the $t\bar{t}t\bar{t}$ production, a very tiny signal must be separated from a large number of background events that have similar but not identical features. MVA techniques use the information from many variables to identify patterns, and, in case of more sophisticated techniques (Deep Neural Networks), to reconstruct their correlations. Boosted Decision Trees (BDTs), that have been extensively used in high energy physics, are also used for this analysis. A decision tree usually targets two types of problem:

- *classification*: the algorithm has to estimate a discrete value from the event information, in other words, it has to classify the event.
- *regression*: the algorithm has to estimate a continuous value from the event information (for example the price of a house from its specifics).

The goal in the presented analysis is to separate the *signal* $t\bar{t}t\bar{t}$ from the *background* (classification problem) events and to do so the BDT produces a continuous score peaking at two different values for the two typologies of events. The produced score is then used as discrimination variable.

The output of the BDT can be interpreted as follows. Every event is described by a set of N input variables and therefore corresponds to a point in an N-dimensional space. The BDT aims at finding patterns in this space to separate the different classes, in this case signal and background. To do so, a *training* of the architecture is needed. The training procedure uses correctly classified events as input to populate the space and learn the patterns of the different classes. The learning process consists of dividing the input variable space in smaller and smaller fragments via recursive cuts on the values of the variables (*nodes*), resulting in a *tree* structure. Every recursive segmentation is

decided by maximising the gain of information, quantified by the *loss function*. The loss function models the improvement of the knowledge of the state from the separation of the considered portion of the space into sub-parts. An example of loss function is the *Gini impurity*, that for a classification in two classes (A and B) is equal to:

$$I_G(p_A, p_B) = \sum_{i \in [A,B]} p_i(1 - p_i) = 1 - p_A^2 - p_B^2 \tag{5.1}$$

where $p_{A/B}$ are the estimated probabilities for being classified as A or B in the given segmentation. From Equation 5.1, I_G is minimised to zero when all the events are assigned to the same label. The iteration stops when either a lower limit of training events is reached in the segmentation (*leaf*), or a maximum depth of the tree is reached or no further cut can improve the knowledge of the state[1]. The evaluation of the BDT associates a score to non-classified input events. This score is the probability for the different classes in the leaf in which the event falls.

The performance of the algorithm is improved further by *boosting*: this indicates that multiple trees are generated by boosting the decision towards a better classification. Typically the boosting of the algorithm is obtained by successive training and evaluation on training events to test the classification of the BDT and to correct it, for example by giving larger weights to events that are more difficult to classify. Different types of boosting are available and usually the final score corresponds to a weighted average of the outputs of all the generated trees. In this analysis the used BDTs provide a continuous output in the range $[-1, 1]$, peaking at one for $t\bar{t}t\bar{t}$ events and the opposite for background events.

Generally, a fine tuning of the BDT parameters allows better and better separation of the classes. The *hyper-parameters* are, for example, the maximum depth, the number of generated trees, the learning rate in the boosting procedure, and the number of scanned cuts for each node. However, this may lead to *overtraining*, meaning that the BDT learns "too well" the patterns in the limited number of training events, resulting in a worse performance when different samples are evaluated. To check overtraining, the training samples are divided in k subsets, and different BDTs are trained on each subset. Then the i-th BDT is evaluated on a j-th subset ($j \neq i$) to check whether compatible responses from all the trained BDTs are obtained. Some parameters of the BDT also help avoiding overtraining, such as the *learning rate* that manages convergence of the boosting procedure. Another way to limit overtraining is the *bagging* technique, causing a drop of a fraction of the training events in different consecutive trainings.

5.2. Profile Likelihood fit

A *profile likelihood fit* [121] is used to extract the best estimation of a *parameter of interest* (POI) of a model from measured data. This method is particularly important in

[1] It occurs when either only a single class is in a leaf or the classes are uniformly distributed.

case that the systematic uncertainties on the model have a large impact on the prediction. The effects on the prediction of the systematic uncertainties are indeed included in the fit as extra model parameters to be measured from data, the *nuisance parameters* (NPs).

To explain the profile likelihood concept, a fit in a single region with a single source of background is considered. A single distribution of a chosen discriminating variable is considered in the selected region. The probability density distribution (*pdf*) for a process X of an event x is denoted as $f_X(x)$. The expected number of events from signal is denoted by S, while B is used for background. The POI is the *signal strength*, that is the ratio of the observed signal on the expectation. The probability of observing events $\{x_1, ..., x_n\}$ given the signal strength μ is:

$$P(\{x_1, ..., x_n\}|\mu) = \text{Poiss}(n|\mu S + B) \left[\prod_{e=1}^{n} \frac{\mu S f_S(x_e) + B f_B(x_e)}{\mu S + B} \right] \tag{5.2}$$

In case of a measurement $\{x_1, ..., x_n\}$ is fixed. Therefore, this probability depends only on μ and it is interpreted as the probability of having μ inducing the measured data. This takes the name *likelihood function* $L(\mu)$. The maximum of $L(\mu)$ provides an estimator for μ given the measurement $\{x_1, ..., x_n\}$. This maximisation problem is usually converted into the minimisation of its negative logarithm:

$$-\ln L(\mu) = (\mu S + B) + \ln n! - \sum_{e=1}^{n} \ln \left(\mu S f_S(x_e) + B f_B(x_e)\right) \tag{5.3}$$

Usually *pdf*s are binned in histograms. Therefore the product over the events is replaced by a product over the histogram bins $b \in [1, N]$, and the *pdf*s are expressed in terms of bin contents ν_b. The modified likelihood is:

$$P(n|\mu) = \text{Poiss}(n|\mu S + B) \left[\prod_{b \in \text{bins}} \frac{\mu \nu_b^S + \nu_b^B}{\mu S + B} \right] = \prod_{b \in \text{bins}} \text{Poiss}(n_b|\mu \nu_b^S + \nu_b^B) \tag{5.4}$$

Moving towards a realistic scenario, a fit with multiple regions (indicated with index r) and multiple models (indicated with index s) is considered. Anything that causes a change of the distributions in the regions is parametrised with a NP. The variation of the NP can be constrained or unconstrained. An unconstrained NP that causes an overall change of a model normalisation, such as the POI, is indicated with ϕ_s. The statistical uncertainties of the model, that are constrained NPs, change the shape of the distribution (bin-by-bin) and are indicated with γ_{brs}. Systematic uncertainties are usually parametrised as constrained NPs ($\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_P\}$) that may cause both normalisation ($\eta_{rs}(\boldsymbol{\alpha})$) and shape ($\sigma_{brs}(\boldsymbol{\alpha})$) variations of the model. The luminosity uncertainty, instead, equally affects all the considered models (λ). In terms of these parameters, the content of the bin b in the region r is written as:

$$\nu_{br}(\phi, \alpha, \gamma) = \sum_s \lambda \gamma_{brs} \eta_{rs}(\boldsymbol{\alpha}) \phi_{rs}(\boldsymbol{\alpha}) \sigma_{brs}(\boldsymbol{\alpha}) \tag{5.5}$$

The likelihood of observing n events (intended as collection of events per bin n_{br}) and measuring $\boldsymbol{a}$ values for the NPs $\boldsymbol{\alpha}$ becomes:

$$P(n, \boldsymbol{a}|\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \prod_r \prod_b \text{Poiss}(n_{br}|\nu_{br})\text{Gauss}(L_0|\lambda, \Delta_\lambda) \prod_{p\in\alpha} f(a_p|\alpha_p) \tag{5.6}$$

To make Equation 5.6 useful for the estimation of the POI, a proper treatment of $\boldsymbol{\alpha}$ is needed. The distribution $f(a_p|\alpha_p)$ is usually assumed to be a normal Gaussian. The effect of the variation of a given α_p 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 of α_p is obtained by linear interpolation. Eventually the likelihood depends only on the parameters of the model, the NPs and the POI, and it is maximised to estimate their values from data.

A final remark is needed for the inclusion of MC statistical uncertainty in the likelihood. In this thesis, a single bin-by-bin γ parameter is used, resulting in:

$$\text{Poiss}(n_b|\gamma_b\nu_b(\boldsymbol{\alpha})) \tag{5.7}$$

where γ_b NP takes into account the overall statistical uncertainty of the prediction in the bin. This is an approximation, since the correct procedure would be to associate a NP to the statistical fluctuation of each model s for each bin b, γ_{brs}. This is avoided to limit the number of NPs used in the fit.

5.3. Single lepton and opposite- sign dilepton channel

The 1LOS channel accepts almost 60% of $t\bar{t}t\bar{t}$ events, but covers a phase-space that is highly dominated by $t\bar{t}$ with many radiated jets. The $t\bar{t}t\bar{t}$ signal has a distinct signature in the single lepton (opposite-sign dilepton) channel: 10 (8) jets are expected in the final state, of which 4 are expected to be b-jets. The high jet multiplicity, rich in heavy-flavour jets, helps in the signal-to-background separation, since background processes can populate these regions only with high jet radiation and the occurrence of $g \to b\bar{b}$ splitting. A scheme of the signature of the different background sources compared to $t\bar{t}t\bar{t}$ is shown in Figure 5.1. In addition, the two neutrinos in the final state in the opposite-sign dilepton channel can give a significant $E_{\text{T}}^{\text{miss}}$, especially with respect to events containing *fake* leptons[2].

5.3.1. Regions definition

A pre-selection is applied to data and simulated events to define the phase-space regions for the analysis. The choice of the trigger is the first step of the selection. In this channel,

[2]Here a fake lepton denotes a jet that is reconstructed as a lepton or as a real lepton that does not come from the hard scattering.

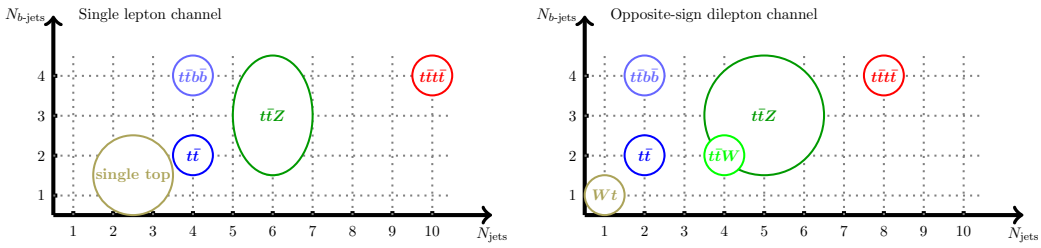


Figure 5.1.: Summary of signatures for $t\bar{t}t\bar{t}$ and main background processes at LO in terms of jets and b-jets multiplicity for the single lepton (a) and opposite-sign dilepton (b) channels.

only events that have fulfilled *single-lepton* trigger conditions are used, requiring at least one lepton in the event. The triggered lepton must also pass certain conditions in terms of p_{T}, isolation and identification imposed in HLT, that can be extracted directly from the trigger name, following the convention:

`HLT_<particle><pT-cut>_<id>_<iso>_<L1>`

where:

- `<particle>` indicates whether the trigger is for muons (`mu`) or electrons (`e`).
- `<pT-cut>` gives the p_{T} threshold (in GeV) applied to the lepton.
- `<id>` stands for the applied identification criteria. For example `lhmedium` indicates a medium identification required on the likelihood discriminant. In case the label `nod0` is present, no requirements on the transverse impact parameter (d_0) are applied[3]. If absent, no identification is applied.
- `<iso>` stands for the applied isolation criteria. If absent, no isolation is required.
- `<L1>` indicates the triggered L1 stream.

Trigger conditions keep a balance between efficiency and rate, in order to stay within the maximum bandwidth and trigger rate allowed by the system. Trigger masks change between data-taking periods to adapt to the different pile-up conditions, and therefore are looser in 2015 than in the other years. In the same way, triggers require a looser isolation or identification the higher the cut on p_{T} of the object is. The list of the used trigger menu is shown in Table 5.1. The events are then categorised into two sub-channels, single lepton (1L) and opposite-sign dilepton (OS2L). Table 5.2 shows a summary of the applied selections. Finally events are divided into regions with different jet and b-jet multiplicities.

[3] d_0 stands for the signed closest distance of the track to the $z-$axis.

2015	2016-2018
HLT_e24_lhmedium_L1EM20VH	HLT_e26_lhtight_nod0_ivarloose
HLT_e60_lhmedium	HLT_e60_lhmedium_nod0
HLT_e120_lhloose	HLT_e140_lhloose_nod0
HLT_mu20_iloose_L1MU15	HLT_mu26_ivarmedium
HLT_mu50	HLT_mu50

Table 5.1.: Summary table of the single lepton triggers used in the analysis in the 1LOS channel for all the data-taking periods of Run 2.

	Single lepton ch.		Opposite-sign dilepton ch.		
	e+jets	μ+jets	$e^{\pm}e^{\mp}$	$\mu^{\pm}\mu^{\mp}$	$e^{\pm}\mu^{\mp}$
Lepton p_{T}	≥ 28 GeV		≥ 28 GeV (leading) ≥ 10 GeV (sub-leading)		
N_{jets}	≥ 5		≥ 2		
Jet p_{T}	≥ 25 GeV		≥ 25 GeV		
$N_{b\text{-jets}}$	≥ 2 (MV2c10 77%)		≥ 2 (MV2c10 77%)		
M_{ll}			$M_{ll} > 15$ GeV $\lvert M_{ll} - m_Z \rvert > 10$ GeV		
Other	Veto on second *loose* lepton		Veto on third *loose* lepton		

Table 5.2.: Summary of the pre-selection conditions applied on data and simulated events for the 1LOS channel. The *loose* leptons are obtained by requiring looser criteria on isolation and identification than in Section 4.4.

A scheme for the region definition is shown in Figure 5.2, inspired by the previous ATLAS analysis, performed on a dataset of 36 fb^{-1} [32]. This definition is motivated by the change of the background composition at different number and flavour of the radiated jets. Regions with low jet and b-jet multiplicity are defined as *background model derivation* regions. The two methods for the background estimation are based on these regions, as described in Chapter 7. The *control regions* are dominated by the $t\bar{t}$+jets contribution and are used to consolidate the background estimation. These are also important in the profile likelihood fit to constrain the background modelling. The $t\bar{t}t\bar{t}$*-enriched* regions are the ones most sensitive to $t\bar{t}t\bar{t}$ signal. More details on the fit strategy are given in Section 5.3.3. The background composition in the *control* and $t\bar{t}t\bar{t}$*-enriched* regions is shown in Figures 5.3a and 5.3b. The dominance of $t\bar{t}$ motivates not only the dedicated effort for its estimation, but also the focus on $t\bar{t}$ features when signal-to-background separation is studied. Indeed, only $t\bar{t}$ is evaluated for the optimisation of the performance of the analysis.

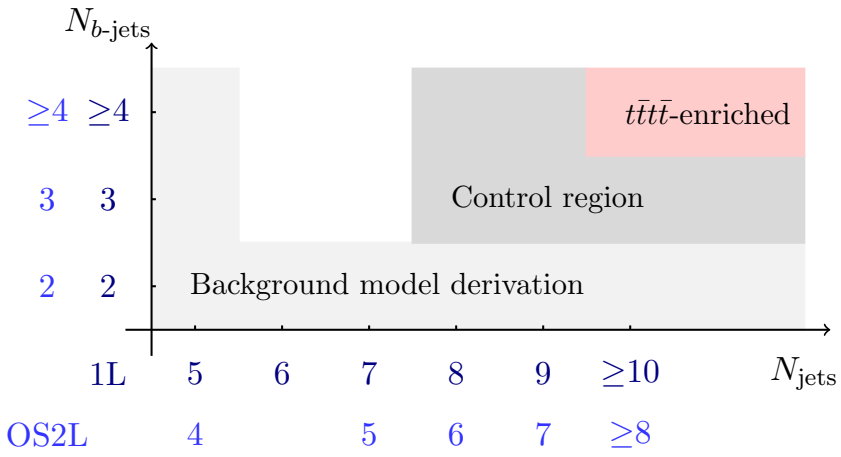
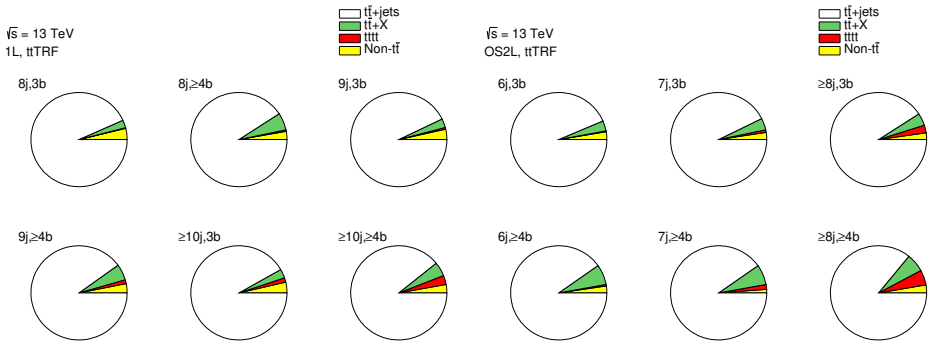


Figure 5.2.: Region categorisation in the 1LOS channel.

5.3.2. Signal discrimination from background

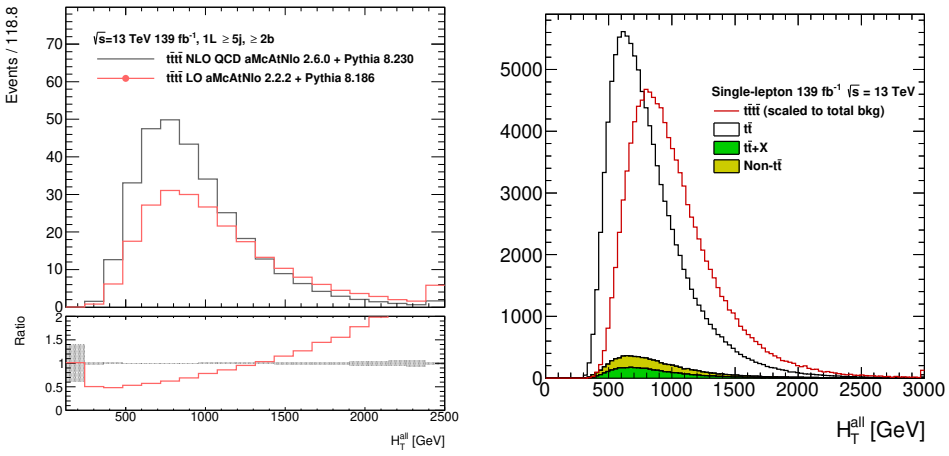
The $H_{\text{T}}^{\text{all}}$ variable was used in the previous $t\bar{t}t\bar{t}$ search [32] to discriminate signal from background. It is defined as the sum of the p_{T} of leptons and jets in the event and provides large separation between $t\bar{t}$+jets and $t\bar{t}t\bar{t}$. This is expected since objects in $t\bar{t}t\bar{t}$ events are expected to be harder and their multiplicity higher. The analysis used a simulation for $t\bar{t}t\bar{t}$ at LO QCD with MadGraph 2.2.2 [75] with the NNPDF2.3LO PDF set [122], interfaced to Pythia 8.186 [79] using the A14 set of tuning parameters [80] and with a different choice of renormalisation and factorisation scales from $H_{\text{T}}/4$. A similar separation is not observed with the setup used in this analysis, as shown in Figures 5.4a and 5.4b. The difference has been found to mainly come from the different PDF settings and software release [123]. Indeed the LO simulation with the current setup results to be much closer to the current NLO prediction with respect to the previous simulation. Given the lower separation with the current setup, this variable has been dropped as discriminating variable.

A more complex discriminant has been created by making use of a Boosted Decision Tree (BDT) available in the ROOT Toolkit for Multi Variate Analysis (TMVA) [116]. The classification consists of separating $t\bar{t}t\bar{t}$ events from $t\bar{t}$+jets events. The training sample for $t\bar{t}t\bar{t}$ is the MC simulation at LO QCD described in Section 4.3.1. The training sample for $t\bar{t}$+jets consists of the simulation described in Section 4.3.1 corrected by the reweighting method (see Section 7.2). The choice of input variables is driven by



(a) (b)

Figure 5.3.: Background composition in the 1LOS channel: six regions at increasing number of jets and b-jets for both single lepton (a) and opposite-sign dilepton (b) channels. Regions are labelled as XjYb, indicating a region that requires for X jets and Y b-jets. The label $t\bar{t}$+X stands mostly for $t\bar{t}$+$Z/W/H$, while *Non-*$t\bar{t}$ events include all the processes without a top quark pair in the final state (single-top production, W/Z+jets, diboson). The estimation of the $t\bar{t}$+jets background here is obtained with the t$\bar{\text{t}}$TRF method, described in Section 7.1



(a) (b)

Figure 5.4.: Comparison of $H_{\mathrm{T}}^{\mathrm{all}}$ spectra between LO prediction used in [32] with NLO QCD prediction used in this thesis after detector simulation (a). Separation of $t\bar{t}t\bar{t}$ from the background with the latest setup (b). Both plots refer to the single lepton channel in a region that requires for five or more jets and two or more b-jets. Both $t\bar{t}$+jets predictions are based on MC simulations.

underlying physics and separation of the distributions[4]. The variables are grouped in the following categories:

- *Activity in the transverse plane*: $t\bar{t}t\bar{t}$ is expected to be more central than $t\bar{t}$, therefore harder transverse kinematics are expected.
- *Flavour-tagging information*: $t\bar{t}t\bar{t}$ has more b-jets in the final state. Therefore, on average, jets from $t\bar{t}t\bar{t}$ are more b-jet-like.
- *Jet information*: the topology of jets in the final state is expected to be different in $t\bar{t}t\bar{t}$, where top-decay topology can be reconstructed, and in $t\bar{t}$+jets where most of the jets come from initial and final state radiation.
- *Reclustered-jets (RC) information*: RC-jets are large-R jets obtained by running the anti-k_t algorithm on jets in the final state by using a larger dumping radius (in this case $R = 1.0$), aiming at the reconstruction of boosted-top topology. Less boosted top quarks are expected in $t\bar{t}t\bar{t}$ events than in $t\bar{t}$+jets events. However, since $t\bar{t}t\bar{t}$ is more central, a larger fraction of them is expected to fall in the detector acceptance. The sum of the two effects results on $t\bar{t}t\bar{t}$ having a larger RC-jet multiplicity than $t\bar{t}$+jets.
- E_T^{miss} *and lepton information*: in the opposite-sign dilepton channel, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is expected to be larger with respect to semileptonic $t\bar{t}$ events with a *fake* lepton. In the single-lepton channel, lepton p_{T} and $E_{\mathrm{T}}^{\mathrm{miss}}$ can be used to reconstruct the transverse mass of the parent top quark.

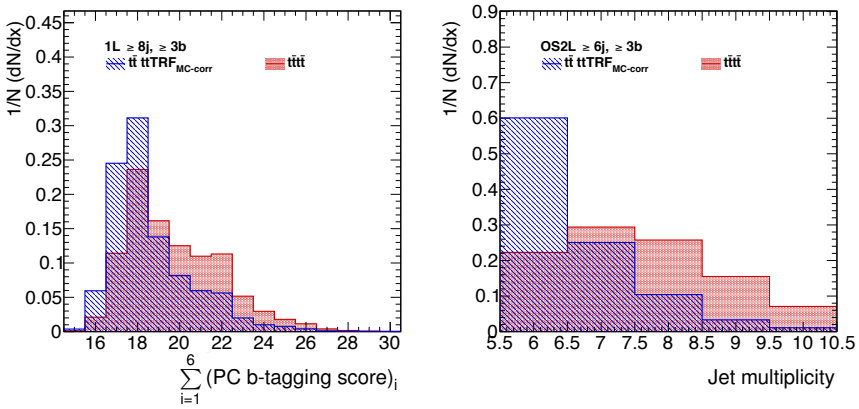
The set of input variables for the BDT is shown in Table 5.3. Every variable has an associated importance that is measured as the drop in the separation of the output distributions when the considered variable is excluded from the training procedure. The distributions for the three highest ranked variables are shown in Figures 5.5a to 5.5f; note that the variables with highest separation are not necessarily more important, since also the correlations among variables play a significant role.

Since $t\bar{t}t\bar{t}$ and $t\bar{t}$+jets processes are decently separated in many of that variables. No refined optimisation of the BDT hyper-parameters has been performed and most of the parameters have been set to default values. It uses *gradient* boosting with a learning rate of 0.05, generating 600 trees. Thirty cuts are allowed in each node and a maximum depth of two is chosen. The BDTs is trained in *semi-inclusive* regions, i.e. a different training is performed at different jet multiplicities, but inclusively in number of b-jets to have enough statistics for the training procedure. A 3-fold validation procedure is used: events are separated into three subsets and a different BDT is trained on each.

[4] The separation S between two distributions, binned in N bins, is

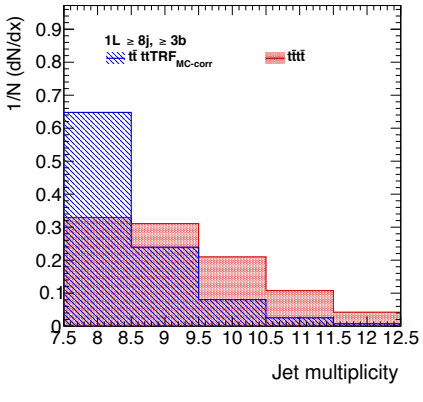
$$S = \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i^A - y_i^B)^2}{y_i^A + y_i^B}$$

where y_i^X is the i-th bin content for the distribution X.

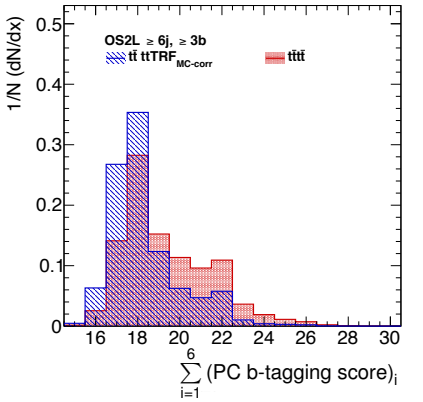


(a) First ranked (1L)

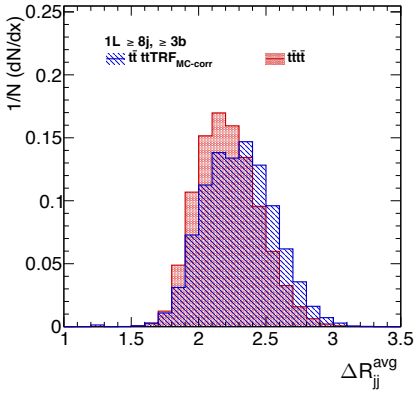
(b) First ranked (OS2L)



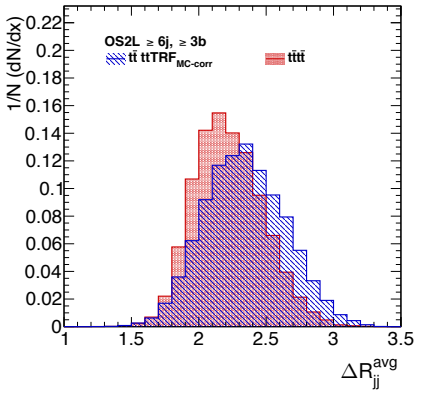
(c) Second ranked (1L)



(d) Second ranked (OS2L)



(e) Third ranked (1L)



(f) Third ranked (OS2L)

Figure 5.5.: Comparison of the distributions for $t\bar{t}$ and $t\bar{t}t\bar{t}$ for the most important variables in the BDT training in single lepton (1L, left column) and opposite-sign dilepton (OS2L, right column) channels. The variables are displayed from top to bottom in order of importance. The distributions for $t\bar{t}$ are obtained with the t$\bar{\text{t}}$TRF method, described in Section 7.1.

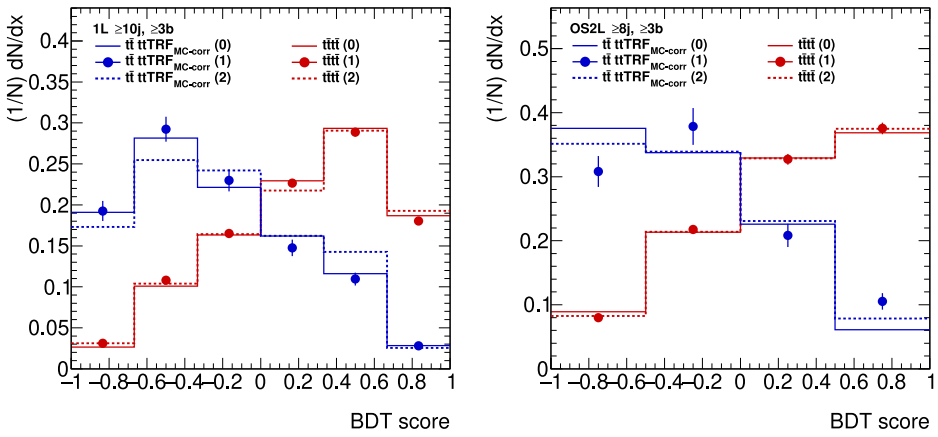


Figure 5.6.: Separation plots for single lepton (a) and opposite-sign dilepton (b) channel. The $t\bar{t}$ estimation is given by t$\bar{\text{t}}$TRF method, described in Section 7.1.

Testing consists in applying a BDT trained on one subset to another one to check possible overtraining.

The distribution of the output variable is shown in Figures 5.6a and 5.6b. In these plots, the BDT is evaluated on t$\bar{\text{t}}$TRF prediction of $t\bar{t}$+jets (see Section 7.1), based on data events and not used in the training procedure. The fluctuations of performances on the $t\bar{t}$+jets are indeed due to the low statistics of the individual set of events. The three distributions are used to quote an uncertainty on the separation: it does not indicate the actual uncertainty on the BDT performance, but rather a limit for the performance optimisation due to data statistics. The separations are 0.20 ± 0.02 and 0.22 ± 0.03 for single lepton and opposite-sign dilepton, respectively.

