Search for the production of four top quarks in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ in the single and opposite sign dilepton channels with the ATLAS detector at the LHC using a Monte Carlo correction method

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades
„Doctor rerum naturalium“
der Georg-August-Universität Göttingen

im Promotionsstudiengang Physik
der Georg-August University School of Science (GAUSS)

vorgelegt von

Fabian Sohns

aus Hameln

Göttingen, 2020

Betreuungsausschuss

Prof. Dr. A. Quadt
Prof. Dr. S. Lai
Prof. Dr. S. Schumann

Mitglieder der Prüfungskommission:

Referent:	Prof. Dr. A. Quadt II. Physikalisches Institut, Georg-August-Universität Göttingen
Korreferent:	Prof. Dr. S. Lai II. Physikalisches Institut, Georg-August-Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

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

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

Prof. Dr. W. Kollatschny
Instit für Astrophysik, Georg-August-Universität Göttingen

Prof. A. Juste, PhD
Institut de Física d'Altes Energies (IFAE), Universitat Autònoma de Barcelona

Tag der mündlichen Prüfung:

Referenz: Diss-2020/02

Search for the production of four top quarks in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ in the single and opposite sign dilepton channels with the ATLAS detector at the LHC using a Monte Carlo correction method

Abstract

The search for the production of $t\bar{t}t\bar{t}$ in pp collisions at $\sqrt{s} = 13$ TeV is presented in this thesis for the full LHC Run II dataset of $139\,\text{fb}^{-1}$ collected by the ATLAS experiment at CERN. With an expected cross-section of $\sigma_{t\bar{t}t\bar{t}} = 11.97^{+18\%}_{-21\%}\,\text{fb}$, the process is one of the rarest being studied at the LHC. This document focuses on high jet and b-jet multiplicities with one or two leptons of opposite charge. A Boosted Decision Tree and the sum of the transverse momenta of several objects are used as discriminant and to constrain uncertainties on the dominant $t\bar{t}$+jets background. To obtain a reliable background prediction, scale factors are derived which are applied to the nominal Monte Carlo prediction in the fitted regions. A profile likelihood fit is used to measure a signal strength of $\mu_{t\bar{t}t\bar{t}} = \frac{\sigma_{t\bar{t}t\bar{t}}}{\sigma^{\text{SM}}_{t\bar{t}t\bar{t}}} = 1.7^{+0.7}_{-0.7}(\text{stat})\,^{+1.0}_{-0.7}(\text{syst})$ with an (expected) significance of 1.76 (1.08) standard deviations. The result is still in agreement with the Standard Model prediction. Due to the excess observed in events with multiple leptons and two leptons of same charge, the combined result of $\mu_{t\bar{t}t\bar{t}} = 2.0^{+0.4}_{-0.4}(\text{stat})\,^{+0.3}_{-0.3}(\text{syst})$ shows an excess and first evidence of the $t\bar{t}t\bar{t}$ process with a significance of 4.71 (2.9) σ.

Suche nach der Produktion von vier Top-Quarks in Proton-Proton Kollisonen mit $\sqrt{s} = 13\,\text{TeV}$ im Kanal mit einem oder zwei Leptonen unterschiedlicher Ladung mit dem ATLAS Detektor am LHC beruhend auf einer Korrekturmethode der Monte Carlo Vorhersage

Zusammenfassung

In dieser Arbeit wird die Suche nach der Produktion von $t\bar{t}t\bar{t}$ in pp Kollisionen des LHC bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV vorgestellt. Hierfür wird der Run II Datensatz des ATLAS Experimentes mit einer integrierten Luminosität von $139\,\text{fb}^{-1}$ verwendet. Momentan ist der gesuchte Prozess mit einem erwarteten Wirkungsquerschnitt von $\sigma_{t\bar{t}t\bar{t}} = 11.97^{+18\%}_{-21\%}$ fb einer der seltensten der am LHC untersucht wird. Ereignisse mit einem oder zwei Leptonen unterschiedlicher Ladung und einer hohen Jet und b-Jet Multiplizität werden in der Suche verwendet. Das Ausgangssignal eines Boosted Decision Trees sowie die Summe der transversen Impulse verschiedener Objekte werden als Diskriminante und zur Bestimmung der Unsicherheiten des dominierenden $t\bar{t}$+jets Untergrundes verwendet. Um eine zuverlässige Vorhersage dieses Untergrundes in dem untersuchten Phasenraum zu erhalten, werden zusätzliche Skalenfaktoren berechnet, welche auf die nominelle Monte Carlo Vorhersage in den Fitregionen angewandt werden. Ein Profile-Likelihood-Fit wird genutzt, um die Signalstärke $\mu_{t\bar{t}t\bar{t}} = \frac{\sigma_{t\bar{t}t\bar{t}}}{\sigma^{\text{SM}}_{t\bar{t}t\bar{t}}} = 1.7^{+0.7}_{-0.7}(\text{stat})\,^{+1.0}_{-0.7}(\text{syst})$ mit einer (erwarteten) Signifikanz von 1.76 (1.08) Standardabweichungen zu ermitteln. Aufgrund der erhöhten Signalstärke in Ereignissen mit mehreren oder zwei Leptonen mit gleicher Ladung, wird eine Abweichung von der Vorhersage des Standardmodells von $\mu_{t\bar{t}t\bar{t}} = 2.0^{+0.4}_{-0.4}(\text{stat})\,^{+0.3}_{-0.3}(\text{syst})$ im kombinierten Ergebniss beobachtet. Dies entspricht einer Signifikanz von 4.71 (2.9) Standardabweichungen womit ein erster Hinweis auf den $t\bar{t}t\bar{t}$ Prozess beobachtet wird.

Acknowledgements

I would like to take this opportunity to thank all the people who helped me to produce this document. First, a big **thank you** to Arnulf and my friends and colleagues of the institute. Without the moral support in endless coffee breaks, feedback and discussion in critical phases my personal contributions to this analysis would not have been possible. Thanks!

In particular I would like to thank the "4-tops Hackathon Survival Crew" and Paolo for an awesome team spirit and working attitude. The many discussions, jokes and chats made this time an incredible once in a lifetime experience. You helped me to develop my skills and knowledge while enjoying it. Thanks!

Last but not least I want to mention my family and girlfriend for moral support outside the crazy world of particle physics. You helped me to keep a clear head and allowed me to focus on the crucial task of optimizing this fit setup while the rest of my life was upside down. Well done!

I will always keep this time in good memory, thanks to all of you!

Contents

CHAPTER 1

Introduction

The discovery of the *electron* by J.J. Thomson [1] marked the beginning of high energy and nuclear physics. While in the first half of the 20th century scientists were mostly focused on the structure of atoms and nuclei, more fundamental structures were studied in the second half. In parallel to more sophisticated experimental measurements, theoretical descriptions progressed as well, ultimately resulting in the Standard Model (SM) of particle physics as we know it today.
A first hypothesis that elementary particles have quantised properties was already made by Max Planck in 1900. This idea was further developed by many physicists, such as Niels Bohr, and resulted in the mathematical description of quantum mechanics which is the foundation of modern quantum field theories.
On the experimental side, after the famous Rutherford experiment in 1911, it became clear that atoms were made of a heavy nucleus which is surrounded by electrons. Protons (1919) and neutrons (1931) were discovered in the following years confirming that the nuclei themselves are not elementary. At this stage, the atomic structure was understood but more and more experimental observations of new particles were made. Muons (1938) and charged pions (1947) were discovered in cosmic rays. Higher energies and therefore particle colliders were needed for further studies. With the new experimental possibilities, many new particles were discovered in a short time but no underlying theory could explain the origin and relation between the particles. In search of a mathematical foundation, the SU(3) group was used to bring structure in the so-called *particle zoo*, which lead to the prediction of quarks in 1964 by Gell-Mann and Zweig. After further developments of the theory, such as the introduction of colour charges, Quantum Chromo Dynamics (QCD) was the accepted theory to describe the strong force.
With the discovery of the Z and W bosons, it was confirmed that the theory of the weak force, developed after the observation of β-decays, and the theory of the electromagnetic force are manifestations of the same more fundamental electroweak theory. With the

discovery of the bottom quark (1977) at FERMILAB, gluons at PETRA (1979) and the top quark by CDF and DØ (1995), all particles required by QCD had been discovered. The last missing pieces were the τ-neutrino (2000), to complete the electroweak sector, and the Higgs boson (2012) [2, 3], to explain the origin of masses in the SM.
While the SM is a well accepted theory, it is known that not all observations can be described by it, as further discussed in a short overview of the SM in Chapter 2.

In this thesis, the search for the production of four top quarks is presented. With a very low cross-section of about 12 fb it is one of the rarest SM processes being studied at the LHC. A challenging phase-space with many jets and b-jets is the key signature of the process. By measuring this rare process, which requires separating it from a large background contribution by top quark pairs, the experimental precision in this phase-space is probed. The available Monte-Carlo generators do not provide accurate predictions of the additional jets in the background processes. Therefore, additional correction factors are derived based on the observed difference between Monte-Carlo and data in background enriched regions. In addition, a variety of theories beyond the SM could have detector signatures similar or identical to the one of the four top-quark production. The measured cross-section provides valuable feedback for theorists regarding the parameters of these new theories.

A more detailed overview on the current knowledge of top-quarks and the properties of the four top quark production is given in Chapter 2.1. In order to measure such a rare process on top of large background contributions, a large amount of data is required. These days, this is only possible with the Large Hadron Collider (LHC) at CERN and modern experiments like the ATLAS experiment which are both introduced in Chapter 3. Subsequently, the focus is placed on the analysis itself starting by introducing the Monte-Carlo samples and object definitions in Chapter 4. Chapter 5 summarises the key aspects of the analysis technique as well as the setup for the profile likelihood fit. Finally, the results are presented in Chapter 6 in which also an outlook is given.

CHAPTER 2

The Standard Model of Particle Physics

The SM [4,5] describes physics on smallest scales, equivalent to high energies, in a locally gauge invariant *quantum field theory* based on the $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry group. It describes the dynamics of all elementary particles and their interactions via three fundamental forces. In the following, it is summarised how the observed objects and their interactions are connected to the mathematical description.
A graphical representation of the particles in the SM is given in Figure 2.1. For each particle shown, a corresponding anti-particle with opposite quantum-numbers exists. The particles can be divided in two groups according to their *spin*. Fermions carry half-integer spins in units of $\hbar$ and can be further split in leptons (blue) and quarks (green). *Gauge bosons* (red and yellow) carry integer spin in units of $\hbar$ and are the mediators of the fundamental forces.

Fermions can be arranged in doublets according to the third component of the weak isospin (I_W^3). The doublets are sorted by increasing mass in so-called generations where all quantum numbers are identical in each generation. Electron- (e), muon- (μ) and tau-lepton (τ) carry an integer charge and a weak isospin of $I_W^3 = -\frac{1}{2}$. They are grouped together with the corresponding neutrinos which carry a weak isospin of $I_W^3 = +\frac{1}{2}$ and are electrically neutral. For quarks, the up-type partners ($I_W^3 = +\frac{1}{2}$) carries an electrical charge of $q = \frac{2}{3}$ while the down-type isospin partner ($I_W^3 = -\frac{1}{2}$) have a charge of $q = -\frac{1}{3}$.
Except for the Higgs boson (spin$=0$), all mediators of the fundamental forces are vector-bosons (spin$=1$). The *photon* (γ) is the massless mediator of the electromagnetic force and acts on all electrically charged particles. In the SM, it is represented by the $\mathrm{U}(1)_{EM}$ [6,7] symmetry group which conserves the charge Q. The only other massless bosons are the eight gluons (g). They only interact with quarks as other fermions do not carry a colour charge C. While γ is a neutral particle w.r.t. all forces, each gluon carries a distinct combination of two colour charges itself and can therefore inter-

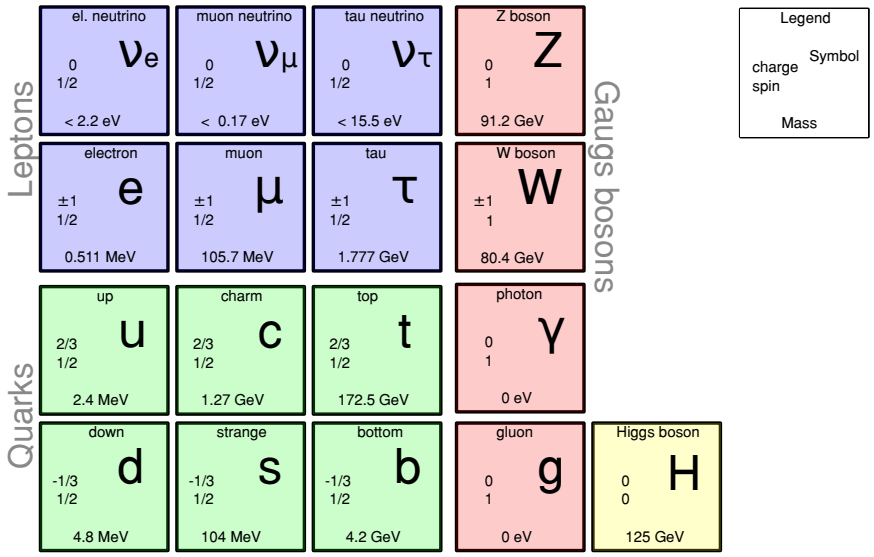


Figure 2.1.: Schematic view of the particles in the Standard Model of particle physics. The particles are grouped in leptons (blue), quarks (green), gauge bosons (red) and the Higgs boson (yellow).

act with itself. QCD is represented by the $SU(3)_C$ [8] symmetry group and conserves C from which the name *quantum-chromo-dynamics* is derived.
For the weak force, of which the Z^0 and $W^{\pm}$ boson are the mediators, the flavour of leptons must be conserved. For quarks, the weak eigenstate do not correspond to the mass eigenstates which are observed. Therefore, quarks can couple among generations which is described by the non-diagonal elements of the CKM matrix [9, 10]. In general, flavour changing *neutral* currents, where a particle keeps the same isospin but changes generation by the interaction with an electrical neutral particle, are suppressed in the SM. The weak interaction conserves the weak isospin I_W. It is maximally parity violating and only acts on left-handed particles. The underlying symmetry group is the $SU(2)_L$ [11].

As mentioned, the mathematical representation of the SM is a quantum field theory. Each particle corresponds to an excitation of a quantum field which satisfies appropriate quantum mechanical field equations: chiral fields describe fermions, vectorial fields describe bosons and scalar fields describe scalar particles. One possible requirement on these fields is *global* gauge invariance under

$$\mathcal{L}(\Psi) \rightarrow \mathcal{L}(\Psi e^{i\theta}) \tag{2.1}$$

of the *Lagrangian density* for a field $\Psi(\mathcal{L}(\Psi))$ under a transformation $e^{i\theta}$. To introduce interactions among the SM particles, invariance under *local* transformations $e^{i\theta(x)}$,

which depend on the space-time coordinates x, is required. Introducing such a local transformation to the fields breaks the symmetry of the Lagrangian density. To restore the symmetry, the derivative on a field has to be replaced by the covariant derivate (D_μ) where the exact form of D_μ depends on the symmetry group. By doing so, additional terms are introduced corresponding to the interaction terms. In the SM, the derivative becomes

$$D_\mu = \partial_\mu + \underbrace{ig_s G_\mu}_{SU(3)} + \underbrace{igW_\mu}_{SU(2)} + \underbrace{ig\tan\theta_W Y_\omega/2 B_\mu}_{U(1)} \tag{2.2}$$

with the coupling constants of the three forces g_s, g and $g\tan\theta_W$, where θ_W is the weak mixing angle, and the generators G_μ, W_μ and B_μ of the respective symmetry group. The dimensionality of the generators corresponds to the number of bosons of the respective force. As discussed above, the generators of the $\mathrm{SU}(3)_C$ group correspond to the strong, of the $\mathrm{SU}(2)_L$ group to the weak and of the $\mathrm{U}(1)_Y$ group to the electromagnetic force.

The Brout-Englert-Higgs mechanism [12–14] introduces mass terms for all massive particles in the SM by introducing a complex doublet scalar field. Due to spontaneous symmetry breaking, the vacuum expectation value of this field is non-zero. As a consequence, massive particles interact with the field. The corresponding *Yukawa coupling* is proportional to the mass of the particle. The Higgs boson is an excitation of this Higgs field and carries no colour or electrical charge but a weak isospin and can interact with itself.
By forming the Lagrange density of all particles and their interactions, the dynamics of the theory can be obtained by the Lagrange formalism. A generalized form of the Lagrange density is [15, 16]

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{Fermion}} + \mathcal{L}_{\mathrm{Gauge}} + \mathcal{L}_{\mathrm{Yukawa}} + \mathcal{L}_{\mathrm{Higgs}} + \mathcal{L}_{\mathrm{Technical}} \tag{2.3}$$

where

- $\mathcal{L}_{\mathrm{Fermion}}$ describes kinematics and interactions of fermions with the respective gauge field
- $\mathcal{L}_{\mathrm{Gauge}}$ describes self interactions of the gauge fields
- $\mathcal{L}_{\mathrm{Yukawa}}$ describes the coupling of fermions to the Higgs field
- $\mathcal{L}_{\mathrm{Higgs}}$ describes the Higgs field, its self-interaction and the interaction between the Higgs and gauge bosons
- $\mathcal{L}_{\mathrm{Technical}}$ contains additional terms such as gauge fixing terms.

With the Lagrangian, the amplitude of the transition from an initial to a final state can be calculated. Typically, this is done in a perturbative approach under the condition of small coupling constants. A typical example of such a process is the interaction of two particles a and b which, via a certain force, form two new particles c and d. From the

amplitude, the cross-section ($\sigma_{ab\to cd}$) can be calculated via Fermi's Golden Rule. For non-physical processes, the amplitude and therefore the cross-section vanishes. With the Feynman rules, a graphical interpretation of the non-vanishing amplitudes can be made. The order of a diagram is given by the number of closed paths (loops) in it. Leading Order (LO) refers to diagrams with the minimum number of vertices and no additional loops. Diagrams with one extra vertex are at Next-to-Leading-Order (NLO) etc..

Although the SM is in very good agreement with current observations from experiments, it has short-comings. Neutrinos are known to undergo flavour-oscillations [17–19] which is not allowed in the SM. There is no explanation for dark matter and dark energy, which make up 95% of the energy density of the universe [20]. Gravitation, which is very weak compared to the three forces discussed above, is not included in the theory either. For this reason, many theories beyond the SM exist, and it is important to probe the SM and its predictions with high precision.

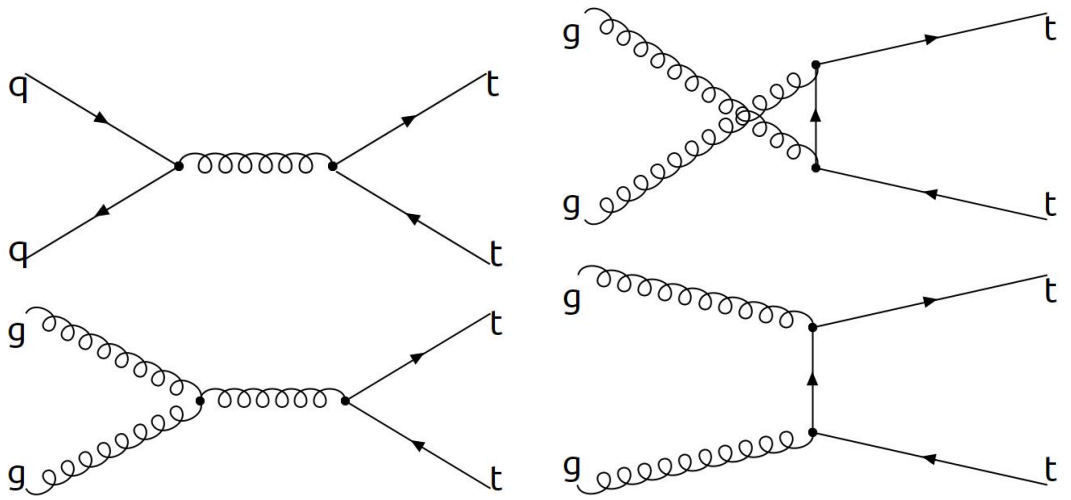
2.1. The Top Quark

The discovery of *CP*-violation in neutral K-mesons [21] could not be explained with only two generations of quarks and leptons considered at that time. The idea of a third generation of quarks [10] could explain the effect and was confirmed by the discovery of the τ-lepton [22] and the *bottom*-quark [23] (b) in the 70's. It took 18 more years until the *top*-quark (t) was discovered by CDF [24] and DØ [25] at the TEVATRON in $p\bar{p}$ collisions.
Its mass is a free parameter in the SM which has to be measured. It has a mass of approximately $172.9\pm0.4\,\text{GeV}$ [26] making it the heaviest particle in the SM. Via loop corrections, the mass is related to other parameters of the SM such as the W-boson and Higgs boson masses.
A consequence of the high top quark mass is the large decay width of $\approx 1.5\,\text{GeV}$ corresponding to a lifetime of $\tau_t \approx 5 \times 10^{-25}\,\text{s}$. This is shorter than the typical timescale of hadronisation $t_{\text{had}} = \frac{1}{\Lambda_{\text{QCD}}} \approx \frac{1}{200\,\text{MeV}} \approx 3 \times 10^{-24}\,\text{s}$. Therefore, on average, top-quarks decay before forming jets in the detector and decay as bare quarks, making it a unique option to study the quark sector.

2.1.1. Top Quark Pairs and Single Top Quarks at Hadron Colliders

At hadron colliders, top quarks are predominantly produced as *$t\bar{t}$-pairs*. As shown in Figure 2.2, several production processes at LO are possible. The probability to find a parton (quark or gluon) p' within a proton p that carries the momentum fraction x of the proton at a certain energy scale Q, is described by parton distribution functions $f_{p'/p}(x, Q^2)$ (PDFs). The cross-section of a given process initiated by the interaction of two partons is then given by the convolution

$$\sigma(x, Q^2) = \sum_{p'} C_{p'} * f_{p'/p}(x, Q^2), \tag{2.4}$$



(a) s-channel (b) u/t-channel

Figure 2.2.: LO production diagrams for $t\bar{t}$-pairs at hadron colliders via (a) s-channel and (b) u (up) or t-channel (down).

where $C_{p'}$ is the matrix-element and can be calculated from theory. At the TEVATRON, the annihilation of a quark and antiquark was the main production mechanism. At the LHC, a pp-collider, the centre of mass energy is higher and a smaller fraction of the proton energy is required to produce top-quarks. Since no anti-quarks are available as valence quarks, the PDFs and therefore the production of $t\bar{t}$-pairs are dominated by gluons. Even though the strong coupling constant α_s becomes smaller at higher energies, resulting in a decrease of the cross-section, overall the $t\bar{t}$ cross-section rises due to the increasing number of gluons. The cross-section has been studied extensively at both colliders and good agreement with SM predictions is observed as summarised in Figure 2.3. The most precise measurement of the cross-section

$$\sigma_{t\bar{t}}(\sqrt{s} = 13\,\text{TeV}) = 826.4 \pm 3.6(\text{stat}) \pm 11.5(\text{syst}) \pm 15.7(\text{lumi}) \pm 1.9(\text{beam})\,\text{pb}$$

by ATLAS was made with the 2015-2016 dataset ($36\,\text{fb}^{-1}$) in the dilepton $e\mu$-channel [27]. The most recent result with the full Run II dataset was made in the lepton+jets channel and $\sigma_{t\bar{t}}(\sqrt{s} = 13\,\text{TeV}) = 830.4 \pm 0.4(\text{stat})^{+38.2}_{-37.0}(\text{syst})\,\text{pb}$ [28] was measured.

Single top quark production has a total cross-section of $\sigma_t = 350\,\text{pb}$ and is another option to study top quarks at the LHC. As shown in Figure 2.4, the production process involves, depending on the channel, the exchange or radiation of a W boson and is therefore a weak process. While in the majority of cases the top-quark is produced in association with one ore two other quarks in the t- (70%) and s-channel (5%), it can also be produced together with a W boson (25%) leading to different signatures in the detector.

The CKM matrix-element for the coupling of top- to bottom quark is about unity and it therefore predominantly decays to a W boson and a $b-$quark. Due to the colour-factor of the strong interaction, quarks are three times more likely to be produced in the W boson decay compared to leptons. Therefore, the branching ratios of the top-quark-decay are [26]

$$t \rightarrow Wb \rightarrow \begin{cases} qq'b & (67.4\%) \\ l\nu_l b & (32.6\%) \end{cases} \tag{2.5}$$

from which for the branching ratios of the $t\bar{t}$-pair decay channels

$$t\bar{t} \rightarrow W^+bW^-\bar{b} \rightarrow b\bar{b} + \begin{cases} qq'q''q''' & (45.7\%) \\ qq'\,l^{\pm}\,\nu_l & (43.8\%) \\ \bar{l}\,\upsilon_l\,l'\,\bar{\nu}_{l'} & (10.5\%) \end{cases} \tag{2.6}$$

follow.
Due to a large amount of data delivered by the LHC, it is possible to study the properties of the top quark in detail. In particular measurements of top-quarks in association with $W/Z/\gamma$ [29, 30] or a Higgs boson [31, 32] were performed to study the couplings of the top-quark. In addition, the cross-section was measured differentially as function of many kinematic variables [33–35] and in particular phase spaces [36] providing consistency checks with theory predictions.

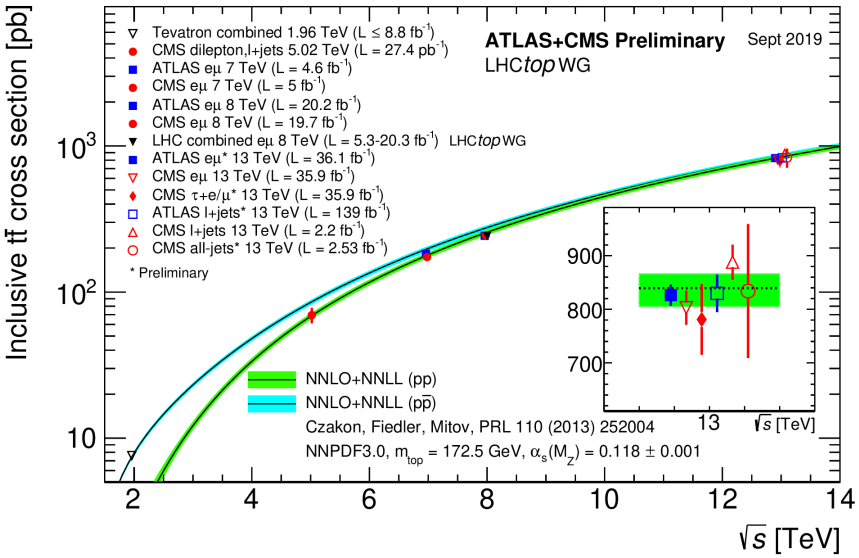
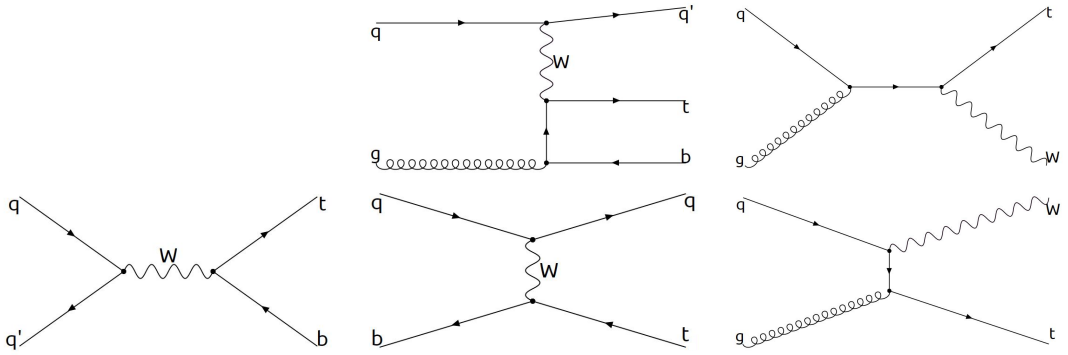


Figure 2.3.: $t\bar{t}$ cross-section measurements made at the TEVATRON and at the LHC compared to predictions from theory [37].



(a) s-channel (b) t-channel (c) W associated

Figure 2.4.: LO production diagrams for single top quarks at hadron colliders via (a) s-channel, (b) t-channel and (c) in association with a W-boson.

2.1.2. Four Top Quarks

Similar to the production of $t\bar{t}$-pairs, the majority of $t\bar{t}t\bar{t}$ events are initiated by gg via the processes at LO shown in Figure 2.5 and only about 6% are produced via $q\bar{q}$ annihilation. The expected SM cross-section for pp collisons at $\sqrt{(}s) = 13\,\text{TeV}$ is $\sigma_{t\bar{t}t\bar{t}} = 11.97^{+18\%}_{-21\%}\,\text{fb}$ @ NLO QCD+EW [38] making it one of the rarest SM processes being studied at the LHC. While always four b-quarks from the top-decay are expected, the decays of the four W bosons give rise to a variety of decay channels which are summarised in Figure 2.6. The single lepton (1L) and the dilepton channel, where the leptons have opposite charges (OS), are both dominated by the $t\bar{t}$+jets background and are therefore grouped to the 1L+OS channel. Similarly, the dilepton channel, where the leptons have the same charge (SS), and the multilepton (ML) channel are dominated by instrumental backgrounds and are therefore grouped in the SS+ML channel.

Not only is the $t\bar{t}t\bar{t}$ cross-section far lower than any other process involving top-quarks measured so far, but the final state is also sensitive to many theories beyond the SM which can be probed simultaneously by measuring the process.
Some production mechanisms of beyond SM (BSM) scenarios are shown in Figure 2.7. They can be grouped into two cases: 1) additional contributions interfering with the SM $t\bar{t}t\bar{t}$ production and therefore altering the cross-section (e.g. Figure 2.7 a)-c)); 2) new contributions with a similar signature in the detector leading to an excess of measured $t\bar{t}t\bar{t}$ events (e.g. Figure 2.7 d)).
Some sources of new particles are listed below:

- Additional electrically neutral heavy Higgs bosons predicted by theories with two Higgs doublet models (2HDM) where the boson decays to $t\bar{t}$ (Figure 2.7 a)) [39,40]

- Gluinos or sgluons are supersymmetric partners of the gluon. If required that they are produced in pairs, each particle decays into $t\bar{t}$ (Figure 2.7 c)) [41]

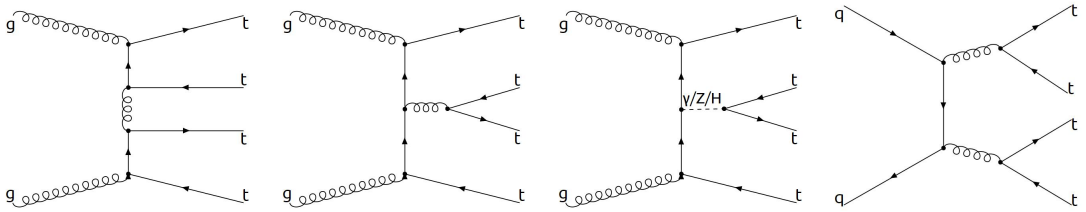


Figure 2.5.: Four of the LO diagrams for the production of four top-quarks at a hadron collider.

- Additional heavy Kaluza-Klein gluons, similar to SM gluons, as a consequence of extra-dimensions [42, 43]
- Top-philic particles as a consequence of wider symmetry groups leading to an excess in the $t\bar{t}t\bar{t}$ phase-space with many jets and b-quarks [44]
- Vector-like quarks (VLQ) which decay into tW resulting in a similar detector-signature, the missing b-quarks can originate from gluon radiation (Figure 2.7 d)) [45].

Furthermore, in case the top-quark is not a fundamental particle but composed of even more fundamental ones, the SM would be an *effective* field theory (EFT). Additional loops by new heavy particles are also possible and both cases can be described by introducing contact interactions to the SM $t\bar{t}t\bar{t}$ production (Figure 2.7 b)) [46]. Measuring the cross-section of the $t\bar{t}t\bar{t}$ production therefore not only probes the SM, it also sets limits on the parameters for theories beyond the SM which can be used by theorists to develop these theories further [47].
The latest searches of Atlas and Cms looked for a $t\bar{t}t\bar{t}$ signature to set limits on the SM cross-section and on BSM signals. In a dataset with $36\,\text{fb}^{-1}$, Atlas analysed the SS+ML channel [48] as well as the 1L+OS channel [49]. In the combined result, an observed (expected) upper limit at 95% confidence level (CL) on the cross section was set at 48.7 (19.3) fb [50]. In a similar search, Cms sets a limit (95% CL) of 33 (20) fb.
Recent results for Atlas and Cms exist for the full Run II datasets, both experiments aim for a 3.0σ deviation from the background-only hypothesis. Cms ($137\,\text{fb}^{-1}$), measures $\sigma_{t\bar{t}t\bar{t}} = 12.6^{+8.8}_{-5.2}$fb, with a signal significance of 2.6 (2.7) σ [51] and Atlas measures $\sigma_{t\bar{t}t\bar{t}} = 24^{+7}_{-5}$fb, corresponding to 4.4 an excess of (2.4) σ [52]. In this thesis, the Atlas measurement in the 1L+OS channel with the full Run II dataset as well as the first combination with the dataset are presented.

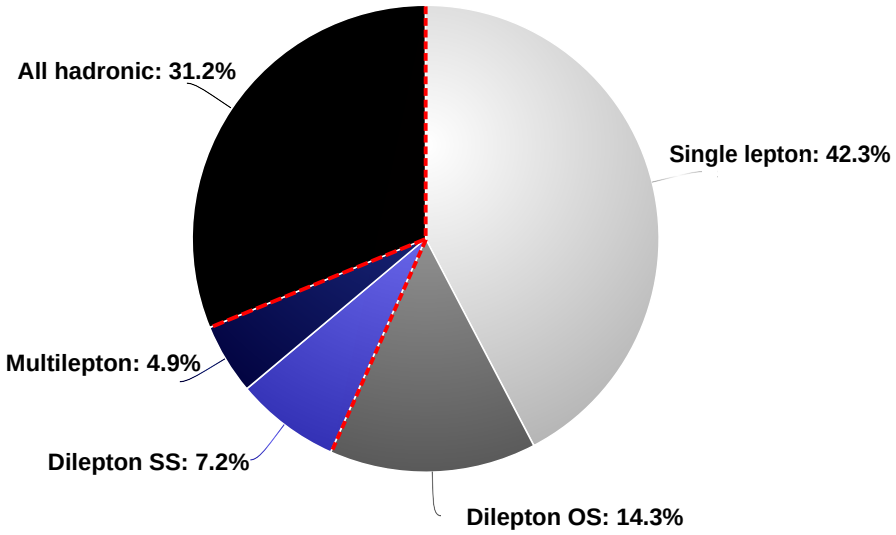
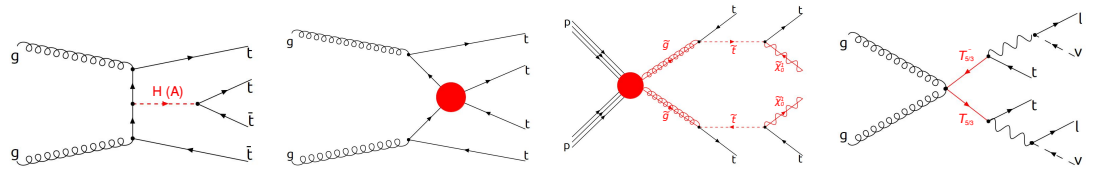


Figure 2.6.: Summary of the decay channels of four top quarks and their branching ratios. The red-dashed line indicates the channels with similar dominant background. OS (opposite-sign) and SS (same-sign) stand for opposite or same charge of the two leptons.



(a) Heavy Higgs boson (b) Contact Interaction (c) Supersymmetry (d) Vector-like Quarks

Figure 2.7.: Processes beyond the Standard Model with a similar signature in the detector. (a) Decay of a heavy Higgs boson (e.g. from 2HDM) in association with a $t\bar{t}$-pair, (b) four-top contact interaction in an effective field theory, (c) production of four top quarks via supersymmetric gluinos, (d) decay of a vector-like quark in a top quark and a boson.

CHAPTER 3

Experimental Setup

The Large Hadron Collider (LHC) [53] has a circumference of 27 km and is the largest and most modern proton-proton and heavy-ion collider on earth. It is operated by the European Organization for Nuclear Research (CERN) in Switzerland. Before the LHC, a variety of hadron and lepton colliders were already built. The Intersecting Storage Ring (ISR, 1971) [54] was the first pp collider at CERN. In order to reach higher-energies in the search for new processes, new hadron colliders such as S$p\bar{p}$S (CERN, 1981) [55] and TEVATRON [56] (Fermilab, 1987) were required. They made use of the annihilation of the valence quarks in proton/anti-proton collisions. Complementary studies were performed at e^+e^- colliders such as LEP (CERN, 1987) [57] and SLC (SLAC, 1991) [58]. Together, the experiments located at these colliders increased the knowledge of particle physics significantly, but it also became clear that a new machine with a higher centre of mass energy ($\sqrt{s}$) was needed. At the beginning of this century, the LHC was installed in the same tunnel that was previously used by LEP. In Section 3.1, more details on the LHC are given followed by an overview of the ATLAS detector in Section 3.2.

3.1. The Large Hadron Collider

The LHC started operating in stable conditions in March 2010 with $\sqrt{s} = 7/8$ TeV until the end of 2012. Within a partial Run I dataset, the Higgs-Boson was discovered in July 2012. After a longer shutdown (LS1), during which upgrades to the accelerator and the detectors have been installed allowing for an increase of $\sqrt{s}$ to 13 TeV, Run II started in April 2015 and lasted until December 2018. After another long shutdown, LHC presumably starts Run III in 2021.
Before protons are collided in the main ring, their energy successively gets increased in the CERN accelerator complex as shown in Figure 3.1. The protons originate from hydrogen gas from which electrons are removed by an electric field. In the linear accel-

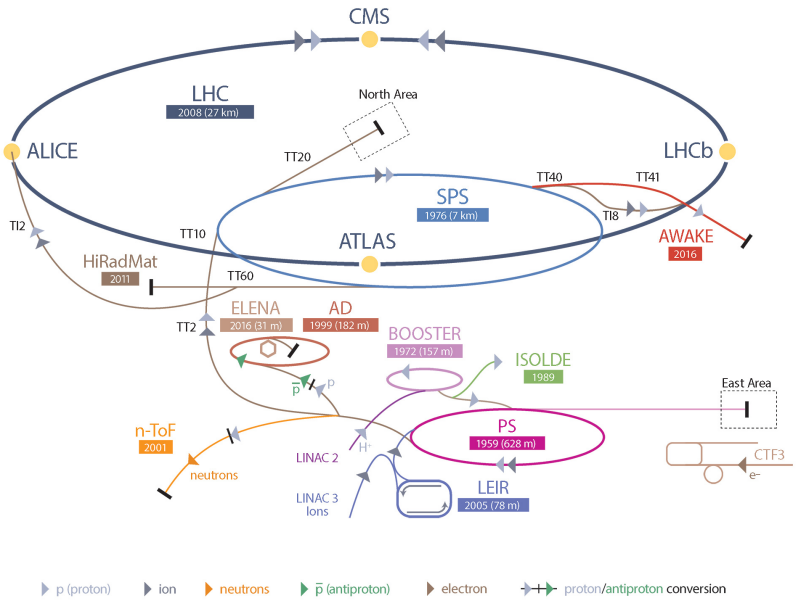


Figure 3.1.: Schematic view of the CERN accelerator complex ©CERN.

erator LINAC2 the particles are boosted to an energy of 50 MeV before the Proton Synchrotron Booster (PSB) accelerates them further to 1.4 GeV. After reaching an energy of 25 GeV in the Proton Synchrotron (PS), the Super Proton Synchrotron (SPS) increases the energy further to 450 GeV. The 2808 proton bunches, consisting of 11.2×10^{11} protons each, are accelerated in the main LHC ring until they reach the final energy of 6.5 TeV in the two counterrotating beams by 8 radio frequency cavities, oscillating at 400 MHz, per beam. The ring consists of 1232 dipole magnets, bending the particles on a circular path, and 392 quadrupole magnets which focus the beam and thereby reduce beam losses. All magnets in the LHC are cooled to 1.9° K in order to maintain the superconducting state which is needed to obtain the necessary field strengths.
Prior to each interaction point, at which one of the four main experiments ALICE [59], ATLAS [60], CMS [61] and LHCb [62] are located, dedicated magnets focus the beam to increase the luminosity. Every 25 ns, corresponding to a frequency of 40 MHz, two proton-bunches are collided and the collisions are observed by the four experiments. While ALICE investigates *quark-gluon plasmas* and LHCb focuses on measuring CP-violating processes in the b-quark sector, ATLAS and CMS are general purpose detectors investigating a variety of topologies which are related to the SM and beyond.

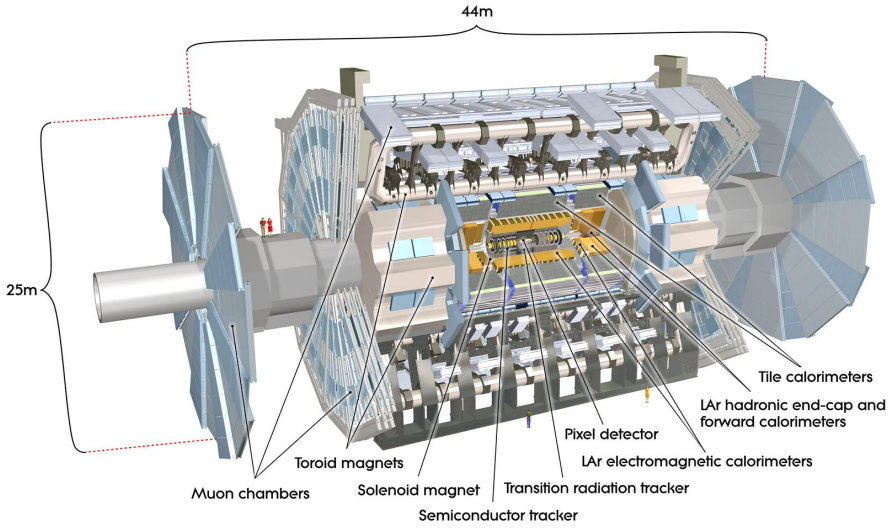


Figure 3.2.: The ATLAS detector at CERN ©CERN.

3.2. The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS) detector [63], shown in Figure 3.2, with the size of $44\,\mathrm{m} \times 25\,\mathrm{m}$, is the largest of the four main experiments at CERN. It is located roughly 100 m below ground to provide shielding from cosmic backgrounds. The physics goals are similar to that of CMS. While the conceptional designs of the detectors are similar, they differ substantially in detail such as material choices for each component. Both detectors need to fulfil a list of requirements like very precise tracking information, good energy resolution, good spatial resolution and precise particle identification. For this reason, the detectors consist of multiple layers.
The following coordinate system is used within the thesis. The z-axis points along the beam-pipe and the origin is placed at the nominal interaction point in the middle of the detector. The x/y direction defines the transverse plane where y is pointing upwards and perpendicular to the earth-surface while x points towards the centre of the LHC ring. Due to the cylindrical shape of the detector the distance (r) from the z-axis and the azimuthal angle (ϕ) can be used.
Furthermore, it is common in high energy physics to define the *pseudo-rapidity* (η) based on the polar angle (θ) as

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \tag{3.1}$$

With this definition, the distance between two objects can be measured by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

The Inner Detector (ID) is located closest to the beam pipe making the environment for its components particularly challenging due to the high radiation. It is embedded by a 2 T magnetic field from the central solenoid which bends charged particles for momentum measurements. Charged particle tracks are reconstructed by combining information from the different layers of the pixel detector, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT).
The pixel detector consists of 4-layers in the central part with a total of 100 million pixels and three disks arranged in an endcap on each side with 13 million pixels per disk per side. It covers the range of $0.0 < |\eta| < 2.5$. The Insertable B-Layer [64] is the innermost layer and was added during LS1. It has a reduced pixel size of $50 \times 250\,\mu\text{m}^2$ compared to $50 \times 400\,\mu\text{m}^2$ in other layers resulting in a better resolution along the beam-pipe (75 µm instead of 115 µm) while the resolution in the transverse plane is similar (10 µm).
The SCT uses eight-layers of silicon microstrips to provide additional information at a larger radius where the combination of two layers provides one space-point measurement. It covers the range of $0.0 < |\eta| < 1.4$ in the barrel and $1.4 < |\eta| < 2.5$ via the end-caps which are located on each side. The strips in a pair of layers are rotated by 40 mrad with respect to each other to avoid ghost hits. With readout strips every 8 µm the resolution is worse than in the pixel detector with 17 µm in the transverse direction and 580 µm along the beam-pipe.
The TRT consists of 4-mm tubes which are located along the beam-pipe in the central part and perpendicular to it in the end-caps. The tubes are filled with a gas-mixture (70% Xe, 27% CO_2, 3% O_2) allowing for an improved recognition of electrons when combining TRT information with the calorimeter. It covers the range of $0.7 < |\eta| < 2.5$ where the range of It covers the range of $0.7 < |\eta| < 2.5$ is covered by the end-caps. The resolution in the transverse plane is larger than for the previous sub-detectors with 130 µm, but the measured path of the traversing particle is longer due to the larger volume. In direction of the beam-pipe, no precise tracking information is available by the TRT.
Due to the curvature of charged particles in the magnetic field the path of the particle can be reconstructed in each layer. Since the curvature depends on the transverse momentum (p_{T}), the resolution is p_{T} dependent. Combining the information of all layers results in a resolution of [65]:

$$\frac{\sigma_{p_{\text{T}}}}{p_{\text{T}}} = 0.04\% \times p_{\text{T}}\,[\text{GeV}] \oplus 2\% \tag{3.2}$$

Tracking information is used to identify the primary vertex of the pp collision in the beam-pipe which is in particular important when multiple pp collisions in the same bunch-crossing occur (pile-up). Secondary vertices from b-hadron decays are also reconstructed with dedicated techniques based on the reconstructed tracks.

The calorimetry system consists of two main parts, the electromagnetic (ECal) [66] and the hadronic calorimeter [67]. In general, particles are stopped in the calorimeters and the deposited energy is recorded from which the energy of the particles and the missing energy in the transverse plane, originating primarily from neutrinos, can be

reconstructed.
ECal is a sampling calorimeter located outside of the TRT. Similar to the inner detector it consists of three central layers ($|\eta| < 1.375$) and two end-caps which provide a larger angular coverage ($|\eta| < 3.5$). In the passive material (lead), photons and electrons undergo pair-production and Bremsstrahlung resulting in an electromagnetic shower. Through ionisation in the active material (liquid argon) an electrical signal proportional to the initial particles energy is obtained. The thickness of the calorimeter was chosen to be > 24 radiation lengths to absorb a large fraction of the electromagnetic shower. The more energy is absorbed, the better the resolution becomes and punch-through in outer detectors is reduced.
Further outside, the Tile calorimeter, also known as hadronic calorimeter, covers a range of $|\eta| < 1.0$ in the central part and $|\eta| < 2.5$ by an extended barrel on each side parallel to the beam-pipe. It is also a sampling calorimeter in which hadronic showers are initiated via the strong interaction in the passive material (iron) which are then measured by scintillators. The scintillators are connected to wavelength shifting fibres which are connected to the readout electronics. The thickness is > 9.7 interaction lengths. For both calorimeters, Tile and ECal, the granularity becomes smaller when getting closer to the beam-pipe. The resolution of ECal for a given energy E [GeV]

$$\frac{\sigma_E^{EM}}{E} = \underbrace{\frac{0.1}{\sqrt{E}}}_{\text{sampling}} \oplus \underbrace{\frac{0.3}{E}}_{\text{noise}} \oplus \underbrace{0.01}_{\text{constant}} \tag{3.3}$$

consists of three terms. The constant term dominates at higher energies ($> 100\,\text{GeV}$), the noise term is dominant at very low energies ($< 10\,\text{GeV}$) and the sampling term dominates in the intermediate range.
For the Tile calorimeter, the resolution for individual hadrons is given by

$$\frac{\sigma_E^{Tile}}{E} = \underbrace{\frac{0.5}{\sqrt{E}}}_{\text{sampling}} \oplus \underbrace{0.03}_{\text{constant}} . \tag{3.4}$$

Here, the sampling term is much larger since not the full shower is contained, and hadronic showers are in itself more complex to reconstruct than electromagnetic ones. For example, due to the decay of charged and neutral pions, hadronic showers contain an electromagnetic component which needs to be calibrated.
In the forward and backward regions of the ATLAS detector, Hadronic Endcaps (copper and liquid argon) overlap with the Forward Calorimeter (FCal, copper/tungsten and liquid argon) to ensure the energy measurements up to $|\eta| < 4.9$.

The Muon Spectrometer (MS) [68] is the outermost part of the ATLAS detector. Since muons are not stopped by the calorimeters, it is possible to perform an additional tracking measurement and use the information for the muon triggers. In the central region ($|\eta| < 1.4$) the magnetic field varies between 0.5-2 T. It is provided by the central toroid

and an overlap of the endcap magnets. These provide a 1-2 T field in the outer region ($1.6 < |\eta| < 2.7$). Monitored drift tubes (MDTs) cover a range of $|\eta| < 2.0$ and consist of three layers. For larger pseudorapidity regions ($2.0 \leq |\eta| < 2.7$) the occupancy is higher. Cathode Strip Chambers (CSCs) are used in this part because of the faster response and better timing resolution. They consist of Multiwire Proportional Chambers (MWPC) combined with a cathode strip readout. The p_{T} resolution degrades from 2-3% for momenta in the order of GeV down to 10% for 1 TeV muons due to the smaller bending radius.
To trigger muon events, Resistive Plate Chambers (RPCs) are used in the barrel region and thin gap chambers (TGCs) in the end-caps. Both provide a well defined p_{T}-threshold and an additional orthogonal measurement of the muon p_{T}.

Luminosity measurements [69, 70] are important to precisely determine the amount of data taken as this is an irreducible uncertainty for all cross-section analyses. By measuring the inelastic pp cross-section in the forward regions, the luminosity can be calculated. The LUCID [70] detector measures the rate of pp events at ± 17 m along the beam-pipe by Cherenkov radiation while ALFA uses scintillators inside Roman pots at ± 140 m.

The trigger and data acquisition system (TDAC) [71] is needed to select and store events. Not all events can be stored due to bandwidth and storage limitations. Therefore, only events that fulfil certain physics criteria, such as having high p_{T} leptons, are kept. The system has to be fast as a decision has to be taken for each bunch-crossing at a frequency of 40 MHz. The trigger-system consists of a hardware (L1) and a software based high level trigger (HLT).
The L1 trigger uses information from the calorimeters and the muon spectrometer to define regions of interest (RoI) in the detector. Only data which comes from an RoI is readout from the on-detector electronics and forwarded to the HLT. Here, information from L1 and all detector components are processed at a rate of 100 kHz. After the HLT, events are selected at a manageable frequency of about 1 kHz and are permanently stored on disks at CERN and are distributed via the world-wide computing grid.

CHAPTER 4

Monte Carlo Simulation and Object Definition

In this chapter, the dataset, simulated samples and physics objects which are used in the analysis are presented. After a summary of the Run II dataset in Chapter 4.1, a brief overview of Monte Carlo (MC) generators and specific choices in this analysis is given in Chapter 4.2. Finally, the physics objects and their reconstruction are summarised in Chapter 4.3.

4.1. Datasets

The full Run II dataset of ATLAS was accumulated between 2015 and 2018 with $\sqrt{s} =$ 13 TeV and has, summed over all years in units of fb^{-1}, a total integrated luminosity of $3.2 + 33 + 44 + 59 = 139$. The integrated luminosity per year increased, after commissioning in 2015, from 33 (2016) to 59 (2018) $\text{fb}^{-1}$by improving the data taking conditions based on the gained knowledge of the previous years. A consequence of this increase is a higher average pile-up, collisions from other protons in the same bunch crossing, which complicates the reconstruction of objects.
Several runs of typically 10-15 hours each, consisting of several luminosity blocks which are around a minute long, were collected each year. Each run has slightly different data taking conditions, for example caused by the injection into LHC. Only luminosity blocks that fulfil certain quality criteria, such as a fully functional detector and at least one clearly identified vertex to reduce pile-up, are listed in the *Good Run List* and are considered in this analysis.

4.2. Monte Carlo Simulation

To interpret the events measured by the detector, they are compared to the SM prediction. Ideally, this is done by an analytical approach taking all effects into account.

However, due to the complexity o a typical particle collision, this is not possible in particle physics. Instead, it is typical to use the Monte Carl (MC) technique, a numerical approach. According to the *theorem of large numbers*, the average of randomly generated events, based on the mathematical problem, converges to the expected value of the problem itself unless it it biased. The difficulty is the exact formulation of the mathematical problem as certain assumptions, further discussed below, have to be made beyond the analytical calculation of the matrix element (ME). In addition, computing resources are often limited which in turn limits the possibilities for further numerical improvements. As further discussed in Section 5, this approach leads to imprecise predictions in the phase-space of this analysis.
For most generators, the simulation is split in several levels, following the factorisation theorem [72], where each level corresponds to a different scale of transferred momenta Q. For each level, different effects are simulated and, therefore, it is common to interface different generators for the ME calculation and the shower generation. Even though the modern generators differ in the exact implementation, they all follow a similar approach [73].
At the LHC, protons are collided but the actual interaction, which leads to the final state of interest, occurs between partons with kinematics given by PDFs which were introduced in Section 2.1. Besides the main interaction, the proton remnants interact as well. This resulting *underlying event* leads to additional final state particles in the detector in the same event, often with a low transverse momentum (soft), which do not originate from the process of interest but cannot be separated. Therefore, this has to be simulated in addition to the main interaction of interest.

The first step of the simulation is the *parton level*. The amplitude of a given process, as discussed in Section 2, of the initial states is simulated at a fixed order in perturbation theory with a ME generator. To avoid ultraviolet divergences from large momentum loops, a reference point for the strong coupling constant α_S at the *renormalisation scale* (μ_R) is chosen. By introducing this scale, the strong coupling constant becomes scale dependent which takes the running with energy into account. Similarly, to avoid infrared divergences due to radiation of massless particles by other massless particles, the *factorisation scale* (μ_F) is chosen to make the PDFs dependent on Q^2. Typically, the scales are related to the masses M of the transferred momenta of the particles being considered and often $\mu_R = \mu_F = Q^2 = M^2$ is chosen. At infinite order of perturbation theory, the final result would not depend on the choice of the scales. Since this is computationally impossible to achieve, uncertainties based on varying the scales have to be taken into account.

Next, a *shower* generator is interfaced with the ME generator to model the radiation of gluons and photons by correspondingly charged particles in the initial and the final states. Since gluons self-interact, they emit further gluon radiation resulting in a *parton-shower* [74]. During this step, the number of particles considered increases while the energy scale decreases which, according to QCD, further increases the gluon radiation.

The process continues until perturbation theory is not applicable anymore due to the growing of the strong coupling constant with decreasing energy. A typical cut-off scale is in the order of $Q^2 = 1\,\mathrm{GeV}^2$. In principle, a parton-shower corresponds to a higher order correction via additional loops to the ME. However, due to the complexity and the missing analytical description, it is not feasible to perform a full calculation of the process like it is done for the ME. Therefore, in an approximation, only the dominant contributions from collinear or low-energy (soft) parton splitting are considered. Technically, this is done by simulating a step-wise Markov chain, based on the cross-section to obtain one additional parton (σ_{n+1}), for n existing partons. The cross-section can be written as splitting of a parton j from type i into two partons $j + k$, where k carries a certain energy fraction z and is radiated with an opening angle (θ), an azimuthal angle(ϕ) corresponding to a splitting function $P_{ij}(z, \phi)$ [73]:

$$d\sigma_{n+1} \approx d\sigma_n \frac{\alpha_s}{2\pi} \frac{d\theta^2}{\theta^2} dz P_{ij}(z, \phi) d\phi. \tag{4.1}$$

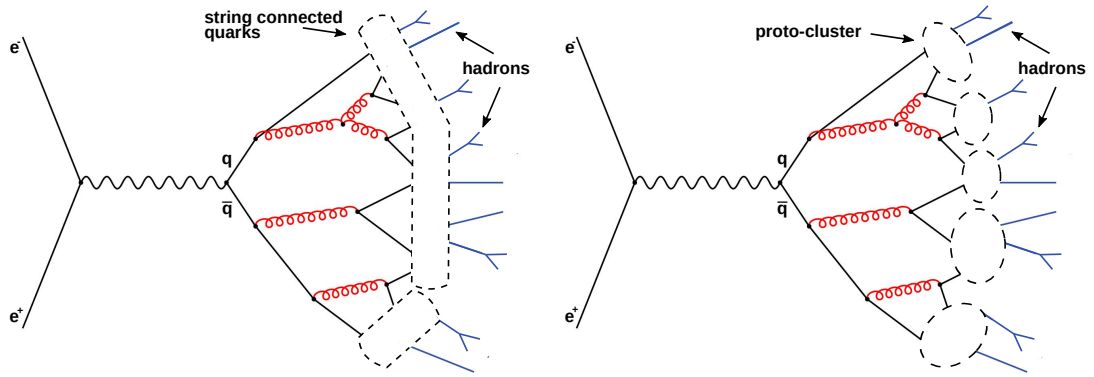
The preferred choice of the HERWIG [75,76] shower generator is to approach Equation 4.1 by evolving in the opening angle θ. This treats the coherent radiation of soft gluons correctly and is therefore referred to as *coherent showering* or *angular-ordered showering*. In this approach, $q\bar{q}$-pairs are considered as a colour singlet which only radiates soft gluons inside a cone of a certain angle. Outside this cone, no significant soft gluon emission occurs and the emission is generated as direct (hard) gluon emission from the parent parton.
An alternative, used by PYTHIA 8 [77] and SHERPA [78] but also available in newer versions of HERWIG [79], is *dipole showering*. Here, Equation 4.1 is evolved w.r.t. the transverse momentum p_T of the emitted particle relative to the mother particle. Assuming that each parton is connected uniquely to another parton by its colour, each parton-pair is considered as an independent dipole which emits further radiation. The radiation of a gluon corresponds to the splitting of one of these dipoles into two with given kinematics. Since emissions with large p_T are generated first, including the recoil to the system, this approach is also referred to as *transverse-momentum-ordered dipole shower*. Despite technical advantages in implementing ME corrections, momentum conservation can be satisfied at all stages of the shower generation while keeping the partons on mass-shell since the process corresponds to a $2 \to 3$ rather than a $1 \to 2$ splitting. After this step, non-perturbative effects become important as the transferred momenta become small ($Q^2 \approx 1\mathrm{GeV}^2$) and the distance scale rises.

The transition to *particle level*, known as *hadronisation*, is the most important non-perturbative effect. It describes the process in which colour-neutral, often unstable hadrons are formed from the coloured partons. This process is required by the *confinement* of the strong interaction which, again, is an unsolved theoretical problem and perturbation theory cannot be applied any more. Therefore, effective models, such as the *string* and the *cluster* [74] model, are used.
The *string model*, sketched in Figure 4.1 (a), is the hadronisation model of the PYTHIA

shower generator. It is most easily described for the $e^{\mp}e^{\pm} \to q\bar{q}$ process. The two final state quarks move in opposite direction and lose energy in the colour field. The colour connection of the quarks is described by so-called strings which, effectively, correspond to gluons. The strings ensure that the colour charge is conserved in each of the subsequent steps. A string has a uniform energy density per unit length which increases, due to the increase of the strong coupling with distance and the self-interaction, as the quarks move apart. A string may be broken and, if the potential energy of the system is high enough, a new $q\bar{q}$-pair is produced afterwards. Alternatively, the string reconnects to another quark or splits into two other strings (self-interaction) which impacts the final angular distribution. This process repeats until the kinetic energy of the initial quarks is transformed into $q\bar{q}$-pairs which are only connected by short strings without sufficient energy to produce further partons.
The *cluster model*, sketched in Figure 4.1 (b), is the hadronisation model of the SHERPA and HERWIG shower generators. The partons are directly grouped into colourless clusters and make use of the preconfinement principle. It implies that the mass distribution of two colour-connected neighbouring partons falls rapidly towards high masses and is asymptotically independent of the momentum scale Q^2. For the same process as discussed previously, this implies that two neighbouring quarks form a so-called *proto-cluster*. By introducing a cut-off energy scale, typically around $Q^2 = 1\,\mathrm{GeV}^2$, the mass of the cluster is regulated to be in the order of a few GeV which can be interpreted as superposition of different mesons. The proto-clusters finally decay into hadrons which ends the hadronisation.
In both models, the final hadrons, shown on the right in Figures 4.1 (a) and (b), decay in long-lived or stable lighter objects, such as $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, K^0_{L}, γ, p and neutrons. The exact decays and branching ratios for each hadron are simulated by sophisticated models based on tables which contain the latest experimental results [26].

In the final step, the interaction of these particles with the detector material and the detector response are simulated. GEANT 4 [80] is the most accurate method used in ATLAS but the method is computing intensive. Therefore, the time needed to produce a certain number of events increases which cannot always be afforded by an analysis. Alternatively, ATLFAST-II [81, 82] can be used which parametrises the response of the calorimeters to the showers and is therefore faster but, depending on the phase-space, not always as accurate and additional uncertainties have to be considered. As a result of the simulation, MC events and data events have exactly the same format and can be processed further to identify objects as discussed in the next chapter.

Analysis Specific Choices Several Monte Carlo and Shower Generators exist and different assumptions or parameter settings in the exact implementation of a given process are made as discussed above. The choices made for this analysis are summarised in Table 4.1, more details are given in the text below. To estimate the uncertainty of the choice, for $t\bar{t}t\bar{t}$ and $t\bar{t}$+jets, *alternative samples* by a different generator are compared to the nominal setup, further discussed in Section 5.5. As the respective cross-sections



(a) String model (b) Cluster model

Figure 4.1.: Schematic view of the effective models for hadronisation for the $e^{\mp}e^{\pm} \to q\bar{q}$ process. In the string model (a), the final $q\bar{q}$ pairs are connected via strings while in the cluster model (b) closest neighbours are grouped into proto-clusters [74].

for other samples are small compared to $t\bar{t}$, the effect of an uncertainty derived by an alternative samples on the result is expected to be negligible as further discussed in Chapter 5.

The $t\bar{t}t\bar{t}$ matrix element is modelled twice, once at NLO and once at LO in QCD, both with MadGraph5_aMC@NLO 2.6.2 [83] (NNPDF3.1NLO [84]). The LO samples are used in the training of the BDT as the NLO samples contain a large number of negative weights which cannot be treated properly in the training of the BDT. The factorisation and renormalisation scales are set to 0.25 times $H_{\mathrm{T}}^{\mathrm{all}}$ which is the sum of transverse momenta of all objects. Pythia 8.230 [77] (NNPDF23LO, A14-tune [85]) is used for generating showers. For hadronisation, MadSpin [86, 87] takes the spin correlations in top-quark decays into account and heavy-flavour hadron decays are modelled by Evtgen 1.6.0 [88].
For systematic uncertainties on the shower, the Herwig 7.04 [75,76] (MMHT2014LO [89], H7UE-tune [76]) was used as replacement for Pythia 8.230. All samples are normalised to NLO in QCD+QED due to corrections from the large tH coupling [38]. The detector response is modelled with AtlFast-II.

The $t\bar{t}$ process is modelled at NLO in QCD with Powheg-Box v2 (NNPDF3.0NLO, $h_{\mathrm{damp}} = 1.5 \times m_{\mathrm{top}}$), where h_{damp} is a parameter that regulates the p_{T} of the first additional jet beyond LO which is needed to cancel a divergence. Uncertaintes on h_{damp} are estimated by varying the parameter to $h_{\mathrm{damp}} = 3.0 \times m_{\mathrm{top}}$. Pythia 8.230 (NNPDF23LO [84], A14-tune) is used for generating showers. For systematic uncertainties on the shower, the Herwig 7.04 (MMHT2014LO, H7UE-tune) was used as replacement for Pythia 8.230. For systematic uncertainties on the matrix-element,

Table 4.1.: Basic MC generator settings for signal and background samples.

Sample	Generator	ME PDFs	Shower
$t\bar{t}t\bar{t}$	MadGraph5_aMC@NLO	NNPDF3.1	Pythia 8
$t\bar{t}$	Powheg-Box v2 +Evtgen	NNPDF3.0	Pythia 8
Single-top (tW)	Powheg-Box v2 +Evtgen	NNPDF3.0	Pythia 8
Single-top (t-chan.)	Powheg-Box v2 +Evtgen	NNPDF3.0	Pythia 8
Single-top (s-chan.)	Powheg-Box v2 +Evtgen	NNPDF3.0	Pythia 8
$t\bar{t}$ +H	Powheg-Box v2 +Evtgen	NNPDF3.0	Pythia 8
$t\bar{t}$ +Z	MadGraph5_aMC@NLO	NNPDF3.0	Pythia 8
$t\bar{t}$ +X	MadGraph5_aMC@NLO	NNPDF3.0	Pythia 8
$t\bar{t}$ +W	Sherpa	NNPDF3.0	Sherpa
Diboson	Sherpa	NNPDF3.0	Sherpa
V+jets	Sherpa	NNPDF3.0	Sherpa

the MadGraph5_aMC@NLO 2.6.0 (NNPDF23NLO [84]) was used as replacement for Powheg-Box v2.
To obtain more events in relevant regions and thereby reduce statistical uncertainties, *heavy-flavour*- (HF) and H_T-filtered samples have been generated. HF-filtered samples are classified at particle level according to the number of heavy-flavour hadrons, H_T-filtered samples cover different regions in H_T. All samples are normalised to NNLO in QCD cross-section. Soft gluon emission at NNLL [90–93] accuracy is calculated by the Top++2.0 [94] package. The detector response is modelled with AtlFast-II.

The single-top *tW* [95] **and *s-channel*** [96, 97] processes are modelled at NLO in QCD with Powheg-Box v2 (NNPDF3.0NLO) with the five-flavour scheme where the b-quark is part of the PDF. Interference terms with $t\bar{t}$ are removed via diagram removal.

The single-top *t-channel* [98] process is modelled at NLO in QCD with Powheg-Box v2 (NNPDF3.0NLOnf4) with the four-flavour scheme where the b- and the c-quarks are part of the PDF.
Pythia 8.230 (NNPDF23LO, A14) is used for generating showers of all single-top samples. All samples are normalised to NNLO cross-section [99, 100].

The $t\bar{t}$ +Z process is modelled at NLO in QCD with MadGraph5_aMC@NLO 2.3.3 (NNPDF3.0NLO). Pythia 8.210 (NNPDF23LO, A14-tune) is used for generating showers. A NNLO QCD+QED cross-section normalisation is applied [101].

The $t\bar{t}$ +W process is modelled at NLO in QCD with Sherpa 2.2.1 (NNPDF3.0NLO) [78]. NNLO QCD+QED cross-section normalisation is applied [101].

The $t\bar{t}+H$ process is modelled at NLO in QCD with POWHEG-BOX v2 (NNPDF3.0 NLO). PYTHIA 8.230 (NNPDF23LO, A14-tune) is used for generating showers.

Other $t\bar{t}+X$ processes (tZ, tWW, ttt) are modelled at LO in QCD with MADGRAPH5_aMC@NLO 2.6.2 (NNPDF3.0NLO). PYTHIA 8.210/8.230 (NNPDF23LO, A14-tune) is used for generating showers for $t\bar{t}$ +X samples.

Diboson and V+jets processes are modelled with SHERPA 2.2.1/2 (NNPDF3.0NLO) at NLO (LO) depending on the number of extra jets. SHERPA shower generation is interfaced by the MEPS@NLO prescription [102–105]. Tuned parameters were developed by the SHERPA authors. V+jets [106] (Diboson [107]) samples are normalised to NNLO (NLO) in QCD cross-section.

4.3. Object Definitions

As discussed previously, only long-lived or stable particles are observed in the detector. In general, it is known which signature can be expected in which part of the detector by a particle, as ATLAS is designed to provide a good particle identification. Figure 4.2 shows the different layers of the ATLAS detector and the interactions of different particles in each layer. For example, an electron results in a track in the inner detector which matches a shower in the electromagnetic calorimeter while a photon only has a shower in the calorimeter and no associated tracks. However, more information from the detector can be used for a better separation and resolution. These choices can significantly impact the result of the analysis and are summarised for the relevant quantities such as electrons, muons, jets, jets originating from b-quarks (b-jets), missing transverse energy and overlap removal in the following.