5.3.3. Fit strategy

The remaining choice for setting up the profile likelihood fit (see Section 5.2) is the definition of the regions. The *control regions* have a dominant contribution from the background and no significant signal contamination. These regions are used in the fit to constrain the NPs associated to the background. The *signal regions* instead have a significant contribution from $t\bar{t}t\bar{t}$ signal and drive the estimation of the signal strength in the fit. Note that the definition of the role of a region in the fit is different from the categorisation of the regions in Section 5.3.1. In case of a distribution that strongly separates signal and background, some bins of the distributions can be classified as control regions, and others as signal regions, independent from their categorisation in Figure 5.2. The prediction of the background in the control regions is checked against

List of the input variables for the classification BDT		
Name	Importance	Short description
Activity in the transverse plan (computed on all the particles in the final state)		
H_T^{all}	8 / 12	$\sum p_T$
Centrality	10 / 9	$H_T^{all}/\sum E$
N_j	2 / 1	Jet multiplicity.
$p_T^{\text{lead. jet}}$	4 / 5	Leading jet p_T.
Flavour- tagging information (computed on jets)		
$\sum_{i=1}^{6}$ PC$_i$	1 / 2	Sum of the six highest pseudo continuous (PC) b-tagging scores in the event.
Jets information		
$\Delta R_{jj}^{\text{avg}}$	3 / 3	Average ΔR between two jets.
$\Delta R_{\text{min}}^{bb}$	11 / 11	Minimum ΔR between two b-jets.
$\Delta R_{\text{min}}^{bl}$	6 / 10	Minumum ΔR between b-jet and lepton.
$M_{jjj}^{\Delta R_3^{\text{min}}}$	5 / 4	Tri-jets mass for ijk triplet of jets such as $\Delta R_3 = \sqrt{\Delta R_{ij}^2 + \Delta R_{ik}^2 + \Delta R_{jk}^2}$ is the minimum.
M_{bbb}^{min}	9 / 8	Minimum mass of three b-jets.
Reclustered-jets (RC-jets) information		
N_{RC}	16/ 14	RC-jets multiplicity with mass above 100 GeV
$\sum_{i\in\text{RC}} d_{12}^i$	15 / 6	Sum of d_{12} splitting scales of all the RC-jets in the event. This indicates the separation of the leading two substructures (if present) in the k_T algorithm: $d_{12} = \min(p_{T1}, p_{T2})\,\Delta R_{12}$.
$\sum_{i\in\text{RC}} d_{23}^i$	12 / 7	Sum of d_{23} splitting scale of all the RC-jets in the event. This indicates the separation of the second and third (if present) substructures in the k_T algorithm.
E_T^{miss} *and lepton information.*		
$E_{\text{T}}^{\text{miss}}$	13 / 15	Missing transverse energy.
M_{T}^W	7 / -	$E_{\text{T}}^{\text{miss}} + m_{\text{T}}^{\text{lepton}}$ (only in single lepton events).

Table 5.3.: Summary of input variables for the BDT. The more important is the variable in the BDT performance, the lower the value in the importance column. The "/" separates the importance in single lepton and opposite-sign dilepton channels.

Variable	Regions		Blinding condition
	1L	OS2L	
BDT	$8j^*b$	$6j^*b$	No blinding
BDT	$9j^*b$	$7j^*b$	BDT ≥ 0.7
BDT	$\geq 10j^*b$	-	BDT ≥ 0.5
BDT	-	$\geq 8j^*b$	BDT ≥ 0.2
Any kinematic variable	$\geq 10j \geq 4b$	$\geq 8j \geq 4b$	$H_{\mathrm{T}}^{\mathrm{all}} > 1$ TeV

Table 5.4.: Blinding strategy for the 1LOS channel. The notation $^*b$ stands for any b-jets multiplicity.

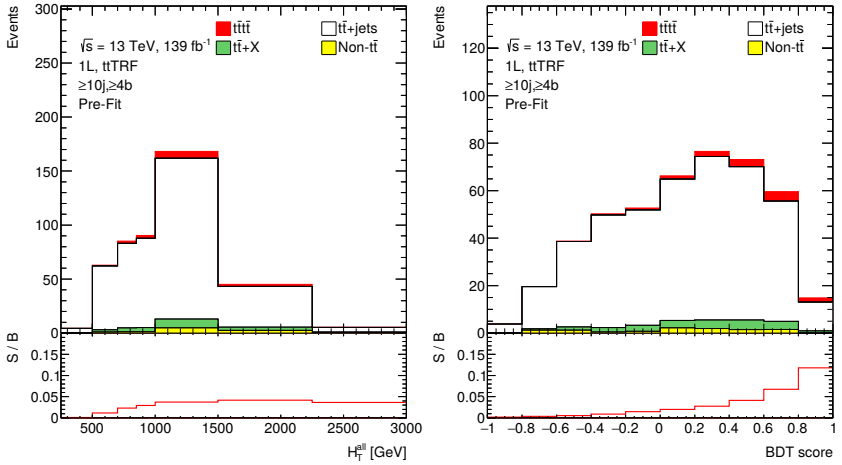
data to validate it. This is not possible in signal regions, that are kept *blinded*, i.e. data are never displayed before the measurement is fully validated. In the following, two conventions are adopted: the b-tagging working point is chosen to be 70% if not indicated otherwise, and the regions are expressed as $XjYb$, indicating the region with X jets and Y b-jets.

The fit is performed on the BDT discriminant in regions with ≥ 8 (≥ 6) jets and ≥ 3 b-jets in the final state (corresponding to *control* and *$t\bar{t}t\bar{t}$-enriched* regions in Figure 5.2). The inclusion of lower jet and b-jet multiplicity regions is motivated by the necessity to keep the background modelling at different jet and b-jet multiplicities under control. Since the BDT output has a larger separation power with respect to any single kinematic variable such as $H_{\mathrm{T}}^{\mathrm{all}}$, as shown in Figures 5.7a to 5.7d, in every region the distribution has no significant signal contamination for values close to -1, and is used as control region. The bins of the distributions close to 1 are instead richer in the signal and are used as signal region. Therefore, a dedicated blinding strategy is defined and shown in Table 5.4.

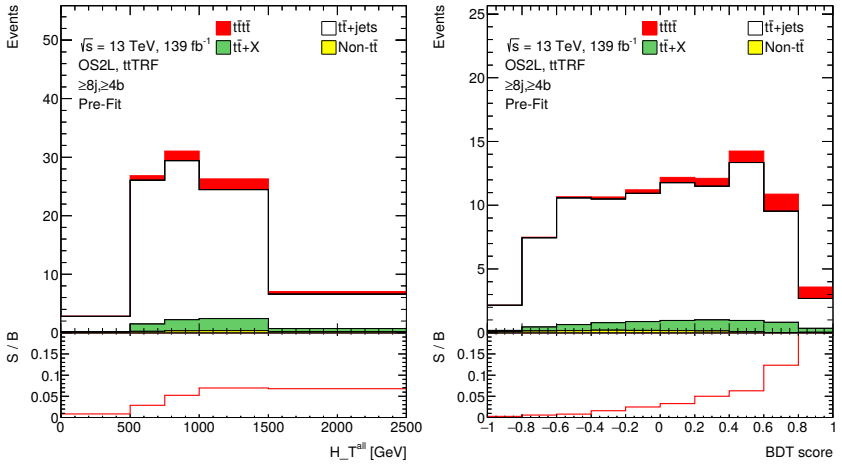
Checking the background prediction in the regions included in the fit is also important for the validation of the $t\bar{t}$ modelling. None of the kinematic variables have large enough separation to motivate a blinding in the non-$t\bar{t}t\bar{t}$-enriched regions. In *$t\bar{t}t\bar{t}$-enriched* regions, the modelling of kinematic variables is checked for events with $H_{\mathrm{T}}^{\mathrm{all}} < 1$ TeV, where the signal-over-background ratio is lower than 5%.

5.4. The dilepton same-sign and trilepton channel

The same-sign dilepton and trilepton channel (SSML) is the most sensitive channel for the $t\bar{t}t\bar{t}$ search. Instrumental background is an important source of background in this final state that is quite rare in SM processes. The main instrumental backgrounds are *charge mis-identification* and *fake or non-prompt leptons*. The former consists of $t\bar{t}$ dileptonic events with a lepton (mostly an electron) whose charge sign is mis-reconstructed.



(a) H_T^{all} (1L) (b) BDT score (1L)



(c) H_T^{all} (OS2L) (d) BDT score (OS2L)

Figure 5.7.: Comparison of signal-over-background (S/B) ratio for $H_{\text{T}}^{\text{all}}$ (left column) and BDT score (right column) variables in the two most signal populated regions in the single lepton (a,b) and opposite-sign dilepton (c,d) channels. The $t\bar{t}$ prediction is from the tt̄TRF method, described in Section 7.1. Uncertainties on the signal and background modelling are not displayed.

The latter arises from events with jets (usually from $t\bar{t}$ single- or di-leptonic events) mis-reconstructed as leptons by the detector (fake) or with real leptons from semileptonic decays (non-prompt). However, some SM processes, such as $t\bar{t}V$ production[5], can have the same signature as the signal, if associated with large jet and b-jet radiation. The MC predictions for these backgrounds are usually not validated in the regions of interest. Therefore, their estimation is also a challenge for the analysis in this channel.

As in the case of the 1LOS channel, the $t\bar{t}t\bar{t}$ signal is expected to have a higher jet and b-jet multiplicity with respect to the backgrounds. Moreover, SM processes with real leptons have a more significant $E_{\mathrm{T}}^{\mathrm{miss}}$ with respect to instrumental background since more neutrinos are produced in the final state.

Preliminary results of this analysis in the SSML channel have been released by the ATLAS Collaboration [36], and this section contains an overview of the analysis strategy. The event selection and categorisation is described in Section 5.4.1. The fit setup is described in Section 5.4.2.

5.4.1. Event selection and categorisation

A pre-selection is performed on data and simulated events to select the interesting phase space for the measurement.

The first applied condition is the trigger condition: the same single lepton triggers as in the 1LOS channel are also used in the SSML channel. The selected events are classified in terms of multiplicity of the leptons in the final state in same-sign dilepton and trilepon events and categorised in different flavour combinations. Trilepton events are vetoed if any reconstructed mass between two opposite-sign, same-flavour leptons fulfils $|m_{ll} - m_Z| < 10$ GeV to reject leptons coming from a Z^0 resonance.

The instrumental backgrounds are categorised based on their source, identified by using the particle level information of the simulated leptons. The instrumental background categories are summarised in Table 5.5. Most of the categories are estimated through MC simulation and their normalisations are measured in dedicated control regions included in the fit (see Section 5.4.2). The background coming from lepton charge mis-identification is estimated with a data-driven method. This method estimates the probability of mis-assigning the lepton charge on a sample of $Z \to ee$ events in data. This sample is obtained by requiring the lepton pair mass to be within a 10 GeV window around the Z boson mass, but no requirements are imposed on the lepton charges, so that the ratio between the yield of events with a same-sign lepton pair and the opposite-sign lepton pair event yield give an estimate of the charge mis-assignment probability. This probability is then used to weight the events with an opposite-sign dilepton pair in data passing the analysis selection, providing the estimate of the background coming from lepton charge mis-assignment.

[5]It stands for $t\bar{t}$ production associated with a vector W or Z boson.

Name	Label	Description
Physics	-	All leptons are prompt, coming from the hard interaction. In this case, the event is labelled as the simulated physics process.
Charge mis-identification	Q mis-id	Dilepton events with one prompt lepton and the other one is either prompt with a flipped charge or trident lepton ($e \to e\gamma(e^+e^-)$) with flipped charge.
Material conversion	Mat. Conv.	One lepton is real material photon conversion and all the others are prompt.
Virtual conversion	Low mass e^+e^-	One lepton is from virtual (internal) photon conversion [124] and all the others are prompt.
Heavy-flavours decays	HF e/μ	One lepton is from heavy-flavour hadrons decays and all the others are prompt. This category is split in lepton flavour.
Other	Others	It contains minor physics processes (such as VVV production) and all the sources of instrumental background not listed above.

Table 5.5.: Event categorisation based on the nature of the leptons in the event.

5.4.2. Fit strategy

A summary of the fit setup, including the definition of control and signal regions and respective discriminating variables, is shown in Table 5.6. In this channel, the b-tagging working point for the region definition is 77%.

A Boosted Decision Tree [116] has been trained in the signal region (SR) to further separate the $t\bar{t}t\bar{t}$ signal from the total background. The chosen input variables show large separation power and are motivated by physics. The final set of 14 input variables can be grouped into different categories:

- *Jet activity*: $t\bar{t}t\bar{t}$ production exhibits larger jet multiplicity and harder p_{T} spectra with respect to total background.
- *Flavour-tagging information*: $t\bar{t}t\bar{t}$ production exhibits a larger number of b-jets in the final state, therefore on average jets in $t\bar{t}t\bar{t}$ events are more b-jet-like than in the background.
- *Angular variables*: $t\bar{t}t\bar{t}$ events have a different topology, especially with respect to $t\bar{t}$ and $t\bar{t}V$ events. Therefore angles between leptons and jets play a key role in separating the $t\bar{t}t\bar{t}$ from the background.
- E_T^{miss} *and lepton information*: $t\bar{t}t\bar{t}$ signal has more leptons and neutrinos from top quark decays with respect to the background. These variables are important to separate $t\bar{t}t\bar{t}$ from instrumental background.

All the input variables have been shown to have good modelling by the adopted background estimation. The BDT uses typical hyper-parameters of the training [116], using same boosting and fold validation as the BDT used in the 1LOS channel(see Section 5.3.2). It is trained on $t\bar{t}t\bar{t}$ LO QCD sample against all the simulated backgrounds. The resulting distribution of the BDT output is shown in Figure 5.8. Data events with a BDT score ≥ -0.2 are blinded.

As mentioned in Section 5.4.1 and shown in Table 5.6, dedicated regions have been defined to control the contributions of different physics background. Unconstrained normalisation factors are included in the fit to adjust the normalisations of the backgrounds from material conversion, low mass e^+e^-, HF e/μ and $t\bar{t}$+W backgrounds, driven by the fit in the corresponding control regions.

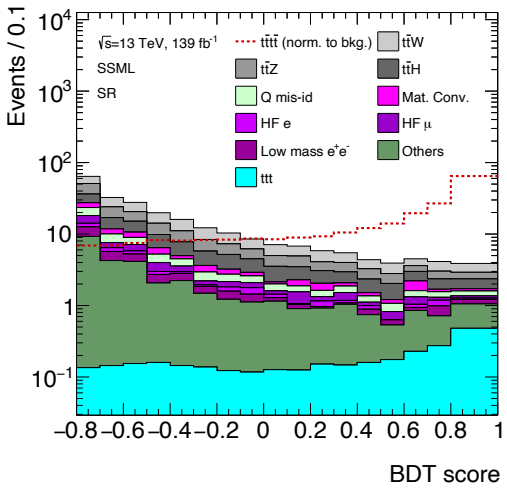


Figure 5.8.: Shape comparison of the BDT score in SSML channel events in the signal region between signal and total background. The event categories are defined in Table 5.5. The $t\bar{t}t\bar{t}$ contribution is normalised to the total background yield.

Name	Target	Variable	Categories	N_b	N_j	Other selections
SR	$t\bar{t}t\bar{t}$	BDT score	SSML	≥ 2	≥ 6	$H_{\mathrm{T}}^{\mathrm{all}} > 500$ GeV
CR$t\bar{t}$CO	Mat. Conv. Low mass e^+e^-	$e^\pm e^\pm$ $e^\pm \mu^\pm$	m_{ee}@PV	≥ 1	$4 \leq N_j < 6$	$0 < m_{ee}$@PV< 0.1 GeV $200 < H_T^{\mathrm{all}} < 500$ GeV
CR1b3Le	HF e	counting	eee or $ee\mu$	$= 1$		$100 < H_T^{\mathrm{all}} < 250$ GeV
CR1b3Lm	HF μ	counting	$\mu\mu\mu$ or $e\mu\mu$	$= 1$		$100 < H_T^{\mathrm{all}} < 250$ GeV
CR$t\bar{t}W$2L	$t\bar{t}W$	$\sum_{\text{leptons}} p_{\mathrm{T}}$	$\mu^\pm\mu^\pm$ $e^\pm\mu^\pm$	≥ 2	≥ 4	m_{ee}@PV< 0 or m_{ee}@PV> 0.1 GeV, $\lvert\eta\rvert < 1.5$ for $N_b = 2$: $H_T^{\mathrm{all}} > 500$ GeV or $N_j < 6$ for $N_b \geq 3$: $H_T^{\mathrm{all}} < 500$ GeV

Table 5.6.: Summary table of all the regions used in the profile likelihood fit for the $t\bar{t}t\bar{t}$ measurement in SSML channel. Each region is used to target a specific process ("Target" column): the signal region (SR) aims for the measurement of $t\bar{t}t\bar{t}$, while the control regions (with CR in the name) measure different background sources. A single variable is fitted in each region ("Variable" column): in case of "counting", only the total yield is used. In the other columns, the regions are defined in terms of lepton flavours ("Categories"), jet and b-jet multiplicities (N_j and N_b, respectively) and additional selections ("Other selection"). The variable m_{ee}@PV corresponds to the invariant mass of the electron and the closest track at the primary vertex [36].

CHAPTER 6

Systematic uncertainties

Several sources of systematic uncertainties on the background and signal predictions are considered. They are classified in two main categories: *experimental* and *modelling* uncertainties. Experimental error sources originate from detector measurements and reconstruction of the physics objects, while modelling uncertainties are associated with the predictions of the physics processes. Since 1LOS and SSML channels have a different composition of the background and a different sensitivity to systematic variations, two systematic models have been adopted for the two channels. The description of the common features is given in Section 6.1. The description of peculiarities of the two channels is given in Section 6.2 (1LOS) and Section 6.3 (SSML).

6.1. Common systematic uncertainties

The experimental systematic uncertainties on reconstructed objects and running conditions are the same for the two channels, since they use the same dataset and objects definition. The description of all the considered instrumental systematic uncertainties is found in Section 6.1.1. Modelling uncertainties on the signal process is also in common and described in Section 6.1.2.

6.1.1. Experimental sources of systematic uncertainties

Luminosity

The uncertainty associated to the luminosity value for the full Run 2 dataset is 1.7%. The dedicated measurement is performed with the LUCID detector [125] as described

in Section 4.1. This normalisation uncertainty applies to the MC predictions of all the considered processes, since the luminosity value is used to predict the number of expected events.

Pile-up reweighting

To correctly take pile-up into account in simulations, a set of minimum-bias events is superimposed to the simulated events in the MC samples. This contribution depends on the measured pile-up distribution. Therefore each event is reweighted to match the observed pile-up profile in data [126]. The systematic uncertainty on this weight applies to all MC simulations.

Charged leptons

Systematic errors associated with charged leptons come in two categories. The first category is connected to the trigger, reconstruction and identification efficiencies. As described in Sections 4.4.1 and 4.4.2, the efficiencies are measured in tag-and-probe analyses in MC and data events and scale factors are derived to make the MC efficiency match the one measured in data. The measured scale factors come with an associated uncertainty that is propagated to the MC prediction. The other sources of uncertainty are momentum scale and resolution applied to MC to match data. These are again measured and validated in tag-and-probe analyses (see Sections 4.4.1 and 4.4.2) and therefore errors on these two are used as additional sources of uncertainty on charged leptons.

Jets

Because of the complex procedure that is needed for reconstruction and calibration of a jet, many uncertainties are associated to this object. As described in Section 4.4.3, many different MC-based and *in-situ* calibrations [110] are applied, targeting different kinematic regions. The systematic uncertainty associated to the jet energy scale (JES) consists of a set of 29 *eigen*-variations, defined such that there is no correlation among them. These variations are separated in sub-categories depending on the source of the uncertainty. Uncertainties from *in-situ* calibrations contribute with 16 NPs in the fit. Other five NPs come from the η intercalibration, four are due to the pile-up component subtraction, one describes the uncertainty on the *punch-through* modelling[1] and one comes from the calibration of high-p_T jets. Two NPs are associated to the uncertainty on jet flavour composition and the detector response to the different flavours. The jet energy resolution (JER) has eight associated *eigen*-variations.

[1] "Punch through" corresponds to the fraction of the jet not contained by the calorimeter system.

Jet vertex tagging

An extra source of systematic errors arises from the use of the jet vertex tagger, calibrated against Z+jets events to extract an associated scale factor. The scale factor uncertainty is dominated by the contribution of the pile-up uncertainty.

Jet flavour tagging

As described in Section 4.4.4, every operational point for b-jet tagging is calibrated to extract the efficiency scale factors. It gives nine systematic uncertainties associated to b-tagging efficiency and other four for each of c-jets and light-jets mis-tagging efficiency. The pseudo-continuous b-tagging calibration uses the calibration of five correlated working points as input, each having its own tagging and mis-tagging efficiency. Therefore, each variation must be varied for each working point.

Missing transverse energy

Since the missing transverse energy depends on all the other reconstructed objects (see Section 4.4.5), uncertainties on those are propagated down to the $E_{\mathrm{T}}^{\mathrm{miss}}$ computation. The uncertainty on the soft-component is evaluated by using $Z \to \mu^{+}\mu^{-}$ events [127]. All the uncertainties are eventually merged in total scale and resolution uncertainties.

6.1.2. Systematic uncertainties on physics process modelling

Signal process

The main uncertainty associated to $t\bar{t}t\bar{t}$ modelling is the uncertainty on the parton shower and hadronisation model, obtained by comparing nominal simulation using aMC@NLO+Pythia 8.230 [79] with a simulation that uses the same generator interfaced with Herwig 7.04 [82, 83] (see Section 4.3.1).

The uncertainty on the choice of the renormalisation scale μ_R is estimated by changing its value in the simulation. The "up" and "down" variations of the scale correspond to $\mu_R^{\mathrm{up}} = 2\mu_R$ and $\mu_R^{\mathrm{down}} = 0.5\mu_R$ respectively, keeping the factorisation scale at the nominal value. The difference between the varied and the nominal predictions is considered as uncertainty. The uncertainty on the factorisation scale μ_F is estimated following the same procedure.

6.2. Systematic uncertainties for the single lepton and opposite-sign dilepton channel

The 1LOS channel is dominated by the $t\bar{t}$+jets contribution, thus the assessment of its modelling uncertainties is crucial for the analysis in this channel and described in the following. However, the two background estimation methods have some differences in the implementation of these uncertainties which are described in Section 7.3. The uncertainties on the modelling of the minor backgrounds is also described in this section.

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

Uncertainties on the flavour composition of $t\bar{t}$ radiation are of major importance in this analysis. To assess the flavour of the radiation, a classification in terms of number of heavy-flavour hadrons in the final state at particle level is performed. A $t\bar{t}$+b-jets event is identified by at least one the particle-level jet that is associated to a B-hadron not originated by the top quark decay chain. In case this condition is not fulfilled, an event is classified as $t\bar{t}$+c-jets in case at least one jet is associated to a C-hadron not coming from the W decay. In all the other cases, the event is classified as a $t\bar{t}$+light-jets event. The uncertainties on the $t\bar{t}$+b-jets and $t\bar{t}$+c-jets rates are set to $\pm 50\%$, based on previous measurements of $t\bar{t}$ production associated with heavy-flavour jets [128]. To avoid a double counting of these uncertainties, all the following variations are reweighted to the same flavour composition of the $t\bar{t}$ radiation as the nominal simulation at parton level.

In the set of systematic uncertainties adopted by the reweighting method for the $t\bar{t}$+jets estimation (see Section 7.2), all the following uncertainties associated to $t\bar{t}$+jets modelling are divided based on the flavour of the $t\bar{t}$ radiation, so in $t\bar{t}$+b-jets, $t\bar{t}$+c-jets and $t\bar{t}$+light-jets. The systematics model used by the t$\bar{\text{t}}$TRF method (see Section 7.1) for the $t\bar{t}$+jets estimation, instead, has no separation of the $t\bar{t}$ modelling uncertainties into different flavours of the radiation. The reason behind this is explained in Section 7.3, after the description of both methods.

The uncertainty on the choice of factorisation scale in the $t\bar{t}$ simulation is estimated by looking at the effect of setting the scale at double and half of the nominal value. The same procedure is used to assess the uncertainty on the choice of the renormalisation scale. The impact of a different contribution from the initial state radiation (ISR) is estimated by changing the *Var3c* tuning parameter of Pythia 8.230 [79] that regulates the α_s value in the ISR simulation. The variations are obtained by using the *Var3cUp* and *Var3cDown* tuning parameters. A similar procedure is followed to estimate the effect of the final state radiation (FSR), by setting to half and double the nominal value the factorisation scales used for the FSR simulation.

The uncertainty associated to the choice of the *hdamp* parameter of Powheg-Box v2 is obtained by comparing the nominal simulation with an alternative one using *hdamp* set

Process	Uncertainty
VV	$\pm 50\%$
V+jets	$\pm 50\%$
Single-top	$\pm 13\%$
$t\bar{t}+X$	$\pm 15\%$

Table 6.1.: Summary of the cross-section uncertainties applied to the minor background processes in the 1LOS channel [129–132].

to double the nominal value (see Section 4.3.1). The difference between the two predictions is symmetrised and taken as uncertainty.

The uncertainty associated to the choice of parton shower and hadronisation model is given by comparing the nominal simulation with an alternative one that uses the same generator as the nominal (POWHEG-BOX v2) interfaced with a different showering algorithm (HERWIG 7.04, see Section 4.3.1). The uncertainty associated to the matching of the generator to the parton shower algorithm is estimated by comparing the nominal simulation with an alternative one provided by a different generator (aMC@NLO 2.6.0, see Section 4.3.1) interfaced with the same parton shower algorithm as the nominal[2]. In both cases, the difference between the nominal and the alternative predictions is symmetrised and considered as uncertainty.

Minor background cross-sections

Uncertainties on the cross-section of the background processes are considered and applied as normalisation uncertainties to the respective MC predictions. A summary of these uncertainties is in Table 6.1. This analysis focuses on a small corner of the phase-space that is far from the one used for reference measurements and theoretical calculations of these background processes. A large extrapolation of the uncertainties is therefore needed, resulting in larger uncertainty than in the corresponding measurements or simulations. The quoted uncertainties include the whole modelling uncertainty on these minor background processes in the considered phase space [129–132].

6.3. Systematic uncertainties for the same-sign dilepton and multilepton channel

The complex background composition in the SSML is reflected in the implemented systematic model. Dedicated uncertainties are associated to the event categories described

[2]Since different matrix element generators are used, the settings of the parton shower to match it have to be changed accordingly.

in Section 5.4.1. Moreover, since $t\bar{t}+X$ processes have a significant contribution in this channel, multiple uncertainties associated to their modelling are introduced.

6.3.1. Uncertainties on reducible backgrounds

Two normalisation factors for the material photon conversion and low mass e^+e^- backgrounds are included in the fit. The remaining shape uncertainty is estimated by comparing the simulation of $Z\to\mu^+\mu^-$+jets provided by POWHEG-BOX v2, interfaced with PYTHIA 8.230 with data in a region enriched of $Z\to\mu^+\mu^-$+jets+γ events. The simulation is normalised to data yields in this region and the remaining shape discrepancy is used as uncertainty in the fit.

As described in Section 5.4.1, the background due to the mis-assignment of the lepton charges is estimated with a data-driven method. The uncertainties on this prediction arise from the statistical uncertainty of the fit used to measure the charge mis-assignment rate, the change of the ranges where the fit is performed and the non-closure between data and MC prediction after applying the rates.

Two normalisation factors for the contribution of non-prompt leptons from heavy-flavour hadron decays are included in the fit and estimated in dedicated control regions. The corresponding shape uncertainty is assessed by comparing the MC prediction with data. The contribution of this type of event in data is estimated by subtracting all the other contributions from the data yields. The MC prediction is normalised to data and the remaining discrepancy is used as estimation of the uncertainty. This uncertainty is extracted separately for muons and electrons in all the regions included in the fit. To increase the statistics of the events with non-prompt leptons for this estimation, the requirements on lepton isolation and identification are loosened. One single nuisance parameter is associated to this variation.

Additional normalisation uncertainties are associated to sub-dominant reducible processes. An uncertainty of 100% is associated to the background coming by from non-prompt leptons from light-flavour hadrons [133], and an ad-hoc uncertainty of 30% is associated to all the other sources.

Modelling uncertainties on $t\bar{t}+X$ processes

The main $t\bar{t}+X$ processes that contribute in the $t\bar{t}t\bar{t}$ phase space are $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}H$ productions. For each of them, the uncertainty associated to the choice of renormalisation and factorisation scale are assessed as for the other processes (see Section 6.1.2).

The uncertainty on the choice of the matrix element generator for the $t\bar{t}H$ simulation is obtained by comparing the nominal simulation (using POWHEG-BOX v2) with an alternative prediction from aMC@NLO 2.3.3 [75]. The alternative and nominal samples use the same PDF set, parton shower algorithm and set of tuning parameters.

The uncertainties on the modelling of $t\bar{t}W$ and $t\bar{t}Z$ processes are estimated by using alternative simulations with SHERPA 2.2.5 [71] at QCD NLO, using the NNPDF3.0 NNLO PDF set. The parameters of the generator are tuned by the authors. For both processes, the difference between the nominal and alternative predictions is symmetrised and used as uncertainty.

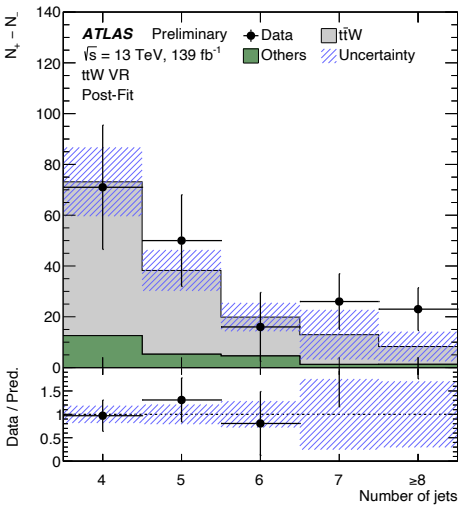


Figure 6.1.: Comparison of the $t\bar{t}W$ modelling with data in the validation region as a function of jet multiplicity [36].

Additional uncertainties are introduced to cover the discrepancies of data with the simulation of these processes associated with the production of heavy-flavour jets. An uncertainty of 50% is associated to $t\bar{t}+X$ production with 3 and $\geq$4 b-jets [128]. A validation region is used to check the $t\bar{t}W$ modelling in different kinematic variables. This region requires at least four jets and at least two b-jets and consists of the difference of the yield of events with positive total lepton charge N_+ and N_-, standing for the yield of events with negative total lepton charge[3]. This procedure removes all the charge-symmetric processes, providing a pure region in $t\bar{t}W$ events. The distribution of jet multiplicity is checked in this region and shown in Figure 6.1. The observed discrepancy of the $t\bar{t}W$ simulation with data at large jet multiplicity is used to quote an uncertainty on the modelling $t\bar{t}W$ process associated with 7 and $\geq$8 jets modelling of 125% and 300%, respectively.

[3]In each bin of the given distribution, the quantity $(N_+ - N_-)$ is evaluated. N_+ (N_-) stands for the number of events having $\sum_l q_l > 0$ ($\sum_l q_l < 0$) where q_l is the charge of the l-th lepton in the event and l runs over all the leptons in the event.

Cross section uncertainties

The uncertainties on cross-section of the background processes are summarised in Table 6.2. The values may differ from the values in Table 6.1, since different jet, b-jet multiplicity and conditions on leptons are required in this channel. Therefore a different extrapolation of the uncertainties is needed [129–133]. Uncertainties on the modelling of processes associated with heavy-flavour jets (e.g. $t\bar{t}$, ttt) are divided into different b-jet multiplicities to account for the $g \to b\bar{b}$ contribution [128]. In the SSML analysis, the uncertainty of 20% on the $t\bar{t}t\bar{t}$ cross-section calculation is also included [21].

Process	Uncertainty	Process	Uncertainty	Process	Uncertainty
Single-top	±30%	VV	±40%	ttt	±100%
V+jets	±30%	$t\bar{t}H$	±20%	ttt ($3b$-jets)	±50%
$t\bar{t}Z$	±15%	$t\bar{t}$ ($3b$-jets)	±30%	ttt ($\geq 4b$-jets)	±50%
$t\bar{t}VV(H)$	±50%	$t\bar{t}$ ($\geq 4b$-jets)	±30%	Other	±50%

Table 6.2.: Summary of the cross-section uncertainties applied to the background processes in the SSML channel [128–133]. No uncertainty is associated to $t\bar{t}W$ cross-section because its normalisation is included in the fit as a free-floating parameter [36].

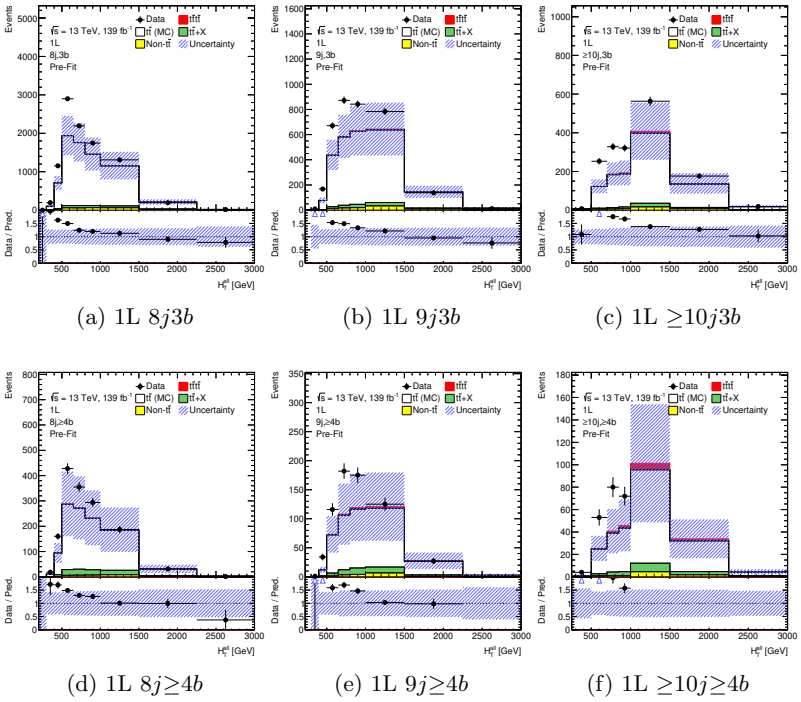
CHAPTER 7

Background estimation in the single lepton and opposite-sign dilepton channel

As extensively discussed in Chapter 5, this analysis aims to measure of a tiny signal in a very challenging phase space, especially in the single lepton and opposite-sign dilepton channel (1LOS). To make this possible, a robust estimation of the background is important, usually provided by MC simulations. In the 1LOS channel, as shown in Figures 5.3a and 5.3b, the background consists mainly of $t\bar{t}$ production associated with many radiated jets and b-jets. The MC simulation for this process, however, is designed to predict the $t\bar{t}$ process in an inclusive phase space dominated by $t\bar{t}$ production with no or one additional jet, and not in an extreme phase space where $t\bar{t}t\bar{t}$ lies. Therefore, the MC simulation is not expected to provide robust predictions under these conditions.

The MC modelling of the $t\bar{t}$ process has been checked in the control regions (see Section 5.3.3) in the single lepton channel, shown in Figures 7.1a to 7.1f, and in the opposite-sign dilepton channel, shown in Figures 7.2a to 7.2f, at different jet and b-jet multiplicities. The plots show a large discrepancy between MC predictions and observed data, with predictions underestimating the data of about 50%, increasing with jet and b-jet multiplicities. A disagreement also in the shapes of the distributions is visible in $H_{\mathrm{T}}^{\mathrm{all}}$, as well as in many other distributions, especially of p_{T}-related variables. Therefore, alternative estimation methods for the $t\bar{t}$+jets contribution have been developed for the analysis in this channel.

The two developed methods are the data-driven MC-assisted $t\bar{t}$ *tag-rate-function* method ($\mathrm{t\bar{t}}$TRF) and the MC-driven data-assisted *reweighting* method. The former uses data events in lower b-jet regions to estimate distributions at higher b-jet multiplicities, corrected by MC-based factors. The latter instead uses MC events, reweighted to data



(a) 1L $8j3b$ (b) 1L $9j3b$ (c) 1L $\geq 10j3b$

(d) 1L $8j\geq 4b$ (e) 1L $9j\geq 4b$ (f) 1L $\geq 10j\geq 4b$

Figure 7.1.: Comparison of MC prediction with observed data in control regions for the single lepton channel (1L) as a function of $H_{\mathrm{T}}^{\mathrm{all}}$. Jet multiplicity increases from the left to the right column. The first row contains the plots for regions with 3 b-jets, while the second one contains the plots for regions requiring ≥ 4 b-jets. The notation used to label the regions is explained in Section 5.3.3.

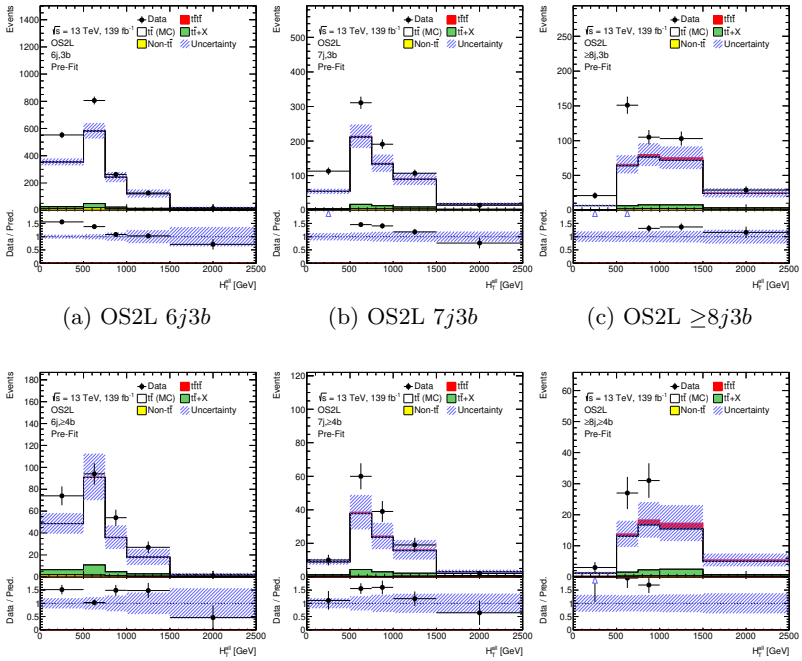
in lower b-jet multiplicity regions to estimate the yields at higher b-jet multiplicities.

The implementation and development of the $t\bar{t}TRF$ represents the core part of this thesis and it is extensively described in Section 7.1. An overview of the *reweighting* method is given in Section 7.2. A comparison of the two methods is given in Section 7.3.

7.1. The $t\bar{t}$ tag-rate-function (t$\bar{\text{t}}$TRF) method

The $t\bar{t}$ tag-rate-function (t$\bar{\text{t}}$TRF) method aims at providing a data-driven estimation of $t\bar{t}$+jets yields and distributions at high b-jet multiplicity. This method is based on the *tag-rate-function* (TRF) method that has been extensively used in several measurements to improve the statistics of the predictions (usually MC) at large b-jet multiplicity.

An example of TRF application can be found in the analysis targeting the $pp \to t\bar{t}H(b\bar{b})$ process with the ATLAS experiment. The TRF method has been used in the hadronic



(a) OS2L $6j3b$ (b) OS2L $7j3b$ (c) OS2L $\geq 8j3b$

(d) OS2L $6j\geq 4b$ (e) OS2L $7j\geq 4b$ (f) OS2L $\geq 8j\geq 4b$

Figure 7.2.: Comparison of MC prediction with observed data in control regions for the opposite-sign dilepton channel (OS2L) as a function of $H_{\mathrm{T}}^{\mathrm{all}}$. Jet multiplicity increases from the left to the right column. The first row contains the plots for regions with 3 b-jets, while the second one contains the plots for regions requiring ≥ 4 b-jets. The notation used to label the regions is explained in Section 5.3.3.

channel [134] to estimate the contribution from the $pp\rightarrow$jets (multi-jet) process, while the single leptonic channel [135, 136] has tested the TRF method to predict the yields from fake leptons. The TRF method is based on the estimation of the probability of a jet of being b-tagged by the tagging algorithm. The probabilities assigned to each jet in the events are combined to reweight the events to different b-jet multiplicities, either those are data or MC events. The probabilities, or *efficiencies*, are extracted in dedicated regions, usually orthogonal to the regions used for the measurement.

The t$\bar{\text{t}}$TRF consists of the TRF method applied to the $t\bar{t}$+jets process. The estimation of the efficiencies is performed on data events in regions that are orthogonal to control and signal regions of the fit. The regions for the extraction need to have a dominant $t\bar{t}$ contribution and a sufficient statistics of data events. The measured efficiencies are then used to reweight data events in regions with 2 b-jets to estimate the $t\bar{t}$+jets con-

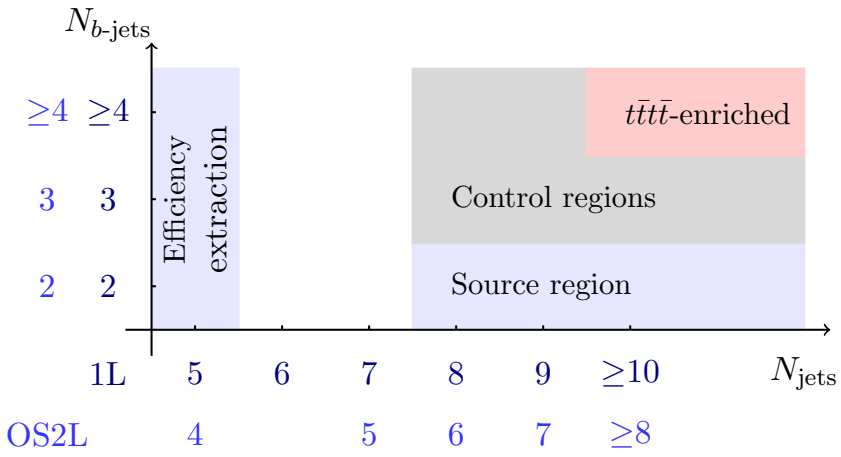


Figure 7.3.: Regions definition for tt̄TRF method derivation and application.

tribution in regions with ≥3 b-jets[1]. The obtained prediction, that is fully data-driven, is eventually corrected by a MC correction factor to compensate for the assumptions of the method.

This method has been already used in a previous search for the $t\bar{t}t\bar{t}$ signal with the ATLAS detector [32], and provided a robust estimation of the $t\bar{t}$+jets background. In this analysis, some improvements have been implemented to fit the new analysis strategy. The description of the method is given in the following sections. The extraction and application of the efficiencies are described in Section 7.1.1 and Section 7.1.2, respectively. The MC correction factor is introduced in Section 7.1.3. The improvements are then presented. The estimation of the pseudo-continuous b-tagging score is described in Section 7.1.4. The procedure to reduce the MC correction factor by fitting the dependency of the efficiencies on jet multiplicity is explained in Section 7.1.5.

7.1.1. Extraction of the efficiencies

This section describes the computation of the efficiencies for the tt̄TRF method. The efficiencies refer to the probability of a radiated jet from $t\bar{t}$ to be tagged as a b-jet by the used b-tagging algorithm. In the following, the tagger is assumed to be MV2c10 at the 70% working point. The regions for the extraction must be orthogonal to the regions used for the measurements, rich in (data and MC) statistics and representative of the regions used in the fit in terms of physics processes and background composition. The regions chosen for the efficiency extractions are shown in Figure 7.3: the efficiencies are extracted in $5j$ ($4j$) region for single lepton (opposite-sign dilepton) channel and applied in $2b$ regions at different jet multiplicities to predict the $t\bar{t}$ yields in $3b$ and $\geq 4b$

[1] Since the estimation regards only $t\bar{t}$ background, data events are used by removing the (minor) non-$t\bar{t}$ contributions estimated with MC.

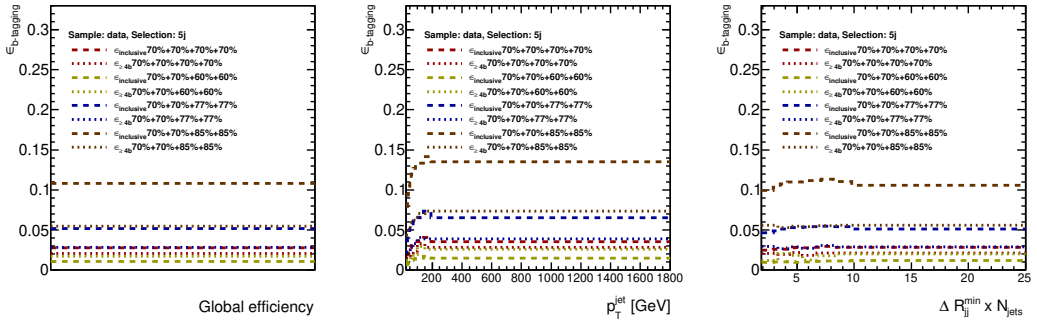


Figure 7.4.: Summary plots for efficiencies on data for 2015/16 data-taking, for the single lepton channel. The overall efficiencies are shown on the left, while their parametrisations as a function of jet p_{T} and $\Delta R_{jj}^{\mathrm{min}} \times N_{\mathrm{jets}}$ are shown in the center and on the right, respectively. Both inclusive ($\epsilon_{\mathrm{inclusive}}$) and dedicated ($\epsilon_{\geq 4b}$) efficiencies are displayed in the plots. Efficiencies using different working points for jet b-tagging are compared.

regions. Section 5.3.3 introduces the conventions on regions names in terms of jet and b-jets multiplicities.

In principle, the efficiencies can be estimated both on data and MC events: since data are not well described by the $t\bar{t}$ simulation even in the extraction region, the efficiencies used for the nominal t$\bar{\mathrm{t}}$TRF prediction are based on data events. However, efficiencies estimated on MC are also useful for the computation of the MC correction factor (see Section 7.1.3).

Because of the definition of efficiencies as a probability for jets radiated from $t\bar{t}$ events to be b-tagged, the b-jets from top quark decay (called *bulk* b-jets) must be excluded from the computation. These are identified as the two b-tagged jets with highest b-tagging score in the event. Event-based cuts are also applied, for instance $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV and $(E_{\mathrm{T}}^{\mathrm{miss}} + M_{\mathrm{T}}^{W}) > 60$ GeV are required in the single lepton channel to suppress the contribution of multi-jet background. After having processed all the jets in all the events in the extraction region, the number $N_{\mathrm{tagged}}^{\mathrm{total}}$ of b-tagged jets over a total number of jets N_j^{total} is measured. In case the efficiency is estimated on data, $N_{\mathrm{tagged}}^{\mathrm{total}}$ and N_j^{total} contain both contributions from $t\bar{t}$ and non-$t\bar{t}$ background sources. Therefore, the contributions coming from the non-$t\bar{t}$ backgrounds are subtracted, based on their MC prediction. The obtained numbers $N_{\mathrm{tagged}}^{t\bar{t}}$ and $N_j^{t\bar{t}}$ are used for the efficiency computation. The efficiencies are computed as a function of jet p_{T} and $\Delta R_{jj}^{\mathrm{min}} \times N_j$ to take into account the tagger response for different jet kinematics and event topologies[2]. Efficiencies are computed separately for each data-taking period, to account for possible effects of pile-up. Typical efficiencies estimated on data are shown in Figure 7.4.

[2]The variable $\Delta R_{jj}^{\mathrm{min}} \times N_j$ stands for the ΔR of the considered jet with the closest one, "normalised" by the number of jets in the event.

Several efficiencies are shown in Figure 7.4, firstly classified in inclusive ($\epsilon_{\text{inclusive}}$) and dedicated ($\epsilon_{\geq 4b}$) efficiencies. The inclusive efficiencies are estimated on all the extra-jets, and used to predict $t\bar{t}$+jets in regions with three b-jets. The dedicated efficiencies are computed in regions with $\geq$3 b-jets by excluding also the third b-tagged jet with the third highest b-tagging score. They account for the change of the average efficiencies due to the contribution of $g \to b\bar{b}$ splitting in regions with $\geq$4 b-jets and are indeed used for the estimation of $t\bar{t}$+jets in these regions. Figure 7.4 shows also a second classification of the efficiencies in different used working points. The efficiencies follow the naming convention `XX+XX+YY+ZZ`, meaning that the first two b-tagged jets (excluded from the computation) are asked to be tagged by the `XX` working point, `YY` is applied for the first extra jet and `ZZ` from the second extra jet onwards. The efficiency computed with a fixed working point (`70%+70%+70%+70%`) is used for the $t\bar{t}$TRF prediction. The efficiencies using floating working points are computed for the estimation of the pseudo continuous b-tagging score of the jets, as described in Section 7.1.4.

The efficiencies tend to increase with jet p_{T}, probably due to lower rejection power of the tagger against light-jets [114, 115]. The small increase in $\Delta R_{jj}^{\text{min}} \times N_j$ can be explained by the actual improvement in b-tagging efficiency in a cleaner environment. The effect of gluon splitting is visible in the efficiencies using the tightest operating point on the extra jets, purer in real b-jets, where $\epsilon_{\geq 4b}$ is higher than the inclusive one. The probability of having a fourth real b-jet once a third real b-jet is indeed expected to be higher than the average because of the contribution of $g \to b\bar{b}$ in the radiation of jets.

7.1.2. Application of the efficiencies

The extracted efficiencies are used to estimate the $t\bar{t}$+jets contribution in 3b and $\geq$4b regions. The events used for the prediction are data events from the *source regions* in Figure 7.3. Therefore, the statistics of the prediction at higher b-jet multiplicities corresponds to data statistics in the 2b region at the same jet multiplicity. To promote events from the source region with N_j jets to 3b and $\geq$4b regions with the same jet multiplicity, the calculation of some auxiliary weights coming from the TRF mathematics is needed.

$$w^{2b}_{\text{TRF,incl.}}(N_j) = \prod_{i=1}^{N_j-2} \left(1 - \epsilon^j_{\text{incl.}}\right) \tag{7.1}$$

$$w^{3b}_{\text{TRF,incl.}}(N_j) = \sum_{i=1}^{N_j-2} \epsilon^i_{\text{incl.}} \prod_{j\neq i} \left(1 - \epsilon^j_{\text{incl.}}\right) \tag{7.2}$$

$$w^{3b}_{\text{TRF},\geq 4b}(N_j) = \prod_{i=1}^{N_j-3} \left(1 - \epsilon^j_{\geq 4b}\right) \tag{7.3}$$

$$w^{\geq 4b}_{\text{TRF},\geq 4b}(N_j) = \sum_{i=1}^{N_j-3} \epsilon^i_{\geq 4b} \prod_{j\neq i} \left(1 - \epsilon^j_{\geq 4b}\right) \tag{7.4}$$

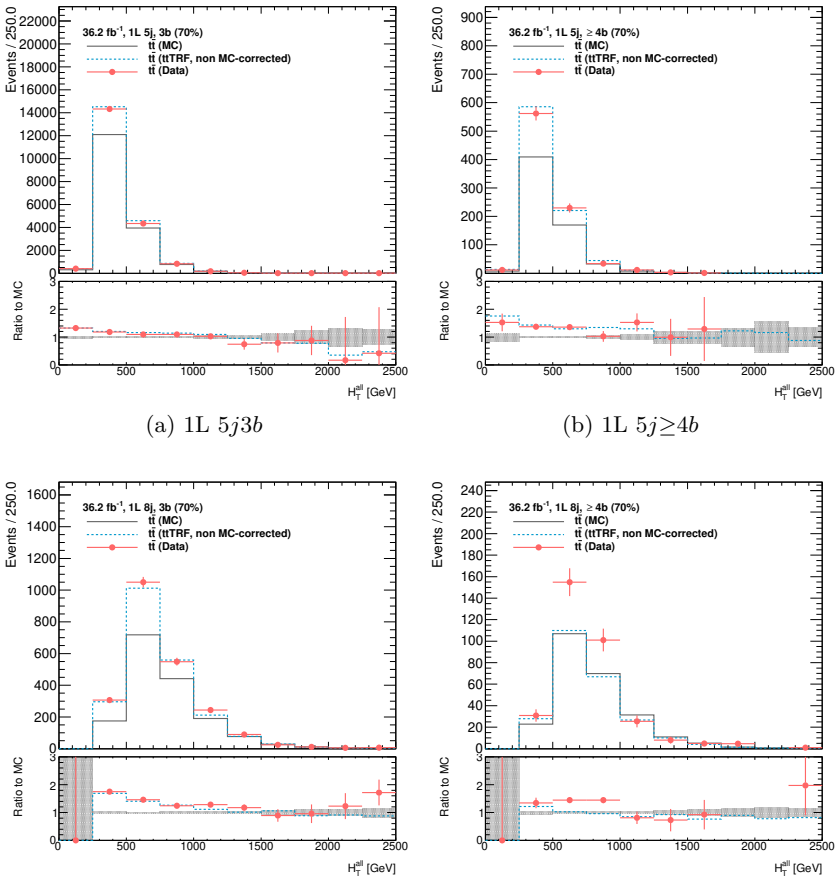
Note that two (Equations 7.1 and 7.2) or three (Equations 7.3 and 7.4) jets are excluded in the application: these are the jets with the highest b-tagging score in the event. The weight in Equation 7.2 (7.4) is the usual TRF weights that, when applied to events in the $\geq 2b$ ($\geq 3b$) inclusive region, predicts the background in the $3b$ ($\geq 4b$) region. In case of the $t\bar{t}$TRF method, the source events are from $2b$ exclusive regions, requiring a proper normalisation of the weights in Equations 7.2 and 7.4. To predict the $t\bar{t}$+jets background in the $3b$ exclusive region, the inclusive efficiencies are used and the $t\bar{t}$TRF weight corresponds to the weight in Equation 7.2 normalised by the probability of the event to be in the $2b$-exclusive region in Equation 7.1:

$$w^{3b}_{t\bar{t}\text{TRF}}(N_j) = \frac{w^{3b}_{\text{TRF,incl.}}(N_j)}{w^{2b}_{\text{TRF,incl.}}(N_j)} \tag{7.5}$$

The prediction of the $t\bar{t}$+jets background in regions with ≥ 4 b-jets is obtained by using the dedicated efficiencies. Since events in $3b$ exclusive region can not be directly accessed, the $t\bar{t}$TRF estimation is used. A further step to reweight those events in $4b$ inclusive regions is therefore needed: the $t\bar{t}$TRF weight for events promoted into $\geq 4b$ regions corresponds to the weight in Equation 7.4, normalised by the probability of being in the $3b$ exclusive region computed with the dedicated efficiencies in Equation 7.3, multiplied by the $t\bar{t}$TRF weight for the prediction in the $3b$ exclusive regionn in Equation 7.5):

$$w^{\geq 4b}_{t\bar{t}\text{TRF}}(N_j) = \frac{w^{\geq 4b}_{\text{TRF},\geq 4b}(N_j)}{w^{3b}_{\text{TRF},\geq 4b}(N_j)} w^{3b}_{t\bar{t}\text{TRF}}(N_j) \tag{7.6}$$

Since a weight is associated to each event, every kinematic distribution can be predicted by the method. Non-b-tagged jets in the event are promoted as b-tagged by the method according to the total event probability of the combination of tagged and non-tagged jets. This allows the computation of b-jet related variables. The promoted combination in $3b$ by $t\bar{t}$TRF is used for the prediction in $\geq 4b$ regions. The expected contributions from the minor backgrounds, as estimated by $t\bar{t}$TRF using the corresponding MC samples, has to be subtracted. A comparison of MC and $t\bar{t}$TRF predictions for $H_{\text{T}}^{\text{all}}$ of the $t\bar{t}$+jets process with data is shown in Figures from 7.5a to 7.5d in the extraction regions and two control regions. As expected, the $t\bar{t}$TRF matches with data in the extraction regions within the data statistical error. However, the agreement degrades with higher jet and b-jet multiplicities, as shown in Figure 7.5d. This is due to the caveats of the methods that are fixed by the introduction of the MC correction factor.



(a) 1L $5j3b$ (b) 1L $5j{\geq}4b$

(c) 1L $8j3b$ (d) 1L $8j{\geq}4b$

Figure 7.5.: Comparison of the $t\bar{t}$ prediction for $H_{\mathrm{T}}^{\mathrm{all}}$ from a data-driven t$\bar{\mathrm{t}}$TRF method with respect to $t\bar{t}$ in data and MC. Distributions are shown for the 2015/16 data-taking period for the single lepton channel, the same as in Figure 7.4. Here, "$t\bar{t}$ (Data)" is the estimation of $t\bar{t}$ in data obtained by subtracting the non-$t\bar{t}$ backgrounds from data yields. The distributions in the efficiency extraction regions are shown in the top row, while the bottom row contains the plots for two control regions, requiring 8 jets in the final state. Three and at least four b-jets are required in the plots on the left and right, respectively. The discrepancies between t$\bar{\mathrm{t}}$TRF prediction and data in the bottom row plots (especially in Figure 7.5d) are addressed by the MC correction factor.

7.1.3. The Monte Carlo correction factor

Two main assumptions are made in the implementation of the $t\bar{t}$TRF as described up to this point:

- *Independence on jet multiplicity*: the efficiencies are assumed to be constant in jet multiplicity, since they are computed in the extraction region and applied at higher jet multiplicity. However, the efficiencies may depend, even strongly, on the number of jets. Moreover, since jet kinematics change with jet multiplicity, an effective dependence of the mean value of the efficiencies occurs. This dependence is not necessarily in the same direction of the intrinsic efficiency dependence on the number of jets, causing even a larger discrepancy of the prediction from data.
- *Correlation between b-jet radiation*: as shown in Equations 7.5 and 7.6, no correlation between b-tagged jets is taken into account. It has to be noted that the largest correlations are expected to be introduced by gluon to $b\bar{b}$ splitting, and this is partially mitigated by the use of dedicated efficiencies. However, other correlations are completely neglected.