4.3.1. Electrons and Photons

The reconstruction of electrons and photons is similar and is therefore done in the same way via a discriminant. Due to the expected signature, only information from the ID and ECal are used [108].
The deposited energy ($E_{\mathrm{T}}^{\mathrm{cell}}$) in a cell of the ECal is compared to the expected noise-level ($E_{\mathrm{T}}^{\mathrm{noise,cell}}$) from which the significance $\epsilon^{\mathrm{cell}} = \frac{E_{\mathrm{T}}^{\mathrm{cell}}}{E_{\mathrm{T}}^{\mathrm{noise,cell}}}$ of a measured signal can be derived. Deposits with $\epsilon^{\mathrm{cell}} \geq 4$ are used as *seed cells* to build a *proto-cluster* by iteratively collecting neighbouring cells with $\epsilon^{\mathrm{cell}} \geq 2$. After a cell is collected, it becomes the seed for the next iteration. If a proto-cluster fulfils $E_{\mathrm{T}}^{\mathrm{cluster}} \geq 400\,\mathrm{MeV}$ and at least 50% of the energy were deposited in the ECal, it is matched with tracks from the ID [109–111]. Matched proto-clusters within 5×3 cells around the seed barycentre are merged to form electron *super-clusters*. Photons originating from Bremsstrahlung are either already merged in a proto-cluster or, if it created its own cluster, are merged in this step.

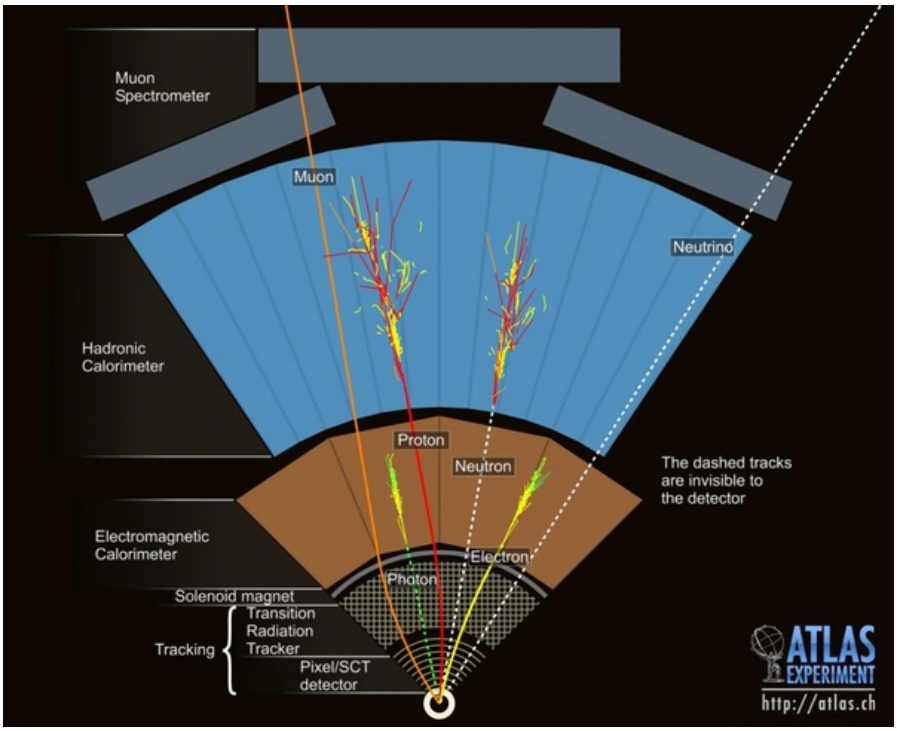


Figure 4.2.: Schematic view for different particles and the layers of the ATLAS detector they interact with. The information is used to identify long-lived particles.

Electrons produced within jets via decays have to be distinguished from final-state electrons. Based on the information of the primary track, the compatibility with the cluster and the shape of the clusters, a discriminant is formed.
In addition, requirements on the kinematics and the isolation of the track are made to define working points. The *tight* working point, used in this analyses, has the strongest cut on the discriminant, it requires $E/p < 10$ and one track with $p_{\mathrm{T}} > 2\,\mathrm{GeV}$.
The isolation of electrons is obtained by summing the energy in ECal cells ($E_{\mathrm{T}}^{\mathrm{cone20}}$) and the p_{T} of tracks ($p_{\mathrm{T}}^{\mathrm{varcone20}}$) in a radius of $\Delta R < 0.2$ around the track while excluding the electron. The *FCtight* working point requires $E_{\mathrm{T}}^{\mathrm{cone20}}/p_{\mathrm{T}} < 0.06$ and $p_{\mathrm{T}}^{\mathrm{varcone20}}/p_{\mathrm{T}} < 0.06$ as well as a loose matching of the track with the primary vertex to suppress pile-up effects. The tight working points are chosen as they provide the best background rejection and a good robustness against pile-up. Additional scale-factors for calibration of MC to data are derived in tag-and-probe [108, 112] analyses.
In addition to these working points, electrons in this analysis are required to have $p_{\mathrm{T}} > 28\,\mathrm{GeV}$ and $|\eta| < 1.37$ or $1.57 < |\eta| < 2.47$.

4.3.2. Muons

No energy reconstruction for muons is possible as they are not stopped in the calorimeters but the momentum can be measured from the track curvature. The reconstruction is track-based and combines tracking by the MS and the ID. Before combining the information, tracks are reconstructed in both sub-detectors individually [113]. In this analysis, only muons that have a matching track between ID and MS, so-called *combined muons*, are considered. Other types, where only a track from the MS is available that loosely matches the primary vertex (extrapolated), where the calorimeter is used for tracking (calorimeter-tagged) or where the track is not fully reconstructed in the MS (segmented-tagged) are not considered as the background contribution, for example from hadron decays, is too large.
Based on the χ^2 agreement between fitted track and measured points and the agreement with the ID track, working points are defined. In this analysis the *medium* working point and an isolation requirement of $\mathrm{p}_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.06$ is used to guarantee a high efficiency ($> 98\%$) and a strong rejection ($> 80\%$) of background events.
The medium working point is the default in ATLAS as it minimises the systematic uncertainties associated with muon reconstruction and calibration [113]. Depending on η, it requires a minimum number of hits in the MDTs for a solid track reconstruction in addition to a loose agreement with the ID track. As for electrons, the primary vertex for muon candidates must be loosely matched to suppress pile-up. Additional scale-factors, derived in tag-and-probe analyses [113] are applied to simulation to reduce systematic uncertainties and to improve the MC modelling. In addition to these criteria, muons in this analysis are required to have $p_{\mathrm{T}} > 28\,\mathrm{GeV}$ and $|\eta| < 2.5$.

4.3.3. Jets and RC-Jets

In general, a jet is defined as a cluster of colour-neutral objects which matches the momentum of a parton from the matrix-element decay. In practice, an algorithm finds the energy deposits in the calorimeters which were created by the same object. Similar to the procedure in Section 4.3.1, topological clusters in the calorimeter are formed by a nearest neighbour algorithm [114,115]. Next, the FastJet 3.2.2 [116] software package, which makes use of the *anti-kt* algorithm [117] with a radius of $R = 0.4$, merges the topological clusters to one object.

In the anti-kt algorithm, for an entity i, for example a cluster, with a transverse momentum k_{ti} and a geometrical distance $\Delta_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ to another object j, the two quantities

$$d_{ij} = \min\left(k_{ti}^{-2}, k_{tj}^{-2}\right) \frac{\Delta_{ij}}{R^2} \tag{4.2}$$

$$d_{iB} = k_{ti}^{-2}, \tag{4.3}$$

where d_{iB} is the distance to the beam, are compared. The smallest d_{ij} for all i,j combinations is taken. The two objects are merged and the iteration continues, otherwise, if d_{iB} is the smallest distance for all objects, these objects are considered jets. The reason this algorithm is used by many analyses in Atlas is the circular shape of the final objects which makes it easier to remove overlaps between jets and calibrations.
Since the properties, in particular the *jet energy scale* (JES), of a jet depends on the algorithm and detector effects, many calibrations are required. First, the four-momenta of jets are calibrated to the energy of particle-level objects using MC events. This takes several effects, such as energy loss in dead material and energy deposits outside of the idealized cone into account. Next, the global sequential calibration takes global jet observables into account to reduce the dependence on the flavour and the type of object (quark or gluon) [118]. Finally, calibration to data is performed to reduce the dependence on η due to detector inhomogeneities. Typically, events with well calibrated objects, such as Z bosons, with additional jets are selected and the recoil of the additional jet is measured. From this measurement, scale factors are derived which are applied to MC to match data. In a similar fashion, the *jet energy resolution* (JER) is derived. Jets in this analysis are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$.
Due to pile-up, additional energy deposits can be merged during the jet reconstruction. For Run II, a *jet-vertex-tagger* (JVT) [119] was developed which matches the tracks of the energy deposits with the primary vertex and provides a likelihood discriminant to reduce pile-up. In the future, *particle-flow* [120] jets will be used which includes tracking information directly in the reconstruction of the jets.

Reclustered Jets (RC-jets) [121] are formed after the regular (small-R) jets are calibrated. In this analysis, they are motivated by the fact that quark-decay products, in particular the ones from the W boson, can have a high p_{T} and are therefore boosted. The resulting jets can merge to one larger jet with a substructure. Due to the overlap

removal, the kinematics would not be correctly reconstructed if two small-R jets were formed. The anti-kt algorithm with $R = 1.0$ is applied on fully calibrated jets and inherits uncertainties and calibration from the small-R jets [122]. Furthemore, small-R jets with a $p_T < 5\%$ of the total p_T are removed from the RC-jet to suppress pile-up and soft gluon radiation [123]. Finally, $p_T > 200$ GeV and $|\eta| < 2.0$ are required in this analysis to identify RC-jets.

4.3.4. Jets Originating from bottom quarks

While the reconstruction of jets originating from bottom quarks (b-jets) is similar to the one for other jets, dedicated tagging algorithms exist to identify this type of jet in the detector. For top-physics and in particular for this analysis, where at least four b-jets are expected, this is important to define signal enriched and background enriched regions.
B-hadrons have a relatively long mean lifetime of 1.5 ps [26] allowing them to travel up to millimetres before they decay. This distance is resolvable by the tracking algorithms and a second displaced vertex can be identified. B-hadron decays contain an electron or muon in $\approx 20\,\%$ of decays. As they are relatively massive compared to e and μ, the decay products have a high p_T with respect to the b-quark direction. In addition, the high mass often results in a chain of decays which is different from the decays of lighter jets.
The *MV2c10* [124, 125] algorithm is based on the *Toolkit for Multi Variate Analysis (TMVA)* [126] in ROOT and combines the outputs of three lower level algorithms (IP2D/IP3D [124], SV1 [127], JetFitter [128]), which make use of the b-jet signature discussed above, in a *Boosted Decision Tree* (BDT). As can be seen from Figure 4.3, a good separation between light- and b-jets is achieved while the c-jet distribution tends to be more flat. From this figure, working points can be derived based on cuts on the MV2c10 output. Only events with a score higher than a cut value are considered for a given working point. The values for the cut are defined by the fraction of b-jets in $t\bar{t}$ events which remain after the cut (60%, 70%, 77%, 85%). A lower fraction corresponds to a better background rejection but also fewer signal events which is why it is not feasible to always use the tightest working point. For *pseudo-continuous* (PC) b-tagging, used in this analyses, bins between the cut values are formed as summarised in Table 4.2. Each bin contains different information about how b-jet-like a certain jet is.

Related to the definition of b-jets is the additional classification of $t\bar{t}$ with additional heavy flavour (HF) jets. In general, jets are classified according to the response of the b-tagger. However, for MC samples it is possible to apply more detailed matching criteria based on the available *truth level* information. Therefore, in this analysis, a matching criteria of $\Delta R < 0.4$ between truth level objects, for example hadrons before they decay, and jets is applied. Based on the number of successful matches, different sub-categories are defined to separate the effects of different radiation sources, for example the radiation of b-jets via additional gluons:

- $t\bar{t}+b$: Exactly one b-hadron from truth level is matched with exactly one particle-

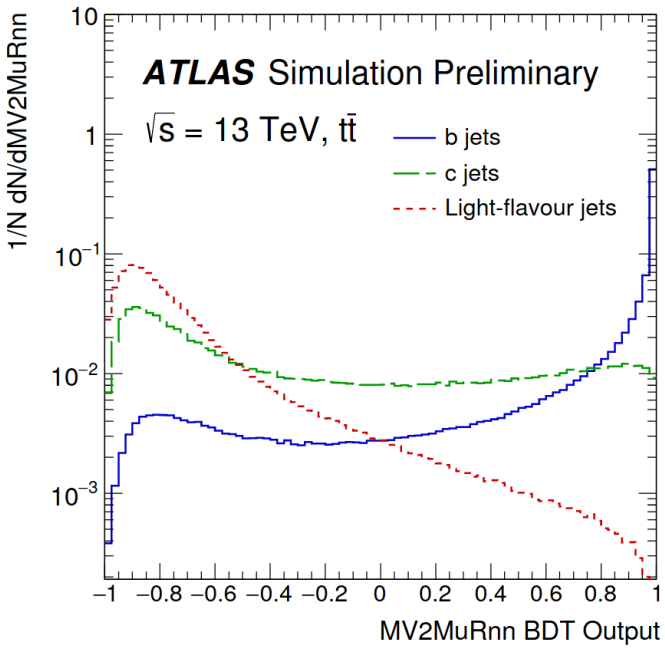


Figure 4.3.: Output of the MV2c10 BDT for light- c- and b-jets. The MV2c10 score can be used to derive working points based on the amount of truth b-jets in $t\bar{t}$ that pass the selection [125].

Table 4.2.: Cuts on the MV2c10 output to define working points and bins for the pseudo-continuous (PC) b-tagging. Working points are defined based on the fraction of truth b-jets that pass the selection. In the regular case, the cut-value corresponds to a lower value. For PC b-tagging, bins between the cuts, where 1.00 is the upper limit, are formed and an integer values is assigned.

Working Point	Cut value	PC value
60%	≥0.92	5
70%	≥0.79	4
77%	≥0.58	3
85%	≥0.05	2
100%	≥0.00	1

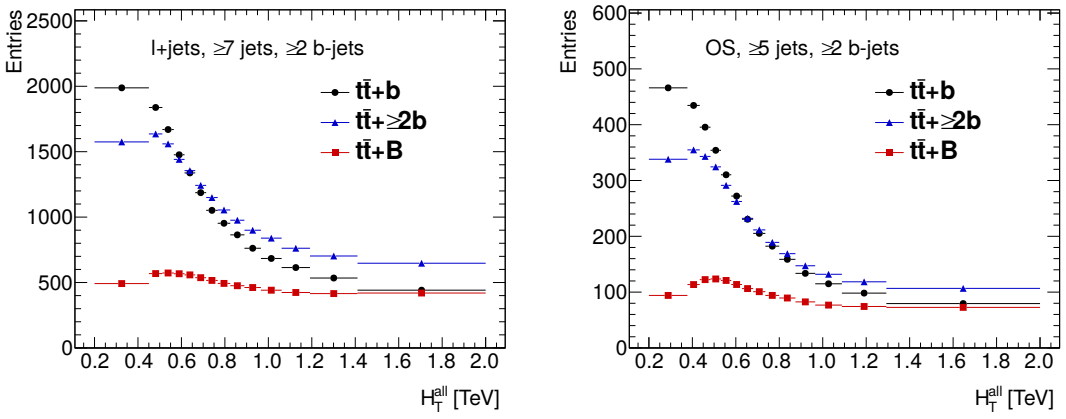


Figure 4.4.: Distributions of $t\bar{t} + b$ sub-categories defined in the text.

level jet.

- $t\bar{t} + B$: Exactly two b-hadrons are matched to exactly one particle-level jet.

- $t\bar{t}+ \geq 2b$: More or equal than two b-hadrons are matched to the same number of particle-jets.

In a similar way, the non-overlapping sub-categories are defined for c-jets. The categories are used for all systematic uncertainties where a difference between the categories is expected as further discussed in Section 5.5. By introducing the additional free parameters for each sub-category, the final fit is sensitive to shapes and normalisation of the different $t\bar{t}$+jets processes without being able to directly measure the processes in the detector. As example, Figure 4.4 shows the different kinematic distributions of $H_{\mathrm{T}}^{\mathrm{all}}$ in the given b-sub-categories.

4.3.5. Missing E_T

In e^+e^--collisions the longitudinal momentum is exactly known since $e^{\pm}$ are elementary particles. In pp-collisons, the momentum fraction of a parton can only be estimated via PDFs as discussed in Section 2.1. As a consequence, momentum and energy conservation in the longitudinal direction cannot be used to reconstruct the energy of particles that do not, or only very weakly, interact with the detector. In the SM, only neutrinos escape the detector undetected. As the direction of the invisible particles cannot be determined, only the sum of the transverse momenta, the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) [129], can be calculated. In the transverse plane, the initial transverse energy (E_{T}) of the partons can bed assumed to be zero, otherwise energy/momentum conservation would

be violated. The sum of transverse momenta has to be conserved which implies that

$$\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}} = \left(E_{\mathrm{x}}^{\mathrm{miss}}, E_{\mathrm{y}}^{\mathrm{miss}}\right) = -\sum_{k} \boldsymbol{p}_{\mathrm{T}}^{k} \tag{4.4}$$

$$E_{\mathrm{T}}^{\mathrm{miss}} = |\boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}|, \tag{4.5}$$

where k can be electrons, muons, jets, hadronic taus [130] and soft components. From Equation 4.4, it becomes clear that all other objects have to be reconstructed and calibrated first. Then, soft components, well-identified tracks not associated to any other object which originate from the underlying event or pile-up, can be measured and $E_{\mathrm{T}}^{\mathrm{miss}}$ be determined. In addition, any overlap between the object has to be removed as discussed in the next Section. The azimuthal angle $\phi^{\mathrm{miss}} = \tan^{-1}\left(\frac{E_y^{\mathrm{miss}}}{E_x^{\mathrm{miss}}}\right)$ is the only directional information that can be derived for the missing energy. The calibration of the detector response to the soft-component makes use of processes like $Z \rightarrow \mu^+\mu^-$, where $E_{\mathrm{T}}^{\mathrm{miss}} = 0$ is expected, and $W \rightarrow e/\mu\nu$ where the neutrino final state is well-defined and $E_{\mathrm{T}}^{\mathrm{miss}}$ can assumed to be the p_{T} of the neutrino.

4.3.6. Overlap Removal

Overlap removal aims to avoid double counting of detector signals. For example, an energy deposit in the calorimeters should only contribute to one jet and electrons should not have shared tracks with other electrons or muons. Following an ATLAS recommendation, an algorithm is used which follows the BOOSTEDSLIDINGDRMU option [131]. It allows to avoid low p_{T} muons from decay chains. Sequentially, it removes

1. lower p_{T} electrons within $\Delta R \leq 0.2$ of another electron
2. calorimeter-tagged muons that overlap with the ID track of an electron
3. electron tracks that overlap with a muon track
4. the closest jet within a cone of $\Delta R \leq 0.2$ from an electron
5. electrons within $\Delta R \leq 0.4$ from a jet
6. jets within $\Delta R \leq 0.2$ from a muon if the jet has less than three tracks or if the p_{T} is compatible with the muon
7. muons within momentum dependent distance of $\Delta R \leq \min\left(0.4, 0.04 + 10\,\mathrm{GeV}\,/p_{\mathrm{T}}\right)$ of a jet.

CHAPTER 5

Analysis Strategy

In this Chapter, the strategy to measure the production of four top quarks ($t\bar{t}t\bar{t}$), introduced in Section 2.1.2, with the full Run II dataset of 139 fb^{-1} is presented.
Only reconstructed events that fulfil certain detector and physics requirements are of interest for this analysis. Hence, a selection, as presented in Section 5.1, on all events is applied before continuing with details about the analysis.
Even though 60% of $t\bar{t}t\bar{t}$ events decay into one or two charged leptons with opposite charge, the 1L+OS channel, which is a focus in this document, is not expected to be the most sensitive [49] part of the combined result due to a large background arising from production of $t\bar{t}$ pairs with additional radiated jets ($t\bar{t}$+jets). While the branching ratio for events with two leptons of the same charge or more than three leptons is only about 13%, the background contributions are small and a higher significance is expected [48]. As the backgrounds in both channels are very different, each channel is studied independently. To make use of the high statistics and the good sensitivity, the analysis combines both channels in the end. The main background of the all-hadronic channel is multi-jet production which is considered too difficult to separate from the signal in this document. Hence, this decay channel is not considered in this analysis.
The quantity to be measured by the *Profile Likelihood fit* on the output of a *Boostes Decision Tree*, while taking into account all uncertainties, is the cross-section of the $t\bar{t}t\bar{t}$ production. It is extracted as the signal strength $\mu_{t\bar{t}t\bar{t}}$, the ratio of the cross-section to theory prediction. This quantity allows to easily access the consistency of the measured $\mu_{t\bar{t}t\bar{t}}$ with the SM prediction of $\mu_{t\bar{t}t\bar{t}} = 1$. Compared to the background processes, more jets and in particular more b-jets are expected. Therefore, the fitted regions, further discussed in Section 5.2, are defined based on (b-)jet multiplicities. Lower multiplicities are used as control regions (CR) to constraint the $t\bar{t}$+jets background. Higher multiplicities are used as signal regions (SR) to extract $\mu_{t\bar{t}t\bar{t}}$ and to constraint signal uncertainties.

Besides the information from b-tagging, signal and background differ in many kinematic distributions which are combined by a *Boosted Decision Tree* (BDT). Due to the high energy threshold of four times the top-mass, the initial partons have, on average, a similar momentum. With the consequently small boost, the decay products are more central in the detector and the transverse kinematics of the decay products are expected to be harder, even though $t\bar{t}t\bar{t}$ events are less energetic due to the mass threshold. The corresponding distribution of particular interest is the sum over the transverse momenta $H_{\mathrm{T}}^{\mathrm{all}} = \sum_k p_{\mathrm{T}}$ where k can be missing transverse energy, leptons (electrons and muons) and jets. In addition, the angular distributions for (b-)jets differ between signal and backgrounds. For examples, two b-quarks from gluon splitting are expected to be more aligned while the b-quarks from the top decays are to first order produced independently.

For the 1L+OS channel, the modelling of the dominant $t\bar{t}$+jets background is crucial. In the phase-space of this analysis, the modelling relies on the shower generator. At least two b-jets on top of the matrix-element calculation are required to reach the (b-)jet multiplicities of the $t\bar{t}t\bar{t}$ signal. However, the agreement with data in CR is observed not to be reliable. For this reason, a reweighting method, discussed in Section 5.3, was developed which corrects the Monte-Carlo prediction and is therefore referred to as *MC-based reweighting method* or *MC-based method*. The *$t\bar{t}$ Tag Rate Function* (ttTRF) method, which was used in the previous publication [49], is a data-driven approach and is used for cross-checks as it provides an independent background prediction. Both methods are applied prior to the training of the BDT, presented in Section 5.4, to provide a reliable input for the $t\bar{t}$ background.
The systematic uncertainties for the 1L+OS channel as well as more details of the fit setup are summarised in Section 5.5 and 5.6. Finally, the setup of the SS+ML channel is briefly summarised in Section 5.7.

5.1. Event Selection

The *pre-selection* is a combination of technical and trigger requirements. In addition, during the *event selection*, additional cuts on objects are applied to obtain events that are enriched in signal. Both selections are applied to data and MC to obtain a consistent comparison.
Triggers are adjusted to keep a balance between efficiency and rate while keeping the bandwidth from saturating. Typically, they combine information from multiple sub-detectors to select events with a signature of interest, for example one or two lepton events. The exact conditions can change between the years as data-taking conditions differ. In the 1L+OS channel, single-lepton triggers are applied as they perform well in both channels. The settings of this analysis, based on the available options [132] by the ATLAS experiment, are summarised in Table 5.1.
The additional requirements of the event selection, listed in Table 5.2, are applied based on the following considerations. Typically, low p_{T} objects tend to originate from additional gluon radiation or sub-leading decays into leptons. As the expected transverse momenta are expected to be softer than for the events of interest, cuts on the object p_{T} are applied. The requirement on the number of (b-)jets was derived to obtain sufficient $t\bar{t}$+jets events for a measurement of the background while cutting regions that do not benefit the analysis. As the final goal is a combination with the SS+ML channel, sub-leading leptons with less strict selection criteria on p_{T} are vetoed to suppress multi-lepton backgrounds and a requirement on the invariant mass of the two leptons in the OS channel is applied to avoid $Z \to l^+l^-$ events. The selection is orthogonal to the SS+ML channel.

Table 5.1.: Trigger menu for the HLT in the 1L+OS channel per year and channel. The menu reads as follows: <particle-type>_<p_{T}-cut [GeV]>_<identification-criteria>_<isolation-criteria>_<data-stream>. More details can be found in Ref. [132].

Type	2015	2016-2018
Single Electron	e24_lhmedium_L1EM20VH	e26_lhtight_nod0_ovarloose
	e60_lhmedium	e60_lhmedium_no0
	e120_lhloose	e14_lhloose_nod0
Single Muon	mu20_iLoose_L1MU15	mu26_ivarmedium
	mu50 mu50	

Table 5.2.: Pre-selection in the 1L+OS channel applied to data and MC. *Loose* leptons are obtained similarly to the *tight* leptons in Section 4.3.1, but with less strict criteria.

	Single lepton ch.	Opposite-sign dilepton ch.
Objects	$e^{\pm}$+jets, $\mu^{\pm}$+jets	$e^{\pm}e^{\mp}$, $\mu^{\pm}\mu^{\mp}$, $e^{\pm}\mu^{\mp}$
Lepton p_{T}	$\geq$ 28 GeV –	$\geq$ 28 GeV (leading) $\geq$ 10 GeV (sub-leading)
N_{jets}	$\geq$ 5	$\geq$ 2
$N_{\text{b-jets}}$	$\geq$ 2 (MV2c10 77%)	
Jet p_{T}	$\geq$ 25 GeV	
M_{ll}	– –	$\geq$ 15 GeV $\lvert M_{ll} - m_Z \rvert > 10$ GeV
Veto	Second loose lepton	Third loose lepton

5.2. Region Definition

Figure 5.1 sketches the main backgrounds and signal sorted in multiplicities of (b-)jets. For all background processes, additional gluons, which cannot be excluded by kinematic cuts, are required to mimic the $t\bar{t}t\bar{t}$ signal. For $t\bar{t}$ decays, at least one additional $b\bar{b}$-pair and two additional $q\bar{q}$-pairs are required in addition to the decays of the top quarks in both channels. While the exact number of jets changes, this requirement is similar for all other backgrounds like $t\bar{t}+W/Z/H$. Most additional radiation of at least three additional b-quarks is required by the single-top quark or W boson production in the 1L, and single-top quark production in association with a W boson for the OS channel. Figure 5.2 compares the b-jet and jet multiplicities (N_b and N_j) for all backgrounds in both channels after the event selection. It is observed that the flavour composition of $t\bar{t}$+jets changes from mainly light-jets at low jet and b-jet multiplicities to almost only $t\bar{t}$ +b-jets at higher multiplicities. All other backgrounds remain small and therefore, for the region definition, it is mostly important to consider the separation between $t\bar{t}$+jets and $t\bar{t}t\bar{t}$.
In the full phase space, a ratio between signal and background events of $R = \frac{\sigma_{t\bar{t}t\bar{t}}}{\sigma_{t\bar{t}}} \approx \frac{12\,\text{fb}}{830\,\text{pb}} \approx \frac{1}{69000}$ is expected. This means that, based on the SM cross-section, for 1 $t\bar{t}t\bar{t}$ event, approximately 69000 $t\bar{t}$ events are produced. Comparing the $t\bar{t}t\bar{t}$ signal normalised to background (red-dashed line) in Figure 5.2, it can be seen that the relative separation increases for higher jet and b-jet multiplicities. Based on this observation, Figure 5.3 compares the ratio of yields for the $t\bar{t}$ background and the $t\bar{t}t\bar{t}$ signal processes for different selection cuts. For this reason, the (b-)jet multiplicities were chosen to define the

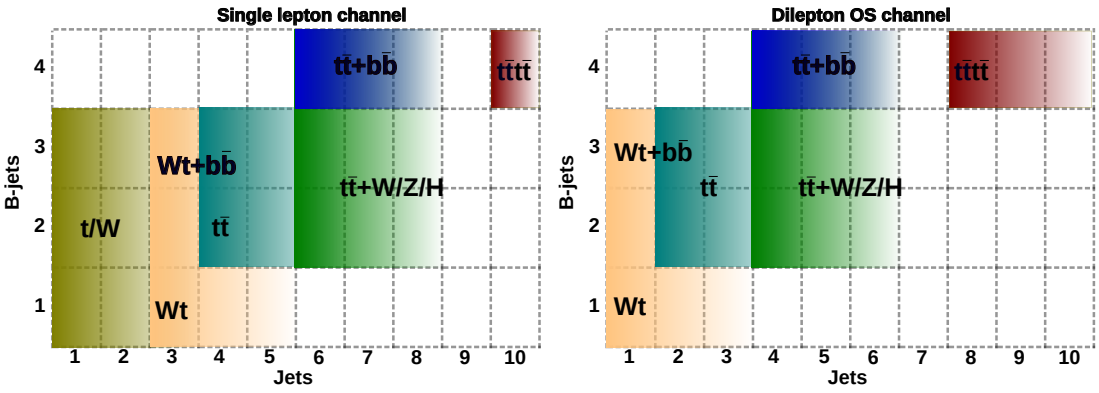


Figure 5.1.: Sketch of (b-)jet multiplicities for selected backgrounds and the $t\bar{t}t\bar{t}$ signal in the 1L (a) and OS (b) channel. In the Figure, up to one additional $g \to qq'$ is indicated by the coloured areas.

regions as shown in Figure 5.4. Lower multiplicities, corresponding to a smaller R, are used to measure the background (blue). This information is then used in higher multiplicities, corresponding to a larger value of R, to extract the signal cross-section (red). In addition, the regions with exactly two b-jets (grey) are used to derive additional MC Scale Factors (SF) to improve the agreement between data and MC in the fit regions as discussed in the next section.
The regions with exactly three b-jets are further divided to separate light- and c-flavoured jets from b-flavoured jets based on the fourth jet in the event. In both cases, exactly three b-jets are required to pass the 70% WP of the MV2c10 tagger. In the $3b^{\geq 4b@85\%}$ (3bH) regions, at least four jets are required to pass the looser 85% WP. The opposite requirement of less than four b-jets passing the 85% WP is applied in the $3b^{<4b@85\%}$ (3bL) regions. By this separation, the $3b^{<4b@85\%}$ regions are enriched in light- and c- flavoured jets while the $3b^{\geq 4b@85\%}$ regions are purer in b-jet events as shown in Figure 5.5 for the $H_{\mathrm{T}}^{\mathrm{all}}$ for inclusive ($\geq$ 75)) jet multiplicities. Consequently, different flavour types and the corresponding uncertainties can be accessed separately.

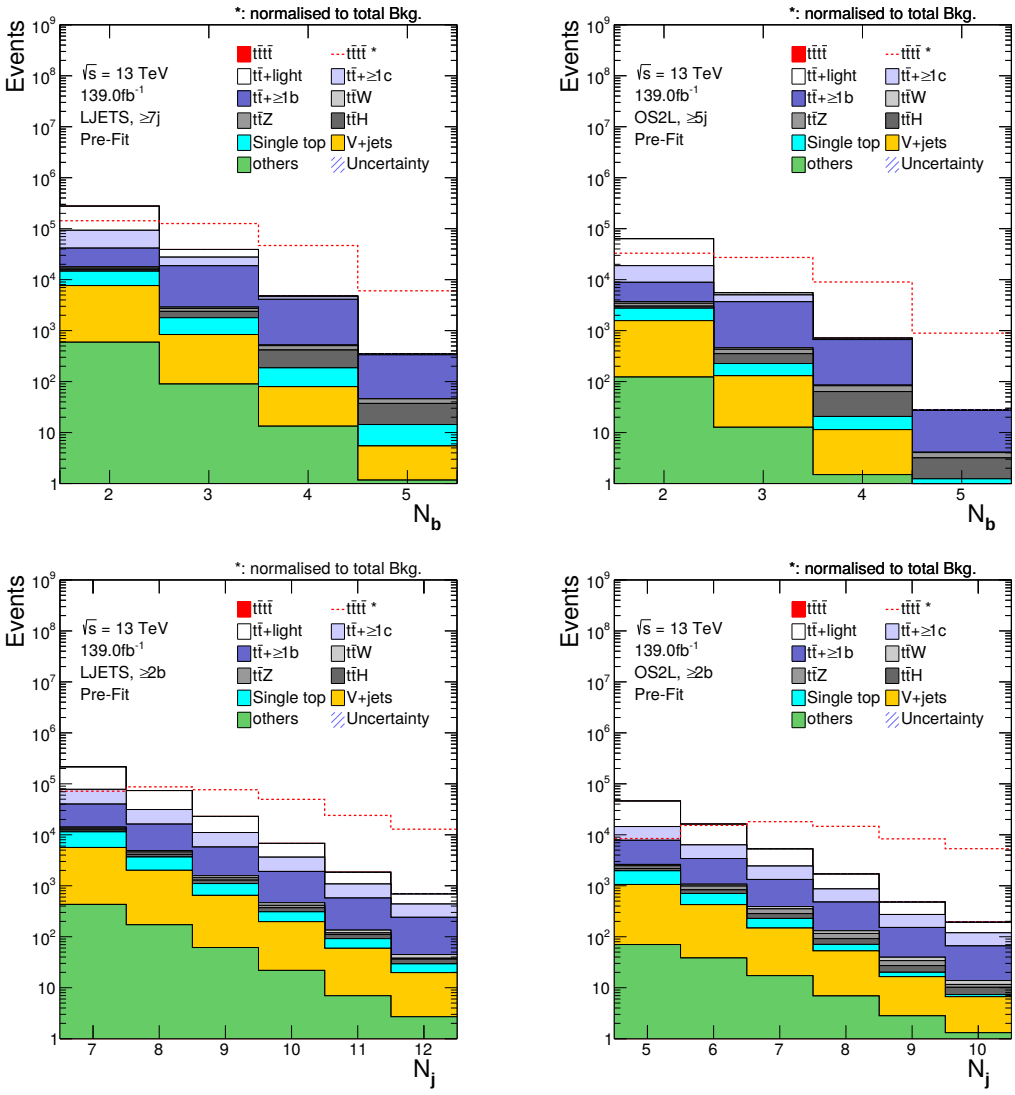


Figure 5.2.: MC predictions for jet and b-jet multiplicities with more than two b-jets or more than seven (five) jets in the 1L (OS) channel for all backgrounds.

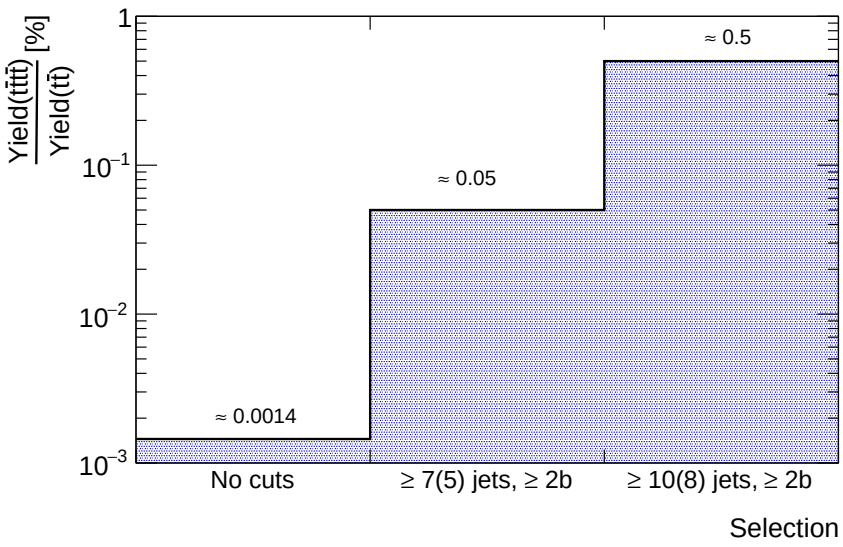


Figure 5.3.: Ratio of yields for $t\bar{t}t\bar{t}$ and $t\bar{t}$+jets when applying different (b)-jet selections.

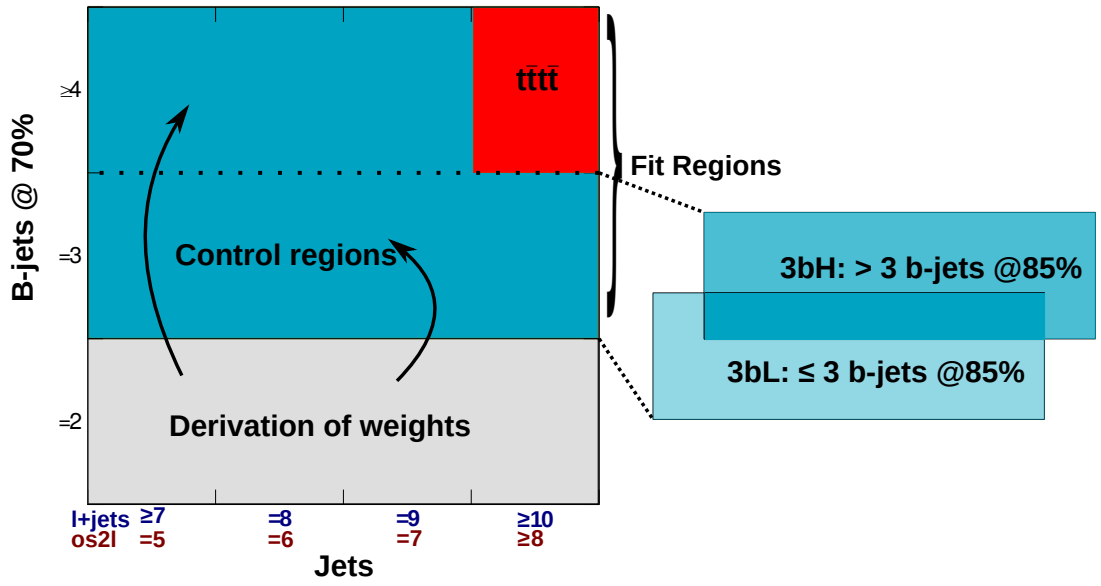
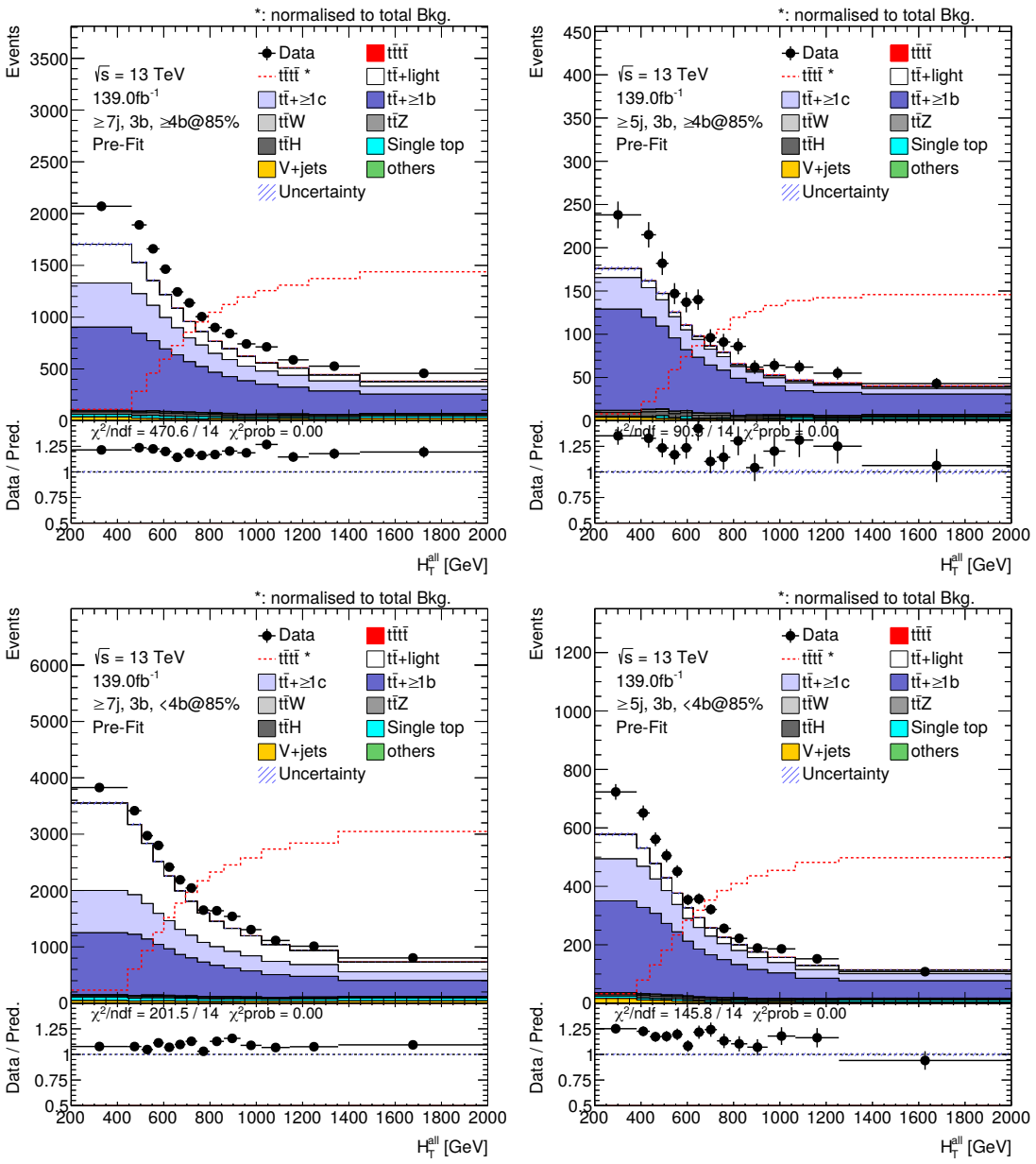


Figure 5.4.: Definition of background enriched (control, blue) and signal (red) regions. The regions for the derivation of the additional scale factors are shown as well.



(a) 1L (b) OS

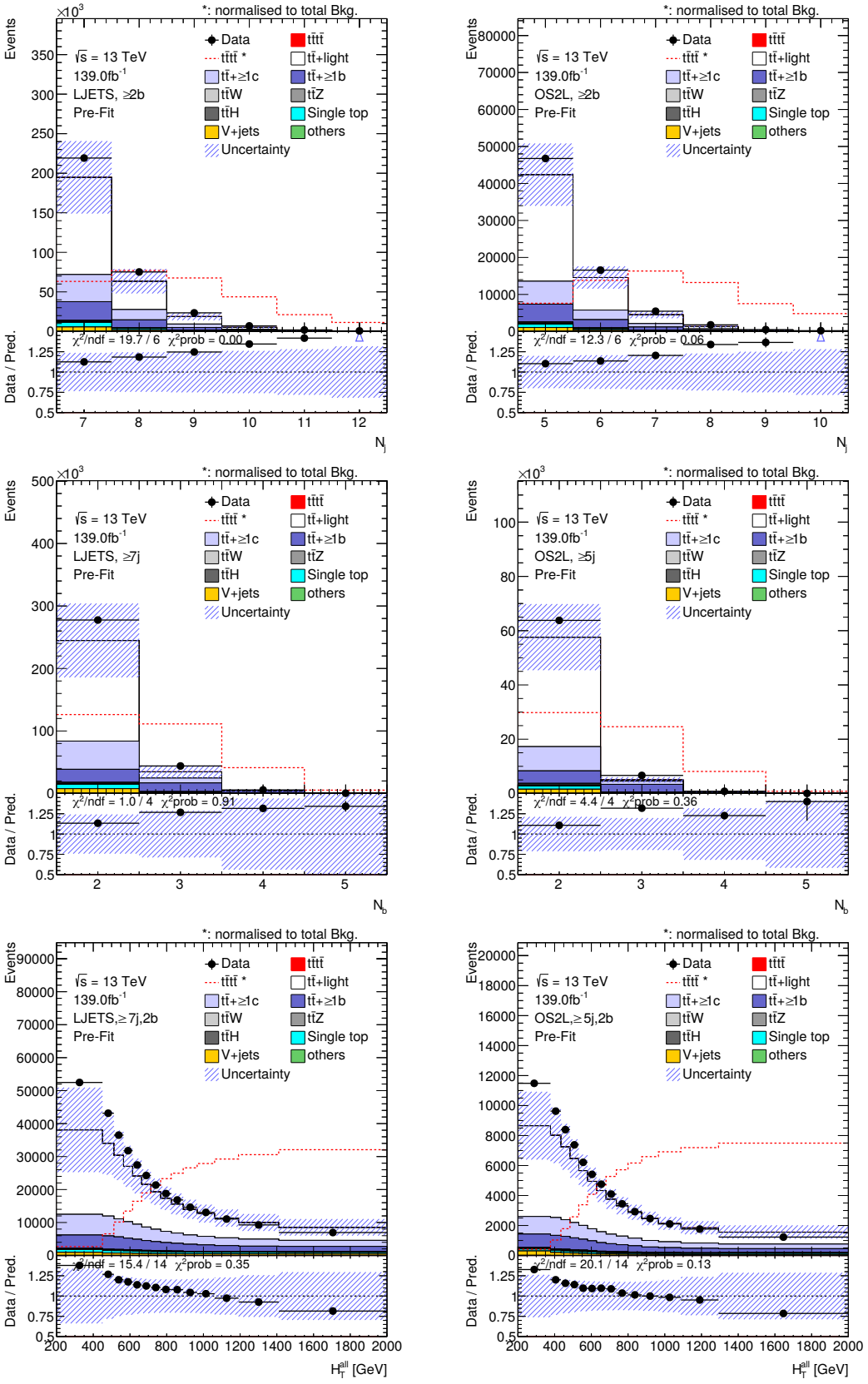
Figure 5.5.: Comparison of the $3b^{\geq 4b@85\%}$ (up) and the $3b^{<4b@85\%}$ (down) region in an inclusive jet multiplicity in the 1L (left) and OS (right) channels. The uncertainties show the statistical component only.

5.3. The Monte-Carlo Based Reweighting Method

A reliable prediction of the large $t\bar{t}$+jets background is needed to precisely measure the small $t\bar{t}t\bar{t}$ signal. As shown in Figure 5.6 for different (b-)jet multiplicities and $H_{\mathrm{T}}^{\mathrm{all}}$ before fitting, this is not the case when using the nominal MC prediction in the given (b)-jet multiplicities compared to data. A discrepancy in the normalisation in all jet regions of more than 20%, increasing with higher multiplicities, as well as a shape effect of 50% from the first to the last bin in the $H_{\mathrm{T}}^{\mathrm{all}}$ distribution is observed. The difference in normalisation shows that the overall rate is not modelled correctly while the kinematic difference shows that the jets from MC are modelled with too high transverse momenta. The effect is barely covered by the systematic and statistical uncertainties and can therefore lead to a biased fit with unphysical pulls and constraints. This behaviour was already observed in previous analyses, for example in the measurement of $t\bar{t}$ production in association with an Higgs boson [133] which operates in a similar phase-space. As discussed in Section 4.2, the shower generation is based on effective models. Furthermore, matrix-element calculations are only available to a certain order in perturbation theory. Therefore, additional jets are based on an approximation which, overall, works reasonably well if only few are considered. However, each additional jet is slightly mismodelled and the effect adds up for higher jet multiplicities and the overall discrepancy increases as observed. Therefore, in the extreme phase space of the $t\bar{t}t\bar{t}$ analysis the disagreement is no surprise.

The MC-based method derives correction factors for the $t\bar{t}$+jets background, referred to as *scale factors* (SF), in regions with two b-jets (2b) but otherwise similar to the fitted regions as already shown in Figure 5.4. The SFs can be considered as an additional experimental calibration of $t\bar{t}$ events which is applied to the nominal MC prediction and the systematic uncertainties. Normalisation differences due to different $t\bar{t}$+jets flavour compositions in the fit regions are determined by including normalisation uncertainties in the fit.
Based on the observed disagreement, the MC-method makes the following assumptions which are further discussed in Section 5.3.1:

1. The disagreement between data and MC in the fit and derivation regions is not caused by an unknown physics process that interferes with the $t\bar{t}t\bar{t}$ signal.
2. The SM is the valid theory which is not properly described, in particular by the shower generator. Therefore, the result can still be interpreted in the context of the SM.
3. The absolute disagreement in other backgrounds is small compared to the disagreement originating from $t\bar{t}$ yields.
4. The mismodelling does not depend on the flavour of the additional jet, except for normalisation effects, and hence not on the b-jet multiplicity. This corresponds to the exchange of $g \to q\bar{q}$ ($q \neq b$) with $g \to b\bar{b}$, where one of the b-jets might not get tagged in the 3b regions, in the parton-shower generation.



(a) 1L (b) OS

Figure 5.6.: The N_j, N_b and $H_{\mathrm{T}}^{\mathrm{all}}$ distributions of the nominal MC prediction compared to data in jet inclusive regions for the 1L (left) and OS (right) channel. The uncertainty band shows the statistical and systematics uncertainties.

With these assumptions, the scale factor $R(x)$ for a given variable x for the $t\bar{t}$+jets background is defined as

$$R(x) = \frac{\text{Data}(x) - \text{MC}^{\text{non-}t\bar{t}}(x)}{\text{MC}^{t\bar{t}}(x)}. \tag{5.1}$$

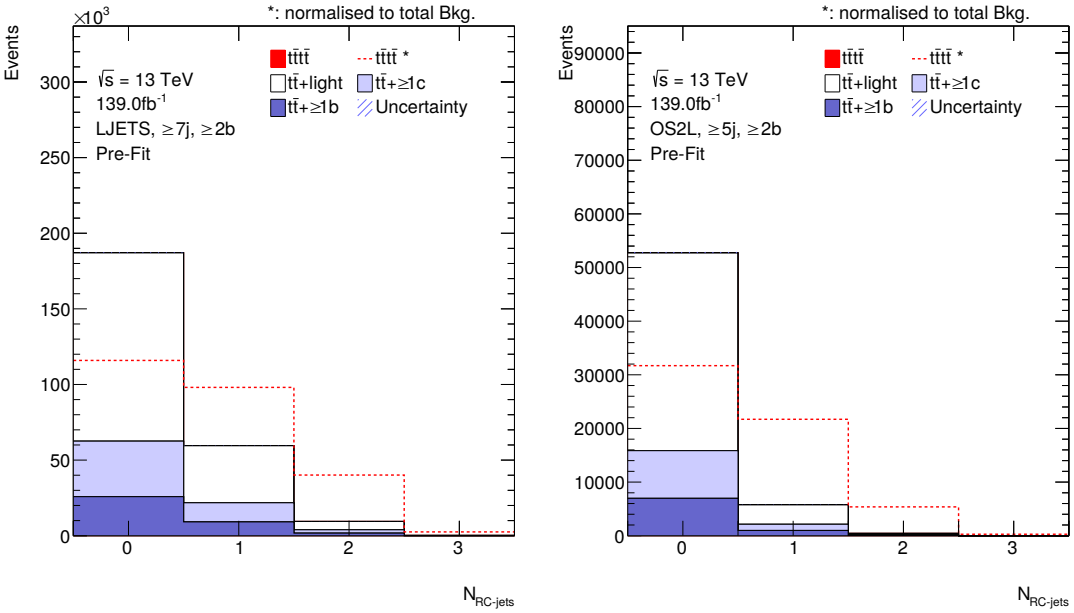
By subtracting the MC prediction for smaller backgrounds ($\text{MC}^{\text{non-}t\bar{t}}$), the MC prediction for $t\bar{t}$ +jets is compared to $t\bar{t}$ +jets events in data. With the same definition, a scale factor is derived for all systematic uncertainties by replacing $\text{MC}^{t\bar{t}}$ with the alternative prediction $\text{MC}^{t\bar{t},\text{alt}}$. The alternative prediction is obtained by using a different set of generators or by applying different scales as further discussed in Section 5.5,
What remains is the exact choice of variables x to derive the SFs. As further discussed in Section 5.4, the analysis uses a BDT for the signal and background separation. The BDT combines many variables and therefore a good MC modelling is required for each variable to not introduce a bias in the training. For the MC-method, this has the consequence that not only one, but three variables are used to extract the SFs. The following variables have been chosen to sequentially derive the SFs in three steps:

1. Number of jets (N_{j}): By choosing this variable, the overall normalisation per jet multiplicity is corrected. Even for only two b-jets, where statistical uncertainties are relatively small, the regions with > 10 (8) jets for 1L (OS) have only few events in individual bins of a given kinematic distribution. A larger binning is one possible solution but limits the description of shape effects. Therefore, these regions are not considered separately in the reweighting. Instead, higher jet multiplicities, up to $\geq$13 (11) jets, are corrected in normalisation only while the shapes are derived inclusively in the consecutive step.

2. Sum of transverse momenta of jets, MET and leptons: $H_{\text{T}}^{\text{all}}$ was the variable used for the fit in the previous analysis. It is correlated to the p_{T} of all objects. Consequently, the SF helps to correct many important kinematic variables for the BDT. As further presented below and defined in Equation 5.3, $H_{\text{T}}^{\text{all,red}}$, which is closely related to $H_{\text{T}}^{\text{all}}$ but independent of the jet multiplicity, is used to derive $R(x)$.

3. The average distance between two jets: $\Delta R_{\text{Avg}}^{jj}$ is sensitive to the $g \to bb$ splitting and is highly ranked for the BDT as shown in Section 5.4. Furthermore, it contains η and ϕ but no p_{T} information and is therefore correlated to other BDT input variables that are not scaled by the second step.

In principle, the order in which the SFs are derived does not impact the outcome. However, the chosen order allows to simplify the derivation of the SFs of the second step as discussed further below. The regions for the derivation are further split in the number of RC-jets ($N_{\text{RC-jets}}$), introduced in Section 4.3.3, where, limited by the number of events, 0, 1 or $\geq$2 RC-jets for 1L and 0 or $\geq$1 RC-jets for OS were chosen. As discussed in Section 4.3.3, RC-jets are motivated in this analysis due to top-decays where one large jet instead of two small-R jets are formed. As shown in Figure 5.7 by



(a) 1l channel (b) 2l OS channel

Figure 5.7.: Reclustered jet multiplicity in the (a) 1l and (b) OS channel.

the background normalised signal (red-dashed), the RC-jets multiplicity shows a good signal to background separation. In principle, this motivates to also split the fit region in multiplicities of RC-jets. However, no significant gain on the expected significance is obtained by applying this additional splitting while making the setup more complex and potentially biased due to overfitting. Nevertheless, the splitting is included in the reweighting as sufficient events are available in the 2b regions. As a dependence of the SFs is observed, an improvement of the modelling is to be expected.

For N_{j} and $\Delta R^{jj}_{\mathrm{Avg}}$, the first and last variable of the sequential scaling, the SFs are derived per bin. For $H_{\mathrm{T}}^{\mathrm{all}}$, a parametrisation f of the ratio $R(H_{\mathrm{T}}^{\mathrm{all}})$, with three free paramaters a, b, and c,

$$f(H_{\mathrm{T}}^{\mathrm{all}}) \;=\; \mathrm{a} + \frac{\mathrm{b}}{H_{\mathrm{T}}^{\mathrm{all\,c}}} \tag{5.2}$$

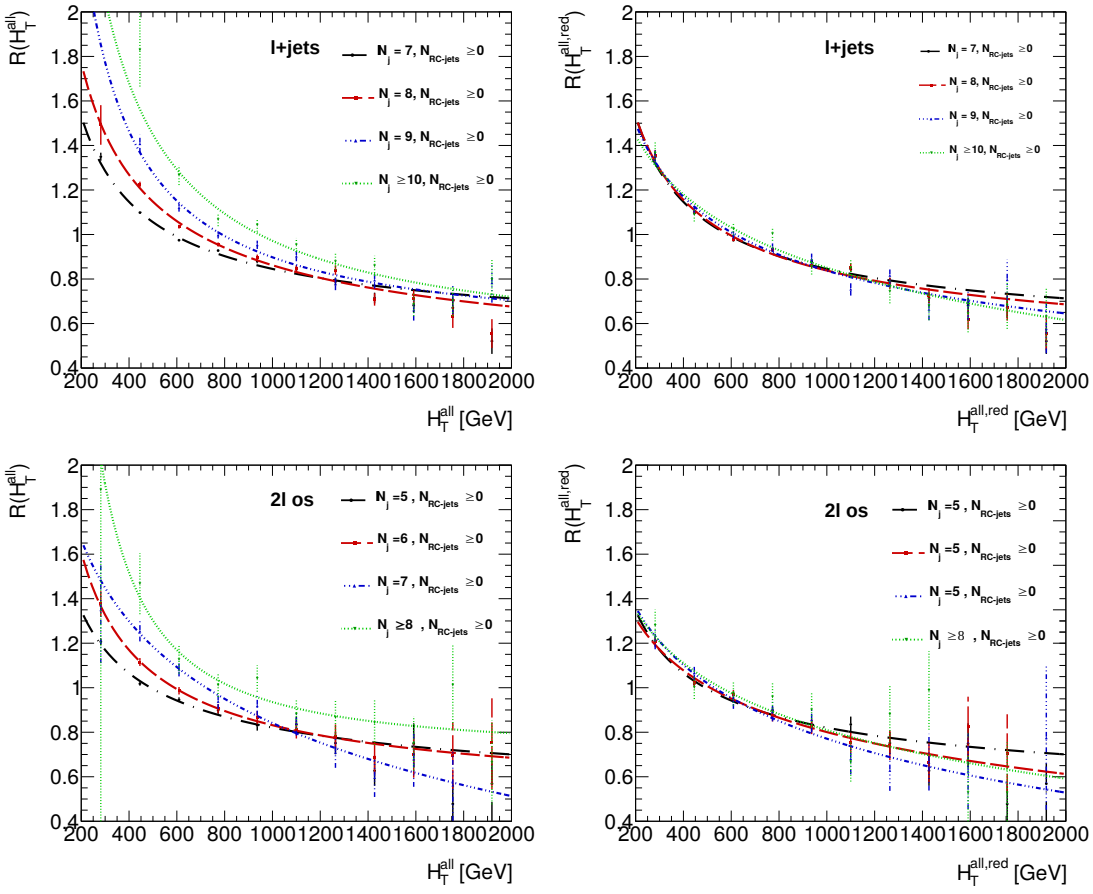
is applied to smooth statistical fluctuations in the $H_{\mathrm{T}}^{\mathrm{all}}$ bins. As shown in Figure 5.8, $f(H_{\mathrm{T}}^{\mathrm{all}})$ is derived by fitting the SFs as function of $H_{\mathrm{T}}^{\mathrm{all}}$. In total 12 (8) parametrisations would be needed (4 N_j regions and 3 or 2 $N_{\text{RC-jets}}$ regions). However, the $H_{\mathrm{T}}^{\mathrm{all}}$ correction is only a shape effect since the normalisations was already corrected in the first step. As a consequence, the scale factors for $H_{\mathrm{T}}^{\mathrm{all}}$ are very similar in each jet multiplicity and only shifted by a constant d for each additional jet, as shown in Figure 5.8.
To simplify the parametrisation, a reduced variable

$$H_{\mathrm{T}}^{\mathrm{all,red}} \;=\; H_{\mathrm{T}}^{\mathrm{all}} \;-\; 90\ \mathrm{GeV}\ \times (N_{\mathrm{j}} - N_{\mathrm{min}}) \tag{5.3}$$

Table 5.3.: Regions and number of scale factors (SF) per channel and step of the three step sequential derivation of additional SFs.

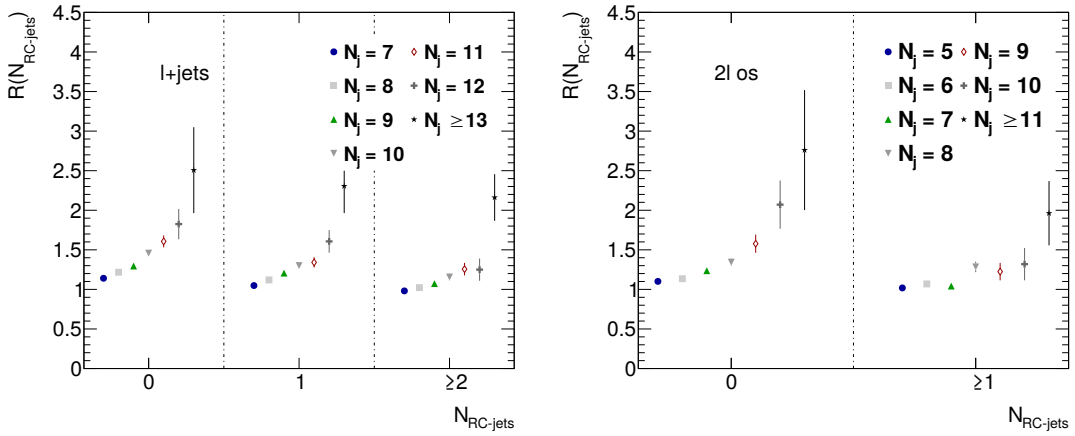
Variable	Single Lepton	# SF	Dilepton OS	# SF
$N_{\text{RC-jets}}$	$7 \leq N_{\text{j}} \leq 12$ or ≥ 13 $N_{\text{RC-jets}} \in [0,1,\geq 2]$	21	$5 \leq N_{\text{j}} \leq 10$ or ≥ 11 $N_{\text{RC-jets}} \in [0,\geq 1]$	14
$H_{\text{T}}^{\text{all,red}}$	$N_{\text{j}} \geq 7$ $N_{\text{RC-jets}} \in [0,1,\geq 2]$	3	$N_{\text{j}} \geq 5$ $N_{\text{RC-jets}} \in [0,\geq 1]$	2
$\Delta R_{\text{Avg}}^{jj}$	$7 \leq N_{\text{j}} \leq 9$ or ≥ 10 $N_{\text{RC-jets}} \in [0,1,\geq 2]$ 6-bins in $\Delta R_{\text{Avg}}^{jj}$	72	$5 \leq N_{\text{j}} \leq 7$ or ≥ 8 $N_{\text{RC-jets}} \in [0,\geq 1]$ 6-bins in $\Delta R_{\text{Avg}}^{jj}$	48
Sum		96		64

was defined where N_{min} corresponds to the smallest jet multiplicity considered for each channel (7 for 1L, 5 for OS). The shift of 90 GeV in Equation 5.3 was derived by parametrising $R(H_{\text{T}}^{\text{all}} - d)$ while fixing a, b and c in Equation 5.2 to the values of the lowest jet multiplicity. By using $\text{R}(H_{\text{T}}^{\text{all,red}})$ instead of $\text{R}(H_{\text{T}}^{\text{all}})$, all regions can be treated inclusively in $N_j \geq 7(5)$ in the 1L (OS) channel and consequently only three (two) parametrisations $\text{f}(H_{\text{T}}^{\text{all,red}})$ are needed. The total number of SFs per channel is summarised in Table 5.3 and the actual SFs are shown in Figure 5.9, 5.10 and 5.11.



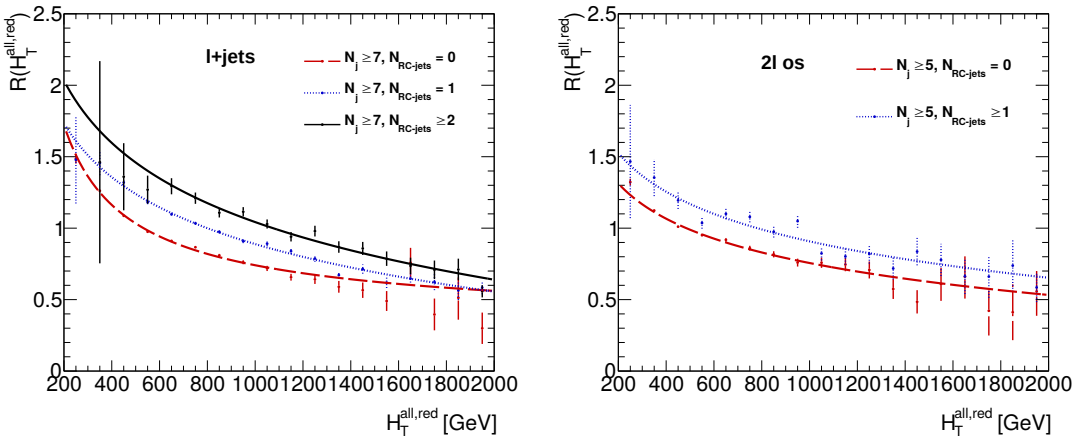
(a) H_{T}^{all} (b) $H_{\mathrm{T}}^{all,red}$

Figure 5.8.: Scale factors in 2b regions for all jet multiplicities with parametrisations for 1l (top) and 2l os (bottom) for H_{T}^{all} (a) and $H_{\mathrm{T}}^{all,red}$ (b). The uncertainties are statistical only.



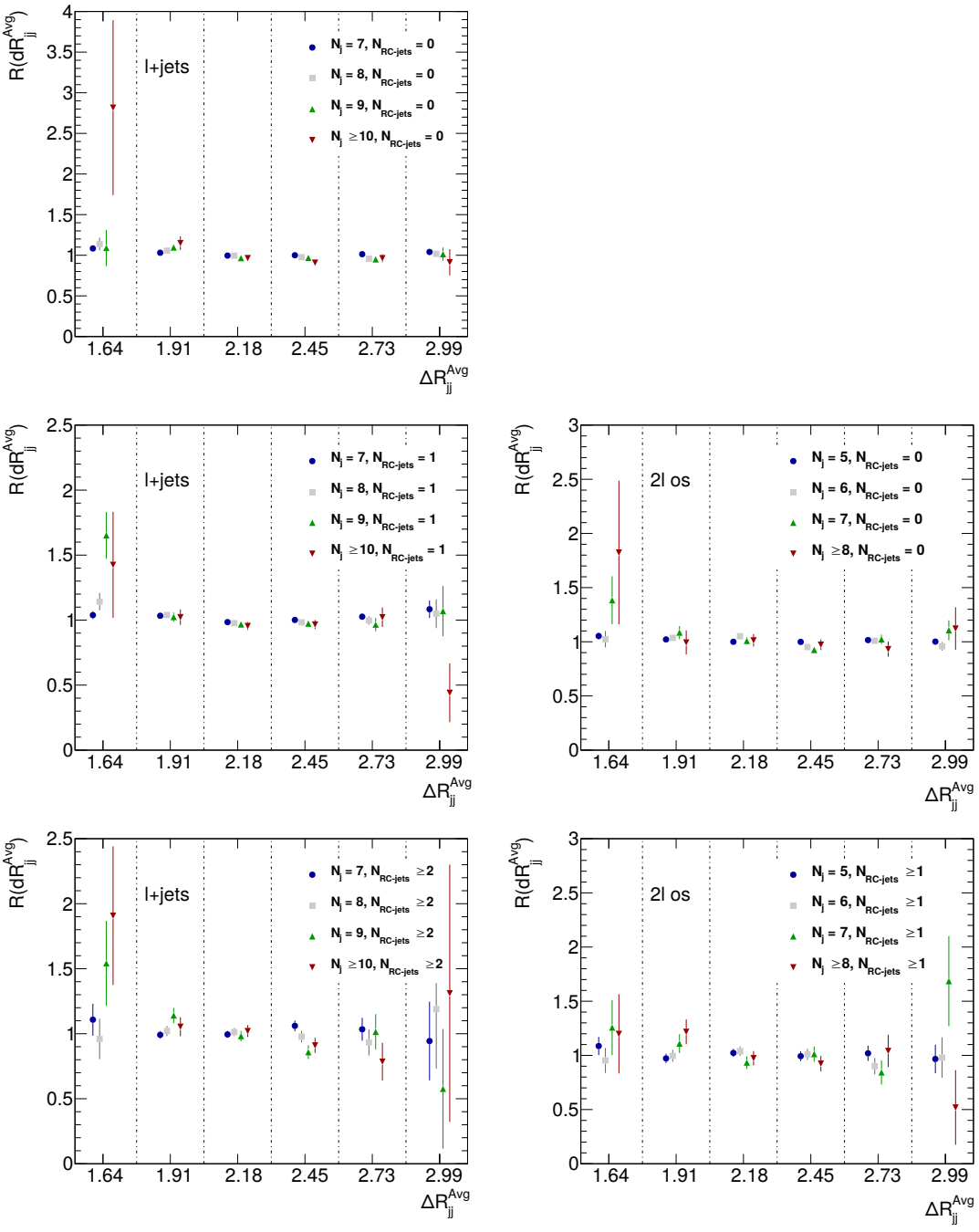
(a) 1L (b) OS

Figure 5.9.: Scale factors of the first reweighting step in jet and RC-jet multiplicities for the 1L channel (a) and the OS channel (b). The uncertainties are statistical only.