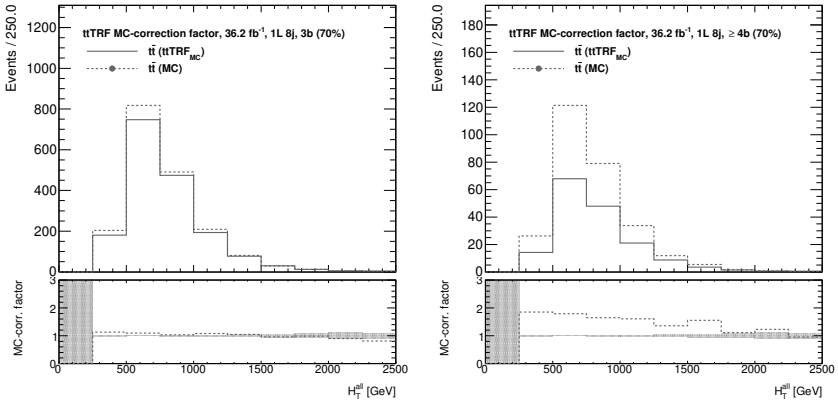
These effects can be cured introducing a MC correction factor. This aims to correct the bias introduced by the $t\bar{t}$TRF assumptions, by comparing the MC prediction with $t\bar{t}$TRF prediction for the same MC sample. The correction is estimated on the nominal simulation of the $t\bar{t}$ process. The expression for the MC correction factor $f^{\text{MC}}_{\text{t}\bar{\text{t}}\text{TRF,i}}$ for a bin i of a given distribution is shown in Equation 7.7.

$$f^{\text{MC}}_{\text{t}\bar{\text{t}}\text{TRF,i}} = \frac{y_{\text{MC, i}}}{y^{\text{MC}}_{t\bar{t}\text{TRF,i}}} \tag{7.7}$$

corresponding to the ratio between the MC yield ($y_{\text{MC, i}}$) and the yield estimated by the $t\bar{t}$TRF method, using efficiencies extracted on the same MC simulation in the extraction region and applied to events from the same MC in the source regions ($y^{\text{MC}}_{t\bar{t}\text{TRF,i}}$). Eventually, the final $t\bar{t}$TRF-estimated yield ($y^{\text{MC-corr.}}_{\text{t}\bar{\text{t}}\text{TRF,i}}$) is given by correcting the data-driven $t\bar{t}$TRF prediction ($y_{\text{t}\bar{\text{t}}\text{TRF,i}}$) by the MC factor:

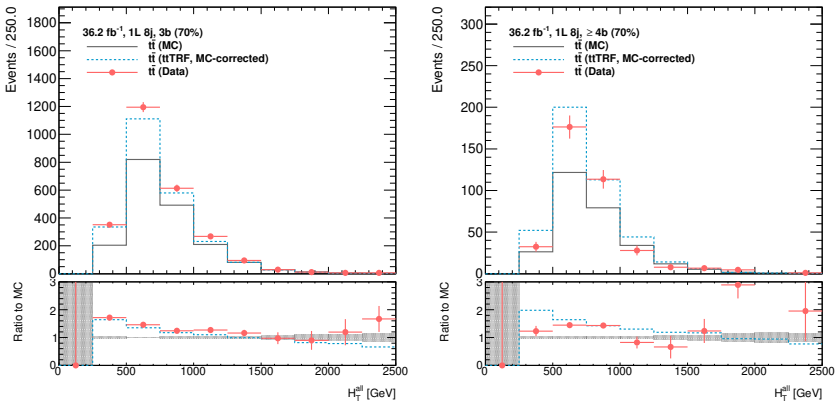
$$y^{\text{MC-corr.}}_{\text{t}\bar{\text{t}}\text{TRF,i}} = y_{\text{t}\bar{\text{t}}\text{TRF,i}} \cdot f^{\text{MC}}_{\text{t}\bar{\text{t}}\text{TRF,i}} \tag{7.8}$$

Two examples of MC correction factors as a function of $H_{\text{T}}^{\text{all}}$ are shown in Figures 7.6a and 7.6b. As expected this becomes larger at larger jet multiplicities which is compatible with the discrepancy seen for non-corrected $t\bar{t}$TRF prediction. As shown in Figures 7.7a and 7.7b, the application of the MC correction factor leads to better modelling of the prediction to data even at large jet and b-jet multiplicites. On the other hand, the application of the MC correction factor makes the method dependent on the MC prediction, therefore all the systematic uncertainties affecting the $t\bar{t}$ MC must be propagated through the $t\bar{t}$TRF method to obtain the uncertainty on the prediction. Since a systematic variation is evaluated as the ratio of the variation and its corresponding estimation



(a) (b)

Figure 7.6.: Comparison of the $t\bar{t}$ MC prediction and tt̄TRF estimation applied to $t\bar{t}$ MC for the MC correction factor extraction in the single lepton channel, as a function of $H_{\mathrm{T}}^{\mathrm{all}}$. The two regions are the same as in Figures 7.5c and 7.5d: $8j3b$ (a) and $8j{\geq}4b$ (b). The two MC correction factors are consistent with the discrepancies between un-corrected tt̄TRF prediction and data.



(a) (b)

Figure 7.7.: Comparison of the tt̄TRF prediction including the MC correction factor, with the $t\bar{t}$ MC and the $t\bar{t}$ in data as a function of $H_{\mathrm{T}}^{\mathrm{all}}$. The contribution of $t\bar{t}$ in data is obtained by subtracting all the minor background contributions from the data distribution. Only statistical uncertainties are shown in the plots. Same regions as in Figures 7.6a and 7.6b are shown for a direct comparison: 1L $8j3b$ (a) and 1L $8j{\geq}4b$ (b).

from t$\bar{\text{t}}$TRF, the systematic uncertainty is smaller than the original MC one. Note that even systematics on the modelling of minor backgrounds must be propagated, since they are used in the subtraction to the data yields in the efficiency computation.

7.1.4. Estimation of the pseudo-continuous b-tagging score

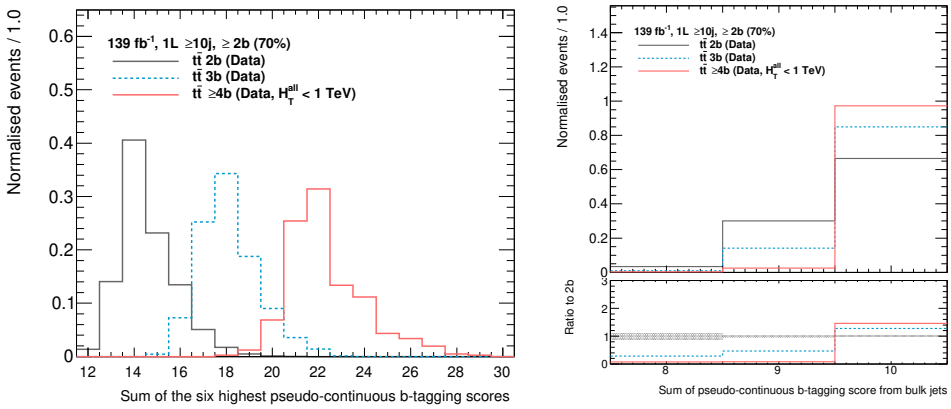
The pseudo-continuous (PC) b-tagging score of the jets is not predicted by the implementation described above. The t$\bar{\text{t}}$TRF method promotes events from the regions with 2 b-jets to higher b-jet multiplicities, in which the jets in these promoted events have a different spectrum of values for the PC b-tagging scores from data events, since they originally belong to a different b-jet multiplicity, as shown in Figure 7.8a. The PC score is however very important in the separation between signal and background. It is used in the BDT implementation as shown in Table 5.3 and is the most important feature. Therefore, making t$\bar{\text{t}}$TRF predictive for those variables is necessary for the current analysis strategy.

The recalculation of the b-tagging score for each jet is performed by using a larger set of efficiencies to properly describe the probability of a jet to be tagged by each working point. Once the promoted jets are calculated by the standard method, the probability ϵ_j^{XX} of each jet to be tagged by a given operating point `XX%` is estimated via `70%+70%+XX%+XX%` and `70%+70%+70%+XX%` efficiencies for $3b$ and $\geq 4b$ regions respectively (see Figure 7.4). For example, the probability of a non-b-tagged jet j at the 70% working point to be tagged by the 77% in the $3b$ region is given by:

$$\epsilon_j^{77\%} = \frac{\epsilon_{\text{incl.}}^{70\%+70\%+77\%+77\%}(p_{\text{T}}^j, \Delta R_{jj}^{\text{min}} \times N_j)}{1 - \epsilon_{\text{incl.}}^{70\%+70\%+70\%+70\%}(p_{\text{T}}^j, \Delta R_{jj}^{\text{min}} \times N_j)} \tag{7.9}$$

According to ϵ_j^{XX}, a new pseudo-continuous score is associated to each jet in the event. The jets from the *bulk*, i.e. the jets excluded from the t$\bar{\text{t}}$TRF weight computation, are also subjected to this recomputation, with special efficiencies. In fact, their scores depend on the b-jet multiplicity, as shown in Figure 7.8b. The t$\bar{\text{t}}$TRF predicted distribution for the sum of the six highest pseudo-continuous b-tagging scores in the event in the $\geq 10j \geq 4b$ region in the single lepton channel is shown in Figure 7.1.4. This variable is the input variable for the BDT.

This new development has two important outcomes: a data-driven prediction of the pseudo-continuous b-tagging scores and an independence of the method on the used b-tagging working point. In fact, the b-jet multiplicity is recomputed for each working point after the reassignment of the pseudo-continuous b-tagging score to all the jets in the events. Therefore, the working point used for the t$\bar{\text{t}}$TRF estimation is independent from the one used for b-jet definition and region classification (if tighter). The only difference in using one working point instead of another is in the statistics of the prediction. As mentioned in Section 7.1.2, data statistics in the $2b$ region are used to estimate $t\bar{t}$+jets at higher b-jet multiplicity.



(a) (b)

Figure 7.8.: Sum of the six highest pseudo-continuous b-tagging scores in data events for different b-jet multiplicity (a). The goal of $t\bar{t}$TRF is to predict the distributions at $3b$ and $\geq 4b$ regions starting from the $2b$ distribution. Sum of the pseudo-continuous b-tagging values of the two bulk jets for different number of b-jets in the event (b). A clear trend of the scores towards the highest value of 5 is visible by requiring larger b-jet multiplicity. Therefore the pseudo-continuous b-tagging score have to be recomputed by $t\bar{t}$TRF even for the jets excluded from the $t\bar{t}$TRF prediction.

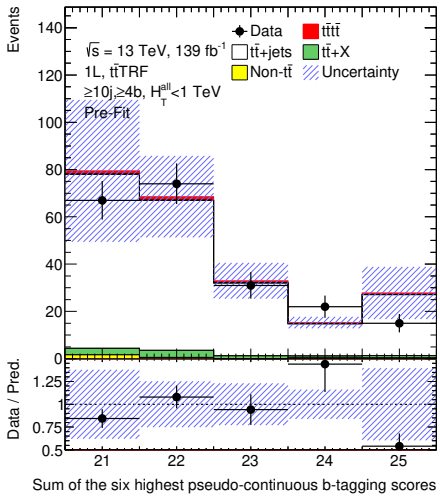


Figure 7.9.: Distribution of the sum of the six highest PC b-tagging scores in single lepton $\geq 10j \geq 4b$ region (c). This variable is used in the BDT training. The displayed uncertainties correspond to $t\bar{t}$ modelling uncertainties on the $t\bar{t}$TRF predicion.

In Tables 7.1 and 7.2, the comparison of the data statistics used for the prediction in $3b$ and $\geq 4b$ regions with a $t\bar{t}$TRF estimation based on 77% working point is shown. Table 7.1 confirms that the statistics from $t\bar{t}$TRF in $3b$ and $\geq 4b$ regions is the same as in the $2b$ region at the same jet multiplicity, since the working point used for the estimation and region definitions are the same. If the statistics of the prediction in $3b$ and $\geq 4b$ regions at 70% working point is compared with the data yields in the corresponding $2b$ region at 70% working point, it is up to five times smaller. Therefore, it is convenient to estimate $t\bar{t}$TRF with the same working point as the one used for the region definition.

$N_{b\text{-jets}}$	1L $8j$	1L $9j$	1L $\geq 10j$	OS2L $6j$	OS2L $7j$	OS2L $\geq 8j$
Nominal data						
2b (77%)	64132	19042	7245	15230	4776	1961
3b (77%)	17458	6013	2805	2670	1033	572
$\geq$4b (77%)	3316	1403	848	418	232	154
$t\bar{t}$TRF data – using MV2c10 77% for the estimation						
3b (77%)	64132	19042	7245	15230	4776	1961
$\geq$4b (77%)	64132	19042	7245	15230	4776	1961

Table 7.1.: Comparison of the yields of data and the $t\bar{t}$TRF prediction (based on MV2c10 77% WP) in control and signal regions for the single lepton (1L $8j$, $9j$ and $\geq 10j$ columns) and the opposite-sign dilepton (OS2L $6j$, $7j$ and $\geq 8j$ columns) channels at different b-jet multiplicities ($N_{b\text{-jets}}$ column), using the MV2c10 77% WP.

$N_{b\text{-jets}}$	1L $8j$	1L $9j$	1L $\geq 10j$	OS2L $6j$	OS2L $7j$	OS2L $\geq 8j$
Nominal data						
2b (70%)	56121	16961	6655	12959	4134	1786
3b (70%)	9920	3496	1654	1565	640	360
$\geq$4b (70%)	1511	649	419	222	117	79
$t\bar{t}$TRF data – using MV2c10 77% for the estimation						
3b (70%)	24546	7389	2749	7253	2222	940
$\geq$ 4b (70%)	12528	3785	1521	5402	1709	746

Table 7.2.: Comparison of the yields of data and the $t\bar{t}$TRF prediction (based on MV2c10 77% WP) in control and signal regions for the single lepton (1L $8j$, $9j$ and $\geq 10j$ columns) and the opposite-sign dilepton (OS2L $6j$, $7j$ and $\geq 8j$ columns) channels at different b-jet multiplicities ($N_{b\text{-jets}}$ column), using the MV2c10 70% WP.

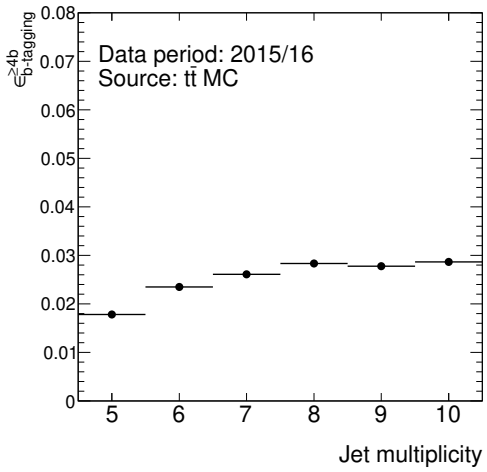


Figure 7.10.: Evolution of $\epsilon_{\geq 4b}$ with jet multiplicity estimated on $t\bar{t}$ MC simulation.

7.1.5. Extrapolation of the dependence on jets multiplicity

The $t\bar{t}$TRF method can be improved by reducing the impact of the MC correction factor. It would give better closure to data of the uncorrected $t\bar{t}$TRF distributions and a consequent reduction of the systematic uncertainty associated with the prediction. A large MC correction factor is symptomatic of the MC correcting for large and/or many effects that are not covered by the data-driven $t\bar{t}$TRF estimation. Systematic variations of the $t\bar{t}$ modelling might describe these effects in different ways, resulting in a correction factor associated to the variation that is far from the nominal one. By evaluating the effects the MC correction factor shold correct in data, the systematic variations would provide a similar MC correction to the nominal prediction, ending up in a smaller systematic uncertainty on the $t\bar{t}$TRF estimation.

The dominant effect that the MC corrects is the evolution of the efficiency with the number of jets, which is significant especially in the $\geq 4b$ efficiencies, as shown in Figure 7.10. The estimation of the evolution of the efficiencies with jet multiplicity would drastically reduce the effect of the MC correction factor. This dependency can be estimated by using the regions with lower jet multiplicity, as illustrated in Figure 7.11: the efficiency extraction region is extended by one more jet and the events are used to fit a functional dependence of the efficiency on jet multiplicity.

This extrapolation is based on the measured fraction $f^{\tilde{b}}_{\tilde{j}}$ of events with $\tilde{j}$ more jets than the extraction region for the *baseline* $t\bar{t}$TRF (described so far) and $\tilde{b}$ radiated b-jets (equal to two b-jets less than the total number of b-jets in the event). The fraction of

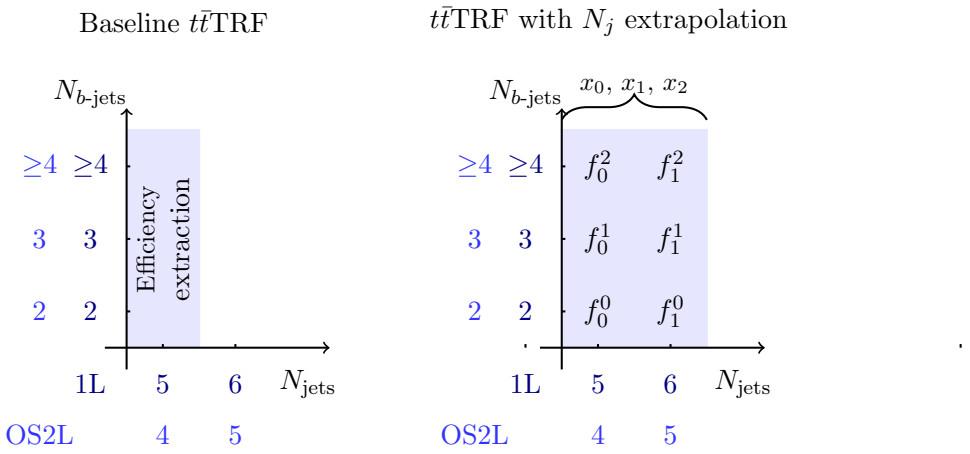


Figure 7.11.: Change of the efficiency extraction regions of the *baseline* t$\bar{\text{t}}$TRF (described so far, in the left scheme) for the t$\bar{\text{t}}$TRF with the jet multiplicity dependence estimation (N_j extrapolation, in the right scheme). Here, f_x^y stands for fraction of the events with x more jets than in the extraction region and y radiated b-jets. In terms for b-jet and jet multiplicity they correspond to the events in $(5{+}x)j(2{+}y)b$ for the single lepton and in $(4{+}x)j(2{+}y)b$ for the opposite-sign dilepton channel. Those values are used to extract the parameters x_0, x_1 and x_2.

events with an additional jet $f^{\tilde{b}}_{\tilde{j}+1}$ is estimated via:

$$f^{\tilde{b}}_{\tilde{j}+1} = x_0 \cdot f^{\tilde{b}}_{\tilde{j}} + x_1 \cdot f^{\tilde{b}-1}_{\tilde{j}} + x_2 \cdot f^{\tilde{b}-2}_{\tilde{j}} \tag{7.10}$$

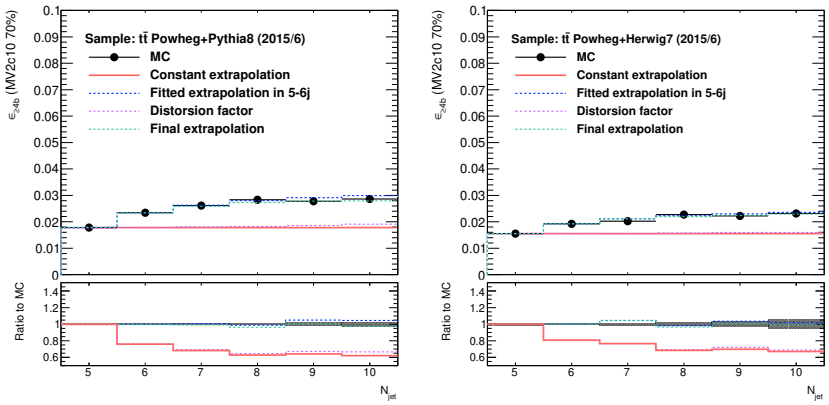
where x_0 stands for the probability of not-tagging the extra jet, x_1 is the probability of tagging the extra jet and x_2 the probability that tagging an extra-jet makes another jet tagged[3]. The expression of the efficiencies as a function of the fractions $f^{\tilde{b}}_{\tilde{j}}$ are:

$$\epsilon_{\text{incl.}}(\tilde{j}) = \frac{1}{\tilde{j}} \sum_{\tilde{b}=1}^{\tilde{j}} \tilde{b}\ f^{\tilde{b}}_{\tilde{j}} \tag{7.11}$$

$$\epsilon_{\geq 4b}(\tilde{j}) = \frac{1}{\tilde{j}-1} \sum_{\tilde{b}=2}^{\tilde{j}} (\tilde{b}-1)\ f^{\tilde{b}}_{\tilde{j}} \tag{7.12}$$

representing the fitting functions in the region shown in Figure 7.11, needed to estimate the probabilities x_0, x_1 and x_2 of Equation 7.10.

[3]This is again to take into account the change of the probability in adding a couple of b-jets due to gluon splitting.



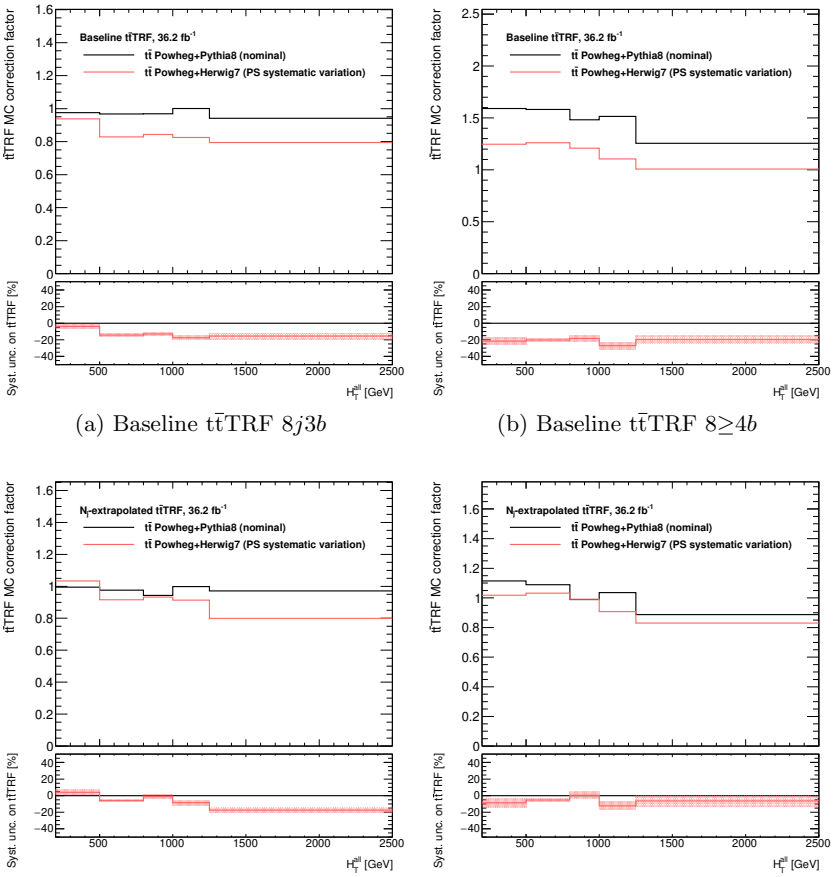
(a) (b)

Figure 7.12.: Comparison of the efficiencies evolution with jets multiplicity with and without $t\bar{t}$TRF extrapolation. This is shown for nominal $t\bar{t}$ simulation (a) and its systematic variation of parton shower and hadronisation (b). Uncertainties associated to the efficiencies and their extrapolation are displayed in the plots and are found to be negligible.

A global picture of the dependence of the efficiencies with the jet multiplicity and its estimation by $t\bar{t}$TRF is shown in Figures 7.12a and 7.12b. The points in the plots correspond to the MC efficiencies extracted at different jet multiplicities. The closer the $t\bar{t}$TRF prediction is to the latter, the smaller the correction factor is. The blue dashed line corresponds to the fitted dependence in the regions with low jet multiplicity, as described above. This has to be further corrected by the violet trend, that is the drift of the average efficiency due to the change of jet kinematics. The result of this is the green solid line. The distance of the green line to the MC points has to be compared with the distance of the red line, that is what is applied without any extrapolation: the much smaller separation indicates a large reduction of the MC correction factor.

Figures 7.12a and 7.12b show a reduction of the MC correction factor both for the nominal prediction and its systematic variations. The actual effect on systematic uncertainty associated to $t\bar{t}$TRF prediction is evident by comparing Figures 7.13a and 7.13b with Figures 7.13c and 7.13d where the MC correction factors for the nominal prediction and its largest systematic variation are compared, with and without the application of the described estimation of the jet multiplicity dependence. The two MC correction factors get closer when this estimation is applied, reducing the error associated with the $t\bar{t}$TRF prediction.

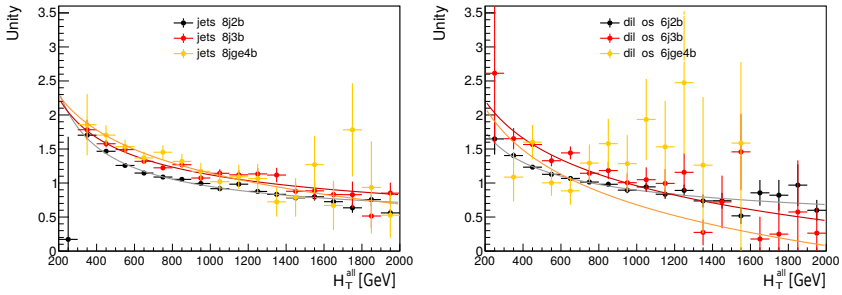
Despite this upgrade being beneficial for the method predictions, it is not used in the following results since it is computationally expensive to combine the jet multiplicity dependence extrapolation with the prediction of b-tagging scores. Since the BDT needs the inclusion of PC b-tagging variables, their estimation has been the priority.



(a) Baseline t$\bar{t}$TRF 8j3b (b) Baseline t$\bar{t}$TRF 8≥4b

(c) N_j-extrapolated t$\bar{t}$TRF 8j3b (d) N_j-extrapolated t$\bar{t}$TRF 8j≥4b

Figure 7.13.: Comparison of MC correction factor as a function of $H_{\mathrm{T}}^{\mathrm{all}}$ for the nominal prediction and the systematic variation associated to the choice of the parton shower and hadronisation model (PS systematic variation). The bottom panel of each plot show their ratio, corresponding to the associated systematic uncertainty on the prediction. Columns show the comparison in different regions in the single lepton channel: 8j3b and 8j≥4b (the same as in Figures 7.6a and 7.6b). In the top row, the t$\bar{t}$TRF prediction is obtained without correcting for jet multiplicity dependence (a,b), while this correction is applied in the bottom row (c,e).



(a) 1L $8j$ (b) OS2L $6j$

Figure 7.14.: Data to MC ratio in single lepton $8j$ regions (a) and opposite-sign dilepton $6j$ regions (b) at different b-jet multiplicities.

7.2. The $t\bar{t}$ reweighting method

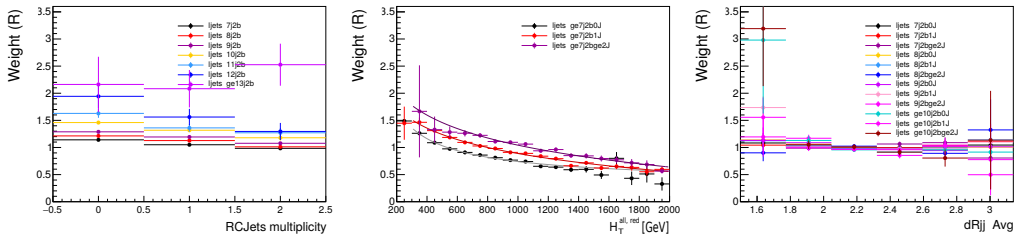
Unlike the t$\bar{\text{t}}$TRF, the $t\bar{t}$ *reweighting method* is a MC-based data-assisted method: it aims to exploit the power of the MC simulation by adjusting it with information from data. The method assumes that the discrepancy between data and simulation (see Figures 7.1a to 7.1f and 7.2a to 7.2f) in the regions used for the fit comes from the modelling of the jet radiation from $t\bar{t}$. The shape discrepancy in p_{T}-related variables is assumed to be independent from the flavour of the radiated jets, since at high jet multiplicity, most of the jet radiation is composed of jets from light-quarks. This assumption supported by Figures 7.14a and 7.14b where the shape in $H_{\text{T}}^{\text{all}}$ of the ratio of $t\bar{t}$ in data and MC is compatible between regions at different b-jet multiplicities. The normalisation discrepancy is eliminated by estimating the rates of the different radiation components in the fit, making use of the dedicated nuisance parameters introduced in Section 6.2.

The correction of the shape mismodelling is derived in $2b$ exclusive regions and then applied to the regions with $\geq$3 b-jets. The reweighting factor $R(x)$, depending on variable x, is defined as:

$$R(x) = \frac{\text{Data}(x) - \text{MC}^{\text{non-}t\bar{t}}(x)}{\text{MC}^{t\bar{t}}(x)} \tag{7.13}$$

It assigns the observed mismodelling in $2b$ regions entirely to $t\bar{t}$ mismodelling. Three different variables have been chosen for the reweighting, addressing different kinematic aspects:

- Jet multiplicity (N_j): this solves the MC underestimation of jets radiation. It fixes the discrepancy in normalisation at each jet multiplicity. Therefore the following steps correct only for the remaining discrepancy in the shapes of the kinematic distributions.
- $H_{\text{T}}^{\text{all,red}}$: it corresponds to $H_{\text{T}}^{\text{all}}$ "normalised" by the number of jets in the event. It



(a) (b) (c)

Figure 7.15.: Reweighting scheme for single lepton channel (ljets): first, the weight as a function of the jet multiplicity is computed (a). Afterwords, the weights as function of $H_T^{\text{all,red}}$ (b) and $\Delta R_{\text{jets}}^{\text{avg}}$ (c) are extracted. A different weight is computed at different RC-jet multiplicities (J). Regions are labelled as $X\text{j}Y\text{b}Z\text{J}$, indicating a region with X jets, Y b-jets and Z RC-jets.

is defined as:

$$H_T^{\text{all,red}} = H_T^{\text{all}} - 90 \text{ GeV} \times N_j \tag{7.14}$$

motivated by an average increase of 90 GeV in H_T^{all} for every additional jet. This reweighting accounts for the observed mismodelling of p_T-related variables. To make $R(H_T^{\text{all,red}})$ a smooth function, it is fitted with an ad-hoc functional model:

$$f(x) = a + \frac{b}{x^c} \tag{7.15}$$

- $\Delta R_{\text{jets}}^{\text{avg}}$: this accounts for angular information in the event and has shown to improve the agreement in the distributions of angular variables. Here, a binned distribution is used for the reweighting.

All the $R(x)$ are extracted separately for different RC-jets multiplicity to eliminate residual observed trends in RC-jet-related variables. This procedure is done in steps, thus correlations between the variables are not taken into account. However, these correlations have been shown to be not significant. The values of $R(x)$ in the single lepton channel for the three variables mentioned above are shown in Figures 7.15a to 7.15c.

After the reweighting, the MC prediction and data agree perfectly in $2b$ regions. A normalisation discrepancy is expected to arise when moving to higher b-jet multiplicity. Systematic variations of the $t\bar{t}$ simulation are also propagated through the method to obtain the reweighted prediction. In principle, a different reweighting should be derived for each systematic variation. However, assuming that the changes of the minor backgrounds are negligible, the reweighting for the $t\bar{t}$ systematic variation $R^{\text{systematic}}(x)$ is:

$$R^{\text{systematic}}(x) \approx R(x) \frac{\text{MC}^{t\bar{t},\text{ nominal}}}{\text{MC}^{t\bar{t},\text{ systematic}}} \tag{7.16}$$

Uncertainties on the reweighting method itself, such as uncertainties of the the fitted parameters of $H_{\mathrm{T}}^{\mathrm{all,red}}$ and statistical uncertainties of the binned distributions will be included in the fit.

7.3. Comparison of the performance of the methods

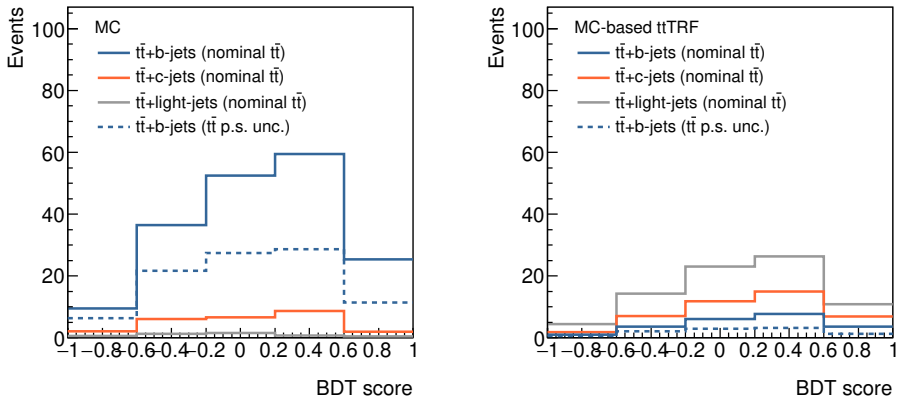
The two methods are compared in two main aspects:

- *$t\bar{t}$+jets modelling*: the predictions of the methods are checked against observed data. Particularly important is the modelling of the variables used as input for the BDT.
- *Fit behaviour and expected performance*: pre-fit and post-fit modelling of the discriminant variable (BDT score) is compared, as well as the fitted values of the nuisance parameters (NPs) and the expected significance of the $t\bar{t}t\bar{t}$ signal.

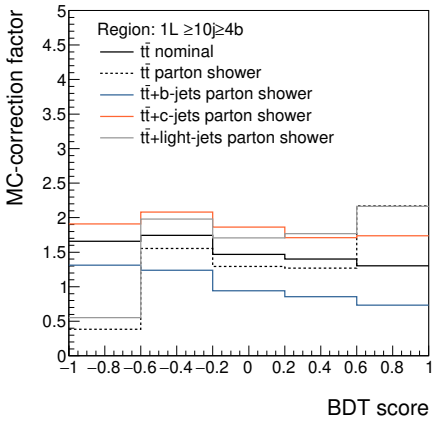
Before going through the comparison, a difference in the treatment of $t\bar{t}$+jets modelling uncertainties should be noted. In the reweighting method, the modelling uncertainties associated to the $t\bar{t}$+jets production are separated into different nuisance parameters depending on the flavour of the $t\bar{t}$ radiation. The description of the flavour classification and related nuisance parameters is given in Section 6.2. This classification and corresponding splitting of the uncertainties is difficult to apply to the t$\bar{\text{t}}$TRF prediction that is based on events in data. However, this classification could be applied through the MC correction factor.

As described in Section 7.1, the uncertainties in t$\bar{\text{t}}$TRF are estimated by comparing the MC correction factor of the variation with the nominal one, allowing the uncertainties to be assessed separately for each flavour of the $t\bar{t}$ radiation by splitting the MC correction factor. However, since the efficiencies on MC are extracted from the inclusive $t\bar{t}$ spectrum and not separately for each radiation component, this leads to a large overestimation of the uncertainties. To better clarify this problem, Figure 7.16a shows the ingredients for the construction of a flavour-splitted uncertainty in the t$\bar{\text{t}}$TRF method, for instance, the parton shower uncertainty on the $t\bar{t}$+b-jets modelling in the single lepton channel in the $\geq$10$j$$\geq$4$b$ region. The key point is the difference in composition of the MC sample and the corresponding t$\bar{\text{t}}$TRF prediction: as expected in a $\geq$4b region, the $t\bar{t}$+b-jets component is the largest in the MC, whereas the corresponding t$\bar{\text{t}}$TRF prediction is dominated by $t\bar{t}$+light-jets events, because it is based on reweighting of events in the 2b-exclusive region.

The parton shower uncertainty on the $t\bar{t}$+b-jets modelling is evaluated by computing the corresponding MC correction factor made of the $t\bar{t}$+b-jets component from the simulation with the alternative parton shower algorithm (dashed blue line in Figure 7.16a) and the other components from the nominal simulation (grey and orange solid lines in 7.16a). The numerator of the MC correction factor is given by the MC estimate, while the denominator is from the corresponding t$\bar{\text{t}}$TRF prediction. The uncertainty



(a)



(b)

Figure 7.16.: Distributions of the BDT score for the components used for the evaluation of the uncertainty associated to the choice of the parton shower and hadronisation algorithm on the $t\bar{t}$+b-jets prediction with the tt̄TRF method (a). The yields of the different flavour components of the $t\bar{t}$ radiation in the MC prediction and in the MC-based tt̄TRF prediction are shown on the left and right, respectively. The figures refer to the $\geq 10j \geq 4b$ region in single lepton (1L) channel. Comparison of the corresponding MC correction factors for "flavour-splitted" and inclusive variations with the nominal one (b).

is given by the comparison of this correction factor with the one based on the nominal simulation, obtained by summing up all the solid-line distributions in Figure 7.16a. Therefore, the variation of the $t\bar{t}$+b-jets component largely affects the numerator, but has a very small effect in the denominator, providing an uncertainty that is very close to the direct comparison of the alternative sample with the nominal one. The power of the t$\bar{\text{t}}$TRF method of reducing the systematic uncertainty with the ratio of the MC and MC-based t$\bar{\text{t}}$TRF predictions is would then not be exploited. Moreover, a similar impact occurs when the $t\bar{t}$+light-jets component is considered, resulting in an overestimation of the uncertainties. The various MC correction factors compared to the nominal one are presented in Figure 7.16b, showing that the inclusive parton shower correction factor is the closest to the nominal in the core of the distribution. This confirms that the uncertainties in the t$\bar{\text{t}}$TRF method must not be separated in radiation flavours to avoid their overestimation.

The general modelling of $t\bar{t}$+jets is discussed in Section 7.3.1 and the fit behaviour and performance are presented in Section 7.3.2. The conclusions are discussed in Section 7.3.3.

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

To fairly compare the two estimation methods, their different goals are mentioned again. The t$\bar{\text{t}}$TRF method provides a robust model of the background already at pre-fit stage and potentially of any variable of interest. The only limitation is given by the strong role of the MC correction factor: limited statistics of simulation (nominal or systematic variation) in a given region may end up in a large variation of the MC correction factor and therefore in mismodelling or overestimation of the systematic variation. This can be anyway cured by smoothing the MC correction factor consistently with the associated statistical uncertainty or by rebinning the distributions. The reweighting method instead does not aim to describe data at pre-fit stage, since normalisation of the flavour components of the $t\bar{t}$ radiation are estimated by the profile likelihood fit. However, it aims to remove any shape discrepancy in any distribution.

An example of the background predictions compared to data in the $H_{\text{T}}^{\text{all}}$ variable close to the most sensitive regions is shown in Figures 7.17a to 7.17h. These regions are used to check the modelling at the highest jet and b-jet multiplicity without any blinding cut applied. A large improvement in the background modelling has to be noted for both methods, if compared with the nominal MC prediction (see Figures 7.1c, 7.1e, 7.2c and 7.2e). The t$\bar{\text{t}}$TRF prediction has no significant bias either in the normalisation or in the shape of the distributions. The reweighting method, as expected, has no trends in the distribution shapes and it underestimates the normalisation of about $15-20\%$, quite constant in the number of b-jets in the event. Concerning systematic uncertainties, t$\bar{\text{t}}$TRF shows smaller uncertainties thanks to the properties of the MC correction factor. This outcomes are valid for all the input variables of the BDT[4].

[4] As mentioned in Section 5.3.2, the input variables of the BDT have been chosen based on the per-

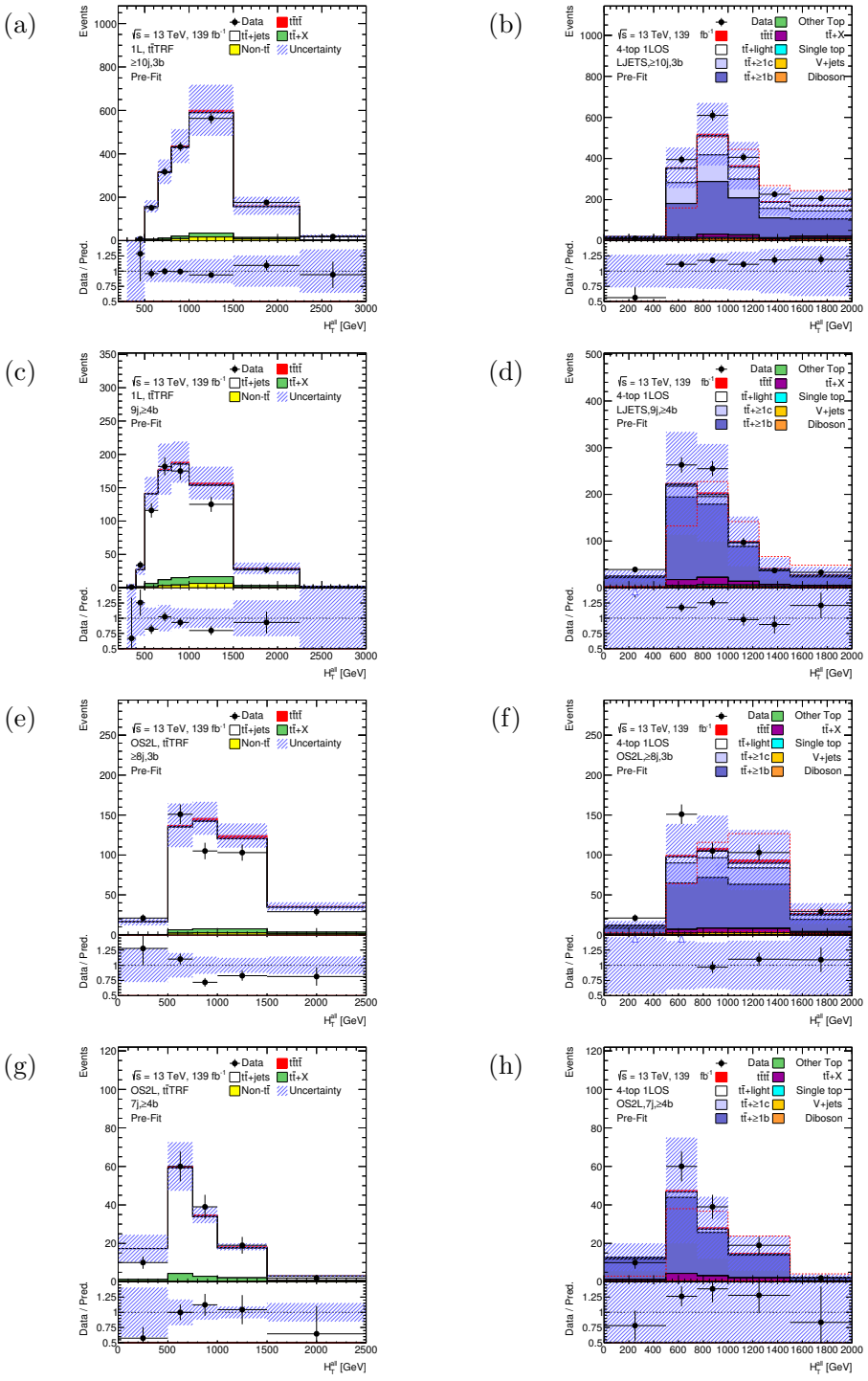


Figure 7.17.: Pre-fit modelling of the $H_{\mathrm{T}}^{\mathrm{all}}$ variable for the t$\bar{\mathrm{t}}$TRF (left column) and reweighting (right column) methods in the closest regions to the most sensitive ones: $9j{\geq}4b$ (a,b) and $\geq 10j3b$ (c,d) regions in the single lepton channel and $7j{\geq}4b$ (e,f) and $\geq 8j3b$ (g,h) regions in the opposite-sign dilepton channel.

7.3.2. Fit behaviour and peformance

As described in Section 5.2, the fit adjusts the model to agree with observed data by moving the NPs. Therefore, checking the control regions is important to understand the fit behaviour, since those regions are the most important ones to constrain the background prediction through fitting the associated NPs. In particular, the regions at lower jet multiplicity, with larger statistics, are most important for the background fitting. The comparison of the predictions in these regions for the two methods is summarised in Figure 7.18. The BDT score is modelled as well as any other variable by the two methods. For the $t\bar{t}$TRF prediction, a trend of larger uncertainties in the high and low tails of the BDT score distributions is evident. Poor statistics is not responsible for this effect, as shown in Figure 7.20. The reason is the selection of a particular corner of phase space in which the b-tagging efficiencies of the systematic variation are significantly different from the average values used for the $t\bar{t}$TRF estimation. In this case, the MC correction factor of the variation is far from the nominal in the considered bin, leading to large systematic uncertainties. This is supported by observing this effect only in the tails of the distributions that do not influence the value of the average efficiencies, and where the composition of the radiation is expected to be different from the average, because $t\bar{t}$+b-jets and $t\bar{t}$+light-jets tend to accumulate at low and high values of the BDT score, respectively. The reweighting method eliminates any shape discrepancy between the prediction and data and show way larger uncertainties than the $t\bar{t}$TRF prediction. This is because the $t\bar{t}$TRF method reduces the impact of the systematics on the prediction. However, the reweighting method needs large uncertainties to cover the normalisation discrepancy, that is expected to be eliminated by the adjustment of the NPs by the profile likelihood fit.

To test the fit behaviour, a fit on data in the control regions is performed. This fit is made by fixing the signal strength of the $t\bar{t}t\bar{t}$ signal to one and by fitting all the remaining NPs. This setup gives information on the pulls and constraints of the NPs that are driven by the control regions. Moreover, it gives information on the expected uncertainty of the background after the fit, and therefore on the signal sensitivity. In the following, the results from the single lepton channel are shown. The results from the opposite-sign dilepton show similar trends, with smaller constraints due to the lower statistical power. More details can be found in Appendix A.

The main differences between the fitted NPs in the two methods are seen in the NPs associated to $t\bar{t}$+jets modelling and jet uncertainties, shown in Figure 7.19. The $t\bar{t}$TRF method yields in general stronger constraints on the NPs associated to $t\bar{t}$ modelling. This is motivated by the lower number of considered NPs, that reduce the degrees of freedom of the fit. In fact, by splitting the $t\bar{t}$ modelling uncertainties into different flavours of the $t\bar{t}$ radiation, the reweighting method makes a single variation separated into three smaller ones. Therefore, the power of the data to strongly constrain each

formance of the reweighting method, selecting only well-modelled variables. The $t\bar{t}$TRF method has instead shown robust predictions without any optimisation of the variables.

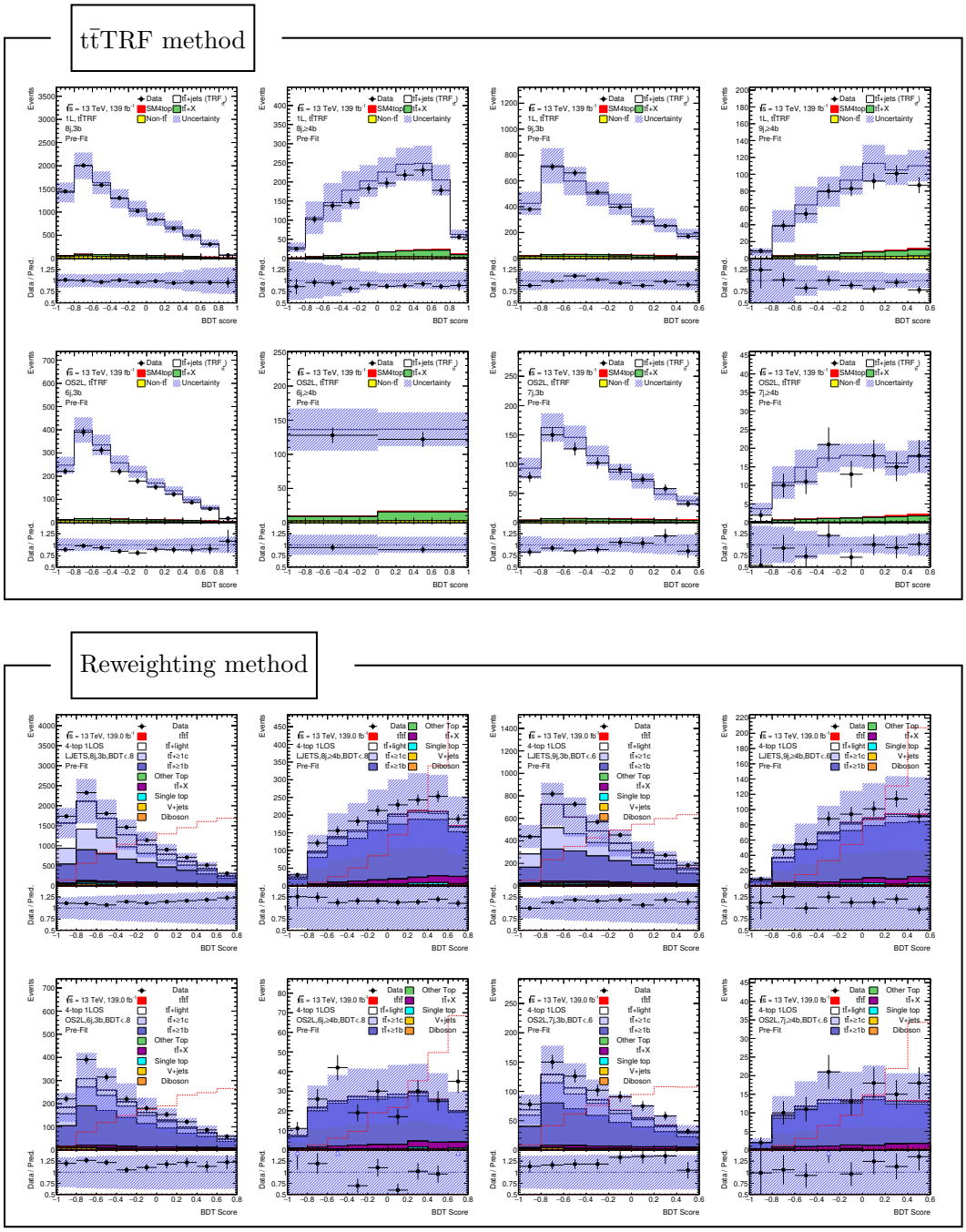


Figure 7.18.: Comparison of the prediction for $t\bar{t}$TRF (top block) and reweighting (bottom block) for the fit control regions at lower jet multiplicity. In each block, the plots for the single lepton channel are in the upper row, while the plots for the opposite-sign dilepton channel are shown in the bottom row. Plots for the regions $8j3b$, $8j{\geq}4b$, $9j3b$, $9j{\geq}4b$ of the single lepton channel are shown from left to right. Plots for the regions $6j3b$, $6j{\geq}4b$, $7j3b$, $7j{\geq}4b$ of the opposite-sign dilepton channel are shown from left to right.

of them becomes lower. The reason of the strong constraints in the $t\bar{t}$TRF prediction for the uncertainties associated to $t\bar{t}$ parton shower algorithm and choice of *hdamp* parameter is their impact on the tails of the BDT distributions, as shown in Figure 7.20. The good pre-fit agreement of the $t\bar{t}$TRF prediction with data even in the tails of the distributions of the BDT score in the control regions with large statistics (see Figure 7.18), strongly constrains these NPs to values close to zero. Therefore, the residual mismodelling is covered by the movement of the other NPs, such as jet uncertainties, as shown in Figure 7.19. Indeed, these NPs are a bit more constrained and pulled for $t\bar{t}$TRF method than in the reweighting. This can again be explained by the presence of more and smaller NPs associated to $t\bar{t}$+jets modelling in the setup for the reweighting method. These NPs play the same role as the NPs associated to jet uncertainties in the $t\bar{t}$TRF method. In the reweighting method, however, pulls in the NPs associated to the $t\bar{t}$+c/b-jets cross-sections are expected to cover the normalisation discrepancy with data observed at the pre-fit level. No significant pulls are instead shown in the values of these NPs (Figure 7.19), while a deviation from the pre-fit value is measured in the NPs associated to the choice of the parton shower and hadronisation model. Therefore, the fit of the model obtained from the reweighting method seems not to behave exactly as expected.

Summarising, in both methods, the fit results are quite healthy and the pulls and constraints of the NPs are mostly understood. The same constraints are observed by substituting the data with an Asimov dataset. The Asimov dataset is a pseudo-dataset that corresponds to the pre-fit predicted model, but with data statistical uncertainty. No pulls of the NPs are expected, while constraints arise from their correlations. In the reweighting method, the Asimov dataset underestimates the background, because the pre-fit model undershoots data in the control regions.

After the fit, the post-fit distributions are compared to assess the expected background uncertainties. The post-fit distributions of the BDT score in the most sensitive regions for the two methods are shown in Figure 7.21. The background uncertainties are similar in the two methods, but slightly smaller for the reweighting method. This is due to the larger statistical uncertainty associated to the $t\bar{t}$TRF prediction. This is further confirmed by comparing the results of a fit of the complete background and signal model on the whole BDT spectrum, using an Asimov dataset. The different impact of the systematics is shown in Figure 7.22: $t\bar{t}$TRF results depend a lot more on the statistical uncertainty of the model than for the reweighting method. This is expected since the statistics of the $t\bar{t}$TRF prediction is given by events in data. The modelling uncertainty plays a dominant role in both methods. The huge impact of the uncertainty associated to the choice of the *hdamp* parameter in the $t\bar{t}$TRF prediction is motivated by the impact of the variation in the high tail of the BDT distribution (see Figure 7.20).

Finally, the expected signal significances for $t\bar{t}t\bar{t}$ using an Asimov dataset are compared. The significance indicates the deviation of the observed data in units of the standard deviation (post-fit background uncertainty) from the background-only hypothesis, and it is reported in Table 7.3. Three different values are quoted for both methods

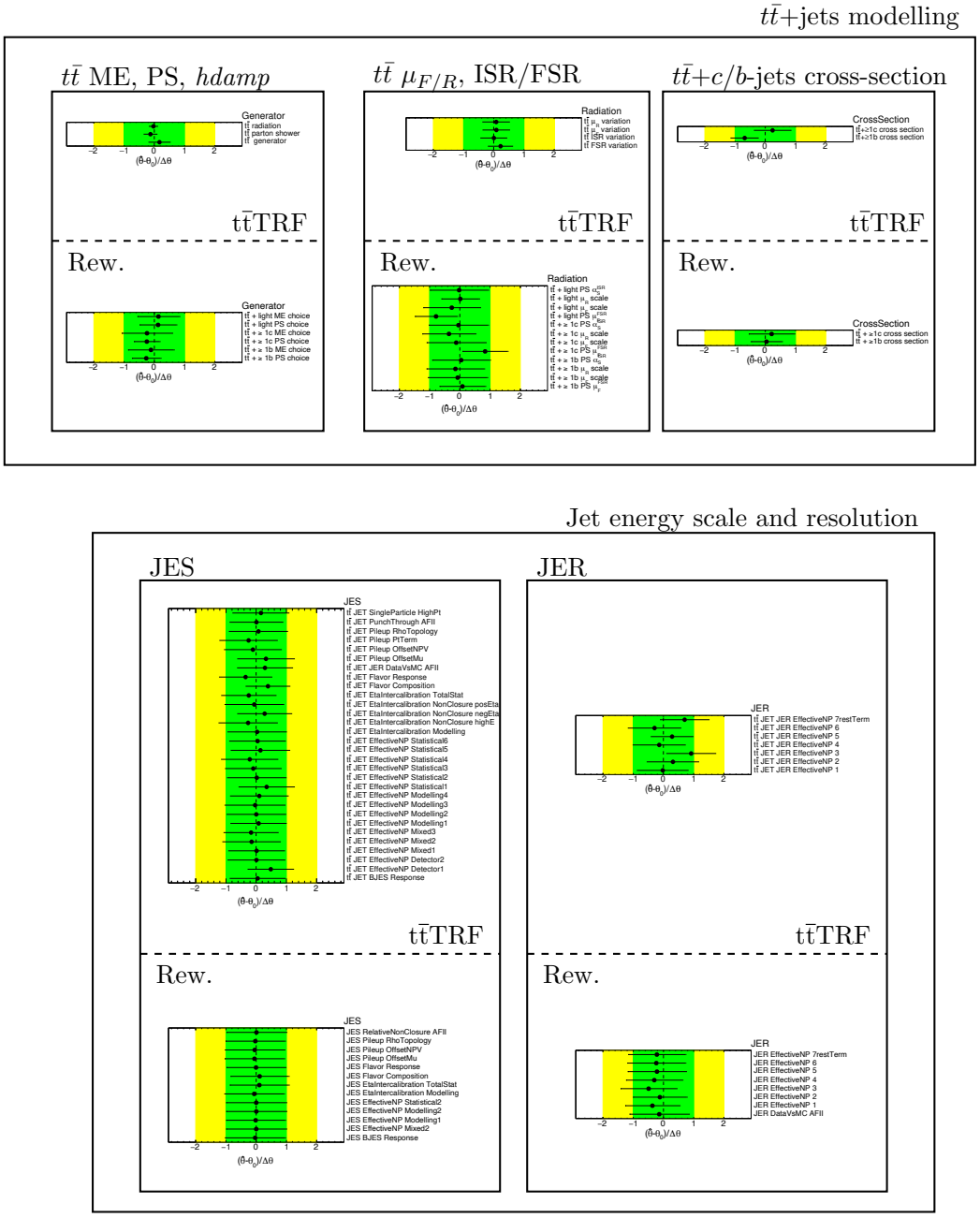
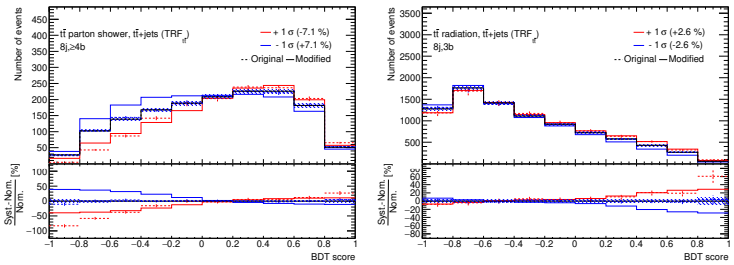


Figure 7.19.: Fitted values of the NPs associated to $t\bar{t}$ modelling (top block) and jet uncertainties (bottom plot) from a background-only fit in control regions for the single lepton channel. Each block is divided in different sources of uncertainties, showing the values for the ttTRF and reweighting (Rew.) predictions. Because of the splitting of the $t\bar{t}$ modelling uncertainties in flavours of the $t\bar{t}$ radiation, the reweighting method has more NPs associated to $t\bar{t}$ modelling. Slightly different naming conventions are used, *ME choice* corresponds to *generator* and *PS* to *parton shower*. The $t\bar{t}$ *radiation* uncertainty, related to the choice of the *hdamp* parameter, is not included for the reweighting method. The green and yellow bands around zero indicate the pre-fit uncertainties at 68% and 95% confidence level. Points stand for the fitted value of the NPs with the associated uncertainties as fractions of the respective pre-fit standard-deviations.



(a) (b)

Figure 7.20.: Pre-fit uncertainties associated to $t\bar{t}$ parton shower and hadronisation modelling in 1L $8j{\geq}4b$ region (a) and *hdamp* choice (b) in 1L $8j3b$ region as a function of the BDT score. 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.

in each channel: "statistical-only" (Stat.), "full systematics" (Full) and "injected NPs" (inj. NPs) significances. The "statistical-only" significance takes only data statistics into account for the estimation. The "full systematics" uses the full systematics model to assess the background uncertainty and thus the significance. However, since the reweighting method underestimates the background in the Asimov dataset, its estimation of these two significances are overestimated. To obtain a realistic value, the fitted nuisance parameters in the control regions are used for the construction of the Asimov dataset, defining the "injected NPs" significance. While no big change is expected to occur in the $t\bar{t}$TRF estimation, this affects significantly the reweighting prediction. In the combination of the two channels, the two methods yield very similar values for the $t\bar{t}t\bar{t}$ significance. The combination in $t\bar{t}$TRF benefits from the constraints in the uncertainty on $t\bar{t}$ parton shower model in the single lepton channel, that increases the sensitivity of the opposite-sign dilepton, in which this uncertainty is has a significant impact on the fitted signal strength.

7.3.3. Conclusions from the comparison of the methods

Neither of the two methods has shown a clear superior performance. Concerning the modelling of the $t\bar{t}$+jets, both methods successfully reach the respective goals: the $t\bar{t}$TRF provides a robust prediction with small uncertainties while the reweighting method provides a prediction that is off the observation only by a normalisation factor. The larger uncertainties allow the method to extract proper normalisation by fitting the distributions to data using the profile likelihood method.