(a) 1L channel (b) 2l os channel

Figure 5.10.: Scale factors and the parametrisation of the second reweighting step in $H_{\mathrm{T}}^{\mathrm{all,red}}$. The uncertainties are statistical only.



(a) 1L channel (b) 2l os channel

Figure 5.11.: Scale factors for the third reweighting step in $\Delta R_{\text{Avg}}^{jj}$, split in jet and RC-jet multiplicities. The uncertainties are statistical only.

5.3.1. Method Validation

In this section, the underlying assumptions of the MC-method are investigated and justified. In addition, since the reweighting is performed sequentially in three steps, it is important to verify that each of the latter steps does not introduce additional effects on already corrected variables which are not taken into account in the final prediction.
For the first and second assumption, that no new physics causes the disagreement and the SM is still the valid theory, the method relies on other analyses which studied the phase space of the derivation regions in the past. As no such effect was observed by neither ATLAS or CMS, any new physics unrelated to the $t\bar{t}t\bar{t}$ production is expected to be negligible in this analysis.

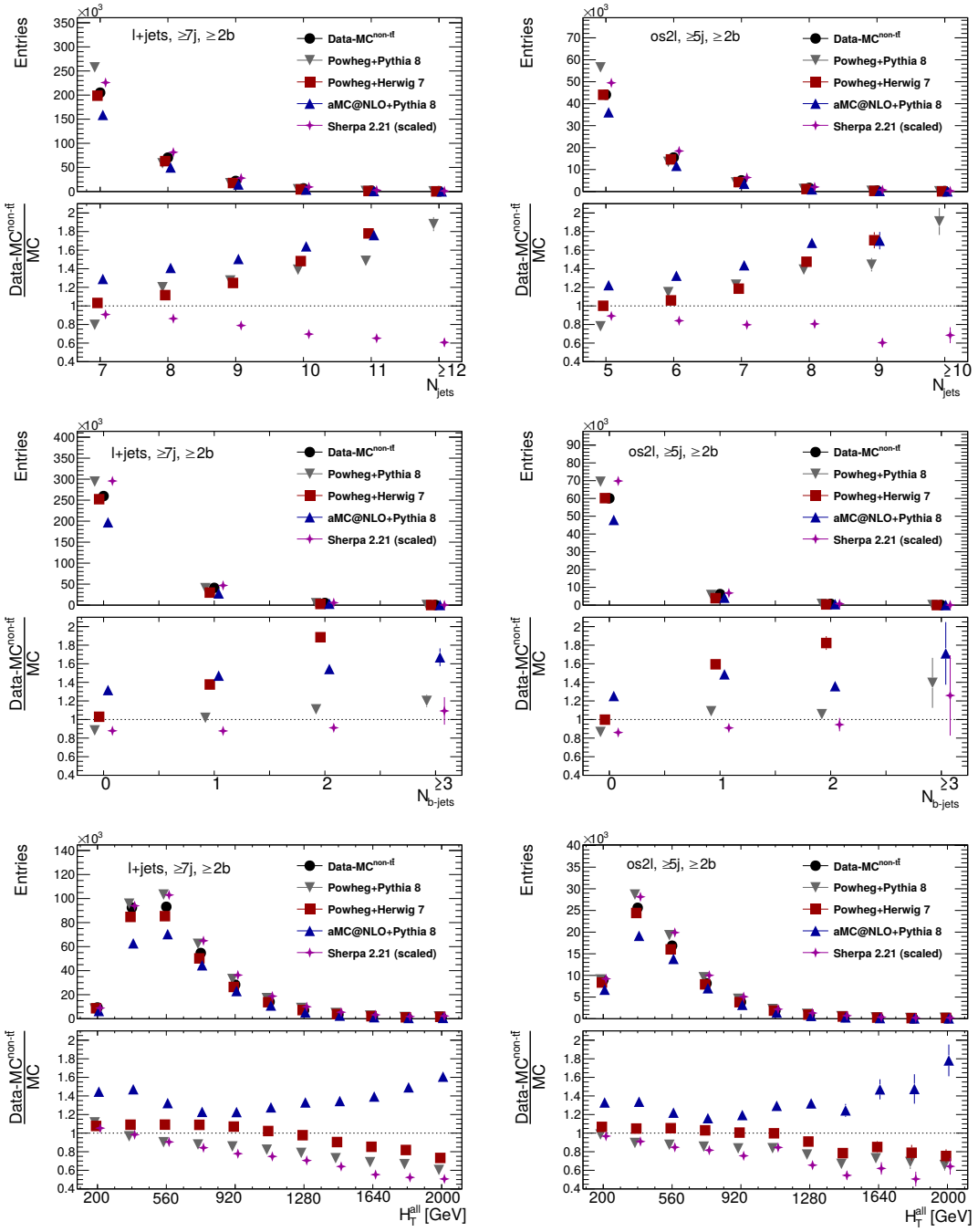
To confirm that the disagreement originates from the assumptions of the shower generators, described in Section 4.2, the agreement between data, subtracted by non-$t\bar{t}$ components, and different $t\bar{t}$+jets predictions is compared in Figure 5.12. For SHERPA, no MC samples were available at the time of the analysis for the 2018 data-taking period. To still allow for a comparison, the available samples (2015-2017) were scaled to match the overall luminosity. The first observation is that all generators show a different level of agreement with data, as expected by the different choices of the generators. Secondly, the disagreement for all generators increases with jet multiplicities (top of Figure 5.12) which confirms that each individual jet is slightly mismodelled and the effect adds up for multiple jets. While for SHERPA the ratio decreases, it increases for all other generator combinations. The increase in the ratio plots with higher b-jet multiplicities (middle) indicates that it is particularly difficult to model additional heavy-flavour jets correctly. The best and almost constant agreement is given by SHERPA which uses its own shower-generator based on the Catani-Seymour [134, 135] dipole formalism. SHERPA is, similar to aMC@NLO, a so-called multi-leg generator. This means that it merges additional parton-showers to NLO ME accuracy, while avoiding double counting of the additional jets. This is of particular importance for many additional jets like in this analysis.
Comparing the generators prediction for the $H_{\mathrm{T}}^{\mathrm{all}}$ distribution (bottom), one can see that also the jet kinematics are modelled differently. For values below approximately 1 TeV the effect is, in an inclusive jet region and mostly a normalisation difference between all generators. By comparing SHERPA with POWHEG, interfaced with PYTHIA 8 and HERWIG, this observation stays true for higher values, however, the ratio to data decreases more steeply in all three cases. For aMC@NLO interfaced with PYTHIA 8, a divergence with the corresponding POWHEG +PYTHIA 8 sample is observed for higher values. This indicates that the choice of the ME and the interference with the shower generator has an effect on the jet kinematics. From the definition of $H_{\mathrm{T}}^{\mathrm{all}}$ as sum of object p_{T} it becomes clear that events above 1 TeV are mostly populated by events with higher (b-)jet multiplicities which explains the difference in shape in the spectrum.
In conclusion, none of the major MC generators gives a satisfying agreement in kinematics and normalisation with data in the given phase space. Due to the observed differences, the disagreement must originate from the approximations. The dependence of the disagreement on the jet multiplicity indicates that in particular the approximations

of the shower generator are of importance. The best agreement in (b-)jet multiplicities is given by SHERPA while the effect in $H_{\mathrm{T}}^{\mathrm{all}}$ is similar to POWHEG samples. However, for technical reasons SHERPA samples were not fully available for this analysis.

The assumption that the total mismodelling is small compared to the $t\bar{t}$+jets mismodelling is shown by Figure 5.6. The total $t\bar{t}$ yields in the 1L region with exactly 7 (2) jets (b-jets) is around 158×10^3 which compares to 188×10^3 data events. From the SFs in Figure 5.9, the additional normalisation factor for $t\bar{t}$ is around 1.15 which means that, after correcting $t\bar{t}$+jets, a remaining normalisation effect of $f = \frac{\mathrm{Events}_{\mathrm{Data}}}{\mathrm{Yields}_{t\bar{t},\mathrm{corr}}} = \frac{188000}{158000 \times 1.15} = 1.04$ remains. This corresponds to the disagreement arising from non-$t\bar{t}$ background. A 4% effect on the normalisation is covered by appropriate systematic uncertainties which are all larger than 10 %, as discussed in Section 5.5. Equally important is the fact that the total shape effects of the non-$t\bar{t}$ backgrounds are small as well.

As already shown in Figures 5.2 and 5.6, the b-jet multiplicities only differ in the relative heavy flavour content. To justify the assumption that the SFs do not depend on the flavour, except for a normalisation effect which is expected, the SFs have been derived in a region with at least three b-jets ($\geq$3b). In Figures 5.13 (number of jets), 5.14 ($H_{\mathrm{T}}^{\mathrm{all}}$) and 5.15 ($\Delta R_{\mathrm{Avg}}^{jj}$), each reweighting step is compared bin-by-bin to the corresponding SFs of the 2b regions. For $H_{\mathrm{T}}^{\mathrm{all}}$, the parametrisation is shown as well. In general, the bin-by-bin comparison shows a good agreement with unity in all plots. Not all bins have a perfect agreement but, as deviations mostly occur in bins with lower statistics typically associated with a higher (RC-) jet multiplicity, the statistical uncertainties are increased as well. In the parametrisation of the $H_{\mathrm{T}}^{\mathrm{all}}$ observable, 1L shows a good agreement in all $N_{\text{RC-jets}}$ multiplicities while for OS a deviation between the curves for $N_{\text{RC-jets}} = 0$ (red and grey) is observed. However, this does not take the uncertainties of the scale factors themselves, which are larger for OS due to the larger statistical uncertainties, into account. Furthermore, the binning and the range of the parametrisation were optimised in the 2b regions. In conclusion, the agreement between the shown b-jet multiplicities is good.

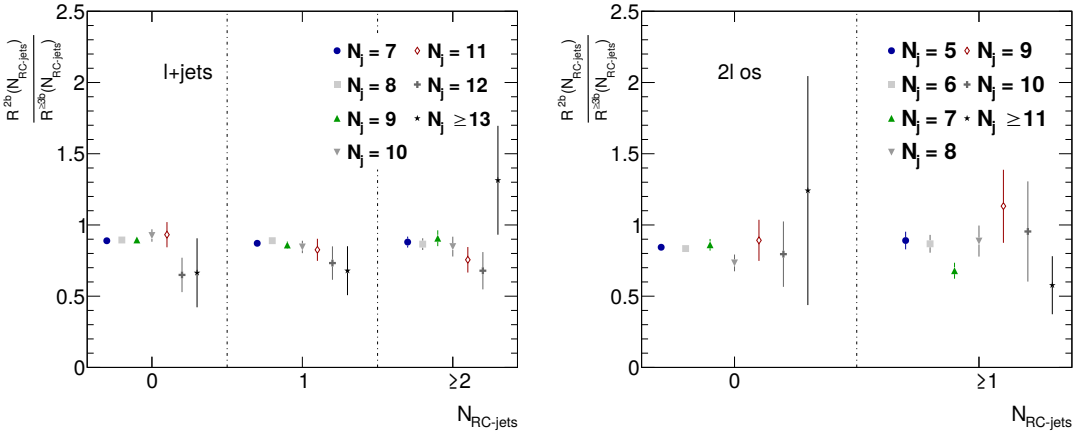
The validation of each reweighting step is shown in Figure 5.16 for the 10 jet inclusive region with 2 b-tags. A more complete validation in all b-jet multiplicities for all jet regions in the derivation variables is shown in Appendix A. As expected, the variables show very good agreement in the regions where the reweighting was derived. Earlier variables, such as the $H_{\mathrm{T}}^{\mathrm{all}}$ distribution, do not change by the second and third reweighting step which shows that the sequential reweighting works as expected. The remaining difference, in particular in individual bins, is a consequence of the parametrisation. $\Delta R_{\mathrm{jj}}^{\mathrm{Avg}}$ exhibits a very good agreement in the bulk region where the scale factors were derived. For $N_{\text{RC-jets}}$, the only difference is observed in bins with three or ≥ 4 RC-jets since these bins were treated inclusively with the two RC-jets bin which dominates.



(a) 1L channel

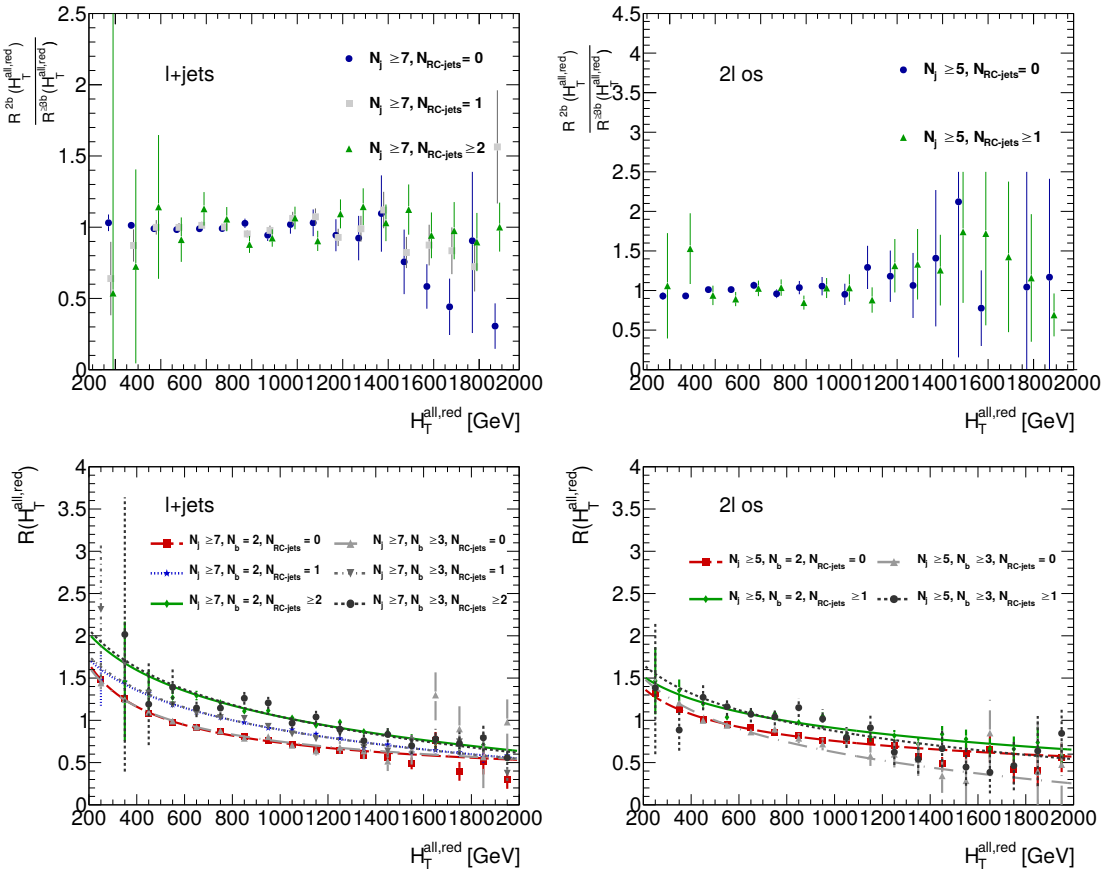
(b) 2l os channel

Figure 5.12.: Jet (top) and b-jet (middle) multiplicities as well as $H_{\mathrm{T}}^{\mathrm{all}}$ (bottom) for the 1L (left) and OS (right) channel for different ME and parton-shower generators. For Sherpa 2.21, no MC samples were available for the 2018 data-taking, therefore, the 2015-2017 Samples have been scaled to match luminosity. The uncertainties are statistical only.



(a) 1L channel (b) 2l os channel

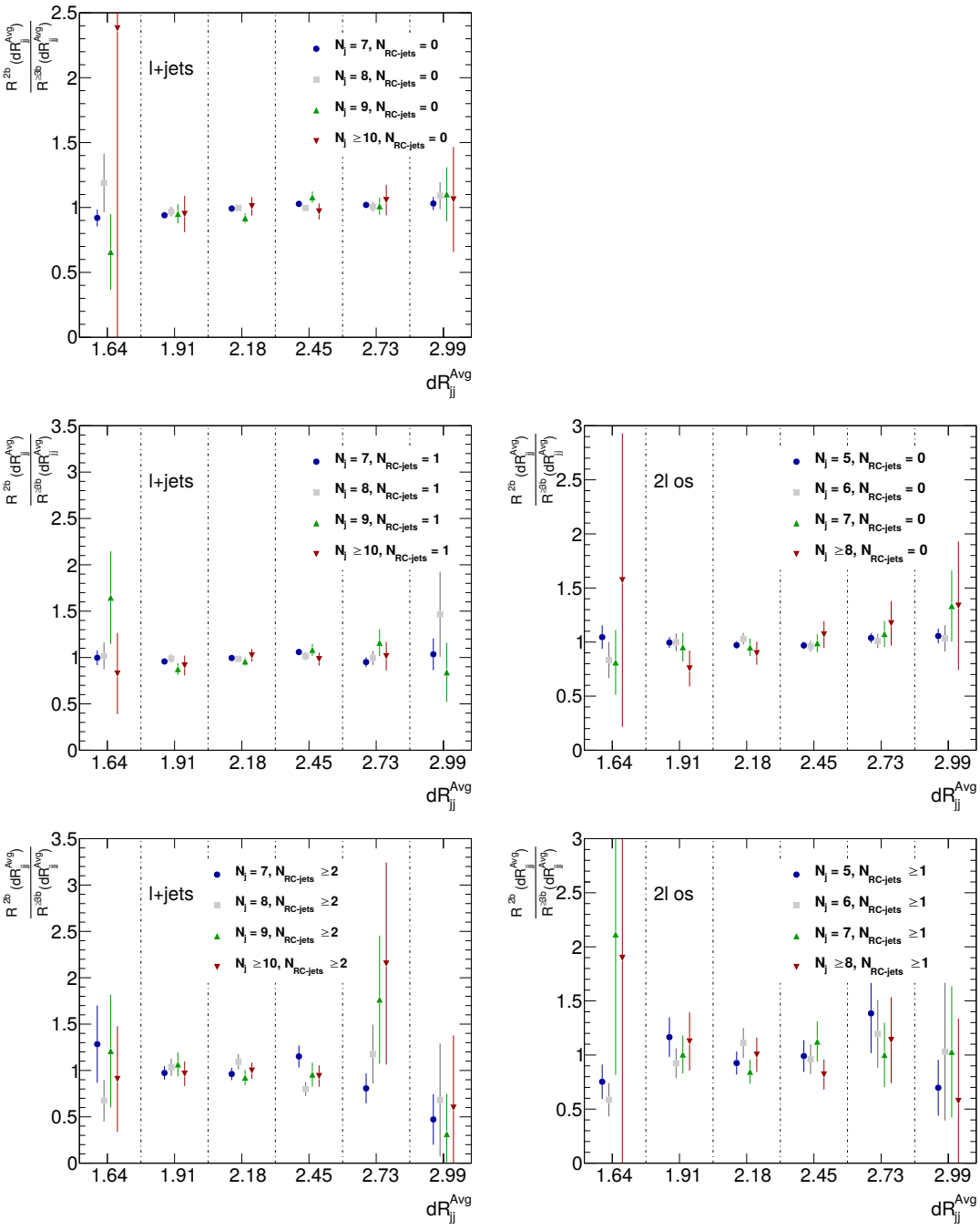
Figure 5.13.: Scale factors in 3b regions compared to the nominal scale factors in 2b regions for the first reweighting step. The uncertainties are statistical.



(a) 1L channel

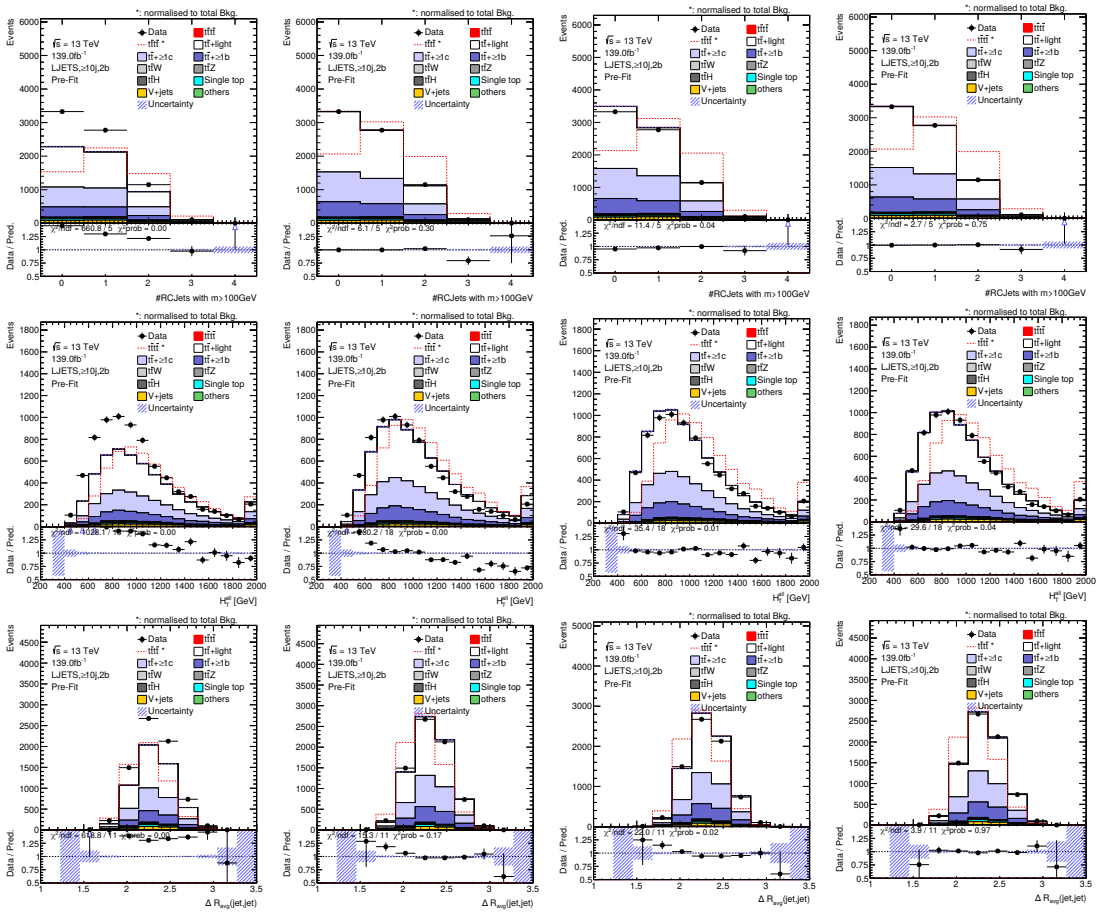
(b) 2l OS channel

Figure 5.14.: Comparison of the scale factors derived in the 2b and in the ≥3b region. The bin by bin ratio is shown in the upper plots and the parametrisations of $H_{\mathrm{T}}^{\mathrm{all,red}}$ are compared in the lower plots. The uncertainties are statistical only.



(a) 1L channel (b) 2l os channel

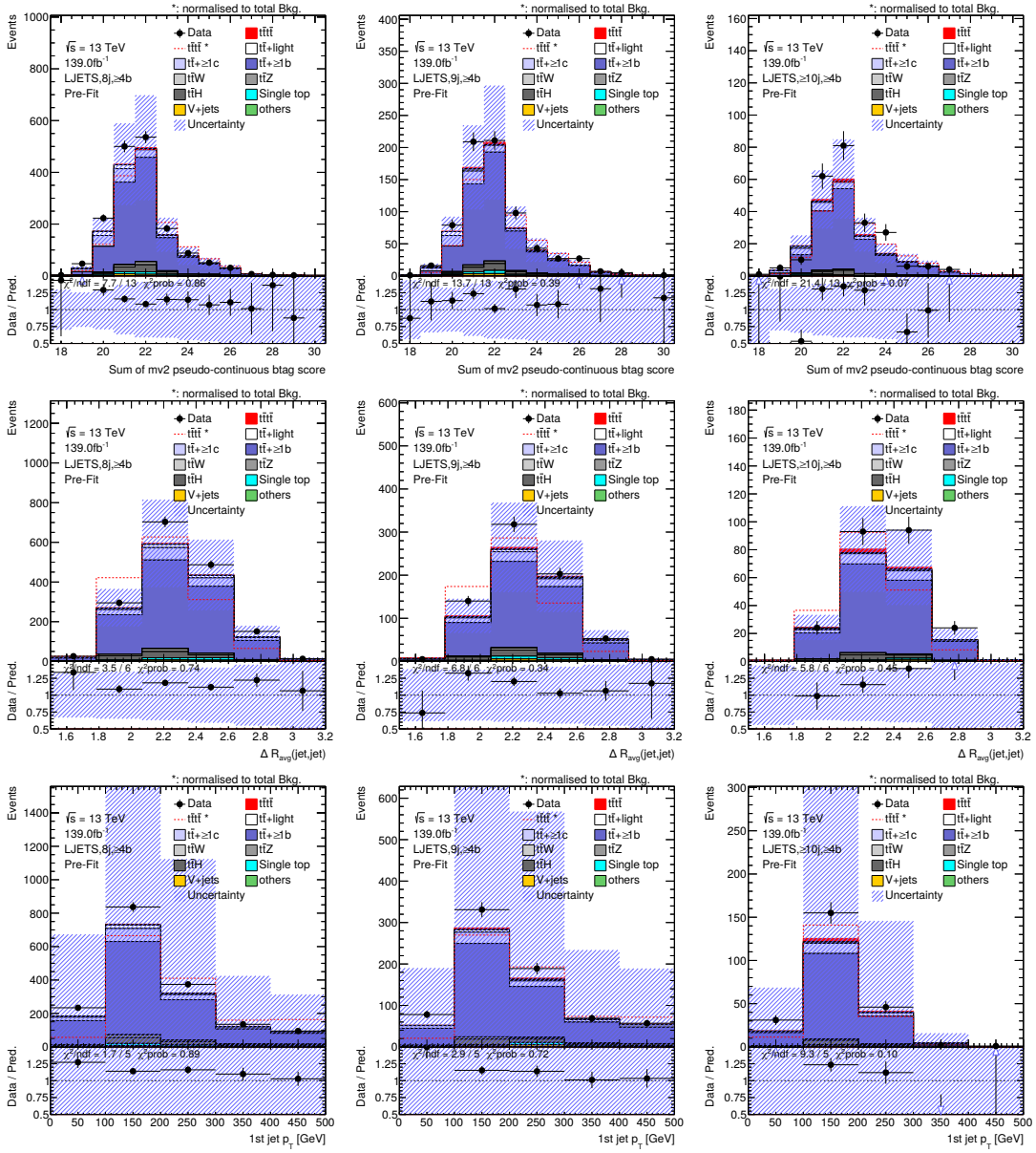
Figure 5.15.: Scale factors in 3b regions compared to the nominal scale factors in 2b regions for the third reweighting step. The uncertainties are statistical.



(a) Unweighted (b) After $N_{\text{RC-jets}}$ (c) After $H_{\text{T}}^{\text{all,red}}$ (d) After $\Delta R_{\text{jj}}^{\text{Avg}}$

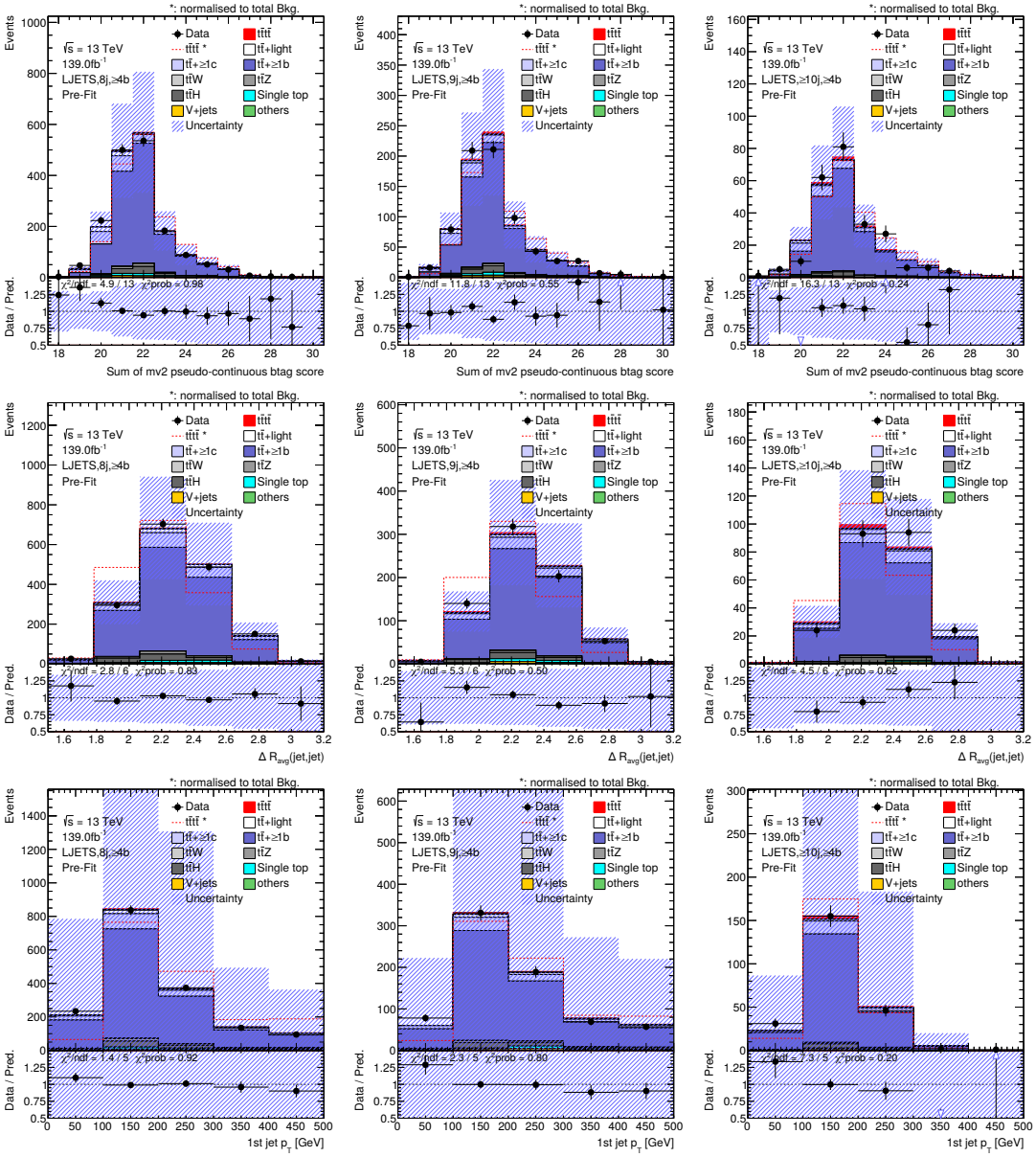
Figure 5.16.: Distributions of reweighting variables in the 10j inclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$,(c) after $H_{\text{T}}^{\text{all,red}}$ and (d) after $\Delta R_{\text{jj}}^{\text{Avg}}$ reweighting. The uncertainty band includes statistical uncertainties only.

Selected variables are shown in Figure 5.17 based on their importance for the BDT as further discussed in Section 5.4. The shown regions correspond to the jet multiplicities of the fit regions with at least 4 b-jets in the 1L channel. In particular, as further discussed in Section 5.5 the $t\bar{t}$+jets cross-section (50%) and modelling uncertainties (typically larger than 10%) add up to a large overall uncertainty before performing the fit at the given multiplicities. In addition, the MC statistics decrease with higher multiplicities. A more complete overview for these variables in both channels is given in Appendix B. In general, all variables used for training the BDT show a good agreement between data and the MC prediction after applying the SFs. An offset due to different flavour compositions is observed in the 3b and $\geq$4b regions. As this offset is assumed to only be a normalisation effect originating from the additional b-jet, the difference in the shape is expected to be negligible. For a better visualisation, Figure 5.18 compares the same distributions after manually fixing the $t\bar{t}$+jets normalisation to match data before fitting (pre-fit). This ad-hoc scaling, which is not applied in the actual fit, corrects any normalisation difference and therefore visualises the shape-only difference. The uncertainties are left unchanged and therefore slightly overestimated. Comparing Figure 5.17 and Figure 5.18, it is observed that shape effects got corrected and the only difference in Figure 5.17 was indeed only a normalisation effect as assumed.



(a) 8 jets (b) 9 jets (c) ≥10 jets

Figure 5.17.: The agreement between data and MC of the most important BDT-input variables in region with at least four b-jets. The uncertainty band includes statistical and systematical uncertainties.



(a) 8 jets (b) 9 jets (c) ≥10 jets

Figure 5.18.: The shape-only agreement between data and MC of the three most important BDT-input variables in regions with at least four b-jets in the (a) 8-jet, (b) 9-jet and (c) at least 10 jet regions. The $t\bar{t}$ normalisation was fixed such that MC and data agree in overall yields. The uncertainty band includes statistical and systematical uncertainties.

5.3.2. Data-driven Tag-Rate-Function Method

The $t\bar{t}$ *tag-rate-function* (ttTRF) provides an alternative prediction of the $t\bar{t}$ background in the fit regions based on data. It was used in the previous $t\bar{t}t\bar{t}$ analysis in ATLAS in the 1L+OS channel [49] and is based on the tag-rate-function (TRF) method which has been used, for example, in the hadronic [136] and the single leptonic channel [133, 137] of the $pp \rightarrow t\bar{t}H\,(b\bar{b})$ analysis. While the signal and control regions are identical to the MC-method, the derivation regions for the background prediction are different, as shown in Figure 5.19, for the following reasons.
The ttTRF method is based on the estimation of the probability that an additional jet, not originating from the direct $t\bar{t}$ decay, is classified as a b-jet by the tagging algorithm. It compares the number of b-tagged jets with the total number of jets to derive a probability p_{b} for a given b-tagging working point,

$$p_{\mathrm{b}} = \frac{N_{\text{b-jets}}}{N_{\mathrm{j}}}. \tag{5.4}$$

It is parametrised by the jet p_{T} and the minimum distance between two jets times the jet multiplicity ($\Delta R_{jj}^{\mathrm{min}} \times N_j$) to take the tagger response for different kinematics into account. The latter variables were chosen as $\Delta R_{jj}^{\mathrm{min}}$ decreases naturally for a higher jet multiplicity which is compensated by multiplying with N_j to be less dependent of the jet multiplicity. The regions to extract this value have to be orthogonal to the fit regions and require a high $t\bar{t}$ contribution. To allow for validation close to the fit regions in the 7 (5) jet regions for 1L (OS), the probabilities in this analysis are derived in the 5 (4) jet regions and are applied to the regions with exactly two b-jets for different jet multiplicities. Non-$t\bar{t}$ contributions based on MC are subtracted from data to obtain a pure $t\bar{t}$ prediction similar to the MC-method. Auxiliary weights $w_{\mathrm{TRF}}^{\text{b-jets}}$ are derived according to the mathematical description of the TRF method

$$w_{\mathrm{TRF,incl}}^{\mathrm{2b}}(N_j) = \prod_{i=1}^{N_j-2} \left(1 - p_{\mathrm{incl}}^{j}\right) \tag{5.5}$$

$$w_{\mathrm{TRF,incl}}^{\geq \mathrm{3b}}(N_j) = \sum_{i=1}^{N_j-2} p_{\mathrm{b}}^{i} \prod_{j\neq i} \left(1 - p_{\mathrm{incl}}^{j}\right) \tag{5.6}$$

$$w_{\mathrm{TRF}\geq \mathrm{4b}}^{\geq \mathrm{3b}}(N_j) = \prod_{i=1}^{N_j-3} \left(1 - p_{\geq \mathrm{4b}}^{j}\right) \tag{5.7}$$

$$w_{\mathrm{TRF}\geq \mathrm{4b}}^{\geq \mathrm{4b}}(N_j) = \sum_{i=1}^{N_j-3} p_{\geq \mathrm{4b}}^{i} \prod_{j\neq i} \left(1 - p_{\geq \mathrm{4b}}^{j}\right). \tag{5.8}$$

$$\tag{5.9}$$

The index "incl" refers to the case when all extra jets are included while the index "$\geq 4b$" indicates that the third highest b-tagged jet is excluded. From these definitions,

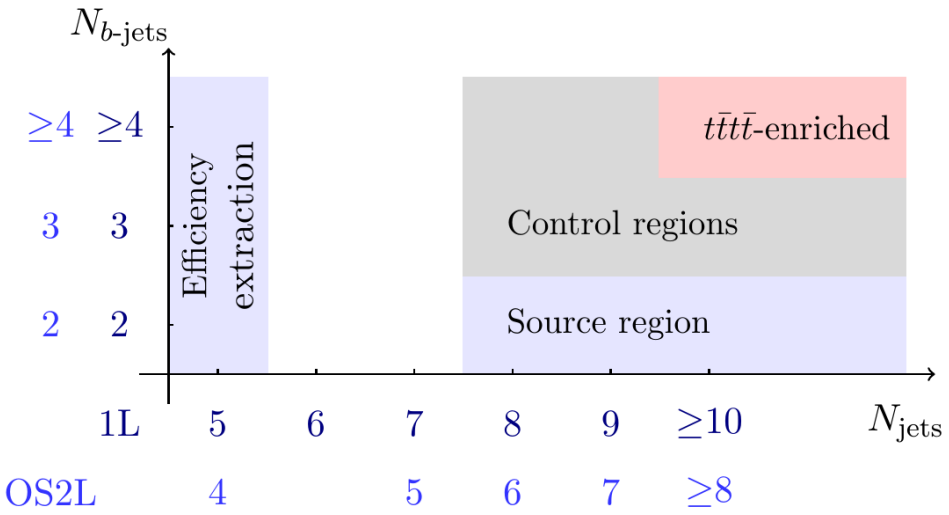


Figure 5.19.: Region definition for the ttTRF method. The efficiencies are derived in 5 (4) jet regions and are applied to the source regions to obtain a prediction in the control and $t\bar{t}t\bar{t}$-enriched regions. Only the control and $t\bar{t}t\bar{t}$-enriched regions are fitted.

the event weight

$$w^{3\text{b}}_{\text{ttTRF}}(N_j) = \frac{w^{\geq 3\text{b}}_{\text{TRF,incl}}(N_j)}{w^{\geq 2\text{b}}_{\text{TRF,incl}}(N_j)} \tag{5.10}$$

is derived. To obtain the weight for the 4b inclusive region, a similar procedure is applied while making use of the ttTRF estimation in the 3b inclusive regions and it follows that

$$w^{\geq 4\text{b}}_{\text{ttTRF}}(N_j) = \frac{w^{\geq 4\text{b}}_{\text{TRF},\geq 4\text{b}}(N_j)}{w^{3\text{b}}_{\text{TRF},\geq 4\text{b}}(N_j)} \times w^{3\text{b}}_{\text{ttTRF}}(N_j). \tag{5.11}$$

The derived weight $w^{3\text{b}}_{\text{ttTRF}}$ ($w^{4\text{b}}_{\text{ttTRF}}$) is applied to regions with exactly 2 (3) b-tags at higher jet multiplicities. A given event from data gets promoted to the higher b-jet multiplicity resulting in a $t\bar{t}$ prediction in the fit regions. Since the weight is applied to the whole event, and not a single distribution, a prediction for all variables is obtained simultaneously and the statistical uncertainty corresponds to the uncertainty in the 2b regions.
This pure data prediction assumes no dependence on the jet multiplicity and no correlation between the b-tags. These assumptions are limited since, for example, events with an additional $g \to b\bar{b}$ can be assigned to the wrong b-jet multiplicity if one b-jet is not tagged. To take this kind of effect into account, correction factors f_i^{MC} are applied to the yields y. The factors are derived by comparing the ttTRF prediction with the

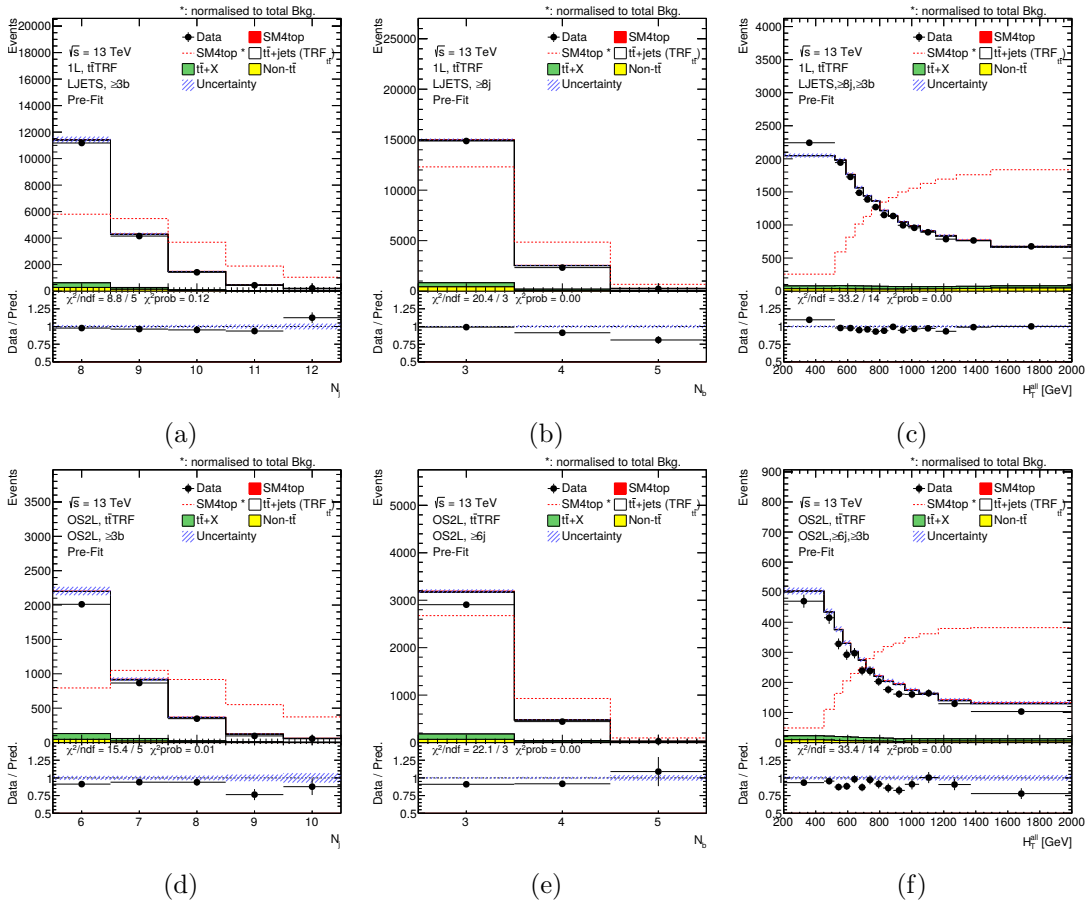


Figure 5.20.: Data MC agreement for the $t\bar{t}$ background prediction in N_j, N_b and $H_{\mathrm{T}}^{\mathrm{all}}$ based on the ttTRF method. (a)-(c) shows the 1L while (d)-(f) shows the OS channel. All but the $t\bar{t}$ background are taken from MC prediction. The uncertainty band shows statistical uncertainties.

nominal MC prediction in each bin i in the extraction region such that

$$y_{\mathrm{ttTRF,i}}^{\mathrm{corr}} = y_{\mathrm{ttTRF,i}} \times f_i^{\mathrm{MC}} = y_{\mathrm{ttTRF,i}} \times \frac{y_i^{\mathrm{MC}}}{y_{\mathrm{ttTRF,i}}^{\mathrm{MC}}}. \tag{5.12}$$

As a result, the method provides a $t\bar{t}$ prediction, as shown in Figure 5.20, which has a good agreement between data and MC.

5.4. Signal and Background Separation

The goal of the analysis is to measure the cross-section of the $t\bar{t}t\bar{t}$ process. If one would do a counting experiment by only using a single bin, no shape information to separate the normalisation for background and signal would be available if there is no prior knowledge

on the background yield. Consequently, the background uncertainty after performing the fit would be large. To gain separation, regions which differ in signal and background composition were already introduced in Section 5.2. In the previous analysis, the variable with the best separation H_{T} was used. In this analysis, many variables are combined to increase the separation as shown in this Section. In addition, the binning was studied to be more sensitive to the shapes of signal and background in the fitted variable as briefly summarised in Section 5.4.2.

5.4.1. Boosted Decision Trees

Figure 5.21 (a) shows the separation between $t\bar{t}$+jets and $t\bar{t}t\bar{t}$ in $H_{\mathrm{T}}^{\mathrm{all}}$ for the 1L channel. While already a difference is visible that can be used by the fit, more features of the signature in the detector can be used by combining multiple variables in a BDT.
The BDT aims to classify events in a single *score* as signal- or background-like based on the input variables and their correlations. While the output of the BDT does not correspond to a physical variable that can be measured, it can be interpreted as follows. Each input variable adds another dimensionality to the BDT. In this N-dimensional space, correlations or patterns among the variables are identified for signal and background. The BDT is trained to learn the patterns by implementing cuts on each variable (node) where multiple cuts on one variable, depending on the previous cuts, are possible. This results in a tree-like structure. The final step on each node, where no further cut is applied based on pre-defined criteria, is called a *leaf*. In this analysis the Toolkit for Multi Variate Analysis (TMVA) provided by the ROOT software package is used for training the BDT. After verifying the agreement of all kinematic distributions between the $t\bar{t}t\bar{t}$ NLO and LO predictions, the LO prediction is used as signal sample even though the NLO prediction is theoretically more precise. This choice is motivated by the 30% of events with negative weights from the NLO samples which statistically limits the BDT during training. The $t\bar{t}$ prediction after applying the additional SFs is used as background sample. The training is done in the $\geq$3b regions but in each jet multiplicity of the fit regions separately.
In more detail, training of the BDT means to iteratively optimise the cut-values in each step in such a way that the overall separation is maximised. In this analysis, a maximum of two consecutive cuts on a node (depth) is used. To quantify the change of the separation after adapting a certain cut while training, a *loss function* is evaluated after each iteration, for example the *Gini impurity*

$$I(p_s, p_b) = \sum_{i \in [s,b]} p_i(1 - p_i) \tag{5.13}$$

where $p_{s/b}$ are the estimated probabilities for a randomly chosen event i to be classified as signal or background for a given set of cuts. The iteration stops when the maximum number of 30 cuts on a single variable, a minimal number of events after a cut in a leaf, the maximum depth or no further improvement of separation by an additional cut is reached.

In this analysis, 600 trees are trained successively on events that have not been correctly classified in the previous steps. Furthermore, *gradient boosting*, which makes use of a loss-function to identify the weaknesses of a single tree before boosting, is applied. For each tree, the correct classification rate is evaluated and a weight is derived. Events that are difficult to separate, based on how often they were classified incorrectly, and trees which correctly classified such events are weighted higher. The final output is the weighted average of all trees which is transformed to a BDT-score $\in$[-1,1] where "1" corresponds to a higher probability to be signal. The final separation between the NLO signal and background in the BDT-score is shown in Figure 5.22.
Besides the optimisation of the cuts, which is done by the training, the *hyperparameters* (depth, number of trees, learning rate) were optimised by a grid-scan. During this fine tuning, the *overtraining* was kept small while improving the separation. Overtraining occurs when the tree learns patterns that originate from statistical fluctuations and are therefore do not correspond to a real separation between signal and background. It is evaluated by dividing the training set into three subsets, training the BDT on each subset and then evaluating the classification input in the other two (3-fold cross validation).

The full list of variables is given in Table 5.4 sorted by their importance which corresponds to the impact of a variable on the separation. In total, 16 (15) variables for the 1L (OS) channel, are combined by the BDT. The three most important variables of this combination are shown in Figure 5.21 (b)-(d). The corresponding plots for OS can be found in Appendix B and look quite similar. The separation between $t\bar{t}t\bar{t}$ and $t\bar{t}$+jets, where $t\bar{t}$+jets is not split in flavours, is given in Table 5.5 for the variables in Figure 5.21 and the BDT-score. The numerical separation S was calculated according to

$$S = \frac{1}{2}\sum_{i=1}^{N}\frac{\left(y_i^{t\bar{t}t\bar{t}} - y_i^{t\bar{t}}\right)^2}{y_i^{t\bar{t}t\bar{t}} + y_i^{t\bar{t}}} \tag{5.14}$$

where y is the respective yield, after applying all scale factors, in each of the N bins. Comparing the importance with the separation shows that the latter is not the only quantity to consider when choosing an input variable. In the case of $H_{\mathrm{T}}^{\mathrm{all}}$, the separation is highest but the importance is quite low. Most of the information gained in the BDT by adding $H_{\mathrm{T}}^{\mathrm{all}}$ is, due to correlations, already available by, for example, considering $p_{\mathrm{T}}^{\mathrm{lead,jet}}$ which is part of the $H_{\mathrm{T}}^{\mathrm{all}}$ definition. For this reason, the variables were chosen based on correlations and separation which reflects the different physical signatures of signal and background in the detector:

- *Flavour tagging*: For $t\bar{t}$+jets at least two b-jets originate from additional radiation while the b-jets from $t\bar{t}t\bar{t}$ decays are products of the top-decay. This leads to a different response in the b-taggers where, on average, signal events are more b-like, for example, due to a higher p_{T} and the corresponding better response of the tagger.

- *Transverse Plane*: Due to the higher centre-of-mass energy required to produce $t\bar{t}t\bar{t}$, the two initial partons are expected to have similar energy. As a consequence,

the $t\bar{t}t\bar{t}$ decay products are more central in the detector and their transverse kinematics are harder. In addition, the relative angular distributions between (b)-jets or leptons are different.

- *Jets*: Similar to the flavour tagging, the $t\bar{t}$ decay requires additional jets via radiation while the $t\bar{t}t\bar{t}$ jets originate from direct decays. As the kinematics for these jets are different, the invariant masses and distances between the jets differ.

- *RC-Jets*: Since, on average, most of the parton energy is needed to produce a $t\bar{t}t\bar{t}$ event, less energy is available in a collision to boost a top quark and the corresponding boost is smaller. On the other hand, as the events are more central, RC-jets from $t\bar{t}t\bar{t}$ events are more likely to pass the acceptance and the additional 100 GeV mass cut. Therefore, not only the number of RC-jets but also the splitting scales are considered.

- *Missing energy*: In the 1L channel, the transverse W mass can be reconstructed for signal and background and only a small separation is gained. In the OS channel, the missing energy is sensitive to the production of fake leptons which slightly differs for $t\bar{t}t\bar{t}$ and $t\bar{t}$+jets.

Table 5.4.: BDT input variables in the 1L+OS channel and their importance, based on the impact on the separation if dropped. The input variables combine flavour tagging, (RC) jet related angular as well es missing energy information.

Variable	Importance		Description	Information
	1L	OS		
$\sum_{i=1}^{6} \mathrm{PC}_i$	1	2	6 highest Pseudo Continuous (PC) b-tagging scores summed	Flavour tagging
N_j	2	1	Jet multiplicity	Transverse Plane
$\Delta R_{\mathrm{jj}}^{\mathrm{Avg}}$	3	3	Average distance between two jets	Jets
$p_{\mathrm{T}}^{\text{lead. jet}}$	4	5	p_{T} of jet with highest p_{T} (leading)	Transverse Plane
$M_{\mathrm{jjj}}^{\Delta R_3^{min}}$	5	4	Mass of the three jets (ijk) which minimise $\sqrt{\Delta R_{ij}^2 + \Delta R_{ik}^2 + \Delta R_{jk}^2}$	Jets
$\Delta R_{\mathrm{min}}^{bl}$	6	10	Minimum distance between b-jet and lepton	Jets
M_{T}^{W}	7	–	Transverse W mass: $E_{\mathrm{T}}^{\mathrm{miss}} + m_{\mathrm{T}}^{\mathrm{lep}}$ in 1L events	Missing energy
$H_{\mathrm{T}}^{\mathrm{all}}$	8	12	$\sum p_{\mathrm{T}}$ for leptons and jets	Transverse Plane
$M_{\mathrm{bbb}}^{\mathrm{avg}}$	9	8	Average mass of three b-jets	Jets
Centrality	10	9	$H_{\mathrm{T}}^{\mathrm{all}} / \sum E$ for leptons and jets	Transverse Plane
$\Delta R_{\mathrm{min}}^{bb}$	11	11	Minimum distance between two b-jets	Jets
$\sum_{i \in \mathrm{RC}} d_{23}^{i}$	12	7	d_{23}^{i} is a splitting scale indicating the separation of the second and (if applicable) third substructures in the k_{T} algorithm. It is summed for al RC-jets.	RC-Jets
$E_{\mathrm{T}}^{\mathrm{miss}}$	13	15	Missing transverse energy	Missing energy
$M_{\mathrm{bb}}^{\mathrm{min}}$	14	13	Minimum mass of two b-jets	Jets
$\sum_{i \in \mathrm{RC}} d_{12}^{i}$	15	6	d_{12}^{i} is a splitting scale indicating the separation of first and second substructures in the k_{T} algorithm. It is summed for al RC-jets	RC-Jets
$N_{\text{RC-jets}}$	16	14	RC-jet multiplicity with $M_{\text{RC-jet}} > 100\,\mathrm{GeV}$	RC-Jets

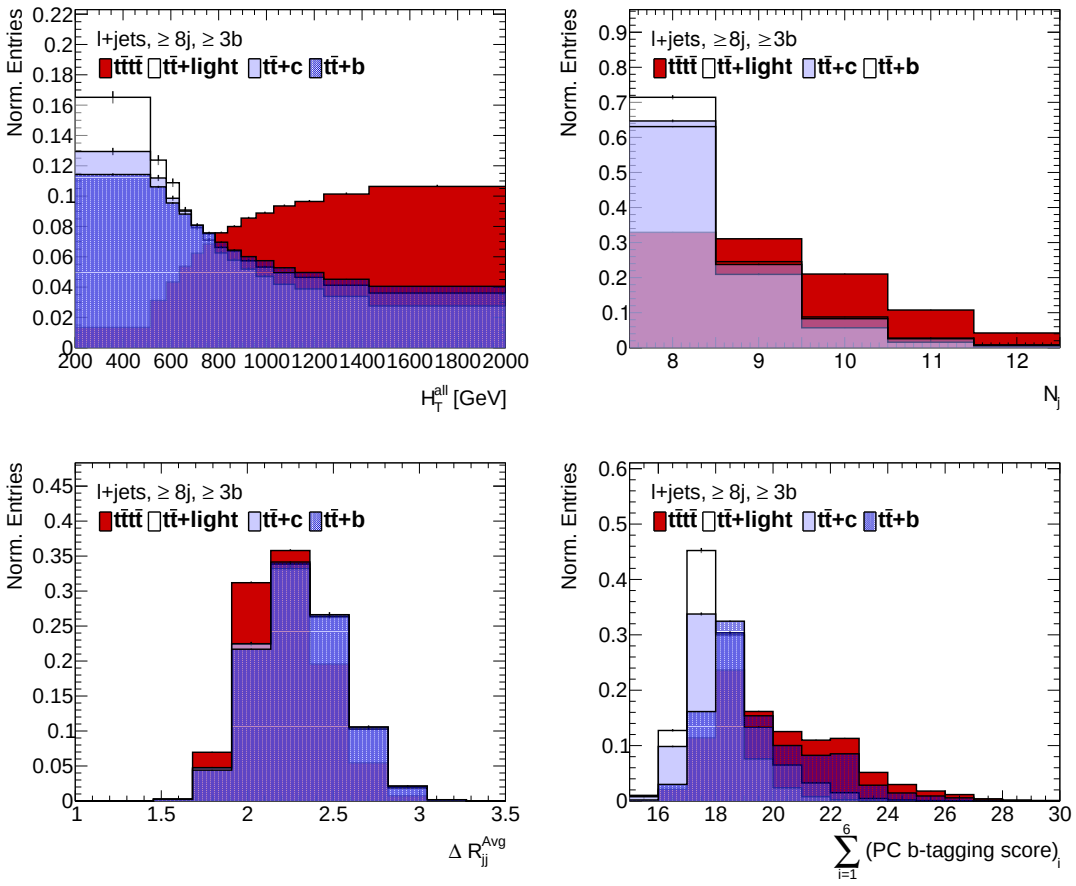
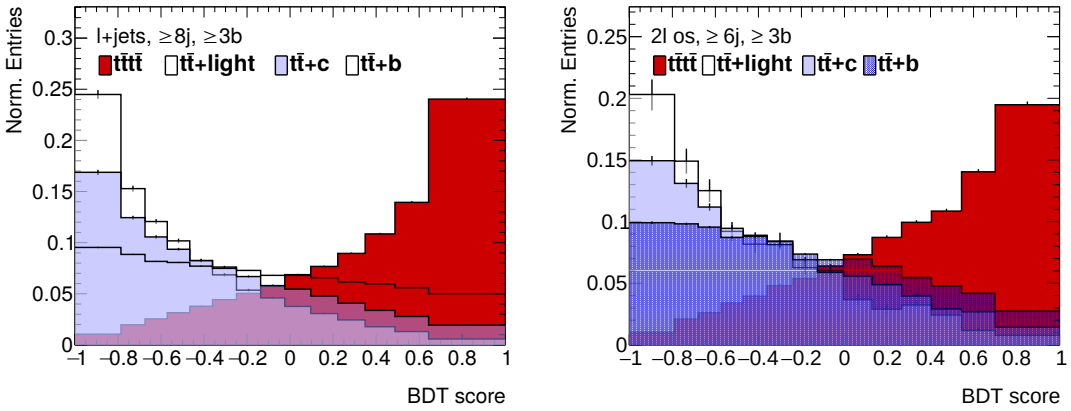


Figure 5.21.: Normalised distributions for $t\bar{t}t\bar{t}$ and the three $t\bar{t}$+jets flavours for the most important BDT input variables and $H_{\mathrm{T}}^{\mathrm{all}}$ in the 1L channel.

Table 5.5.: The separation of $t\bar{t}t\bar{t}$ and $t\bar{t}$+jets inclusive in all jets flavours, based on the plots in Figure 5.21 and 5.22, compared to the importance.

Variable	Separation		Importance	
	1L	OS	1L	OS
BDT score	0.27	0.25	–	–
$H_{\mathrm{T}}^{\mathrm{all}}$	0.14	0.15	8	12
N_j	0.13	0.19	2	1
$\sum_{i=1}^{6}\mathrm{PC}_i$	0.11	0.08	1	2
$\Delta R_{\mathrm{jj}}^{\mathrm{Avg}}$	0.03	0.05	3	3



(a) 1L (b) OS

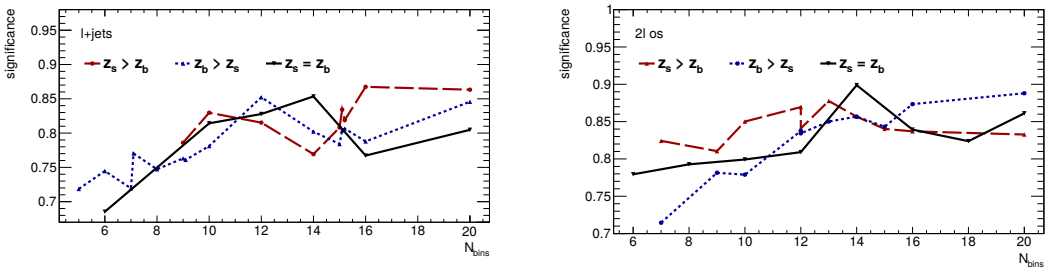
Figure 5.22.: Normalised distributions for $t\bar{t}t\bar{t}$ and the three $t\bar{t}$+jets flavours for the BDT-score in both channels.

5.4.2. Binning Studies

The BDT-score is one of the variables used in the final profile-likelihood fit. Considering again the example of only a single bin, it becomes clear that the binning of the BDT score has an impact on the *significance* to claim the observation of the signal+background (s+b) hypothesis compared to the background only (b-only) hypothesis. In the regime of limited data statistics, too many bins would cause the fit to not be sensitive to the shape differences between signal and background or artificial shapes from statistical fluctuations might get fitted. In particular this is relevant when alternative samples are used to derive systematic uncertainties as also the shape of these systematics are impacted by the binning.
For this reason, the significance was derived for different binnings by comparing s+b and b-only fits using *Asimov* datasets. Such a dataset corresponds to the exact SM prediction based on MC samples and with all corrections applied. To obtain different binnings without empty bins, an algorithm of the TRexFitter software package [138], which is also used to perform the fit, was used. The algorithm has two tunable parameters z_s and z_b where $z_s + z_b = N_{\text{bins}}$ correspond to the total number of bins. The algorithm calculates

$$Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s} \tag{5.15}$$

where $N_x = \sum_{i=1}^{N_{bins}} n_x^i$ is the total number of signal and background events and $n_x = \sum_{i=a}^{b} n_x^i$ the number of entries in the current bin with boundaries a and b. The algorithm stops to increase the bin-width when $Z > 1$ is reached. The significance for a set of choices on z_s and z_b are compared in Figure 5.23. For both channels, a rise depending on the relation between z_s and z_b in significance is observed up to around 12 bins.



(a) 1L (b) OS

Figure 5.23.: Significances to measure signal plus background compared to the background only hypothesis for different binnings calculated by the TransfoD(z_s,z_b) algorifthm.

Afterwards, more fluctuations appear because more choices become possible to obtain a certain amount of bins and the ratio between z_s and z_b impacts the significance. As discussed, choosing too many bins can lead to artificial statistical effects. Therefore, 14 bins with $z_s = z_b = 7$ have been chosen, corresponding to the highest point of the black curve. The change of the significance by other effects, such as the smoothing of systematics to reduce statistical fluctuations of the alternative samples which are discussed in Section 5.5, is in the same order as the difference between the curves in Figure 5.23 for a given number of bins. Therefore, the exact choice of z_s and z_b is not as important as the total number of bins and even for this choice a certain variation gives similar results.

5.5. Systematic Uncertainties

The MC modelling relies on certain choices and implementations of the generator, as discussed in Section 4.2, which leads to *modelling* uncertainties. In addition, reconstruction and identification of each object defined in Section 4 is based on measurements with dedicated uncertainties of the corresponding sub-detector. In this section, the systematic uncertainties in the 1L+OS channel are summarised.

5.5.1. Signal Uncertainties

The uncertainty of the $t\bar{t}t\bar{t}$ ME generator is estimated by varying the renormalisation and factorisations scale simultaneously by a factor of 2 (0.5) to obtain an up (down) variation of the scale applied in the MadGraph5_aMC@NLO prediction. The parton-shower uncertainty is estimated by comparing the nominal MadGraph5_aMC@NLO +Pythia 8 prediction with MadGraph5_aMC@NLO +Herwig 7.

5.5.2. $t\bar{t}$ Modelling Uncertainties

The uncertainties of the $t\bar{t}$ modelling are expected to have the largest impact on the final result as $t\bar{t}$+jets itself is the dominating background. For $t\bar{t}$+jets, the uncertainties are split by the effect of the different jet-flavours on the b-tagging algorithm leading to a total of 23 parameters. In addition, there are 102 nuisance parameters (NPs) representing the uncertainties of the reweighting (RW) in the fit as discussed in Section 5.6.

Matrix-element The uncertainty on the choice of the ME generator is evaluated by comparing the nominal prediction of POWHEG with MADGRAPH5_aMC@NLO where both generators are interfaced with PYTHIA. In addition, the factorisation and renormalisation scale in POWHEG are varied by 2 (0.5) leading to a total of three parameters for each $t\bar{t}$+jets flavour.

Parton-shower The uncertainty on the choice of the parton-shower generator is evaluated by comparing the nominal prediction of POWHEG+PYTHIA 8 with POWHEG +Herwig 7 leading to one parameter for each $t\bar{t}$+jets flavour. In addition, the parton-shower uncertainty for b-jets is split into $t\bar{t}+b$, $t\bar{t}+\geq 2b$ and $t\bar{t}+B$.

Additional Radiation The uncertainty on the additional jets is split in uncertainties for the initial state (ISR), the final state (FSR) and the choice of h_{damp}. For ISR, α_s is varied at the Z-boson mass scale around the tuned values in PYTHIA 8. For FSR, the NPs are obtained by varying the factorisation scale by 2 (0.5) in PYTHIA 8. For h_{damp}, the nominal prediction of $h_{\text{damp}} = 1.5 \times m_{\text{top}}$ is compared with $3.0 \times m_{\text{top}}$. In total, there are three parameters for each $t\bar{t}$+jets flavour.

Flavour Composition The MC-method does not predict the normalisation of $t\bar{t}$ +≥1b and $t\bar{t}$ +≥1c jets. Therefore, a conservatively flat uncertainty, based on recent measurements [139], of ±50% was chosen for these two flavours essentially making the normalisations a free parameter in the fit. In addition, the parton-shower uncertainty for c- and b-jets is split into $t\bar{t}+b/c$, $t\bar{t}+\geq 2b/c$ and $t\bar{t}+B/C$.

Reweighting Factors For each scale factor derived by the MC-method, the statistical uncertainty is propagated as NP. In addition, for the parametrisation of $H_{\text{T}}^{\text{all,red}}$, the three fit parameters are decorrelated. In total, there are 102 parameters for the 1L- and 68 parameters for the OS-channel. Most of the statistical uncertainties are relatively small and are not expected to significantly impact the fit.

5.5.3. Smaller Background Uncertainties

Based on the treatment of the modelling uncertainties, all other backgrounds are grouped in $t\bar{t}+W/Z/H$, single-top, and diboson + minor backgrounds (triboson, VH, $t\bar{t}+WW$, tritop).

$t\bar{t}$ + W/Z/H The choice of the simulation is evaluated by comparing the nominal SHERPA prediction with MADGRAPH5_aMC@NLO for $t\bar{t}+W$, the nominal MADGRAPH5_aMC@NLO prediction with SHERPA for $t\bar{t}+Z$ and the nominal POWHEG prediction with MADGRAPH5_aMC@NLO for $t\bar{t}+H$. For all three samples, the renormalisation and factorisation scales are varied by a factor 2 (0.5). A cross-section uncertainty of 15% is applied on $t\bar{t}+W/Z$ and of 20% on $t\bar{t}+H$. Further normalisation uncertainties for the production with additional jets are derived based on the mismodelling in 2b regions before reweighting. They correspond to 10%, 20% and 30% in the 9 (7), 10 (8) and ≥ 11 (≥ 9) for the 1L (OS) channel.

Single top The t-, s- and W-associated channel are treated similarly by comparing the nominal production of POWHEG with MADGRAPH5_aMC@NLO to obtain a ME uncertainty. For parton-shower, PYTHIA 8 is compared to Herwig 7. For the W-associated channel, the interference schemes of diagram removal and subtraction are compared. Similar to other processes, the renormalisation and factorisation scales are varied by a factor of 2 (0.5). A 30% cross-section uncertainty is applied as well [140, 141].

Minor Processes No dedicated generator comparison is done for the V+jets and Diboson backgrounds. Instead, the cross-section uncertainty is conservatively chosen to be 50%.

5.5.4. Reconstruction Uncertainties

Reconstruction uncertainties are based on the uncertainties of the calibration of each object. Typically, dedicated measurements are performed to extract the calibration and to determine the uncertainties.

Data Taking Conditions As discussed in Section 3.2, the luminosity in ATLAS is measured by LUCID-2. Since all MC samples are scaled by the measured luminosity, the uncertainty of 1.7% for the full Run II dataset propagates to all MC samples. Differences coming from pile-up are taken into account by reweighting MC events to data. The weights are derived from a dedicated simulation during which events are overlayed to match the expected luminosity profile in data (minimum-bias) [142]. The uncertainties for each weight is propagated to all MC samples.

Electrons and Muons The tag-and-probe method is used to derive scale factors in the calibration of the reconstruction, identification, isolation and trigger performance. They cover the differences between data and MC for a variety of object dependent effects such as energy and momentum scales and charge dependencies in the reconstruction for muons. For each of the 7 (13) scale factors for electrons (muons), the uncertainties are propagated to MC simulated events.

Jets Jets are the most complicated objects to reconstruct and hence the total amount of calibration uncertainties is largest compared to other objects. The calibration, and therefore the uncertainties, is a mix of in-situ and MC based correction factors. Each of the 20 jet energy scale (JES) uncertainties has an independent up/down variation while the nine parameters for the jet energy resolution (JER) are one-sided. The JES parameters cover detector-, modelling- and statistical-related uncertainties as well as η, jet-flavour and pile-up dependencies. Furthermore punch-through, single-particle high p_{T} and non-closure of the fast simulation are taken into account. The JER parameters were transformed to be uncorrelated and are therefore referred to as effective NPs.
During the reconstruction of jets, the JVT tags a jet to originate from the primary vertex and a corresponding calibration is applied. An uncertainty of the tagger, mainly coming from pile-up, is extracted from Z+jets and semileptonic $t\bar{t}$ events.