The fit behaviour is quite different because of the different numbers of included NPs

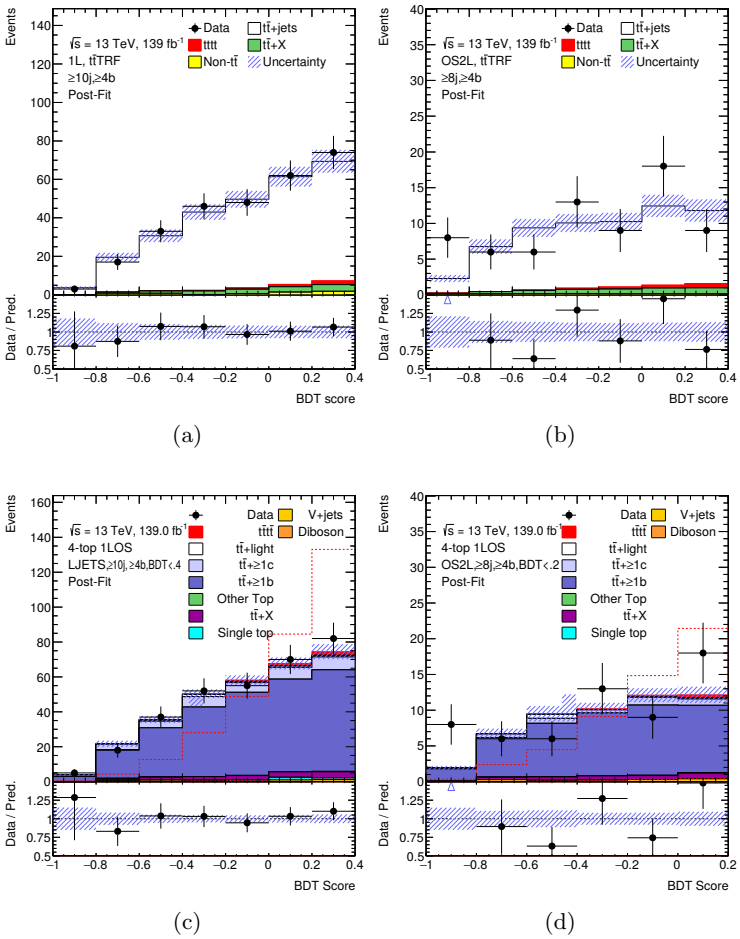
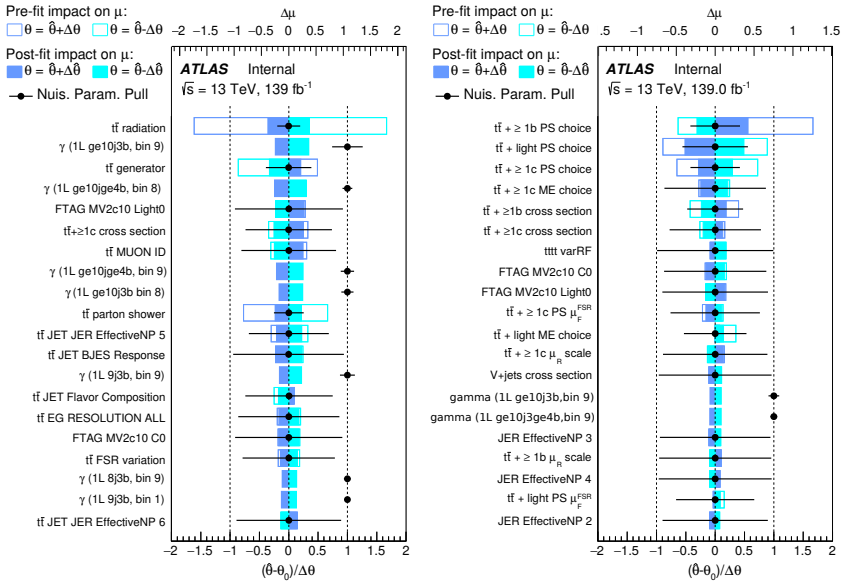


Figure 7.21.: Post-fit distributions in the unblinded bins for the most sensitive regions in the 1LOS channel for tt̄TRF (a,b) and reweighting (c,d) methods.

	1L			OS2L			Combination		
	Stat.	Full	inj. NPs	Stat.	Full	inj. NPs	Stat.	Full	inj. NPs
tt̄TRF	1.50	0.70	0.74	1.20	0.61	0.62	1.90	1.00	1.04
Rew.	1.80	0.87	0.77	1.40	0.83	0.76	2.20	1.19	1.06

Table 7.3.: Values for expected significance from Asimov fit. The *Stat.* values, take into account only data statistical uncertainty for the significance estimation. The *Full* values instead also consider the full post-fit systematic uncertainty. The *inj. NPs* values, consider the full post-fit systematic uncertainty from a fit on an Asimov dataset modelled by the injection of the fitted NPs in the control regions on data.



(a) (b)

Figure 7.22.: Ranking of the most important systematics in a fit on an Asimov dataset in the single lepton channel. The systematics are ranked in terms of the impact of the post-fit systematic on the fitted value of $\mu_{t\bar{t}t\bar{t}}$. The ranking with the t$\bar{t}$TRF model is in (a), while the same plot for the reweighting method is in (b). Points refer to the fitted values of the NPs with the corresponding uncertainty as fraction of the pre-fit standard-deviation of the assumed Gaussian distribution. The filled (empty) bands indicate the impact on the results $\Delta\mu$ for a variation of the NPs of one post-fit (pre-fit) standard deviation. The γ parameters refer to statistical uncertainty on the background model and are assumed to be Poisson-distributed NPs.

and statistics of the predictions. The splitting in flavours of the $t\bar{t}$ radiation for the reweighting method gives more freedom to the fit, resulting in less pulls and constraints of the fitted NPs. The t$\bar{t}$TRF does not use such splitting, constrains more the dominant NPs that contribute especially in the high and low tails of the BDT distributions. Therefore, the remaining non-strongly-constrained NPs compensate for the remaining small mismodelling, resulting in an overall larger number of slightly pulled NPs. It should be mentioned again that the estimation of the t$\bar{t}$TRF uncertainties is suboptimal and can be improved by following the procedure described in Section 7.1.5. The same systematic model as the reweighting method could be also implemented by splitting the efficiency extraction on MC sample in flavours of the $t\bar{t}$ radiation. Finally, the t$\bar{t}$TRF model yields larger statistical uncertainty than the reweighting methods since based on data events. These differences end up anyway in similar performances between the two methods in

terms of post-fit uncertainties and sensitivity.

From a philosophical point of view, the $t\bar{t}$TRF has a more solid background and it is mathematically rigorous. This is not the case for the reweighting method, in which the procedure and the parametrisation are quite arbitrary. On the other hand, $t\bar{t}$TRF is more complex and requires larger computing and storage capacities than the reweighting method, that is also more flexible.

In the ATLAS collaboration, a management decision based on the timeline and available personpower has been taken to adopt the reweighting method as the baseline method for the $t\bar{t}t\bar{t}$ signal measurement. The $t\bar{t}$TRF method is being used as a cross-check and will help to solidify the results. The decision is however not based on any clear physics motivation, as shown in this section.

CHAPTER 8

Results

The results from the fit in the single lepton and opposite-sign dilepton channel described in this thesis are obtained by making use of the $t\bar{t}$TRF method for the background estimation, and presented in Section 8.1. The results from this channel are then combined with public results in the SSML channel [36]. A description of the combined results is given in Section 8.2.

8.1. Measurement in the single lepton and opposite-sign dilepton channel

The fit setup for the 1LOS channel has been already described in Section 5.3. However, after studies of the fit, described in Appendix B.2, a different binning in the opposite-sign dilepton channel has been adopted to reduce the impact of the statistical uncertainty of the $t\bar{t}$TRF prediction. An overview of the results from the two channels is given in this section. More details on the fit results are in Appendix C.

The pre-fit and post-fit distributions for the most sensitive regions of the two channels are shown in Figures 8.1a and 8.1b. The post-fit distributions, at low values of the BDT score, are similar to the results from the fit in the control regions (see Figure 7.21). In fact, nearly identical results are obtained in terms of the fitted NPs, shown in Figure 8.2. A slight difference in pulls and constraints of some NPs is visible with respect to the fit in the control regions. The motivation is the different smoothing of the systematic uncertainties. The smoothing algorithm merges neighbouring bins until the statistical uncertainty of the systematic variation is reasonable with respect to the variation itself,

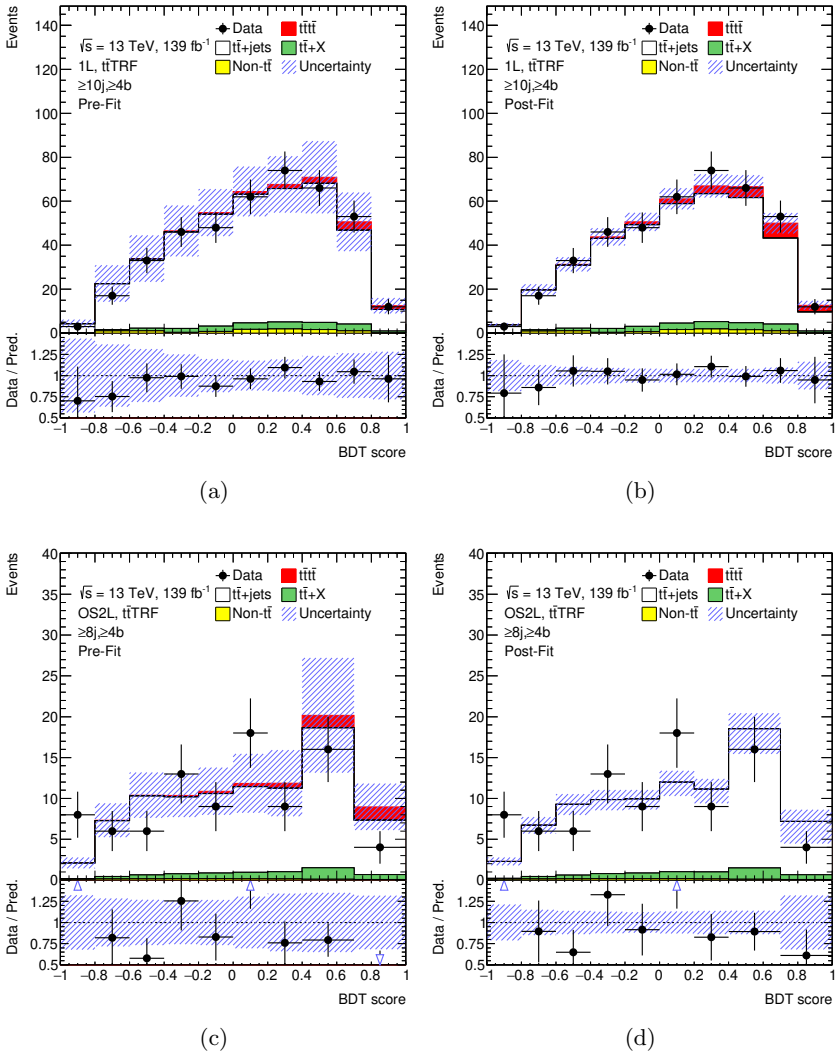
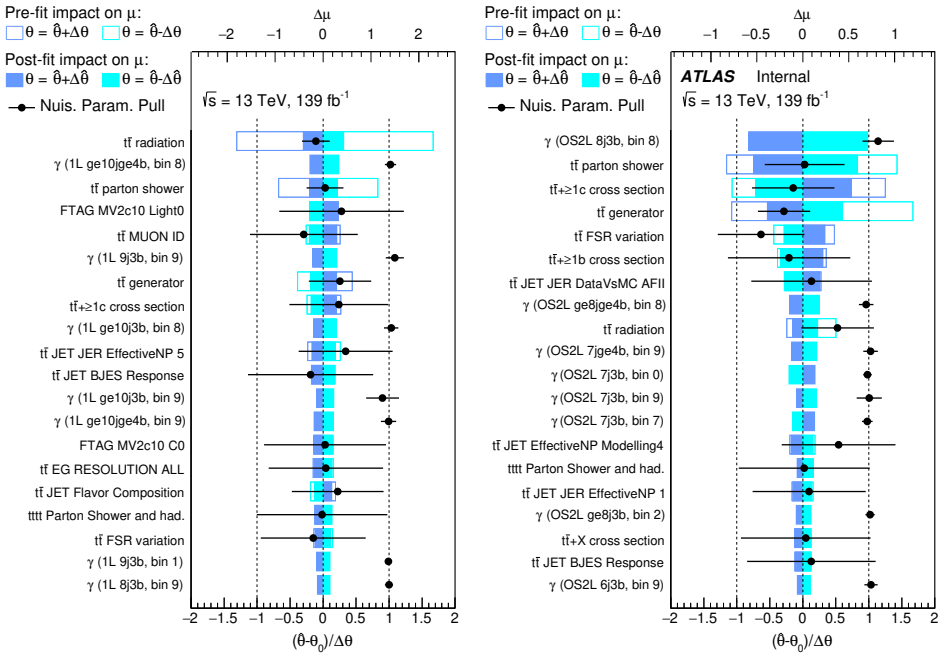


Figure 8.1.: Pre-fit and post-fit distributions of the BDT score for the most sensitive regions in the single-lepton (a,b) and opposite-sign dilepton channel (c,d). Since the the fit in the opposite-sign dilepton channel yield a negative value for the signal strength of the $t\bar{t}t\bar{t}$ process, the $t\bar{t}t\bar{t}$ is not displayed in (d).



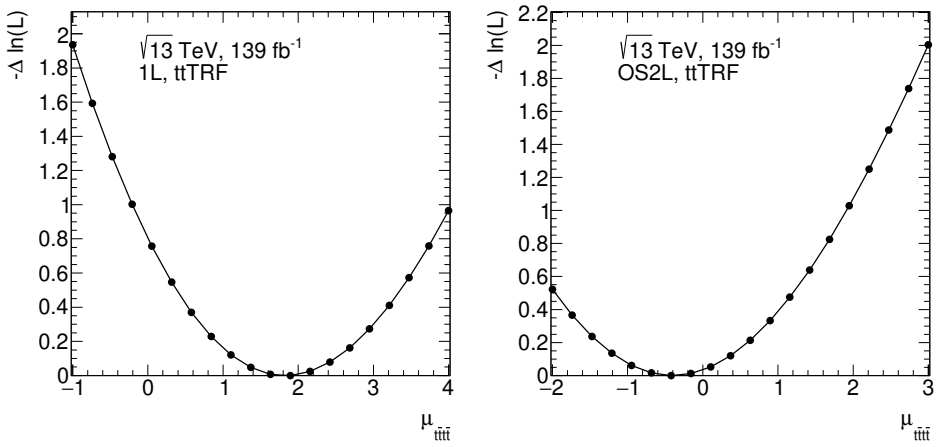
(a) (b)

Figure 8.2.: List of the twenty NPs most important NPs for the results in the single lepton (a) and opposite-sign dilepton (b) channels. The points refer to the fitted value of the NPs with the associated uncertainties. The bands instead refer to the impact on the fitted signal strength $\Delta\mu$ for a variation of the NP of one post-fit standard deviation. The empty bands show the expected impact with pre-fit uncertainties. Note that γ NPs, associated to statistical fluctuations of the model, are expected to be centred around one because they are Poisson-distributed NPs.

and then smooths the histogram of the variation with a quadratic interpolation. Fitting the full spectrum of the BDT adds bins to the control regions, therefore the outcome of the smoothing is different. The constraints observed on this fit are indeed compatible with the expectations from Asimov datasets in the same regions. In general, a dominant contribution of the $t\bar{t}$ modelling uncertainties and statistical uncertainty of the prediction is observed, as expected from tests on Asimov datasets.

The post-fit distribution of the single lepton channel in Figure 8.1c show larger yields for the signal with respect to the pre-fit prediction in Figure 8.1a. The signal strength is indeed fitted to values larger than one. The profile of the likelihood as a function of the signal strength for the single lepton channel is shown in Figure 8.3a. It gives an observed signal strength in the single lepton channel of:

$$\mu_{t\bar{t}t\bar{t}}^{1L} = 1.8^{+1.5}_{-1.4} \tag{8.1}$$



(a) (b)

Figure 8.3.: Likelihood value as a function of the signal strength $\mu_{t\bar{t}t\bar{t}}$ for the single lepton (a) and opposite-sign dilepton (b) channels.

This translates into an observed (expected) limit at 95% CL on the cross-section of the $t\bar{t}t\bar{t}$ production is 54 (49) fb. The observed (expected) significance of the $t\bar{t}t\bar{t}$ signal over the background-only hypothesis is 1.3σ (0.7σ). The most important systematic uncertainties (see Figure 8.2a) are consistent with the expectations. It should be noted that the uncertainty on the signal parton shower algorithm becomes important in this fit on data due to the larger measured signal strength than expected.

Negative statistical fluctuations of the data in the most significant bins of the BDT distributions have been observed in the opposite-sign dilepton channel. This made the observed data consistent with the background-only hypothesis, making the $t\bar{t}t\bar{t}$ contribution disappear in the post-fit distributions in Figure 8.1d. The shape of the likelihood profile as a function of the signal strength for the opposite-sign dilepton channel, shown in Figure 8.3b. It gives a convergence of the fit for an observed signal strength of:

$$\mu_{t\bar{t}t\bar{t}}^{\mathrm{OS2L}} = -0.4^{+1.6}_{-1.6} \tag{8.2}$$

This value for the signal strength has no physical meaning, but indicates that the fit in this channel is not sensitive to the signal. The most important systematic sources for the fit results are consistent with the expectations, indicating a proper behaviour of the fit. The statistical uncertainty (see Figure 8.2b) associated to the most sensitive bin is the most important uncertainty, due to the observed downwards fluctuation of data in that bin. Uncertainties on the signal modelling become more important also in this channel because of the fitted negligible contribution of the signal, that can be significantly be increased through variation of the associated systematics.

The combination of the channels is still beneficial to constrain systematics and provide

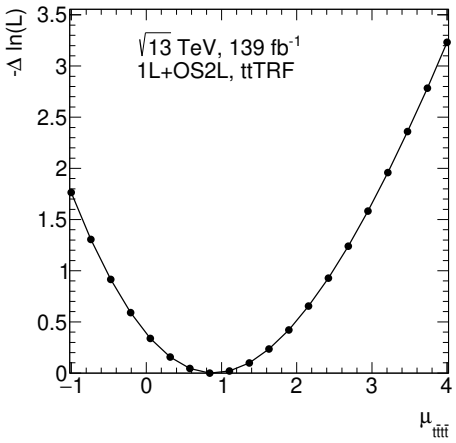


Figure 8.4.: Likelihood value as a function of the signal strength $\mu_{t\bar{t}t\bar{t}}$ for the combination of the single lepton and opposite-sign dilepton channels.

a more precise measurement of the signal strength. The corresponding likelihood scan is shown in Figure 8.4. The measured value of the signal strength is:

$$\mu_{t\bar{t}t\bar{t}}^{\text{1L+OS2L}} = 0.9\,^{+0.5}_{-0.5}\,(\text{stat.})\,^{+1.0}_{-0.9}\,(\text{syst.}) = 0.9^{+1.1}_{-1.0} \tag{8.3}$$

The observed limit at 95% CL on the cross-section of the $t\bar{t}t\bar{t}$ production is 35 (35) fb. The observed (expected) significance of the $t\bar{t}t\bar{t}$ signal over the background-only hypothesis is 0.9σ (1.0σ).

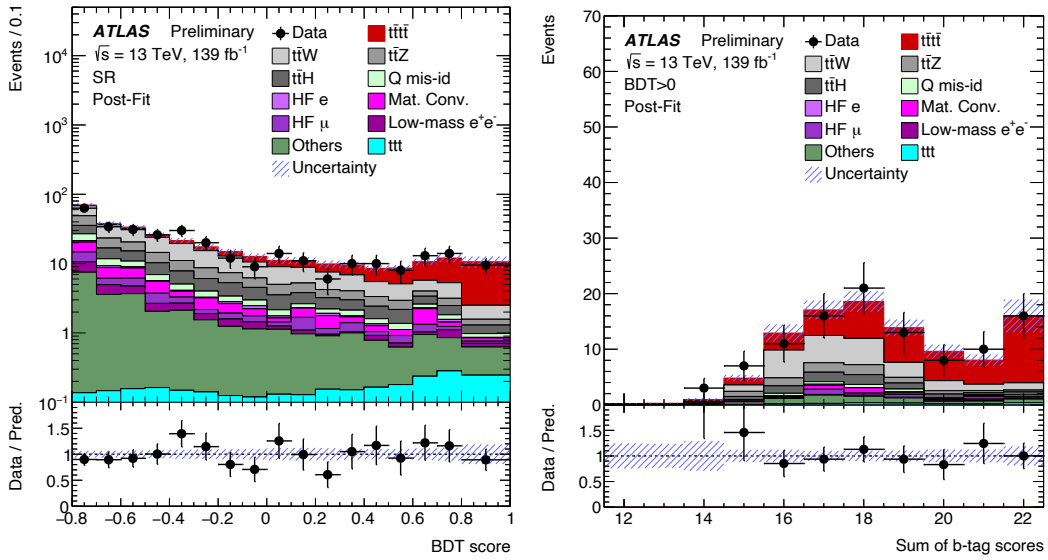
All the results are shown to be compatible with the expectations from the Standard Model predictions. This measurement for the signal strength of $t\bar{t}t\bar{t}$ production is currently the most precise measurement performed in the 1LOS channel.

8.2. Combination with the same-sign dilepton and trilepton channel

The ATLAS Collaboration released preliminary results for the analysis in the SSML channel described in this thesis [36], measuring a signal strength of the $t\bar{t}t\bar{t}$ production:

$$\mu_{t\bar{t}t\bar{t}}^{\text{SSML}} = 2.0\,^{+0.4}_{-0.4}\,(\text{stat.})\,^{+0.7}_{-0.5}\,(\text{syst.}) = 2.0^{+0.8}_{-0.6} \tag{8.4}$$

This value of the signal strength corresponds to an observed (expected) significance of the $t\bar{t}t\bar{t}$ signal over the background-only hypothesis of 4.3 (2.4) standard deviations, claiming the evidence of the $t\bar{t}t\bar{t}$ production process, with a cross-section of 24^{+7}_{-6} fb. This result, despite the measurement of a higher cross-section than the expectations, is compatible with the SM prediction within 1.7 standard deviations.



(a) (b)

Figure 8.5.: Post-fit distribution of the BDT score in the signal region in the SSML channel (a) [36]. Post-fit distribution of the sum of the b-tagging scores of all the jets in the event in the signal region requiring for a positive BDT score, to select a purer sample of the $t\bar{t}t\bar{t}$ signal (b) [36].

The post-fit distribution of the BDT score in the signal region agrees well with the data, as shown in Figure 8.5a. The excess of the data with respect to the background prediction has been found to be compatible with the $t\bar{t}t\bar{t}$ signature in all the considered variables, as, for example, the sum of the b-tagging scores of all the jets in the event, shown in Figure 8.5b, that is the most important feature in the BDT training.

The normalisation factor for the $t\bar{t}W$ background has been measured to be 1.6 ± 0.3, compatible with other measurement of the process in a similar phase space [133]. The other normalisation factors included in the fit (see Section 5.4.2) show reasonable agreement with the expectations [36].

By comparing these results with the measurement of the CMS Collaboration [35], a difference in both expected and observed significance should be noted. The expected significance in the analysis conducted by the CMS Collaboration is higher than in the analysis presented here. The most important systematic uncertainties in the analysis presented here, responsible for the smaller value of the expected significance, are the additional uncertainties on $t\bar{t}W$ modelling associated with large jet multiplicity (see Section 6.3), that are not included in the model used by the CMS Collaboration. Moreover, the value refers to the uncertainty on the background model after the fit: the excess of data due at high values of the BDT score yields to pulls of NPs associated to

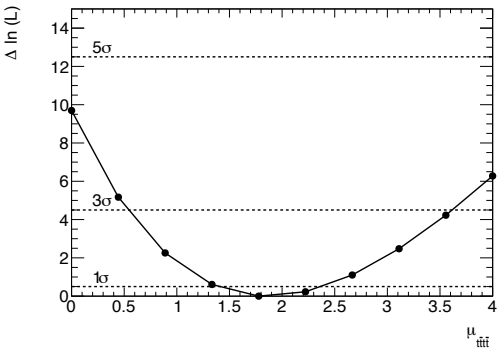
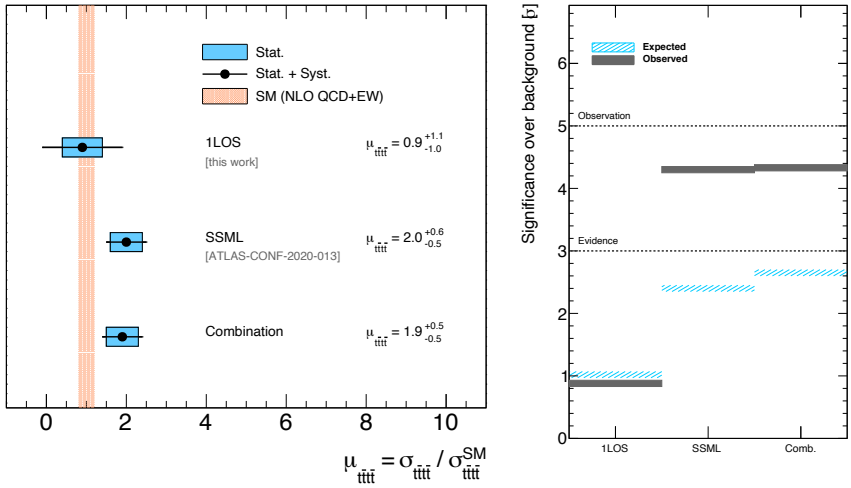


Figure 8.6.: Scan of the likelihood values as a function of the $t\bar{t}t\bar{t}$ signal strength. The lines are in correspondence of the values of the likelihood at $\mu_{t\bar{t}t\bar{t}} = 0$ at which different values of the signal significances are reached in terms of standard deviations (σ).

uncertainties responsible of significant change at high BDT score, increasing the overall background in the most sensitive region. In fact, by using a pseudo-data sample created by injecting the NPs fitted in the control regions, a higher significance is obtained. This value of significance of the $t\bar{t}t\bar{t}$ signal in the SSML was below three standard deviation, but close enough to surpass it in the combination with the measurement in the 1LOS channel, that was then expected to be crucial to reach the evidence of the $t\bar{t}t\bar{t}$ production. Concerning the observed significance and signal strength, the measured values by the two collaborations are different, but still in agreement with the SM hypothesis. However, a similar analysis strategy would be beneficial for a more consistent comparison.

A combination of the 1LOS and SSML channel results still to be beneficial to obtain a more precise and robust measurement and produce a measurement of the $t\bar{t}t\bar{t}$ production process using almost the all the $t\bar{t}t\bar{t}$ events collected by the detector. The combination is performed by correlating the common sources of systematic uncertainties. For this work, the experimental uncertainties and the uncertainties on the cross-sections of the background processes are totally correlated between the two channels[1]. More channel-specific sources of uncertainties are instead treated as independent sources. In some cases, this leads to a double-counting of uncertainties as, for instance, for the uncertainties related to $t\bar{t}$ modelling. Therefore, this approach would result in a overestimation of the uncertainties on the measured $t\bar{t}t\bar{t}$ signal strength and a conservative value of the $t\bar{t}t\bar{t}$ significance. As a final remark, in order to make the systematics model more coherent across the channels, the uncertainty on the $t\bar{t}t\bar{t}$ cross-section is not included for the results in the SSML channel presented in the following. Therefore, deviations of the results shown here from the public preliminary results [36] are due to this change in the analysis setup.

[1] It means that the variations of the distributions in the two channels are associated to a single NP.



(a) (b)

Figure 8.7.: Comparison of the measured signal strength in the individual channels and in the combination (a). Comparison of the significance of the $t\bar{t}t\bar{t}$ signal over the background-only hypothesis in the individual channels and in their combination (b).

The scan of the likelihood as a function of the $t\bar{t}t\bar{t}$ signal strength for combined fit in the 1LOS and SSML channels is shown in Figure 8.6, corresponding to the measurement of the $t\bar{t}t\bar{t}$ signal strength of:

$$\mu_{t\bar{t}t\bar{t}}^{\text{1LOS+SSML}} = 1.9\,^{+0.3}_{-0.3}\,(\text{stat.})\,^{+0.4}_{-0.4}\,(\text{syst.}) = 1.9^{+0.5}_{-0.5} \tag{8.5}$$

This turns into an observed (expected) significance of the signal over the background-only hypothesis of 4.4 (2.8) standard deviations. The corresponding measured cross-section for the $t\bar{t}t\bar{t}$ production is 23 ± 6 fb, compatible with the SM value within 1.8 standard deviations[2]. A summary of the individual results with their combination in terms of $t\bar{t}t\bar{t}$ signal strength and significance are shown in Figures 8.7a and 8.7b, respectively. The results of the combination are totally driven by the result in the SSML channel, since more sensitive to the $t\bar{t}t\bar{t}$ signal. Normalisation factors for the backgrounds in the SSML in the combination are measured to be compatible with the individual channel results [36]. The NPs are moved consistently with the expectations, merging the features of the two individual channel. For instance, NPs associated to instrumental sources of uncertainty are moved similarly to the 1LOS channel, due to the larger statistics. Channel-specific NPs are measured to be consistent with the results from

[2]This value includes the uncertainty on the cross-section $t\bar{t}t\bar{t}$ production by adding it in quadrature to the measured cross-section uncertainty.

the individual channels. The most important sources of uncertainty for this result are identical to the limiting factors of the SSML analysis, since it is the most sensitive: understanding of the $t\bar{t}W$ and $t\bar{t}t\bar{t}$ processes modelling and uncertainty on mis-tagging efficiencies for light-quark jets. This latter has a large impact on the analysis of the SSML channel and it is slightly pulled by the 1LOS channel, ending up with large impact on the final measurement. This behaviour has been already observed in the combination of these channels performed by the ATLAS Collaboration with partial Run 2 dataset [32].

CHAPTER 9

Conclusions and outlook

The success of a physics project of an experiment is given by a performant data-taking of the detector and a sophisticated analysis of the measured data. In this thesis both sides have been studied. The performance of the Pixel Detector has been analysed and investigated, with a look to the expected running conditions of the next years. However, the analysis of the data represents the core of the thesis, focused on the search of one of the most challenging processes to be measured, the $t\bar{t}t\bar{t}$ production.

The conclusions and plans for the coming Run 3 for the detector studies have been already given in Section 3.4. This chapter focuses on the results from the analysis of data: the summary of the obtained results is presented in Section 9.1, the perspective of the analysis in the next years is described in Section 9.2.

9.1. Summary

The production of $t\bar{t}t\bar{t}$ in pp collisions at $\sqrt{s}$ =13 TeV is a process that is predicted by the Standard Model but no observation has been reached yet. It is an interesting process from a theoretical and an experimental point of view. It is sensitive to many new-physics scenarios: its small cross-section of about 12 fb can be strongly enhanced by several Beyond Standard Model theories. Experimentally, it is instead a very challenging process, featured by a large multiplicity of jets and b-jets in the final state. Because of many top quarks in the final state, many combinations of top quark decay modes are possible, giving many different signatures in the detector.