Jet Flavour Tagging B-tagged jets were introduced in Section 4.3.4 and are based on working points of the *MV2c10* output. Each working point uses dedicated efficiency scale factors with different uncertainties depending on the flavour, as miss-tagged jets need a different calibration. For the pseudo-continuous b-tagging, all five working points are used and therefore 5 times more uncertainties have to be considered. In total, there are 45 parameters for b-jets and 20 parameters for c and light jets each. The NPs are obtained after diagonalising the error matrix and are therefore uncorrelated.

Missing Transverse Energy Since $E_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed based on all previous reconstructed objects, the uncertainties of these objects are propagated. In addition, the soft component, which is not associated to any other object, is taken into account by three uncertainties covering the energy scale and the directional resolution.

5.6. Statistical Analysis

The *parameter of interest* (POI) in this analysis that is used used to extract the cross-section is $\mu_{t\bar{t}t\bar{t}}$ where $\mu_{t\bar{t}t\bar{t}} = 1$ corresponds to the SM prediction. The parameter is extracted by fitting the SM MC predictions of all backgrounds and the signal to data after applying all corrections for $t\bar{t}$+jets. The only fully free parameter without prior constraints is $\mu_{t\bar{t}t\bar{t}}$ itself. The TRexFitter software package is used to perform the *Profile Likelihood Fit*. The likelihood to observe n events in a bin b of a region r is given by the Poisson likelihood for signal S and background B by

$$\mathrm{Poiss}(n|\nu) = \mu_{t\bar{t}t\bar{t}}\,(\nu^S + \nu^B) = \prod_r \prod_b \mathrm{Poiss}(n_{br}|\nu_{br} = \mu_{t\bar{t}t\bar{t}}\,\nu_{br}^S + \nu_{br}^B) \tag{5.16}$$

for a certain number of ν expected events. NPs that affect all background models λ equally, referred to as *overall uncertainties*, are considered by an additional Gaussian term. One example is the luminosity. Statistical uncertainties on the MC predictions γ are taken into account by another Poisson term in which the prediction is varied by the uncertainty. The advantage of a profile likelihood fit is to constrain the systematic

Table 5.6.: Fitted variables in each of the jet regions (rows) for the different b-jet multiplicities (columns) and both channels.

Regions (1L/OS)		7j/5j	8j/6j	9j/7j	≥10j/8j
$3b^{70\%}$	$=3b^{70\%}$	$H_{\mathrm{T}}^{\mathrm{all}}$	$H_{\mathrm{T}}^{\mathrm{all}}$	$H_{\mathrm{T}}^{\mathrm{all}}$	BDT-score
	$\geq 4b^{85\%}$	$H_{\mathrm{T}}^{\mathrm{all}}$	$H_{\mathrm{T}}^{\mathrm{all}}$	$H_{\mathrm{T}}^{\mathrm{all}}$	BDT-score
$\geq 4b^{70\%}$		$H_{\mathrm{T}}^{\mathrm{all}}$	$H_{\mathrm{T}}^{\mathrm{all}}$	BDT-score	BDT-score

uncertainties, introduced in the previous section, as additional NPs simultaneously with $\mu_{t\bar{t}t\bar{t}}$. To do so, an additional term $f(\boldsymbol{a}|\boldsymbol{\alpha})$ is introduced in the likelihood to measure $\boldsymbol{a}$ values for NPs $\boldsymbol{\alpha}$, where $\boldsymbol{a}$ and $\boldsymbol{\alpha}$ are vectors of parameters. By combining these considerations, the final likelihood to maximise in the fit becomes

$$P(n, \boldsymbol{a}|\boldsymbol{\varphi}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \underbrace{\prod_r \prod_b \mathrm{Poiss}(n_{br}|\nu_{br}(\alpha))}_{\text{Poisson Likelihood}} \times \underbrace{\mathrm{Gauss}(L_0|\lambda, \Delta_\lambda)}_{\text{Overall uncertainties}} \times \underbrace{\prod_{p\in\alpha} f(a_p|\alpha_p)}_{\text{Model uncertainties}} \times \underbrace{\prod_r \prod_b \mathrm{Poiss}(n_{br}|\gamma_{br}\nu_{br}(\boldsymbol{\alpha}))}_{\text{Statistical uncertainties}} . \tag{5.17}$$

In this analysis, $f(a_p|\alpha_p)$ is assumed to be Gaussian for all NPs and each NP only affects some background models as discussed in the previous section. One example are the NPs for the additionally derived SFs.
As discussed in Section 5.2, the fit regions r require ≥ 7 jets for 1L and ≥ 5 jets where $= 3$ or ≥ 4 jets have to be b-tagged in each jet multiplicity. The 3b regions are further split in $3b^{<4b@85\%}$ and $3b^{\geq 4b@85\%}$ as discussed before. The lower N_j regions are included to constrain and measure the background and the corresponding uncertainties. In principle, the BDT-score should be fitted in all regions to obtain the best possible separation. However, due to the construction of the BDT-score, the modelling uncertainties are, more or less, flat in the BDT-score. Therefore, they cannot easily be distinguished by the fit from a normalisation uncertainty, for example the $t\bar{t}+ \geq 1b$ cross-section uncertainty, which might cause unwanted behaviour such as a bias in the fitted signal strength. To constrain the correct shapes for these uncertainties it is beneficial to use $H_{\mathrm{T}}^{\mathrm{all}}$ as fit variable in the lower regions where the background contribution is measured as further discussed in Section 6.1. The higher multiplicities are then used for the actual signal extraction with the BDT-score. In total, there are 24 regions of which 8 use the BDT-score as fit variable as summarised in Table 5.6.

The pre-fit distributions of $H_{\mathrm{T}}^{\mathrm{all}}$ and the BDT-score are shown in Figure 5.24 for 1L and in Figure 5.25 for OS. Prior to the final fit, the fit model and the background prediction was validated in blinded regions, based on a signal to background ratio of $< 5\%$, as

Table 5.7.: Blinding strategy in the 1L-OS channel.

1L	OS	Blinding Condition
8j*b	6j*b	unblinded
9j*b	7j*b	BDT-score > 0.7
$\geq$10j*b	-	BDT-score > 0.5
–	$\geq$10j*b	BDT-score > 0.2

summarised in Table 5.7 where the naming convention $X\mathrm{j}\,Y\mathrm{b}$ stands for an event with X-jets and Y-bjets. During this validation, the binning in the 6j$\geq$4b region was changed to only two bins, without loss of significance, due to statistical fluctuations in the MC to data agreement. This was observed by the ttTRF and the MC-method and lead to an unsatisfying post-fit agreement in this particular region. If not stated otherwise, unblinded distributions are shown in this document. As introduced in Section 4.3.4, the parton-shower and cross-section uncertainties for $t\bar{t}$+b-jets are further split into $t\bar{t}+b$, $t\bar{t}+B$ and $t\bar{t}+\geq 2b$ sub-categories. All other uncertainties are applied to the three $t\bar{t}$+jets flavours or inclusively to all smaller backgrounds.

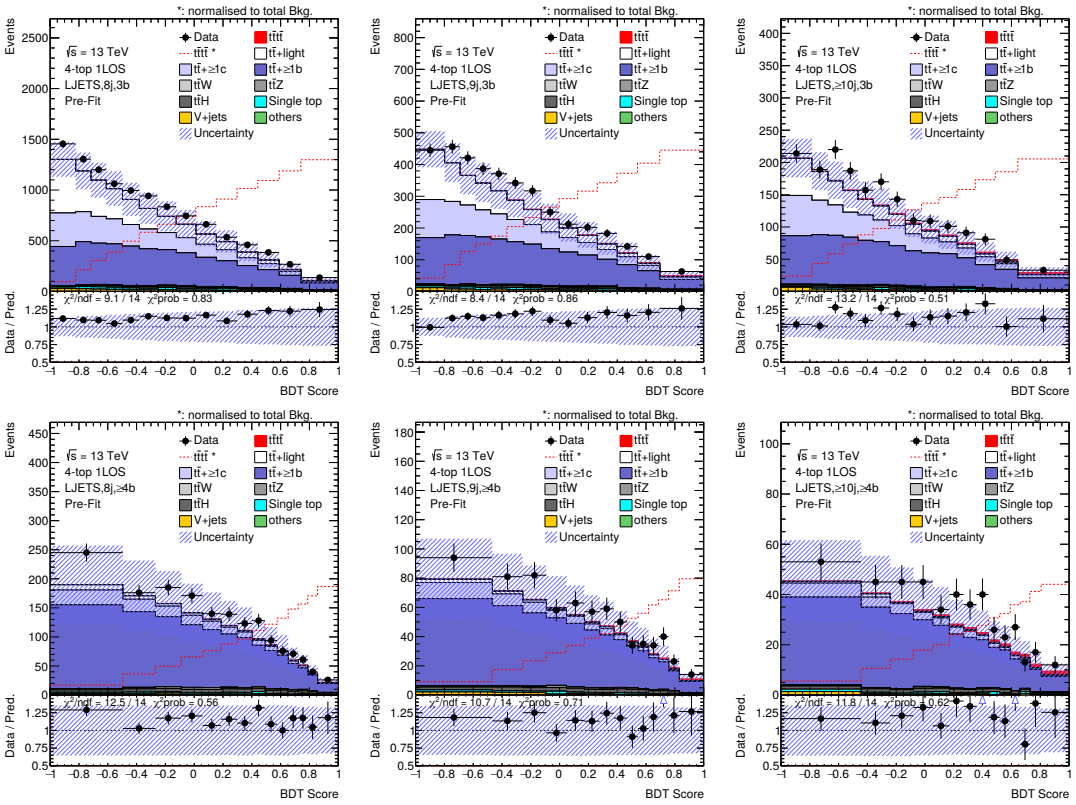


Figure 5.24.: BDT score in the 1L channel for 3b regions (up) and ≥4b regions (down) before fitting for the jet multiplicities indicated in the plots. The uncertainty band includes statistical and systematic uncertainties.

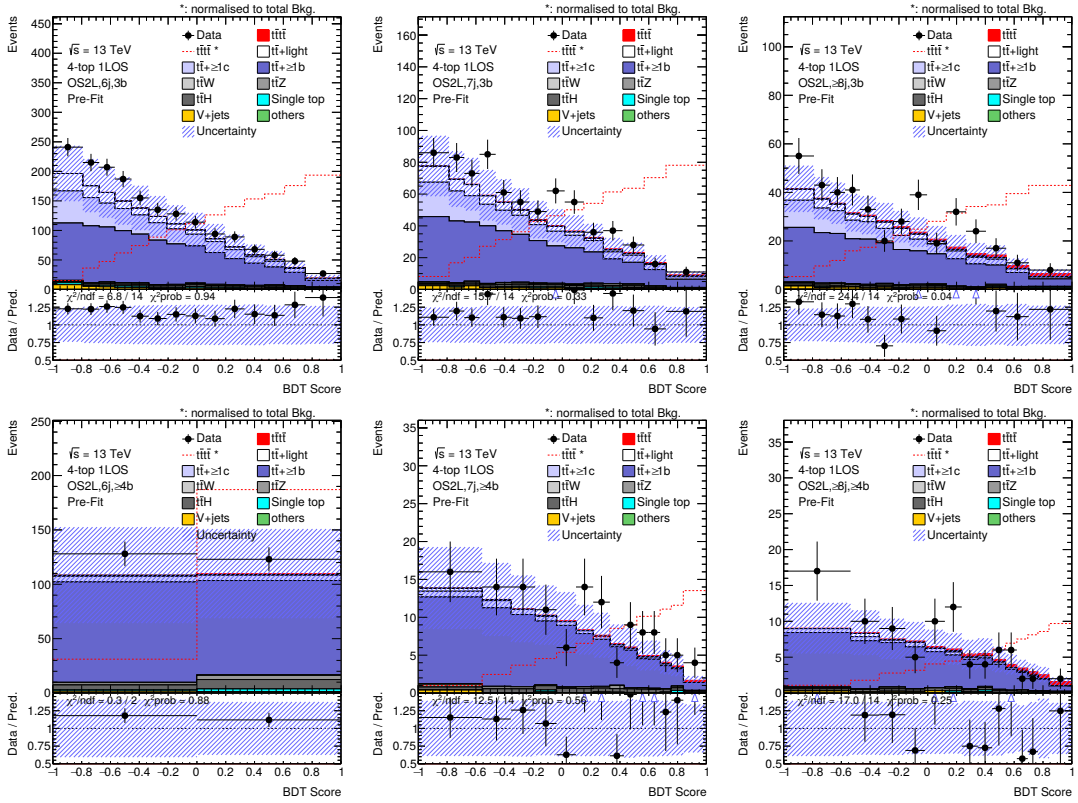


Figure 5.25.: BDT score in the OS channel for 3b regions (up) and ≥4b regions (down) before fitting for the jet multiplicities indicated in the plots. The uncertainties are stat+syst.

5.7. Dilepton Same-Sign and Trilepton Channel

While the focus of this document is the 1L+OS channel, the SS+ML channel is the most sensitive channel since only very few processes exist in the SM to give rise to the required leptons. Therefore, despite the small branching ratio, the signal to background ratio is quite high. This part of the analysis is summarised briefly based on the most recent public result [52].
The dominant physics background originates from $t\bar{t}$ +X where X is an additional boson. Similarly to 1L+OS, the (b)-jet multiplicities are important for signal and background separation. The main difference in the background contributions compared to the 1L+OS channel is the non-negligible instrumental background which, for example, gives rise to additional leptons via conversion of other objects. Since leptons from a direct decay are associated by a corresponding neutrino, the missing transverse energy is more distinct and therefore more important in the SS+ML channel. Due to the different backgrounds, the fit strategy of the profile likelihood fit differs quite substantially as these backgrounds have to be measured in dedicated regions. Still, a BDT is trained to separate signal from the total background and the results of the SS+ML and 1L+OS channel will be combined.

5.7.1. Event Selection

The event selection uses dilepton triggers which, similar to the 1L+OS channel, differ per data taking period as summarised in Table 5.8. Afterwards, an additional event selection, as summarised in Table 5.9, is applied to define dilepton same-charge and multilepton regions.

Table 5.8.: Trigger menu for the HLT in the SSML channel per year and channel. The menu reads as follows: <particle-type>_<p_{T}-cut [GeV]>_<identification-criteria>_<isolation-criteria>_<data-stream of L1>. More details can be found in [132].

	2015	2016-2018
Di-electron	2e12_lhloose_L12EM10VH	2e17_lhloose_nod0 (2016) 2e24_lhloose_nod0
Electron muon	e17_lhloose_mu14 e7_lhmedium_mu24	e7_lhloose_nod0_mu14 e7_lhmedium_nod0_mu24
Di-muon	2mu10 mu18_mu8noL1	2mu14 mu22_mu8noL1

Table 5.9.: Pre-selection in the SS+ML channel applied to data and MC. *Tight* leptons pass the tight cut on the discriminant for the tracking quality and the FCTight cut for isolation as discussed in Section 4.3.1.

	Dilepton	Multilepton
Objects	$e^{\pm}e^{\pm}$, $\mu^{\pm}\mu^{\pm}$, $e^{\pm}\mu^{\pm}$	$N_e + N_\mu \geq 3$
Lepton p_{T}	$\geq$ 28 GeV	$\geq$ 28 GeV
Quality	$\geq$3 *loose* leptons =2 *tight* leptons	$\geq$3 *tight* leptons
N_{jets}	$\geq$ 1	
$N_{\text{b-jets}}$	$\geq$ 1 (MV2c10 77%)	
Jet p_{T}	$\geq$ 25 GeV	
M_{ll}	$\geq$ 15 GeV $\lvert M_{ll} - m_Z \rvert > 10$ GeV	

5.7.2. Signal and Background Separation

A BDT is used to separate signal and the total background. The signal region, defined by having more than 6 (2) (b)-jets and $H_{\mathrm{T}} > 500$ GeV, was used for training. The LO $t\bar{t}t\bar{t}$ sample was used while the NLO sample is used in the fit. Via a grid scan, the optimal hyperparameters were determined, for example to use 600 trees for boosting, to increase the separation while keeping the overtraining small. The 12 input variables were chosen based on a good modelling and to increase the separation. The actual variables are mostly different from the 1L+OS channel as summarised in Table 5.10. In particular, $E_{\mathrm{T}}^{\mathrm{miss}}$ is more relevant, as for $t\bar{t}t\bar{t}$ it mainly arises from real neutrinos while for the instrumental backgrounds, which give rise to fake or misidentified leptons, no missing energy is expected.

5.7.3. Background Estimation

The analysis uses the 77% working point to identify b-jets. The fit regions are defined based on the main backgrounds which can be grouped into three categories.

- Contributions from the production of $t\bar{t}$ with an additional boson (W, Z, H) or the production of multiple bosons. This type of background requires the bosons to decay leptonically and is evaluated based on MC predictions. The normalisation of the $t\bar{t}W$ process is left as a free parameter while appropriate normalisation uncertainties are applied for the others.

Table 5.10.: Input variables for the BDT in the SSML channel with the importance obtained by the change of separation when not including a given variable.

Variable	Importance	Description
Jet activity		
$p_{\mathrm{T}}^{1,jet}$	9	leading jet p_{T}
$p_{\mathrm{T}}^{2,jet}$	12	second leading jet p_{T}
$p_{\mathrm{T}}^{6,jet}$	5	6th leading jet p_{T}
$H_{\mathrm{T}}^{no-leading}$	7	p_{T} sum of all objects but the leading jet
b-tagging information		
$\sum w_{\mathrm{MV2C10}}$	1	sum of the PC b-tagging score over all jets
$p_{\mathrm{T}}^{1,MV2C10}$	11	p_{T} of the highest b-tagged jet (77%WP)
Angular Variables		
$\Delta R_{\mathrm{ll}}^{\mathrm{min}}$	4	closest distance between any lepton pair
$\Delta R_{\mathrm{lb}}^{\mathrm{max}}$	6	maximum distance between a lepton and b-tagged jets (77%WP)
$\Delta R_{\mathrm{jb}}^{\mathrm{min}}$	10	minimum distance between jets and b-tagged jets (77%WP)
$\sum \Delta R_{\mathrm{ll}}^{\mathrm{Avg}}$	8	sum of distances between leading and sub-leading leptons including 3rd leading leptons for the multilepton channel
Lepton and event variables		
$E_{\mathrm{T}}^{\mathrm{miss}}$	3	Missing transverse energy
$p_{\mathrm{T}}^{0,l}$	2	p_{T} of leading lepton

- Events from the OS channel can be observed in the SS channel if the charge of one of the leptons is not reconstructed correctly. In addition, due to the radiation of photons from electrons via bremsstrahlung where the photon can decay into an e^+e^--pair, the wrong electron can be selected. The corresponding charge mis-identification (QmisID) is estimated by a data-driven approach and is only relevant for the SS channel. The probability is estimated by using a sample of dielectron events with an invariant mass close to the Z-mass. The same selection as for the actual analysis is applied but inverting the requirement of the lepton charges to be opposite. Comparing this number to the total number of events gives the mis-identification rate. An event weight is then derived which is used to predict the total mis-identification background. For muons, the rate for bremsstrahlung and therefore charge mis-identification is negligible.

- Due to heavy-flavour (HF) decays or the conversion from photons in the detector material (Mat CO and γ^*), leptons can be produced in the decay-chain. Additionally, quark and gluon jets can under certain conditions be reconstructed as lepton which is referred to as fakes. This background is estimated by the so-called template fit method. It uses pre-defined templates with a fixed shape but a free-floating normalisation parameter for each of the background categories.

Due to the different background sources, in addition to the BDT-score which is only fitted in one region, four more regions, each of them sensitive to one of the background contributions, are used in the fit. While for the electron and muon fake regions, only the total number of events is fit, the $t\bar{t}$ +W and the conversion control regions use the distributions of the sum of the lepton p_{T} and the invariant mass M_{ll} of two leptons at the primary vertex, respectively. Furthermore, a region to monitor the $t\bar{t}$ +Z background was defined. The exact region definitions are based on cuts on H_{T}, (b-)jet multiplicities, lepton flavours and pseudo-rapidty. They are summarised in Table 5.11 and the corresponding distributions are shown in Figure 5.26.

Table 5.11.: Definition of the fit regions in the SSML channel. The SR was defined to measure the $t\bar{t}t\bar{t}$ signal strength. Regions with CR in the name were defined to measure one particular background. The VR $t\bar{t}$ + Z region is used to check the validity of the fit outcome for the $t\bar{t}$ + Z background. HF refers to the electron/muon originating from a heavy flavour jet. M_{ll}^{PV} is the invariant mass of two leptons at the primary vertex.

Region	Motivation	Fit Variable	Channel	Selection		
				N_b	N_j	other
SR	$t\bar{t}t\bar{t}$	BDT-score	SSML	≥ 2	≥ 6	$H_{\mathrm{T}} > 500\,\mathrm{GeV}$
CR$t\bar{t}$CO	photon conv. material conv.	M_{ll}^{PV}	$e^{\pm}e^{\pm}\,\|\|\,e^{\pm}\mu^{\pm}$	≥ 1	≥ 4 & ≤ 6	$200\,\mathrm{GeV} < H_{\mathrm{T}} < 500\,\mathrm{GeV}$ $0 < M_{ee}^{PV} < 0.1$
CR1b3Le	HF electrons	Yield	$eee\,\|\|\,ee\mu$	$= 1$	–	$100\,\mathrm{GeV} < H_{\mathrm{T}} < 250\,\mathrm{GeV}$
CR1b3Lm	HF muons	Yield	$e\mu\mu\,\|\|\,\mu\mu\mu$	$= 1$	–	$100\,\mathrm{GeV} < H_{\mathrm{T}} < 250\,\mathrm{GeV}$
CR$t\bar{t}$ W2L	$t\bar{t}$ +W	$\sum_{\mathrm{leptons}} p_{\mathrm{T}}$	$e^{\pm}\mu^{\pm}\,\|\|\,\mu^{\pm}\mu^{\pm}$	≥ 2	≥ 4	$M_{ee} < 0$ or $M_{ee} > 0.1$ $\|\eta\| < 1.5$ for $N_b = 2$: $H_{\mathrm{T}} > 500\,\mathrm{GeV}$ or $N_j < 6$ for $N_b \geq 3$: $H_{\mathrm{T}} < 500\,\mathrm{GeV}$
VR $t\bar{t}$ +Z	$t\bar{t}$ +Z	N/A	SSML	≥ 2	≥ 6	$H_{\mathrm{T}} > 500\,\mathrm{GeV}$, BDT$< 0.0$

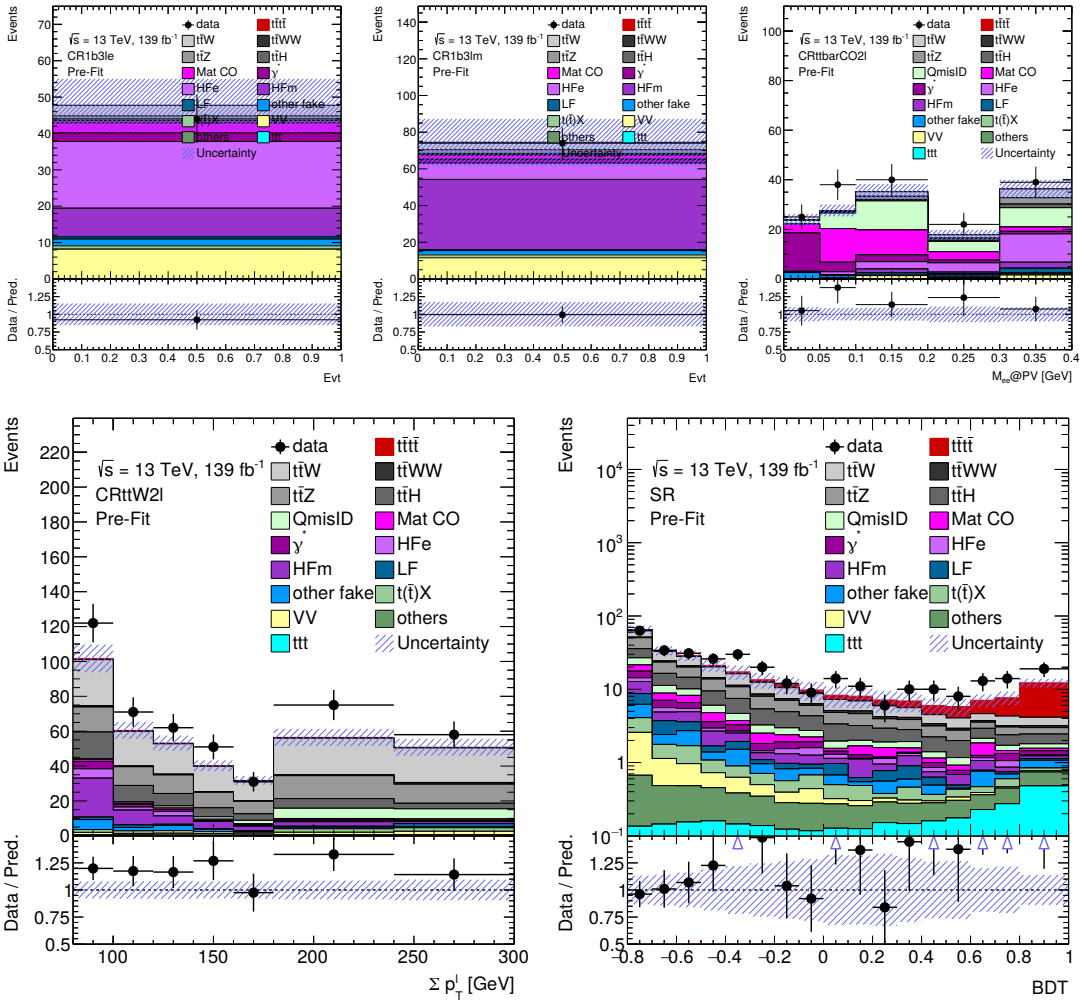


Figure 5.26.: Fit regions in the SS+ML channel before performing the fit. Besides the BDT-score (bottom right) four more regions are used to constraint certain backgrounds. The uncertainty band includes statistical and systematic uncertainties.

5.7.4. Systematic Uncertainties

The systematic uncertainties for the luminosity, charged leptons, jets, pile-up, flavour tagging and the missing transverse energy, discussed in Section 5.5, are similar in both channels. The additional uncertainties are listed below.

$t\bar{t} + X$ Modelling Renormalisation and factorisation scale are varied in the nominal predictions for $t\bar{t}$ +Z/W/H by 2 (0.5) to obtain an up (down) variation, similar to the 1L+OS channel. For $t\bar{t}$ +W/Z, an alternative simulation with SHERPA 2.2.5 at QCD NLO is compared to the nominal prediction. For $t\bar{t}$ +H, POWHEG is compared to MADGRAPH5_aMC@NLO while using the same PDF set, parton-shower and tuning parameters. A 50% uncertainty is applied to events with three or at least four b-jets on truth level. An additional 125% (300%) uncertainty has been added for $t\bar{t}+W$ in regions with 7 (more than 8) jets to take the observed difference in validation regions between data and MC into account. In addition, the $t\bar{t}+W$ normalisation is a free parameter in the fit. For $t\bar{t}+Z/H$, a 1% uncertainty for the PDFs is applied.

Cross-Section Cross-section uncertainties for processes that also occur in the 1L+OS channel are similar. Furthermore, a 30% uncertainty for regions with exactly three or more than four b-jets to estimate the $g \to b\bar{b}$ splitting is added [139]. The production of three top quarks (ttt) has a cross-section uncertainty of 100% as it is an unmeasured process and an uncertainty for ttt events with four b-jets of 50% is added as one additional NP. A 20% theory uncertainty for the $t\bar{t}t\bar{t}$ process is included which does affect the measured cross-section but the expected significance. For $t\bar{t}+Z/H$, a 15/20% uncertainty is applied.

Charge Mis-Identification The charge mis-identification rate is estimated from data in this analysis. The corresponding uncertainties arise from the measured rates. In total, three NPs are used to estimate the statistical uncertainty of the likelihood fit to extract the rate, the difference between the fit and the true rate from MC and the selection of the Z-mass window.

Photon Conversion For each of the sources of photon conversion (material and virtual), a free normalisation factor is included. The template fit method is used which only predicts the shape. $Z \to \mu\mu + jets + photon$ events predicted by POWHEG+PYTHIA are used to derive the normalisation by matching the shape to data. The extra jets are required to be closer to the expected $t\bar{t}t\bar{t}$ signature in which at least two b-jets are expected. An uncertainty on the shape is estimated by extracting the remaining difference between data and MC in control regions.

Non-prompt Leptons and Heavy-Flavour Hadron Decays Similarly to photon conversion, two normalisation factors, one for non-prompt leptons and one for heavy-flavour

decays, are included in the fit. The shapes are accessed by comparing the MC prediction with data events, where the events are obtained by subtracting all other MC contributions from data. The shape is derived for electrons and muons separately in all regions where the requirements for lepton isolation and identification have been loosened to increase statistics.

Light-jets Fake Leptons Fakes can originate not only from non-prompt leptons or heavy-flavour decays but can also come from decays in flight in light jets. However, this process is not well known and therefore, following recent studies [143], a 100% uncertainty is assigned.

CHAPTER 6

Results

In this Chapter, the results of the profile likelihood fit with the MC based method are presented. Prior to the final results, fits to pseudodata with a partial fit model are performed to study the fit-setup in detail and to check the robustness of the fit with respect to a particular model. Next, the results of the SS+ML channel are summarised and a combination of the results is performed. Finally, the results are compared to the corresponding CMS measurements before concluding and giving an outlook in the next Chapter.

6.1. Results from Fits to Pseudodata in the 1L+OS Channel

Using *pseudodatasets* has the advantage that the sample composition and therefore the ideal fit outcome is exactly known. This allows to study a possible bias of the fit. In a fit to data, the exact distributions and uncertainties have to be extracted and unknown effects can be present. Therefore, pseudodatasets allow for a better judgement of the fit-setup and help to understand the pulls, constraints and correlations between uncertainties. Furthermore, no blinding cut needs to be applied and the fit-setup can be validated in the full phase-space. This is advantageous since the blinded regions with real data, by definition, do not contain many signal events as they are designed to study the background in the control regions only.
In total, three pseudodatasets have been generated by replacing the nominal $t\bar{t}$ prediction from POWHEG + PYTHIA with the alternative prediction ($\mathrm{MC}^{t\bar{t},\mathrm{alt}}$) of

- POWHEG + HERWIG (parton-shower uncertainty)
- MADGRAPH5_aMC@NLO + PYTHIA (matrix-element uncertainty)
- ttTRF (alternative data prediction).

Each dataset is generated according to

$$\mathrm{Data}_{\mathrm{pseudo}} = \mathrm{MC}^{\text{non-}t\bar{t}} + \mathrm{MC}^{t\bar{t},\mathrm{alt}} \tag{6.1}$$

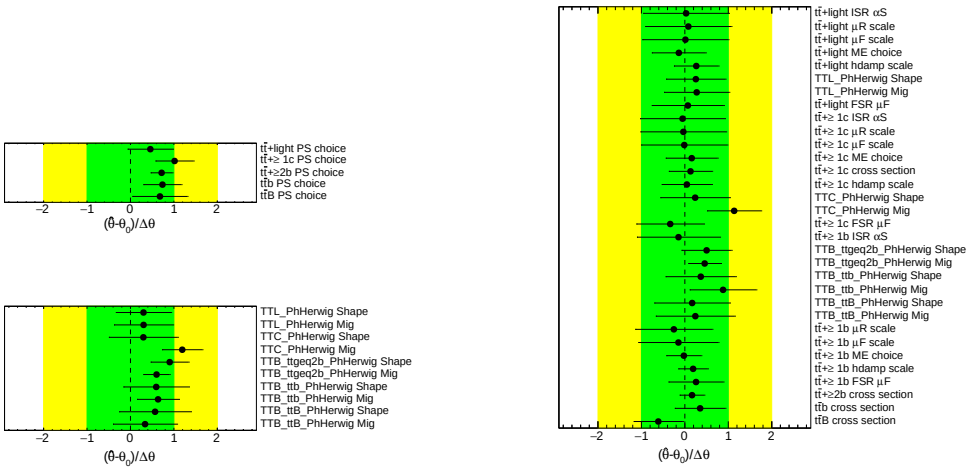
where $\mathrm{MC}^{\text{non-}t\bar{t}}$ contains the smaller backgrounds and the SM prediction of the $t\bar{t}t\bar{t}$ signal. As discussed in Section 5.5, modelling uncertainties, in particular the parton-shower, are expected to have the largest impact on the final result which motivates the first two pseudodatasets. By using the ttTRF prediction after fitting it to data in blinded regions (post-fit), the full unblinded phase-space is studied with an alternative dataset while assuring a consistent behaviour to the blinded regions.
As mentioned in Section 5.6, $H_{\mathrm{T}}^{\mathrm{all}}$ is chosen in regions with three b-jets as discriminant. This choice is motivated by the fit to Powheg + Herwig pseudodata as presented in Section 6.1.1.
Finally, each pseudodataset is varied within the uncertainties of the MC prediction per bin to generate additional *toy experiments*. This allows for estimating how large the uncertainty of the prediction itself on the signal strength is and if a bias occurs in the fit.