The analysis aims at a measurement of the $t\bar{t}t\bar{t}$ process in a combination of two de-

tection channels: the 1LOS and the SSML channels. The first is obtained by selecting events with either one lepton or two leptons with opposite-sign charges. The second is obtained by requiring either two leptons with same-sign charges or three leptons in the events. This thesis focuses on the measurement in the 1LOS channel and provides the combined results of the two channels.

The events in the 1LOS channel are dominated by the contribution of $t\bar{t}$ production associated with a large number of jets. In the selected phase space, the prediction from Monte Carlo simulations for this background do not describe the data well. Therefore, two estimation methods for the $t\bar{t}$+jets background have been implemented: the t$\bar{\text{t}}$TRF and the reweighting methods. The main results of this thesis have been obtained using the t$\bar{\text{t}}$TRF method, that provides robust prediction of the background and comparable sensitivity to the $t\bar{t}t\bar{t}$ signal with the reweighting method.

To improve the sensitivity, a Boosted Decision Tree has been trained to separate the $t\bar{t}$+jets background from $t\bar{t}t\bar{t}$ signal. The output of this algorithm has been used as a discriminant variable since it has shown better separation with respect to any other kinematic variable.

A profile likelihood fit has been used to estimate the $t\bar{t}t\bar{t}$ signal strength in the full Run 2 dataset. The low BDT score region at different jet and b-jet multiplicities have been used as control regions to constrain the background prediction. The measurement of the $t\bar{t}t\bar{t}$ signal is driven by the higher BDT score region, blinded for the validation and testing of the fit setup. The measured value for the signal strength of the $t\bar{t}t\bar{t}$ production in the 1LOS channel is:

$$\mu_{t\bar{t}t\bar{t}}^{\text{1LOS}} = \frac{\sigma_{t\bar{t}t\bar{t}}^{\text{obs.}}}{\sigma_{t\bar{t}t\bar{t}}^{\text{SM}}} = 0.9\,^{+0.5}_{-0.5}\,(\text{stat.})\,^{+1.0}_{-0.9}\,(\text{syst.}) = 0.9^{+1.1}_{-1.0} \tag{9.1}$$

corresponding to a limit on the $t\bar{t}t\bar{t}$ cross-section of 35 fb at 95% CL. The observed (expected) significance of the $t\bar{t}t\bar{t}$ signal over the background-only hypothesis is 0.9σ (0.9σ). This result is more precise than the past measurements performed by both collaborations in the same channel [31, 32].

This measurement in the 1LOS channel is statistically independent from the measurement in the SSML channel and therefore the combination of the two channels is possible. Preliminary results of the analysis in the SSML channel have been released by the ATLAS Collaboration [36], observing double the signal strength than the SM expectation, corresponding to an observed (expected) significance of the $t\bar{t}t\bar{t}$ signal over the background-only hypothesis of 4.3 (2.4) standard deviations. The excess of the data over the background is compatible with the features of the SM signal, therefore this gives the first evidence of the SM $t\bar{t}t\bar{t}$ production process. In this thesis, the combination of this results with the 1LOS channel is performed, providing the most complete measurement of the $t\bar{t}t\bar{t}$ process in terms of the detection channels. The combination of the 1LOS and

SSML channels yields a measurement of the $t\bar{t}t\bar{t}$ signal strength of:

$$\mu_{t\bar{t}t\bar{t}}^{\text{1LOS+SSML}} = 1.9\,^{+0.3}_{-0.3}\,(\text{stat.})\,^{+0.4}_{-0.4}\,(\text{syst.}) = 1.9^{+0.5}_{-0.5} \tag{9.2}$$

corresponding to a $t\bar{t}t\bar{t}$ cross-section of 23 ± 6 fb, compatible with the SM prediction within 1.8 standard deviations. The measured (expected) significance of the $t\bar{t}t\bar{t}$ signal over the background of 4.4 (2.8) standard deviations. The results of the combination are mostly influenced by the analysis SSML channel, that is the most sensitive channel, but still more precise than the measurement performed in the individual channels.

9.2. Outlook

The presented analysis would benefit both from a more optimised analysis and a larger dataset.

Despite the analysis being already quite complex, many optimisations are still possible, addressing different aspects of the analysis. A change of the jet definition to *particle flow* (PFlow) jets may be beneficial for the measurement. This type of jet is constructed by combining both calorimetric and tracking information, providing smaller uncertainties for jets with a p_{T} below 200 GeV. This improvement may sensibly help the analysis since many low-p_{T} jets are present in the selected phase space, and jet uncertainties have a significant impact on the measurement.

Concerning the $t\bar{t}$TRF method, the estimation of the efficiency dependence on the jet multiplicity has been shown to significantly reduce the uncertainties associated to the prediction and improve the closure with data (see Section 7.1). This would reduce the pulls and constraints on the fitted nuisance parameters associated to the $t\bar{t}$+jets background, giving an even more robust background prediction. A change of the regions definition for the fit may also increase the sensitivity to $t\bar{t}t\bar{t}$ signal. For example, the use of different working points of the b-jets tagger for different jets in the event for the b-jet identification, has shown to improve the significance by around 10% [123]. However these studies should be repeated with the current fit setup.

The training of the Boosted Decision Tree, used to separate the signal from the background, could be further improved by a different set of input variables and by a finer tuning of the hyper-parameters. Hints that the current setup may be suboptimal can be ascertained from the shape of the background distributions in high b-jet multiplicity, that tend still to peak towards the signal region. More sophisticated MVA techniques could be also used for the reconstruction of the $t\bar{t}t\bar{t}$ final state, that would be interesting for the construction of more complex variables.

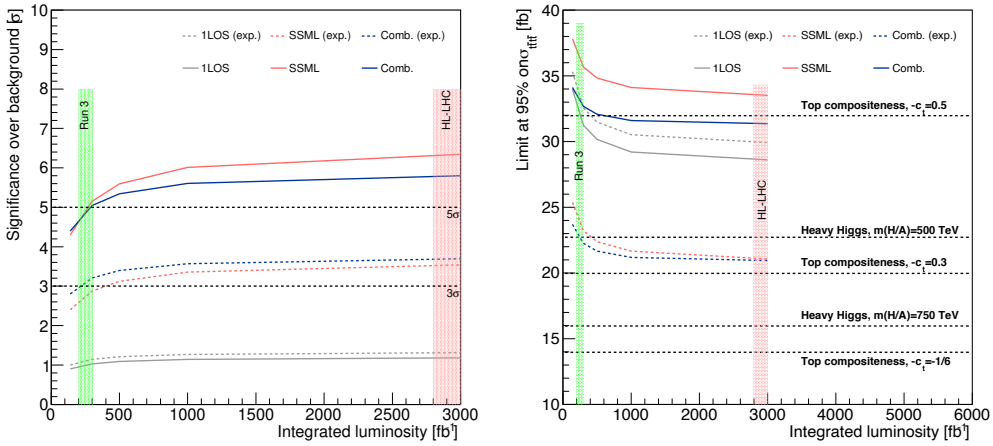
Finally, a more accurate systematic model for the uncertainties on $t\bar{t}$+jets modelling would impact significantly the results of the measurement. For example, the Sherpa algorithm [71] could be investigated for assessing the generator and the parton shower

uncertainties. However, Monte Carlo simulations are extremely complex algorithms, therefore the direct comparison of two predictions would probably end up in double counting of some of the uncertainties. Moreover, the NPs associated to these uncertainties are treated as Gaussian-distributed variables in the profile likelihood fit, that is a reasonable choice and commonly accepted, but still quite debatable.

An increase of the size of the analysed dataset would also significantly improve the measurement in both 1LOS and SSML channels. An improvement of the sensitivity in the analysis in the SSML channel is expected since limited by statistics, even though not excessively over the contribution of systematic uncertainties. The analysis in the 1LOS channel would benefit from a larger dataset, especially in case the $t\bar{t}$TRF method is used. The results from this method that is based on data events, are strongly influenced by the statistical uncertainty on the background prediction, that has the second highest impact on the measurement. The change of the significance of the $t\bar{t}t\bar{t}$ signal as a function of the integrated luminosity is shown in Figure 9.1a. The expected significances in the SSML and combined measurements are always lower than the observed significances, because the latter ones refer to the higher signal strength measured in the SSML measurement. A significant change of significances are observed for luminosities up to 1 ab^{-1} for the SSML and combined measurements, more limited by statistics. In the 1LOS channel a saturation of the sensitivity is reached already with the Run 3 statistics. In case this higher value of the $t\bar{t}t\bar{t}$ signal strength is measured with a larger dataset as well, the observation of the process may be achieved with the Run 3 dataset. This result would be still compatible with the SM prediction within 2 deviations. An observation of the SM $t\bar{t}t\bar{t}$ process would be a strong test on the validity of the SM, even in an extreme environment in terms of jet and b-jet multiplicity, that has not been tested so far at a comparable precision.

Furthermore, a more precise measurement of the $t\bar{t}t\bar{t}$ process cross-section would constrain further the parameters of BSM theories. An extrapolation of the limits on the $t\bar{t}t\bar{t}$ cross-section at 95% CL for the different channels and their combination as function of the integrated luminosity of the dataset is shown in Figure 9.1b. The higher measured signal strength leads to looser limits on the $t\bar{t}t\bar{t}$ cross-sections than expected, in the SSML and combined measurements, that improve their sensitivity up to luminosities of 1 ab^{-1}. After that point systematic uncertainties dominate and an increase of the statistics has no longer any effects. Therefore, the analysis could benefit of the expected dataset of 3000 ab^{-1} at the end of the HL-LHC only in case the systematic uncertainties are significantly reduced. This estimation of the limits is however conservative, since more statistics in the control regions of the fit would give stronger constraints of the background.

Two BSM models have been considered to have an idea of the limits this analysis could set on the parameters of the theory. The 2HDM Type II scenario has been studied, for small values of $\tan\beta$ [26]. Under these assumptions, the heavy-mass Higgs boson is strongly coupled with the top quark, resulting in a large impact on $t\bar{t}t\bar{t}$ production cross-section. No strong limits in this scenario have been set yet by existing analyses [137].



(a) (b)

Figure 9.1.: Significance of the $t\bar{t}t\bar{t}$ signal over the background in terms of number of standard deviations (σ) as a function of the integrated luminosity (a). Limits on the $t\bar{t}t\bar{t}$ production cross-section ($\sigma_{t\bar{t}t\bar{t}}$) at 95% CL as a function of the integrated luminosity. Five predictions for the $t\bar{t}t\bar{t}$ cross-section are shown from two BSM scenarios for different values of the respective parameters (b). The green and pink bands indicate the expected integrated luminosity of Run 3 and HL-LHC, respectively.

As shown in Figure 9.1b, the measurement in the 1LOS channel has no sensitivity to this source of new physics, and the measured higher value of the signal strength in the SSML channels does not help to set more stringent limits in the combination. However, the expected combination with the SSML could set first limits on the mass of the heavy Higgs at more than 500 GeV with the Run 3 dataset[1]. The second considered model assumes the top quark to be composite of more fundamental fields, inducing a four-top quarks contact interaction term [27]. The c_t parameters handle the strength of this term. Limits for $|c_t| < 0.5$ could be put from the analysis in the 1LOS channel with double the data statistics. To reach a sensitivity to higher masses of the heavy Higgs or to smaller strength of the contact interaction, a significant reduction of the systematic uncertainties is needed.

[1] Two different types of Higgs bosons are predicted by the model: CP-even (H) and CP-odd (A). Since their effects on the $t\bar{t}t\bar{t}$ cross-section is not significantly different, here the average is considered.

Bibliography

[1] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nucl. Phys. B **126** (1977) 298 – 318.

[2] V. N. Gribov and L. N. Lipatov, *Deep inelastic ep scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450.

[3] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641–653.

[4] J. C. Collins, D. E. Soper, and G. Sterman, *Factorisation of hard processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1–91.

[5] S. W. Herb, et al., *Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions*, Phys. Rev. Lett. **39** (1977) 252–255.

[6] M. L. Perl, et al., *Evidence for Anomalous Lepton Production in e^+e^- Annihilation*, Phys. Rev. Lett. **35** (1975) 1489–1492.

[7] Pluto Collaboration, *Observation of a narrow resonance formed in e^+e^- annihilation at 9.46 GeV*, Phys. Lett. B **76** 243.

[8] Jade Collaboration, *A measurement of the electroweak induced charge asymmetry in $e^+e^-\to b\bar{b}$*, Phys. Lett. B **146** 437.

[9] CDF Collaboration, *Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab*, Phys. Rev. Lett. **74** (1995) 2626–2631.

[10] DØ Collaboration, *Observation of the Top Quark*, Phys. Rev. Lett. **74** (1995) 2632–2637.

[11] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001.

[12] ATLAS Collaboration, *ATLAS Top Working Group Summary Plots*, `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/TOP/`. Accessed: 15/09/2019.

[13] ATLAS Collaboration, *Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, Phys. Rev. D **98** (2018) 012003.

[14] CMS Collaboration, *Measurements of $t\bar{t}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons*, JHEP **2019** (2019) 149.

[15] ATLAS Collaboration, *Measurements of top-quark pair differential and double-differential cross-sections in the ℓ+jets channel with pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, Eur. Phys. J. C. **79** (2019) 1028.

[16] ATLAS collaboration, *Measurements of differential cross sections of top quark pair production in association with jets in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, JHEP **2018** (2018) 159.

[17] ATLAS Collaboration, *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **99** (2019) 072009.

[18] ATLAS Collaboration, *Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}\gamma$ production in leptonic final states at $\sqrt{s} = 13$ TeV in ATLAS*, Eur. Phys. J. C **79** (2019) 382.

[19] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784** (2018).

[20] CMS Collaboration, *Observation of $t\bar{t}H$ Production*, Phys. Rev. Lett. **120** (2018) 231801.

[21] 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 **2018** (2018) 31.

[22] T. Plehn and T. M. P. Tait, *Seeking sgluons*, Journal of Physics G: Nuclear and Particle Physics **36** (2009) 075001.

[23] M. Guchaita, F. Mahmoudib, and K. Sridharc, *Associated production of a Kaluza-Klein excitation of a gluon with a tt pair at the LHC*, Phys. Lett. B **666** (2008) 347 – 351.

[24] G. Cacciapaglia, et al., *Four tops on the real projective plane at LHC*, JHEP **2011** (2011) 42.

[25] N. Craig, et al., *The hunt for the rest of the Higgs bosons*, JHEP **2015** (2015) 137.

[26] N. Craig, et al., *Heavy Higgs bosons at low tan β: from the LHC to 100 TeV*, JHEP **2017** (2017) 18.

[27] A. Pomarol and J. Serra, *Top quark compositeness: Feasibility and implications*, Phys. Rev. D **78** (2008) 074026.

[28] A. De Simone, et al., *A first top partner hunter's guide*, JHEP **2013** (2013) 4.

[29] C. Kilic, T. Okui, and R. Sundrum, *Vectorlike confinement at the LHC*, JHEP **2010** (2010) 18.

[30] C. Degrande, et al., *Non-resonant new physics in top pair production at hadron colliders*, JHEP **2011** (2011) 125.

[31] 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 **2018** (2018) 39.

[32] 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. D **99** (2019) 052009.

[33] 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. C **78** (2018) 140.

[34] CMS Collaboration, *Search for the production of four top quarks in the single-lepton and opposite-sign dilepton final states in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, J. High Energ. Phys. **2019** (2019) 82.

[35] CMS Collaboration, *Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **80** (2019) 75.

[36] 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*, Tech. Rep. ATLAS-CONF-2020-013, CERN, Geneva, Jun, 2020. http://cds.cern.ch/record/2719519.

[37] L. Evans and P. Bryant, *LHC Machine*, JINST G **3** (2008) S08001–S08001.

[38] K. Johnsen, *CERN Intersecting Storage Rings (ISR)*, Proc. Natl. Acad. Sci. U. S. A. **70** (1973) 619–626. 8 p, National Academy of Sciences Autumn Meeting (October 1972): Symposium: New Accelerators.

[39] K. Cornelis, *Beam-Beam Effects in the SPS Proton-Anti Proton Collider*, http://cds.cern.ch/record/1955087.

[40] S. Holmes, R. S. Moore, and V. Shiltsev, *Overview of the Tevatron collider complex: goals, operations and performance*, JINST **6** (2011) T08001–T08001.

[41] W. Schnell, *Design Study of a Large Electron-Positron Colliding Beam Machine - LEP*, Nuclear Science, IEEE Transactions on **26** (1979) 3130 – 3134.

[42] C. Beckmann, *Proceedings of the 1990 linear accelerator conference*, Linear

Accelerator Conference 03/1990, LA-12004-C.

[43] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1 – 29.

[44] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30 – 61.

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

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

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

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

[49] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1997. https://cds.cern.ch/record/331063.

[50] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996. https://cds.cern.ch/record/331061.

[51] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996. https://cds.cern.ch/record/331062.

[52] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997. https://cds.cern.ch/record/331068.

[53] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** (2017) 317.

[54] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007–P07007.

[55] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, ATLAS-TDR-19 (2010).

[56] ATLAS Collaboration, *Production and integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008–T05008.

[57] ATLAS Collaboration, *ATLAS Pixel Calibration Public Results*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsPixel, Accessed: 15/10/2019.

[58] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007–P07007.

[59] ATLAS Collaboration, *ATLAS Data Quality Summary tables*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RunStatsPublicResults2010, Accessed: 30/07/2019.

[60] ATLAS Collaboration, *ATLAS Pixel Public Results for Run 2*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/PixelPublicResults, Accessed: 15/10/2019.

[61] P. Sabatini, J. Grosse-Knetter, and A. Quadt, *Charge calibration for the FEI4 chip in the Insertable B-Layer*, Tech. Rep. ATL-COM-INDET-2018-052, CERN, Geneva, Aug, 2018. https://cds.cern.ch/record/2636167.

[62] M. Engels, *Spektroskopiemessungen mit dem ATLAS FE-I4 mit planarem n-in-n Silizumsensor*, MSc Universität Bonn (2014).

[63] ATLAS Collaboration, *Study of threshold drop in FE-I4 by SEU*, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/PIX-2018-006/, Accessed: 31/07/2019.

[64] M. Mironova, *Implementation of a Threshold Tuning Algorithm for Silicon and Diamond Pixel Detectors*, BSc Georg-August-Universität Göttingen, II. Physik-UniG-BSc-2017/02.

[65] ATLAS Collaboration, *Modelling radiation damage to pixel sensors in the ATLAS detector*, JINST **14** (2019) P06012–P06012.

[66] ATLAS Pixel Collaboration, *Pixel Radiation Damage Measurements with ATLAS*, https://indico.cern.ch/event/769192/contributions/3295236/.

[67] ATLAS Collaboration, *Radiation damage to the ATLAS pixel*, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/PIX-2017-005/, Accessed: 31/07/2019.

[68] 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.

[69] ATLAS Collaboration, *Luminosity determination in pp collisions at* $\sqrt{s} = 13$ *TeV using the ATLAS detector at the LHC*, https://cds.cern.ch/record/2677054.

[70] ATLAS Collaboration, *ATLAS Luminosity Public Results for Run 2*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2, Accessed: 01/08/2019.

[71] T. Gleisberg, et al., *Event generation with SHERPA 1.1*, JHEP **2009** (2009) 007–007.

[72] S. Agostinelli, et al., *Geant4 – a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250 – 303.

[73] E. Richter-Was, D. Froidevaux, and L. Poggioli, *ATLFAST 2.0 a fast simulation package for ATLAS*, Tech. Rep. ATL-PHYS-98-131, CERN, Geneva, Nov, 1998. https://cds.cern.ch/record/683751.

[74] W. Lukas, *Fast Simulation for ATLAS: Atlfast-II and ISF*, Tech. Rep. ATL-SOFT-PROC-2012-065, CERN, Geneva, Jun, 2012. https://cds.cern.ch/record/1458503.

[75] 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 **2014** (2014) 79.

[76] NNPDF collaboration, *Parton distributions for the LHC run II*, JHEP **2015** (2015) 40.

[77] S. Frixione, et al., *Angular correlations of lepton pairs from vector boson and top quark decays in Monte Carlo simulations*, JHEP **2007** (2007) 081–081.

[78] P. Artoisenet, et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **2013** (2013) 15.

[79] T. Sjötrand, et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159 – 177.

[80] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, ATL-PHYS-PUB-2014-021 (2014).

[81] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152 – 155, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[82] M. Bähr, et al., *Herwig++ physics and manual*, Eur. Phys. J. C **58** (2008) 639–707.

[83] J. Bellm, et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196.

[84] L. A. Harland-Lang, et al., *Parton distributions in the LHC era: MMHT 2014 PDFs*, Eur. Phys. J. C **75** (2015) 204.

[85] S. Frixione, G. Ridolfi, and P. Nason, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **2007** (2007) 126–126.

[86] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, JHEP **2004** (2004) 040–040.

[87] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **2007** (2007) 070–070.

[88] S. Alioli, et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **2010** (2010) 43.

[89] M. Cacciari, et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612 – 622.

[90] M. Czakon, P. Fiedler, and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through* $\mathcal{O}(\alpha_S^4)$, Phys. Rev. Lett. **110** (2013) 252004.

[91] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **2012** (2012) 54.

[92] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **2013** (2013) 80.

[93] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Computer Physics Communications **185** (2014) 2930 – 2938.

[94] R. Frederix, E. Re, and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **2012** (2012) 130.

[95] S. Frixione, et al., *Single-top hadroproduction in association with aWboson*, JHEP **2008** (2008) 029–029.

[96] R. Frederix, E. Re, and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, JHEP **2012** (2012) 130.

[97] S. Alioli, et al., *NLO single-top production matched with shower in POWHEG:s-andt-channel contributions*, JHEP **2009** (2009) 111–111.

[98] S. Alioli, et al., *Erratum: NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*, JHEP **2010** (2010) 11.

[99] N. Kidonakis, *Next-to-next-to-leading logarithm resummation for s-channel single top quark production*, Phys. Rev. D **81** (2010) 054028.

[100] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503.

[101] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, `https://e-publishing.cern.ch/index.php/CYRM/issue/view/32`.

[102] J. M. Campbell and R. K. Ellis, *Update on vector boson pair production at hadron colliders*, Phys. Rev. D **60** (1999) 113006.

[103] ATLAS Collaboration, *Measurement of W and Z-boson production cross sections in pp collisions at s=13 TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601 – 621.

[104] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, JINST **14** (2019) P12006–P12006.

[105] T. Cornelissen, et al., *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007. https://cds.cern.ch/record/1020106.

[106] R. Frhwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **262** (1987) 444 – 450.

[107] T. G. Cornelissen, et al., *The global χ^2 track fitter in ATLAS*, Journal of Physics: Conference Series **119** (2008) 032013.

[108] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}$ = 13 TeV*, Eur. Phys. J. C **79** (2019) 639.

[109] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292.

[110] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, https://cds.cern.ch/record/2669655.

[111] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-ktjet clustering algorithm*, JHEP **2008** (2008) 063–063.

[112] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896.

[113] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2014-018, CERN, Geneva, May, 2014. https://cds.cern.ch/record/1700870.

[114] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, https://cds.cern.ch/record/2273281.

[115] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J C **79** (2019) 970.

[116] A. Hoecker, et al., *TMVA - Toolkit for Multivariate Data Analysis*, arXiv:0703039 [physics.data-an].

[117] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, Tech. Rep. ATL-PHYS-PUB-2017-011, CERN, Geneva, Jun, 2017. https://cds.cern.ch/record/2270366.

[118] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, Tech. Rep. ATL-PHYS-PUB-2018-025, CERN, Geneva, Oct, 2018.

https://cds.cern.ch/record/2645405.

[119] B. Brunt, et al., *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-COM-PHYS-2016-407, CERN, Geneva, Apr, 2016. https://cds.cern.ch/record/2149445.

[120] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATLAS-CONF-2017-029, CERN, Geneva, May, 2017. http://cds.cern.ch/record/2261772.

[121] ROOT Collaboration, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, Tech. Rep. CERN-OPEN-2012-016, New York U., New York, Jan, 2012. https://cds.cern.ch/record/1456844.

[122] R. D. Ball, et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244 – 289.

[123] ATLAS Collaboration, *Search for* $t\bar{t}t\bar{t}$ *Standard Model Production in proton-proton collisions with the ATLAS Detector*, https://cds.cern.ch/record/2670354.

[124] P. Onyisi and A. Webb, *Impact of rare decays* $t \to l\nu bll$ *and* $t \to qqbll$ *on searches for top-associated physics*, JHEP **2018** (2018) 156.

[125] ATLAS Collaboration, *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017–P07017.

[126] ATLAS Collaboration, *Simulation of Pile-up in the ATLAS Experiment*, https://cds.cern.ch/record/1616394.

[127] ATLAS Collaboration, E_T^{miss} *performance in the ATLAS detector using 2015-2016 LHC p-p collisions*, https://cds.cern.ch/record/2294891.

[128] ATLAS Collaboration, *Measurements of inclusive and differential fiducial cross-sections of* $t\bar{t}$ *production with additional heavy-flavour jets in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **46** (2019).

[129] ATLAS Collaboration, *ATLAS simulation of boson plus jets processes in Run 2*, Tech. Rep. ATL-PHYS-PUB-2017-006, CERN, Geneva, May, 2017. https://cds.cern.ch/record/2261937.

[130] ATLAS Collaboration, *Modelling of the* $t\bar{t}H$ *and* $t\bar{t}V$ $(V = W, Z)$ *processes for* $\sqrt{s} = 13$ *TeV ATLAS analyses*, Tech. Rep. ATL-PHYS-PUB-2016-005, CERN, Geneva, Jan, 2016. http://cds.cern.ch/record/2120826.

[131] ATLAS Collaboration, *Multi-Boson Simulation for 13 TeV ATLAS Analyses*, Tech. Rep. ATL-PHYS-PUB-2017-005, CERN, Geneva, May, 2017. https://cds.cern.ch/record/2261933.

[132] ATLAS Collaboration, *Simulation of top quark production for the ATLAS*

experiment at sqrt(s) = 13 TeV, Tech. Rep. ATL-PHYS-PUB-2016-004, CERN, Geneva, Jan, 2016. http://cds.cern.ch/record/2120417.

[133] ATLAS Collaboration, *Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2019-045, CERN, Geneva, Oct, 2019. https://cds.cern.ch/record/2693930.

[134] ATLAS Collaboration, *Search for the Standard Model Higgs boson decaying into $b\bar{b}$ produced in association with top quarks decaying hadronically in pp collisions at $\sqrt{s}$=8 TeV with the ATLAS detector.*, JHEP **05** (2016) 160. 53 p.

[135] ATLAS Collaboration, *Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **97** (2018) 072016.

[136] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. B **784** (2018) 173–191. 19 p.

[137] W. Su, et al., *Exploring the low $\tan\beta$ region of two Higgs doublet models at the LHC*, arXiv:1909.09035 [hep-ph].

Appendices

APPENDIX A

Comparison of the methods in the opposite-sign dilepton channel

In this appendix, the comparison of the performance of the $t\bar{t}$TRF and reweighting method is shown in the opposite-sign dilepton channel. The pre-fit distributions and the post-fit distributions are shown in Section 7.3.2. In the following the behaviour of the fit is compared.

Firstly, the background is fitted to data in the control regions. In Figure A.1, the fitted nuisance parameters relative to $t\bar{t}$ modelling and jet reconstruction and calibration are shown. As expected, the constraints are more relaxed than in the single lepton channel due to the lower statistical power of the control regions. The $t\bar{t}$TRF still shows larger constraints than the reweighting method. Moreover, the $t\bar{t}$ modelling systematics in $t\bar{t}$TRF are found not to be significantly large in the lower tail of the distributions as in the single lepton channel and comparable to the statistical uncertainty of the prediction (Figure A.2).

A fit on the full spectrum of the Asimov dataset is performed to study the impact of the different systematics. The ranking of the systematics in the two methods is shown in Figure A.3. In this channel, the importance of the statistical uncertainty of the prediction for the $t\bar{t}$TRF method is even more evident. In Figure A.2 shows the two dominant systematics: in both cases, the variations are quite small in the control regions, and therefore are not constrained much by the fit. However they have a large impact in the higher tail of the distributions and therefore on the signal strength. Indeed, the constrain from the single-lepton helps in the combination of the two channels. Differently from the single lepton channel, here the large increase of the systematics in the last bins seem to be due to poor statistics, thus the constrain of the variation from the propagation of the constraint in the single lepton channel is acceptable.

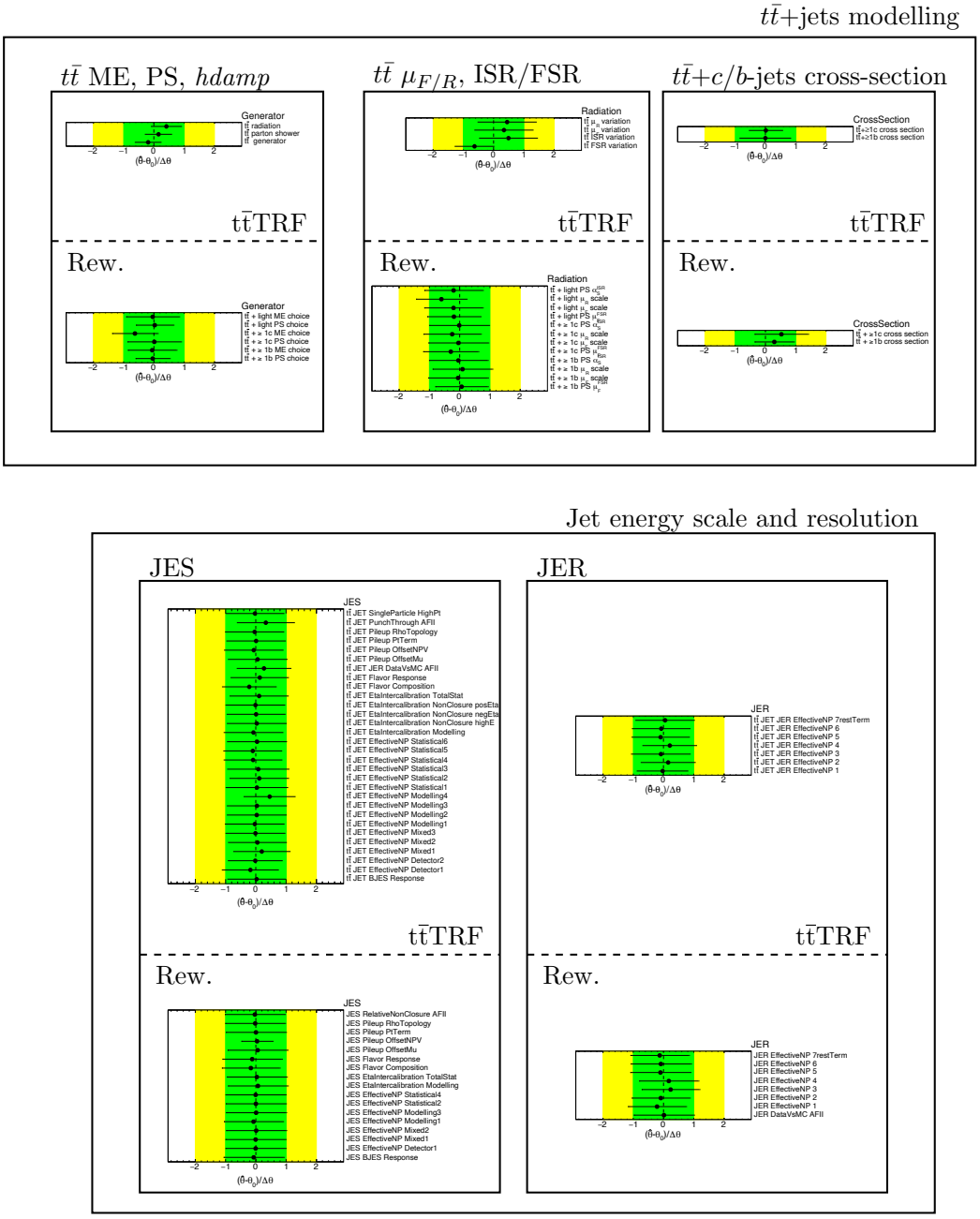
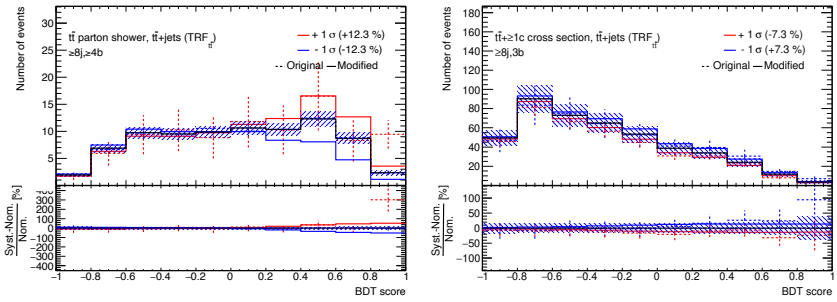
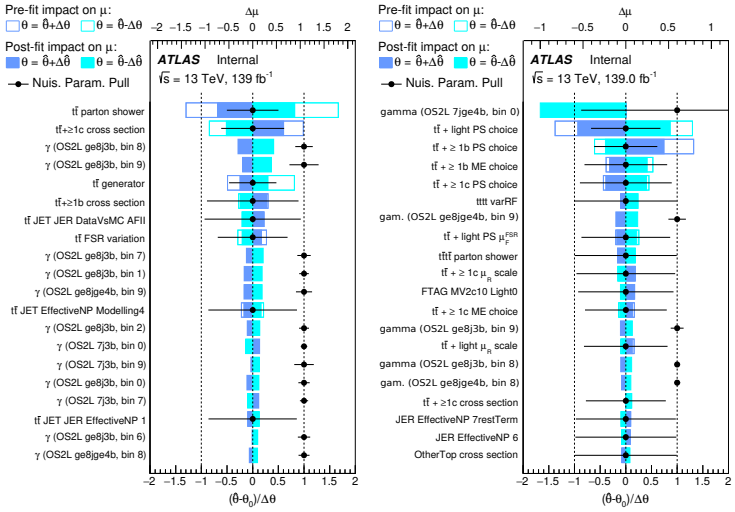


Figure A.1.: Fitted values of the NPs associated to $t\bar{t}$ modelling (top block) and jet uncertainties (bottom plot) from a background-only fit in control regions for the opposite-sign dilepton channel. Each block is divided in different sources of uncertainties, showing the values for the $t\bar{t}$TRF and reweighting (Rew.) predictions. Because of the splitting of the $t\bar{t}$ modelling uncertainties in flavours of the $t\bar{t}$ radiation, the reweighting method has more NPs associated to $t\bar{t}$ modelling. Slightly different naming conventions are used, *ME choice* corresponds to *generator* and *PS* to *parton shower*. The $t\bar{t}$ *radiation* uncertainty, related to the choice of the *hdamp* parameter, is not included for the reweighting method. The green and yellow bands around zero indicate the pre-fit uncertainties at 68% and 95% confidence level. Points stand for the fitted value of the NPs with the associated uncertainties as fractions of the respective pre-fit standard-deviations.



(a) (b)

Figure A.2.: Pre-fit uncertainties associated to $t\bar{t}$ parton shower and hadronisation modelling in OS2L $\geq 8j \geq 4b$ region (a) and $t\bar{t}$+c-jets normalisation (b) in 1L $\geq 8j3b$ region. Red points refer to the original variation, with associated statistical error. The red solid line is obtained by smoothing the uncertainty consistently with the statistics 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.



(a) (b)

Figure A.3.: Ranking of the most important systematics in a fit on an Asimov dataset in the opposite-sign dilepton channel. The systematics are ranked in terms of the impact of the post-fit systematic on the POI $\mu_{t\bar{t}t\bar{t}}$. The ranking from the fit on the t$\bar{\text{t}}$TRF model is in (a), while the reweighting method is in (b). Points refer to the fitted values of the NPs with the corresponding uncertainty as fraction of the pre-fit standard-deviation of the Gaussian prior. The filled (empty) bands indicate the impact on the results $\Delta\mu$ for a variation of the NPs of one post-fit (pre-fit) standard-deviation. The reweighting method showed some issues in the fit convergence with variations of the γ NPs, therefore their ranking is not considered valid.

APPENDIX B

Validation of the fit setup

Some tests have been performed on the fit setup. An effort to reduce the constraints in the fit for the single lepton channel is presented in Appendix B.1. A quick binning optimisation of the signal regions is shown Appendix B.2. The linearity and convergence of the fit has been tested and described in Appendix B.3.

B.1. Decorrelation of nuisance parameters in the single lepton channel

The strong constraints in the $t\bar{t}$ modelling uncertainties in the single lepton channel are motivated by the large shape effect in the tails of the distributions. In case the NP is constraint in a specific control region, far from the signal region, the application of the same constraint in the signal region is an aggressive approach for the evaluation of the post-fit systematics. A more conservative approach is to decorrelate the NP in different regions, i.e. different NPs are associated to different region. However, this also limits the power of the fit to constrain the uncertainties and may cause a reduction of the sensitivity. The decorrelation of the uncertainties associated to the $t\bar{t}$ parton shower, matrix element and *hdamp* choices has been applied in the single lepton channel, and the fitted NPs on the Asimov dataset are shown in Figure B.1. No constraints are observed in any NP, despite the fitted uncertainty on the $\mu_{t\bar{t}t\bar{t}}$ is reduced from $1.00^{+1.49}_{-1.44}$ to $1.00^{+1.40}_{-1.34}$. This effect is explained by looking at the correlations between the fitted nuisance parameters. The correlation of the considered NPs with the signal stregth are shown in the two scenarios in Table B.1: the decorrelation of the nuisance parameters among regions breaks the correlations with the signal strength. This gives to the signal

Uncertainty	NPs correlated among regions	NPs decorrelated among regions
$t\bar{t}$ parton shower	-19%	$(3-4)\%$
$t\bar{t}$ generator	18%	$(3-4)\%$
$t\bar{t}$ radiation	-26%	$(3-4)\%$

Table B.1.: Table showing the correlation of the NPs with the signal strength in case of correlation and decorrelation of the NPs in different regions.

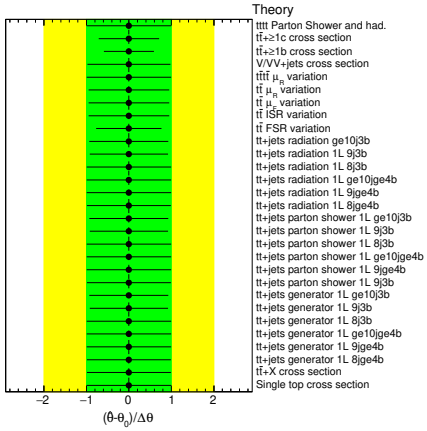


Figure B.1.: Fitted NPs associated to $t\bar{t}$ modelling in case of the decorrelation. Only the NPs associated to $t\bar{t}$ generator, radiation and parton shower have been decorrelated region-by-region.

strength more freedom to vary, resulting in smaller uncertainties. However, this is not a desired behaviour and gives an underestimation of the uncertainty. This also indicates that the observed constraint without decorrelating the NPs is not driven by a specific region, but from all the regions. A single region therefore has not enough statistical power to constrain much the NPs, that can be constrained only if a coherent behaviour is observed in all the regions. No decorrelation among regions is used for the presented results.

B.2. Binning optimisation in the signal regions

The impact of different sources of systematics is evaluated by excluding the considered group of nuisance parameters from the fit and creating an Asimov dataset by setting the remaining NPs to the post-fit values from the nominal fit. The decrease of the uncertainty with respect to the nominal fit corresponds to the contribution to the post-fit uncertainty of the excluded group of systematics and used to rank the importance.

Uncertainty	Contribution to $\mu_{t\bar{t}t\bar{t}}$ uncertainty	
	1L	OS2L
$t\bar{t}$ matrix element, parton shower, radiation	0.51	0.87
$t\bar{t}$ $\mu_{F/R}$ scales, ISR/FSR	0.15	0.21
$t\bar{t}$ $+b/c$-jets cross-sections	0.26	0.62
$t\bar{t}t\bar{t}$ modelling	0.12	0.15
Flavour tagging	0.32	0.08
JER	0.32	0.17
JES	0.37	0.33
Other instrumental	0.40	0.15
Other background modelling	0.05	0.07
Prediction statistics	0.58	0.62

Table B.2.: Impact of different groups of systematics on the uncertainty of the signal strength.

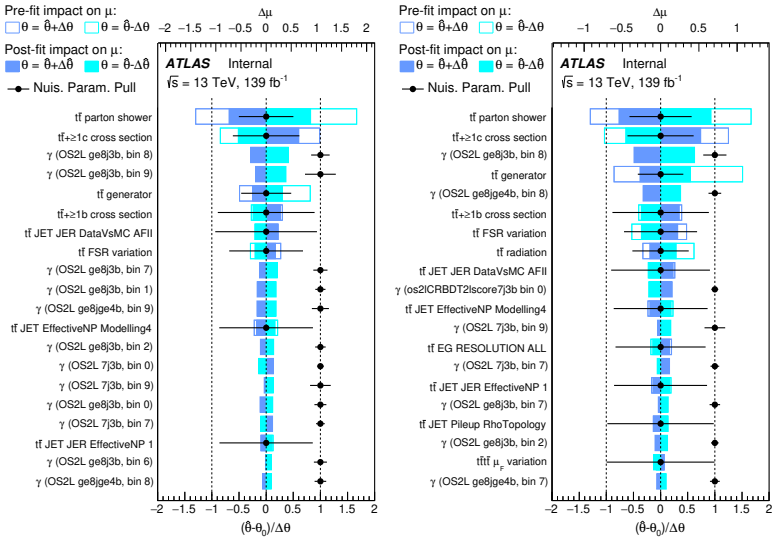
Table B.2 shows the importance of the different groups of systematics in the final fit results. The statistics of the prediction play a quite important role in both channels. This effect can be mitigated by rebinning the low populated bins in the higher tail of the distribution. However, rebinning those bins reduce the separation power of the BDT and affects the significance. A quick study of different binning of the signal regions (higher tail of the BDT spectrum) is performed by evaluating the significance of $t\bar{t}t\bar{t}$ in an fit of the background and signal model on the Asimov dataset, by taking into account only data and background model statistical uncertainties. This gives an indication of the best compromise between prediction statistics and significance. The results are shown in Tables B.3 and B.4, the *standard* binning corresponds to the one used in the results shown for the comparison of the methods (Section 7.3), the *chosen* is instead the binning used for the measurement in Chapter 8. The standard binning in the single lepton channel shows already the best significance among the tested ones. This gives indications that the used binning is at least close to the optimal point. In the opposite-sign dilepton channel, very small differences are shown among the tested setups, but a slightly coarser binning at the largest jet multiplicity is eventually chosen for the measurement to reduce the impact of the statistics in those bins. As expected from this studies, the change of the binning slightly increases the significance in a fit of the Asimov dataset without (with) the injection of the parameters in the OS2L channel to 0.64 (0.62) and to 1.02 (1.05) in the combination with the same setup. The fit behaviour is not significantly affected except for the reduction of the impact of the statistics of the prediction, as shown in Figure B.2.

Single lepton channel				
$9j$	$\geq 10j$	Significance	Standard	Chosen
$[0.6, 1]$	$[0.4, 1]$	1.22		
$[0.6, 0.8, 1]$	$[0.4, 0.7, 1]$	1.26		
$[0.6, 0.8, 1]$	$[0.4, 0.6, 0.8, 1]$	1.30	*	*
$[0.6, 0.7, 0.8, 1]$	$[0.4, 0.5, 0.6, 0.7, 0.8, 1]$	1.15		

Table B.3.: Significance of $t\bar{t}t\bar{t}$ signal computed by taking into account only data and background model statistical uncertainties for different binnings of the signal regions in the single lepton channel.

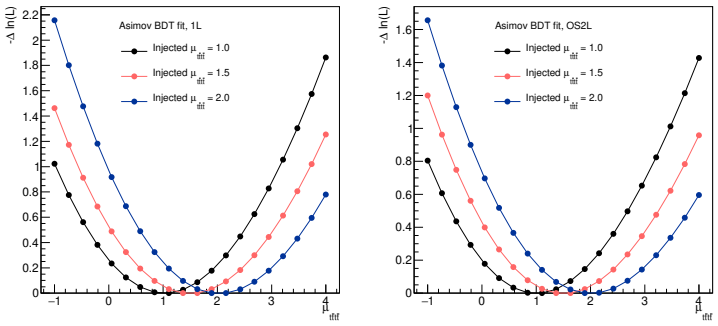
Opposite-sign dilepton channel					
$7j$	$\geq 8j3b$	$\geq 8j \geq 4b$	Significance	Standard	Chosen
$[0.6, 1]$	$[0.4, 1]$	$[0.4, 1]$	0.96		
$[0.6, 0.8, 1]$	$[0.4, 0.7, 1]$	$[0.4, 0.7, 1]$	1.05		*
$[0.6, 0.8, 1]$	$[0.4, 0.6, 0.8, 1]$	$[0.4, 0.6, 0.8, 1]$	1.02	*	
$[0.6, 0.8, 1]$	$[0.4, 1]$	$[0.4, 0.6, 0.8, 1]$	1.02		
$[0.6, 0.7, 0.8, 1]$	$[0.4, 0.5, 0.6, 0.7, 0.8, 1]$	$[0.4, 0.5, 0.6, 0.7, 0.8, 1]$	1.03		

Table B.4.: Significance of $t\bar{t}t\bar{t}$ signal computed by taking into account only data and background model statistical uncertainties for different binnings of the signal regions in the opposite-sign dilepton channel.



(a) (b)

Figure B.2.: Ranking of the most important systematics in a fit on an Asimov dataset in the opposite-sign dilepton channel with the original binning of the signal region (a) and the one chosen for the final measurement (b).



(a) (b)

Figure B.3.: Likelihood scans for different injected signal strength $\mu_{t\bar{t}t\bar{t}}$ for the signal lepton (a) and opposite-sign dilepton (b) channels.

B.3. Check of the linearity of the fit

Some nuisance parameters may mimic the signal contribution at high values of the BDT distribution, resulting lower values of the fitted $\mu_{t\bar{t}t\bar{t}}$. It is especially important to check in the $t\bar{t}$TRF prediction where the larger impacts of systematics are usually the tails of the distribution (see Section 7.3.2), exactly where the signal lies. To check the fit response, different values of the signal strength are injected for the creation of the Asimov datasets, and the fit of the complete background and signal model is performed. A scan of the likelihood is performed for each value of signal strength to check the likelihood profile and the minimum position, shown in Figure B.3. Different initial points for the NPs have been tested and are found not affecting the final result. Therefore the convergence of the fit has proven to be stable and the response to be linear with the injected signal strength.

APPENDIX C

Detailed fit results

This section contains detailed results from the fit of the single lepton and opposite-sign dilepton channels and their combination. The full collection of pre-fit and post-fit plots are shown in Figures C.1 and C.2. In every region, the uncertainties on the model prediction are reduced by the fit and the agreement with data is improved. The post-fit distributions describe the data within the associated post-fit uncertainties.

To achieve this goal the NPs are moved around their priors: a Gaussian prior for modelling and experimental uncertainties, while a Poissonian prior is used for the γ NPs associated to the statistical fluctuations of the prediction. The pre- and post-fit values are compared in the Figure C.1. Here the NPs from the fits in the individual channels and in the combination are compared. The green and yellow bands of in the plot for the fitted values of the NPs indicate the pre-fit uncertainty at 68% and 95% CL. The points indicate the fitted values with the respective uncertainties as fractions of the pre-fit standard-deviation. No particular deviation of the combined fit with respect to the individual ones is observed. Generally, it tends to merge the features of the two individual fits, obtaining stronger constrain of the background prediction. The correlation matrix of the combined fit is also shown. No suspiciously high correlations ($\geq 50\%$) are observed.

The behaviour of the individual and combined fits is resulted to be healthy and compatible with the expectations from the fit on Asimov datasets. The observed constraints and pulls have been understood and many tests on the fit robustness and validity have been performed as described in Appendix B.

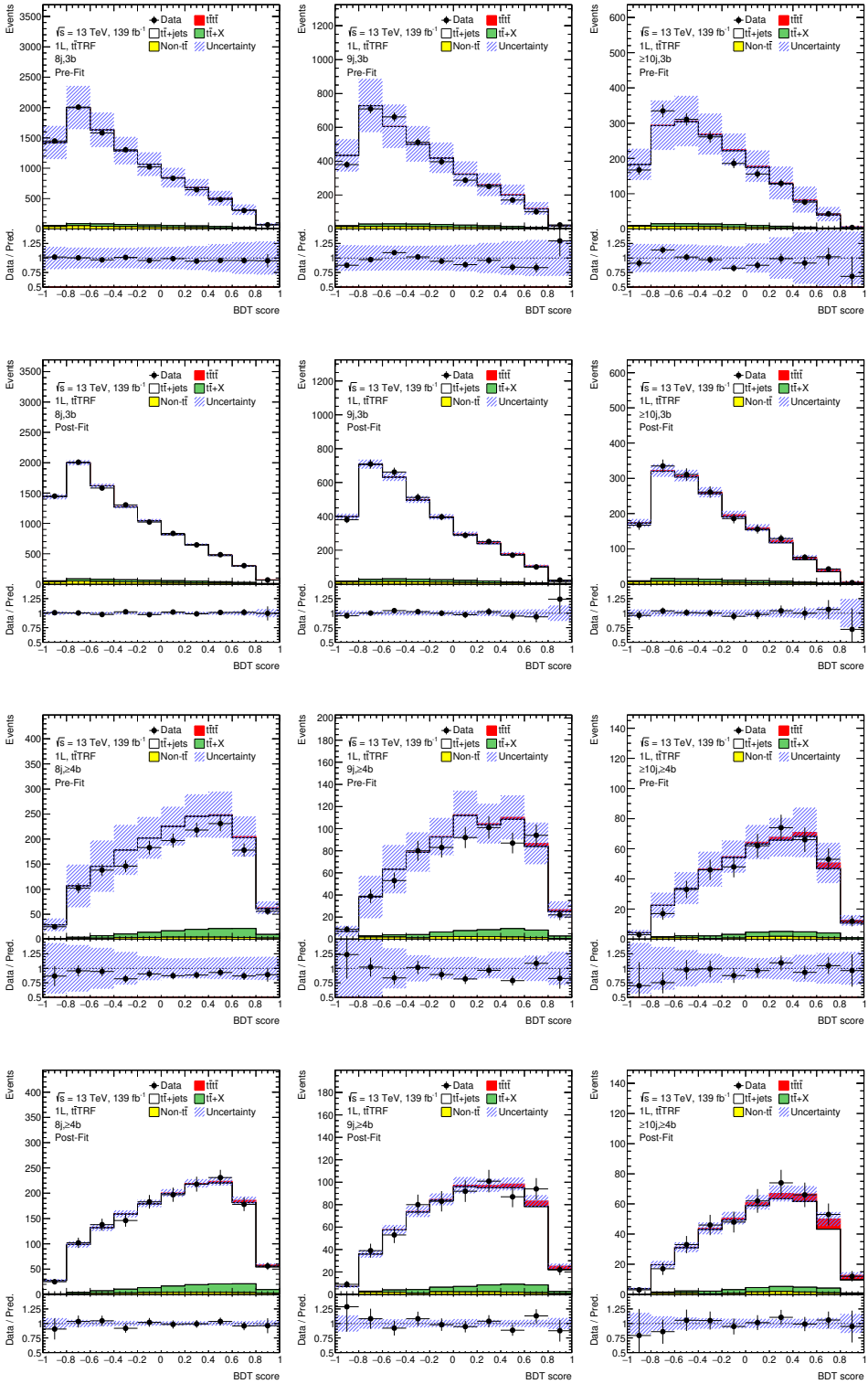


Figure C.1.: Pre-fit and post-fit distribution of all the regions included in the fit for the single lepton channel. In the first and the third rows the pre-fit distributions at different jet and b-jet multiplicities are shown. In the second and fourth rows the corresponding post-fit distributions are shown.

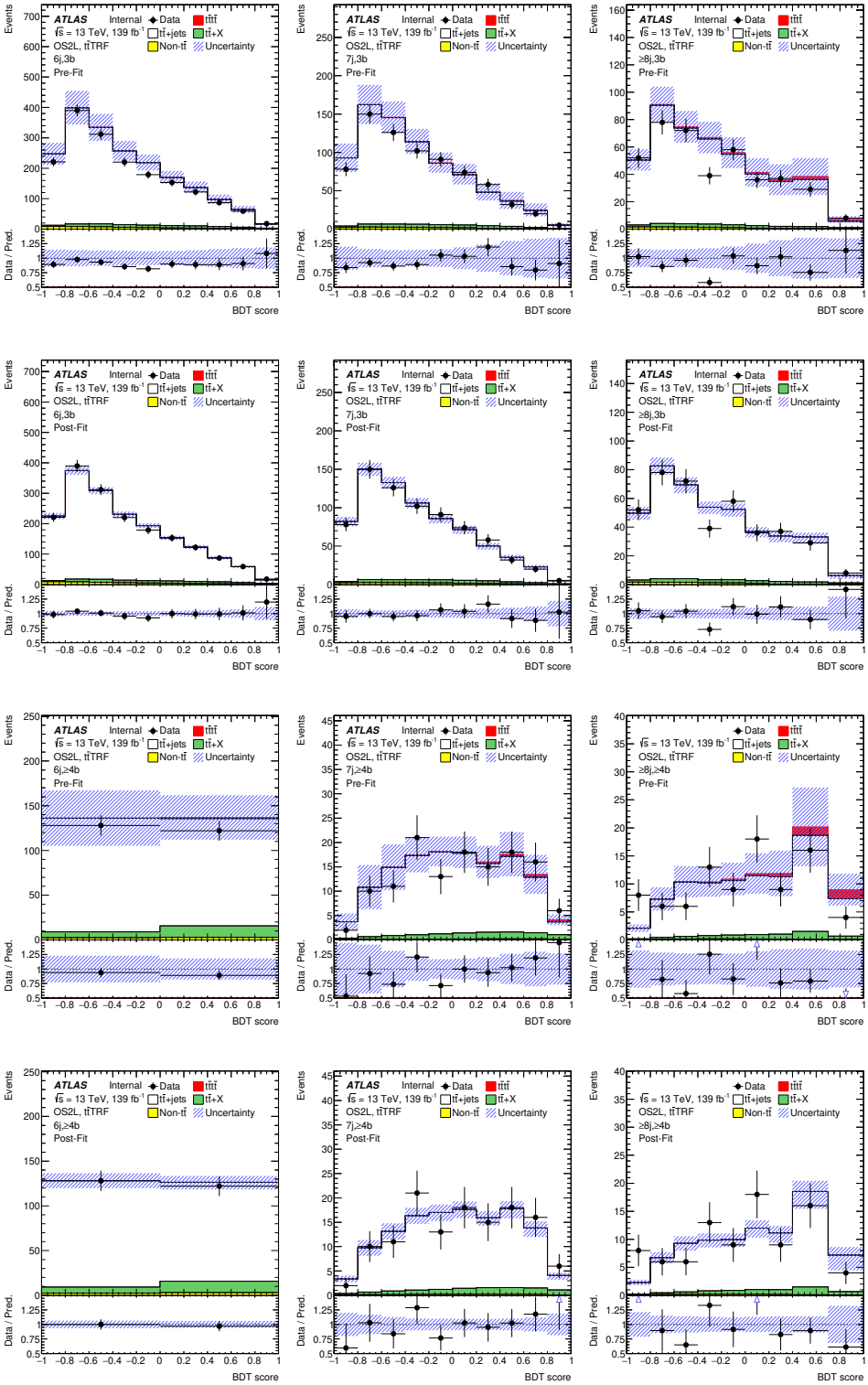


Figure C.2.: Pre-fit and post-fit distribution of all the regions included in the fit for the opposite-sign dilepton channel. In the first and the third rows the pre-fit distributions at different jet and b-jet multiplicities are shown. In the second and fourth rows the corresponding post-fit distributions are shown.

C. Detailed fit results

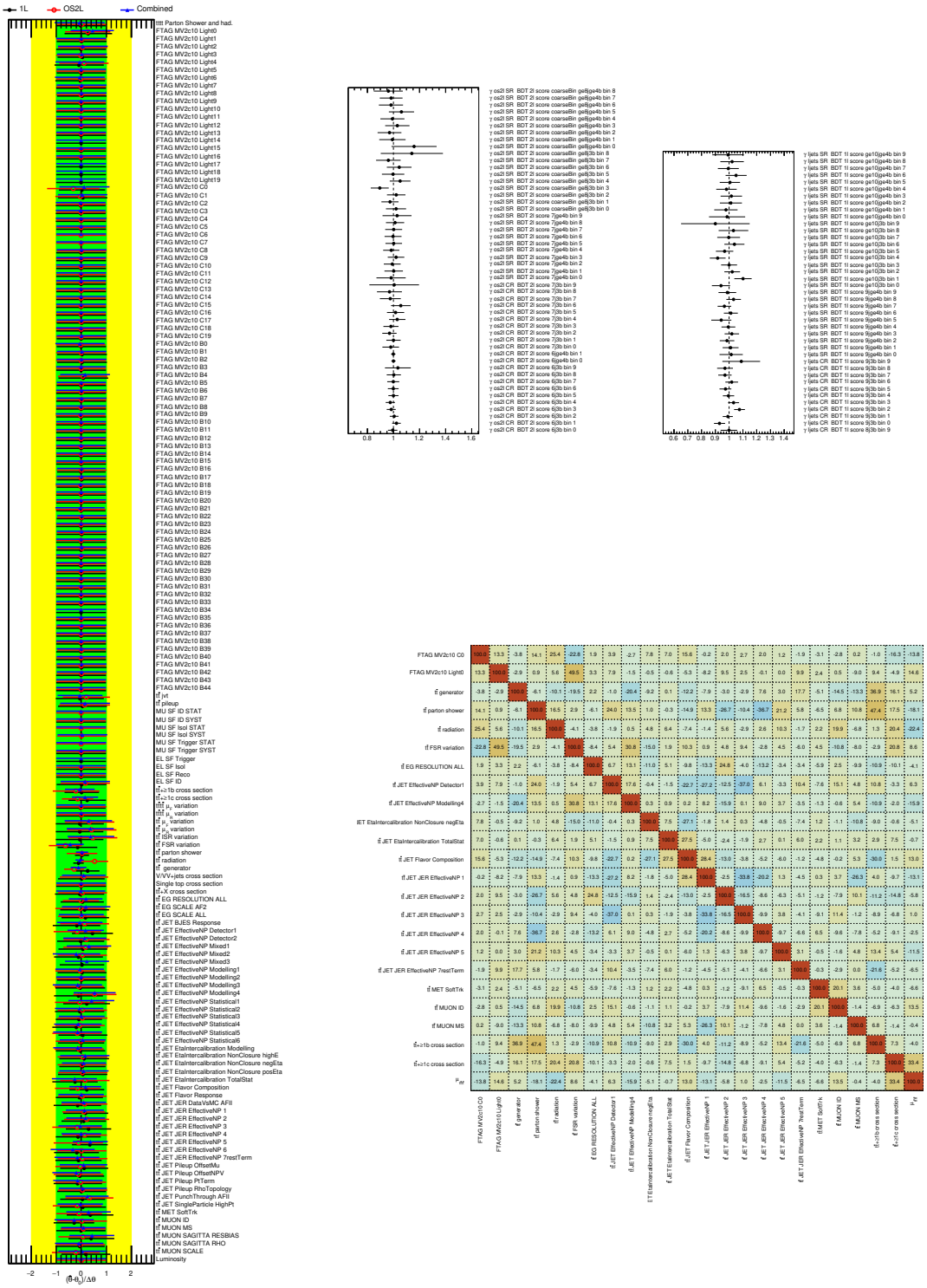


Table C.1.: Values of the fitted NPs for the single channels and their combination (left). Fitted γ NPs for the individual channels (top right) and correlation matrix between the NPs in the combined fit (bottom right).