6.1.1. Fit to Powheg + Herwig Pseudodata

In this setup, the parton-shower (PS) uncertainty and its correlations to other uncertainties are studied in detail. To better understand the output of the fit, the difference between the value that was estimated by the fit ($\hat{\Theta}$) and the input value (Θ), normalized to the uncertainty ($\Delta\Theta$) of the input value, is compared as shown for example in Figure 6.1 for three sets of uncertainties that were used in three different fits. The uncertainty of the estimated value is given by the black line around each data point. In the scenario when only the parton-shower uncertainty is included, by definition, the difference between the nominal prediction and the pseudodataset corresponds exactly to the PS uncertainty. Therefore, as shown in the upper plot in Figure 6.1 (a), the *pulls* on the PS uncertainties are close to $1\,\sigma$ (green area) away from the nominal prediction which is centred around 0. Furthermore, the uncertainties are *constrained* to less than $1\ \sigma$ (black error bars) since the fit is able to determine the uncertainty more precisely than the uncertainty that is provided by the comparison of the nominal and the alternative samples. From the difference in the constraints it is concluded that the fit is most sensitive to the $t\bar{t}+ \geq 2b$ component of the PS uncertainty.
In the lower part of Figure 6.1 (a), the PS uncertainties have been decorrelated by splitting each uncertainty in migration (Mig), corresponding to the difference in yields between different regions, and shape effects, corresponding to the slope of each uncertainty for a given variable, in each region. This allows to alter the normalisation while keeping the shape constant and vice versa for each component and hence adds more degrees of freedom to the fit. However, too many degrees of freedom can lead to overfitting which means that smaller parameters get adjusted to describe statistical fluctuations which can lead to a bias of the parameter of interest (POI) in the final fit to data. By studying this behaviour, it is found that the decorrelation in migration and shape is only

required for the PS uncertainty. In general, the pulls are compatible to the upper case where shape and migration are correlated. The differences in the constraints indicate that the fit is more sensitive to the normalisation than to shape effects. For $t\bar{t}+ \geq 1c$, the difference between migration and shape indicates that mostly the migration is relevant for this component.
In the next setup, shown in Figure 6.1 (b), all other modelling systematics have been included in the fit. It is observed that the constraints slightly decrease while the pulls stay similar but not identical. This is explained by the correlations among the systematics which are given in the correlation matrix in Figure 6.2. For example, as indicated by the red-dashed line, by increasing the $t\bar{t}+ \geq 1c$ cross section a similar effect, corresponding to a correlation of 45%, can be achieved as by altering the $t\bar{t}+ \geq 2b$ migration. What is not shown in Figure 6.1 are the absolute values of the pulls which have to be taken from the corresponding configuration. For example, the pull on the $t\bar{t}+ \geq 1c$ cross section corresponds to 0.13. However, the initial value of the cross-section uncertainty is approximately 40% (50% before applying additional SFs). Therefore, the absolute pull is 0.13×0.4=5% of the nominal prediction. However, the absolute values are typically not as relevant as the relative pulls and constraints to judge the stability and outcome of the fit.
In conclusion, the PS systematic uncertainties cover the difference to the pseudodataset as expected. However, due to correlations they are not observed as 1σ pull when systematic uncertainties are decorrelated or more uncertainties are added to the fit. Therefore, correlations among systematics are monitored in detail while studying the setup as these might hide or introduce additional pulls and constraints that alter the fit outcome.

An important information that is gained from the correlations is that the fit is, if at all, only slightly biased towards larger values of the signal strength. Since exactly the SM prediction is used in the fit, ideally a signal strength of $\mu_{t\bar{t}t\bar{t}} = 1$ is obtained. However, due to statistical uncertainties, certain values around unity are acceptable. The measured signal strengths for the three scenarios in Figure 6.1 are shown in Table 6.1. While all values are reasonably close to unity within uncertainties, the absolute value of $\mu_{t\bar{t}t\bar{t}}$ increases by 0.3 when more modelling uncertainties are included. This indicates that, due to the correlations to the systematic uncertainties, the signal strength is increased to compensate effects that occur only by adding more uncertainties to the fit. To further study the bias, 500 toy experiments (TE) are performed in the setup with all modelling uncertainties. For each TE, the Powheg + Herwig pseudodata is varied within its uncertainties. The obtained distributions for the signal strength (left) and its uncertainty (right) are shown in Figure 6.3. They are parametrized by a Gaussian to extract the mean value and the width. The mean of $\mu_{t\bar{t}t\bar{t}} = 1.08$ is $1\,\sigma$ away from an unbiased result when only the statistical uncertainties of the pseudodataset are considered. However, considering the full uncertainties of about $\Delta\mu_{t\bar{t}t\bar{t}} = 0.77\,\sigma$, which is in good approximation constant for all TEs, the bias is considered small in this setup.
From these studies the following conclusions are drawn for the fit to data. It is important to measure the background distributions and related uncertainties independent of



(a) PS uncertainty (b) Modelling uncertainties

Figure 6.1.: The estimated ($\hat{\theta}$) value for the fit to the POWHEG + HERWIG pseudodataset compared to the input value (θ) for each nuisance parameter. The green (yellow) area corresponds to a 1 (2) σ deviation from θ. Figure (a) compares the pulls and constraints when only the parton-shower (PS) uncertainties are used in the fit while in the fit for figure (b) other modelling uncertainties have been included as well.

the signal strength to reduce correlations and hence a possible bias. This is achieved by decorrelating the shape and migration of the PS uncertainty and by splitting the $t\bar{t}$ $+\geq 1b$ jet component into sub-categories. By choosing fit regions with only few signal events to measure background contributions and their uncertainties, the signal can be extracted in parallel in regions with a better separation. Furthermore, it is motivated to use $H_{\mathrm{T}}^{\mathrm{all}}$ as fit variable in the lower (b-)jet multiplicities: the PS uncertainties appear rather flat in the BDT-score while they show a trend in $H_{\mathrm{T}}^{\mathrm{all}}$ as representatively shown in Figure 6.4 where the difference in event yields (y-axis) for the BDT-score and $H_{\mathrm{T}}^{\mathrm{all}}$ are compared. The range of the variables (x-axis) has been normalized in both cases to allow for a better comparison. By treating the PS uncertainties correlated among all regions, assuming that there is no additional underlying effect related to the jet multiplicity, the shape effect in $H_{\mathrm{T}}^{\mathrm{all}}$ is measured in lower jet multiplicities and is then propagated to the signal enriched regions allowing for a better estimate of the shape of the parton-shower uncertainty.

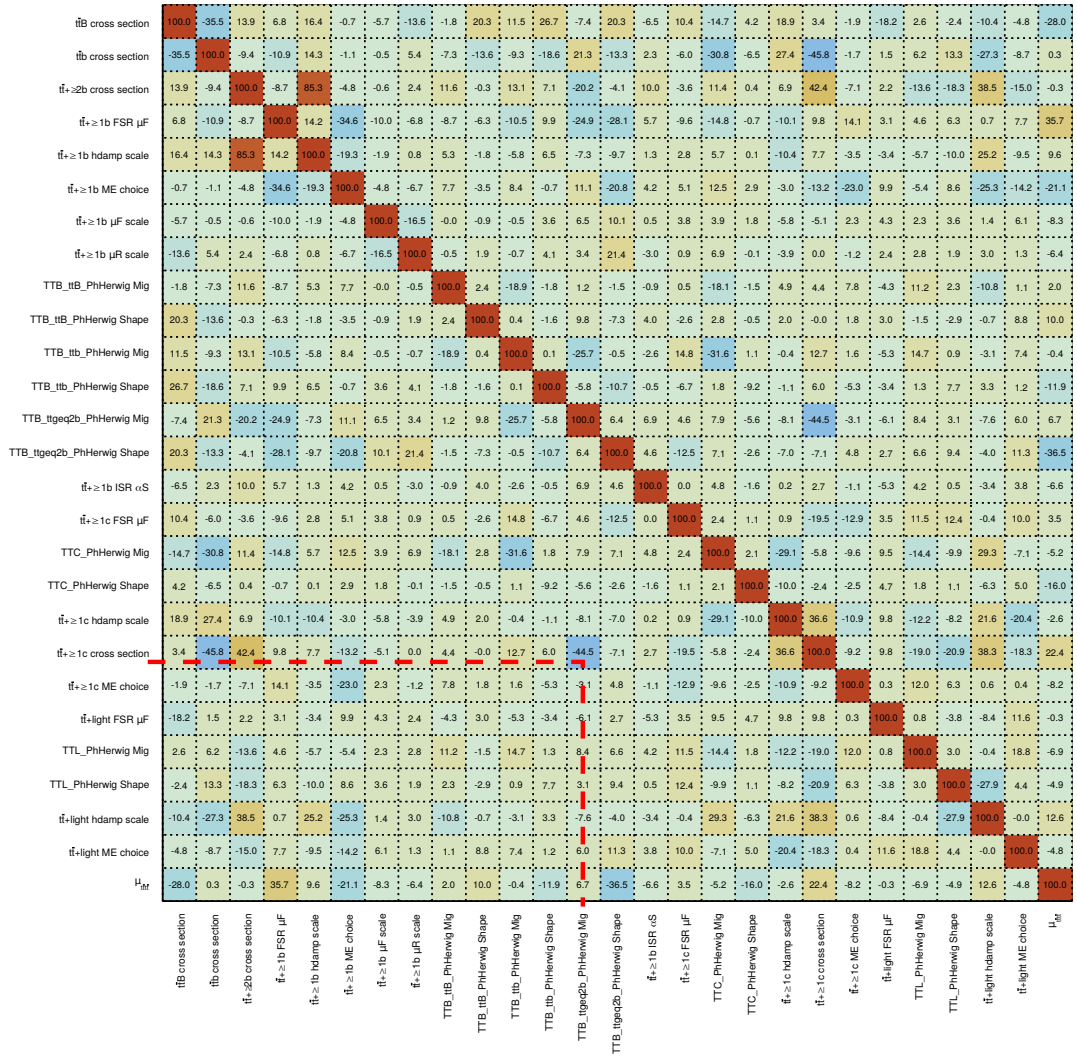


Figure 6.2.: Correlation matrix for the fit to the POWHEG + HERWIG pseudodataset with all modelling systematics. The red-dashed line indicates one of the highest correlations between the $t\bar{t}+ \geq 1c$ cross section and the $t\bar{t}+ \geq 2b$ migration.

Table 6.1.: Signal strength μ and the total fit uncertainty for three fits to POWHEG + HERWIG.

Systematics	$\mu_{t\bar{t}t\bar{t}}$	$\Delta\mu_{t\bar{t}t\bar{t}}$
PS corr.	0.87	0.53
PS decorr.	0.79	0.60
PS+modelling	1.09	0.77

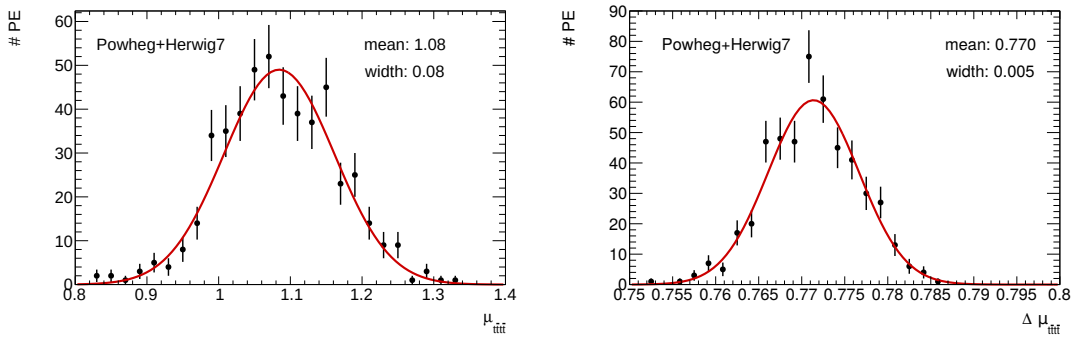


Figure 6.3.: Distribution of the mean (left) and uncertainty (right) of 500 toy-experiments for the PS+modelling setup. The uncertainties are statistical.

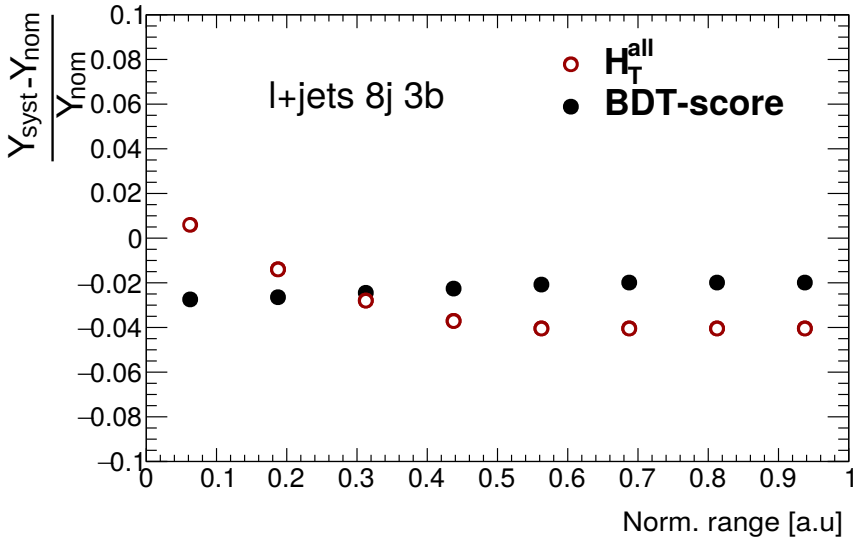
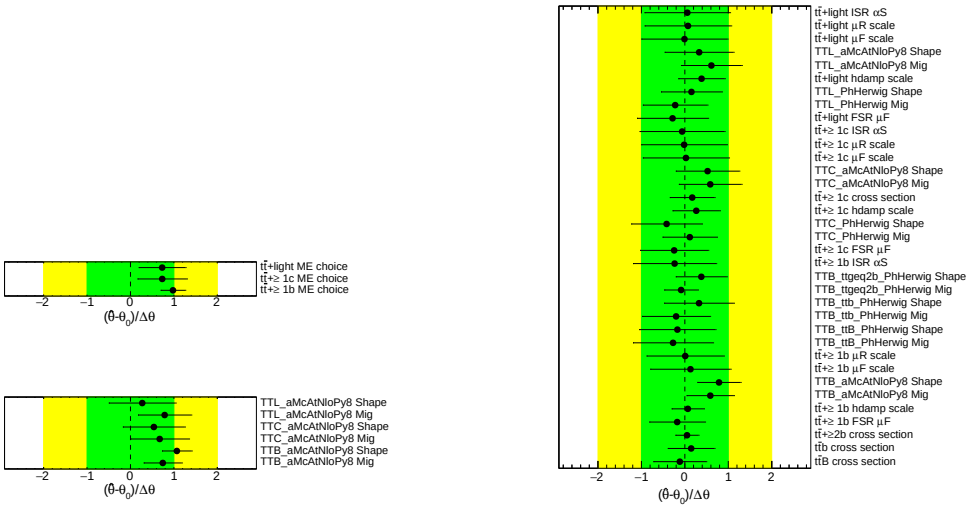


Figure 6.4.: Comparison of the Yields (Y) for the PS systematic (syst) with the nominal (nom) sample for the distribution of the BDT-score and the $H_{\mathrm{T}}^{\mathrm{all}}$ variable. Both variables are normalised to the range [0,1] to allow for a better comparison where the origin value of the BDT-score ($H_{\mathrm{T}}^{\mathrm{all}}$) was [-1,1] ([0-2000]).

6.1.2. Fit to MadGraph5_aMC@NLO + Pythia Pseudodata

In this setup, the matrix element uncertainty and the correlations to other uncertainties are studied in detail. The pseudodataset reflects the difference between the nominal prediction and the the matrix element uncertainty as the nominal $t\bar{t}$ prediction was replaced by the MadGraph5_aMC@NLO + Pythia prediction. In general, similar conclusions as for the fit to the Powheg + Herwig pseudodataset can be made for the fit setup as shown in Figure 6.5 which again shows the difference $\hat{\Theta}$ and Θ with the respective uncertainties. Figure 6.5 (a) shows the fit when only the ME uncertainties are used. The pulls in the upper plots correspond to a $1\,\sigma$ deviation, as expected. No clear tendency between shape and migration effects, shown in the lower plot, is observed. The largest difference occurs for the $t\bar{t}$+light component. When adding more uncertainties, shown in Figure 6.5 (b), correlations among the systematics reduce the pulls and constraints. A more complete overview of the correlations is given in Figure D.1. The largest correlations occur between the cross-section, parton-shower and the signal strength, similar to the correlations shown before in Figure 6.2. While other systematics are not pulled significantly, some of the PS and cross-section uncertainties are constrained. The $t\bar{t}$ +light/c/b cross-section uncertainties were introduced to cover the normalisation difference which is not treated by the MC-method. Therefore, it is expected that the fit setup is sensitive to the difference in flavour normalisation and the overall uncertainty of approximately 40% is reduced. Another point worth mentioning is the constraint of the $t\bar{t} \geq$ 2b PS uncertainty. It exhibits high correlations to the signal strength (40%) and the shape effect of the $t\bar{t}+ \geq$ 1b ME uncertainty (29%). This is in agreement with the fit to the Powheg + Herwig pseudodataset. However, no significant effect of the parton-shower uncertainty is expected as the matrix-element uncertainty should cover all effects. This underlines the importance of the PS uncertainty compared to the ME uncertainty.
The measured signal strengths for the three scenarios are shown in Table 6.2. Similar to the previous fit, 500 TEs are performed where for each TE the MadGraph5_aMC@NLO + Pythia pseudodataset is varied within its uncertainties. The obtained distributions and the Gaussian parametrisations are shown in Figure 6.6. Similar to the Powheg + Herwig TEs, the mean value of 1.09 is in agreement with an unbiased result when considering the width of the distribution which is interpreted as uncertainty of the prediction on the measured signal strength. The full uncertainties, shown on the right in Figure 6.6, are slightly larger than for the previous setup. During the studies, it became clear that the matrix element uncertainty is less important for the fit than the parton-shower uncertainty which is, for example, confirmed by the high correlations of the $t\bar{t} \geq$ 2b PS uncertainty of the signal strength. No clear effect when decorrelating the ME uncertainties has been observed. Therefore, to reduce the chance of overfitting, the matrix element uncertainty is not decorrelated in shape and migration nor $t\bar{t}+ \geq$ 1b sub-components for the final fit to data.



(a) ME uncertainty (b) ME+modelling uncertainties

Figure 6.5.: The estimated ($\hat{\theta}$) value for the fit to the MadGraph5_aMC@NLO + Pythia pseudodataset compared to the input value (θ) for each nuisance parameter. The green (yellow) area corresponds to a 1 (2) σ deviation from θ. Figure (a) compares the pulls and constraints when only the matrix element uncertainties are used in the fit while in the fit for figure (b) other modelling uncertainties have been included as well.

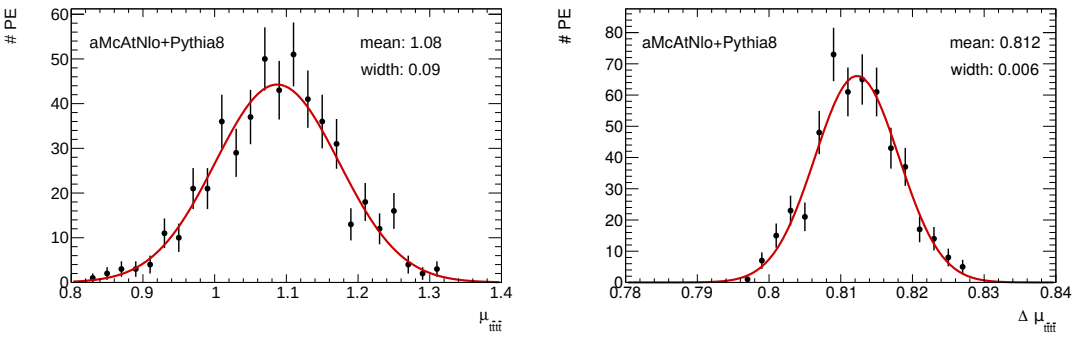


Figure 6.6.: Distribution of the mean (left) and uncertainty (right) of 500 pseudo-experiments for the ME+modelling setup. The uncertainties are statistical.

Table 6.2.: Signal strength $\mu_{t\bar{t}t\bar{t}}$ and its uncertainty for three fits to MadGraph5_aMC@NLO + Pythia.

Systematics	$\mu_{t\bar{t}t\bar{t}}$	$\Delta\mu_{t\bar{t}t\bar{t}}$
ME corr.	1.08	0.50
ME decorr.	1.12	0.50
ME+modelling	1.09	0.81

6.1.3. Fit to Pseudodata Generated by ttTRF

The ttTRF method provides an alternative and independent prediction of the $t\bar{t}$+jets background based on data. As presented in Section 5.3.2, the agreement with data is reasonable in all regions before performing the fit. To study the fit setup and a possible bias, the ttTRF data-prediction is fitted to data, including all systematics, as shown in Appendix C, while applying the blinding cuts presented in Table 5.7. While the regions were chosen to be the same, the systematic uncertainties do not exactly correspond to the ones presented for the MC-based method as before because the ttTRF method does not predict individual jet flavours. Hence, the $t\bar{t}$+jets background is not split in flavours and consequently the PS uncertainty is not split in sub-categories. Nevertheless, the setup corresponds to a valid approach to measure the $t\bar{t}t\bar{t}$ signal and was studied intensively. The fitted values of the NPs in the blinded regions are applied to the unblinded distributions to obtain a prediction in the full phase-space. From this prediction, the $t\bar{t}$ component is summed with the non-$t\bar{t}$ backgrounds to obtain the pseudodataset. In contrary to the PS and the ME uncertainties, a fit to this pseudodataset is performed including *all* uncertainties of the MC-based method. By using the $t\bar{t}$ post-fit prediction, the agreement with data is improved, as shown in Figure 6.7 and in Appendix C.

The advantage, compared to the previous pseudodatasets, is that the outcome is expected to be closer to real data and therefore this test is more realistic. However, the exact pulls are a priori not clear which makes it more difficult to judge if a discrepancy originates from the fit itself or from the ttTRF prediction.

Figure 6.8 (black) shows the NPs for the generator uncertainties. The highest correlation occur between the $t\bar{t}$ +b and $t\bar{t}$ +B cross-section (64%) and between the signal strength an the $t\bar{t}$ +$\geq$ 2b parton-shower shape (40%), the full correlation is given in Figure D.2. In general, the pulls are within $1\,\sigma$ which confirms that the fit gives a reliable result. As before, the strongest constraints of around 25% occur for the $t\bar{t} \geq$ 2b component for the cross-section and parton-shower uncertainties. A single deviation is observed for the $t\bar{t}$ +Z generator systematic which is pulled to $1.6\,\sigma$ of its original prediction, it is correlated to the signal strength by -27.5%. The fitted signal strength is

$$\mu_{t\bar{t}t\bar{t}} = 1.70 \pm 1.10 \tag{6.2}$$

which, within the overall uncertainties, corresponds to an unbiased result. To estimate the impact of the $t\bar{t}$ +Z generator systematic, this particular uncertainty is fixed ad-hoc

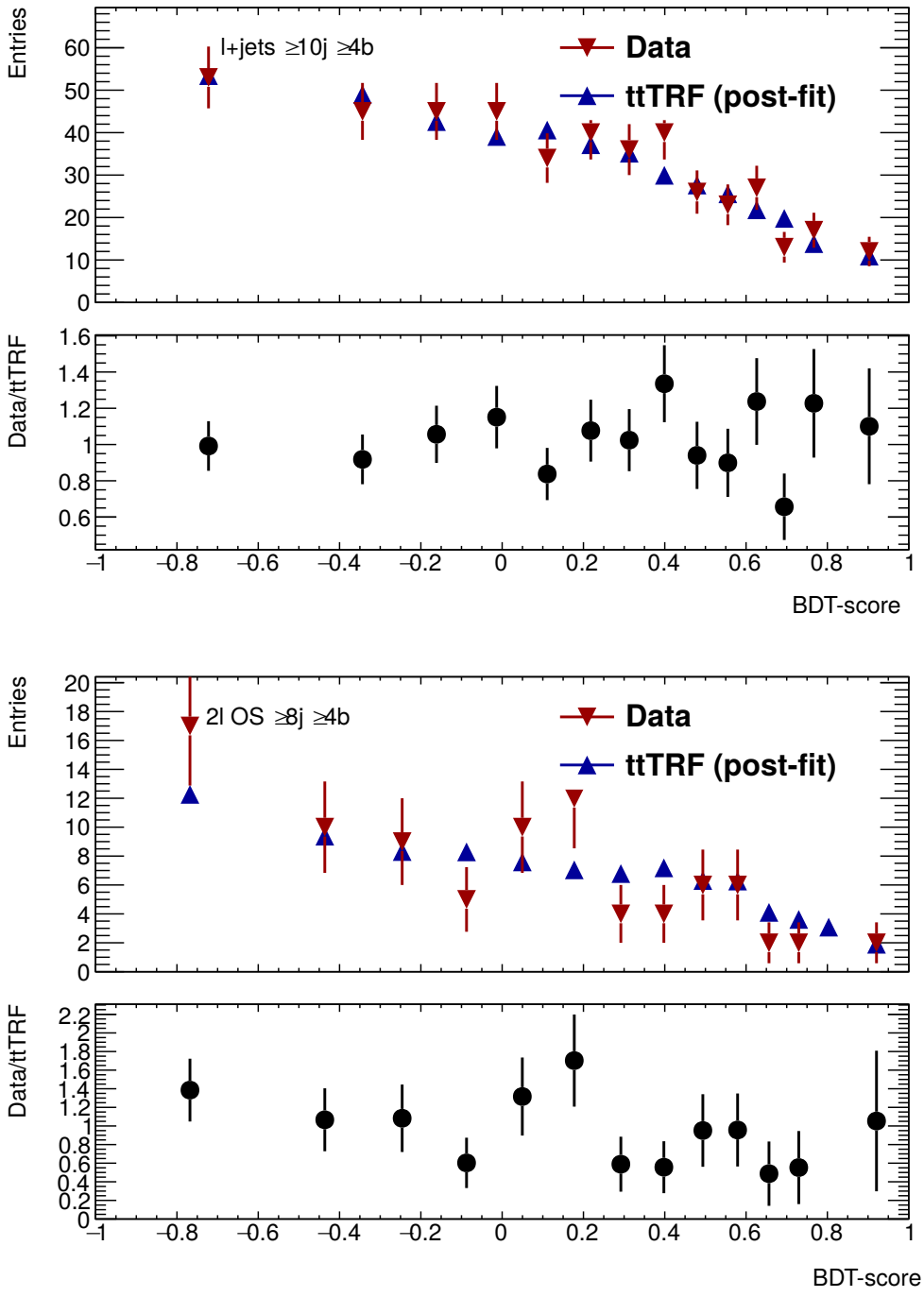


Figure 6.7.: Agreement between the ttTRF pseudodataset, based on the $t\bar{t}$ post-fit prediction, and real data in the highest (b-)jet multiplicities for the 1L and OS channel

in a separate fit to 1.5 times of the predicted value corresponding to a 50% pull. The measured signal strength is $\mu_{t\bar{t}t\bar{t}} = 1.75 \pm 1.06$, the corresponding pulls for the generator systematics are compared in Figure 6.8 (red). In general, they are very similar and agree within a few percent. For this particular study, the large pull is therefore considered negligible but has to be monitored in the fit to data as it is unexpectedly large.

A better judgement of a possible bias is given by comparing the distribution of $\mu_{t\bar{t}t\bar{t}}$ when fluctuating the pseudodataset within the statistical uncertainties, corresponding to a fluctuation in each bin within the Poissonian error, as shown in Figure 6.9 on the left. The width of the distribution of 0.99 around a mean of 0.91 indicates that the measured signal strength is just within the statistical error of an unbiased result. Furthermore, the measured value of $\mu_{t\bar{t}t\bar{t}} = 1.70$ does not correspond to the central value of the distribution and might therefore be an outlier. Fluctuating the distribution within the statistical uncertainties corresponds to the treatment that would be applied to real data. However, this fluctuation is overestimating the actual effect since the statistical fluctuation does not correspond to the uncertainty of the actual ttTRF pseudodataset. A more realistic estimate is obtained by varying the pseudodataset within the uncertainties of the ttTRF prediction as shown in Figure 6.9 on the right. A central value of $\mu_{t\bar{t}t\bar{t}} = 1.71$, corresponding to the measured value of the measured signal, and a width of 0.18 indicate that a bias of around 0.5 in the signal strength exists. This is explained by the generation of the ttTRF pseudodatase.
Even though the fit to generate the post-fit distributions was performed in blinded regions with a signal to background ratio of less than 5%, based on the MC prediction, some signal events are still expected. In this fit, $\mu_{t\bar{t}t\bar{t}}$ is fixed to unity as the goal is to extract the SM background prediction. However, since the fit to generate the post-fit prediction is done to real data, the real signal to background ratio is unknown and it is possible that more events than predicted by the SM are present. This assumption is confirmed by repeating the fit to generate the ttTRF post-fit prediction but without fixing the signal strength $\mu_{t\bar{t}t\bar{t}} = 1$. Furthermore, the signal strength is measured per channel (1L and OS) in a combined fit to judge if the effect might originate from the combination itself. The obtained values are

$$\mu^{\text{ttTRF}}_{t\bar{t}t\bar{t},\,\text{1L}} = 1.48 \pm 2.8 \tag{6.3}$$

$$\mu^{\text{ttTRF}}_{t\bar{t}t\bar{t},\,\text{OS}} = 1.77 \pm 3.6 \tag{6.4}$$

from which it is concluded that the effect is similar in both channels. This confirms the hypothesis that, when fixing the signal strength, the ttTRF post-fit prediction for the $t\bar{t}$ background contains a certain amount of signal events, roughly corresponding to the measured bias of approximately 0.5.
To further check this hypothesis, the linearity of the fit is studied by injecting a scaled SM $t\bar{t}t\bar{t}$ signal and repeating the measurement of $\mu_{t\bar{t}t\bar{t}}$ in the modified ttTRF pseudodataset. Figure 6.10 shows the measured signal strength for five different injections. As reference, the ideal unbiased result with uncertainties corresponding to the width of the distributions in Figure 6.9 are shown. For all cases, the result is unbiased within the to-

tal uncertainties of the measurements itself. Within the statistically fluctuated dataset, no bias is observed and an almost constant bias between approximately 0.5 and 0.7 is seen when only considering the variation of the ttTRF prediction itself. It is concluded that the response of the fit scales linearly with the signal. Of particular interest is the point when no signal is injected (injected μ=0). In principle, this should correspond to a background only fit and no shape or normalisation effect of the signal should be present. However, a small signal of $_{t\bar{t}t\bar{t}} = 0.53 \pm 0.80$ is fitted which is in agreement with $\mu^{\mathrm{ttTRF}}_{t\bar{t}t\bar{t},\,\mathrm{1L}}$ and $\mu^{\mathrm{ttTRF}}_{t\bar{t}t\bar{t},\,\mathrm{OS}}$ indicating that the ttTRF pseudodataset itself is biased and not the fit response.

In summary, the fits to pseudodatasets show that the measured signal strength tends to be slightly higher than expected but, in the fits to Powheg + Herwig and MadGraph5_aMC@NLO + Pythia, is within uncertainties of an unbiased result. In the fit to the alternative ttTRF prediction, a small bias is observed when fluctuating the dataset within the uncertainties of the ttTRF prediction. However, it cannot be concluded that this solely originates from the fit setup as the ttTRF prediction contains more signal than predicted by the SM and therefore $\mu_{t\bar{t}t\bar{t}} \approx 1.5$ is expected. Furthermore, the statistical uncertainties, based on Poissonian fluctuations, are much larger. Consequently, when varying the prediction within these uncertainties, no bias of the fit is observed.

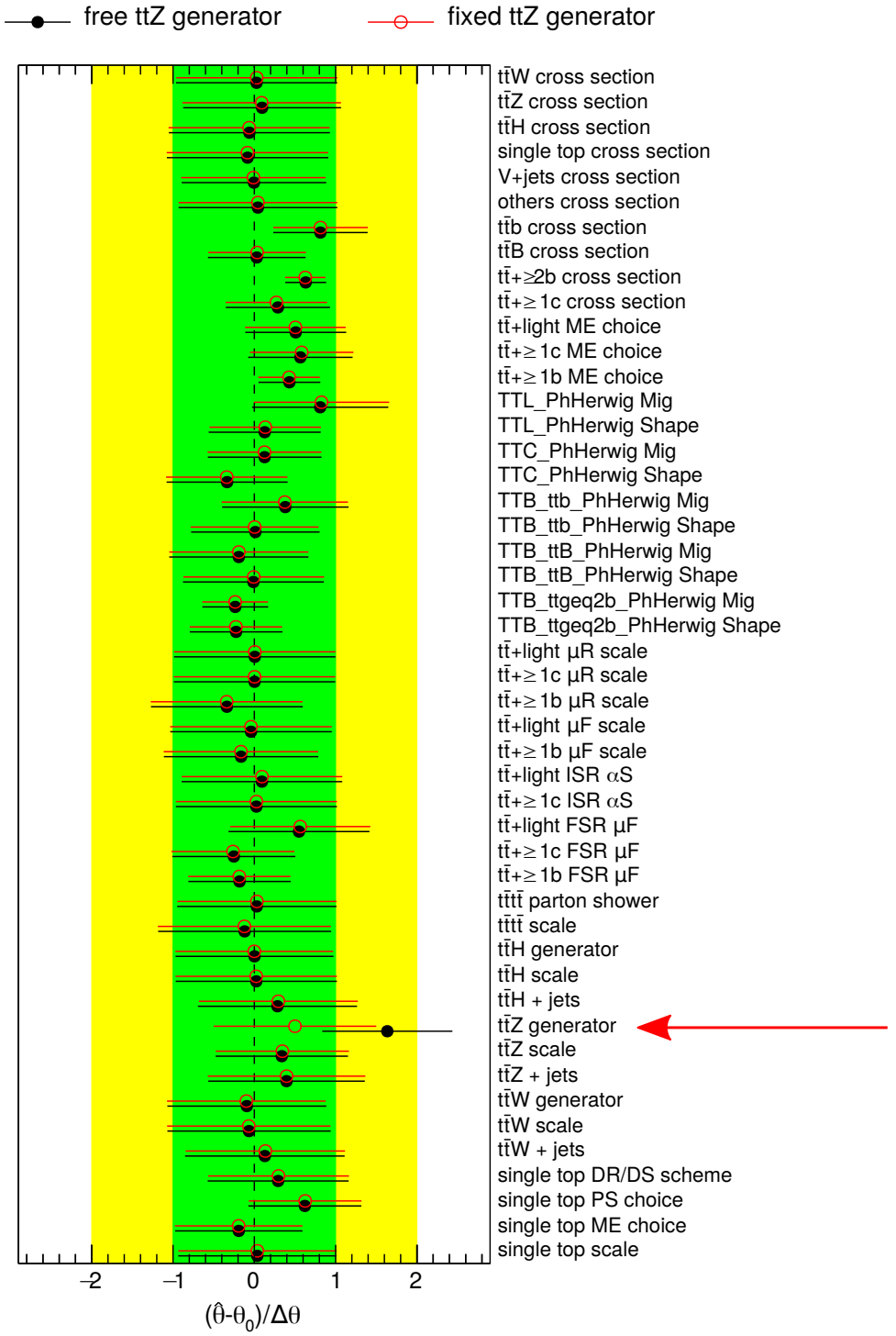


Figure 6.8.: Pulls and constraints for systematic uncertainties affecting the theory prediction in the fit to the ttTRF pseudodataset (black). They are compared to the scenario where the $t\bar{t}$ +Z generator systematic is fixed to a 0.5 σ pull.

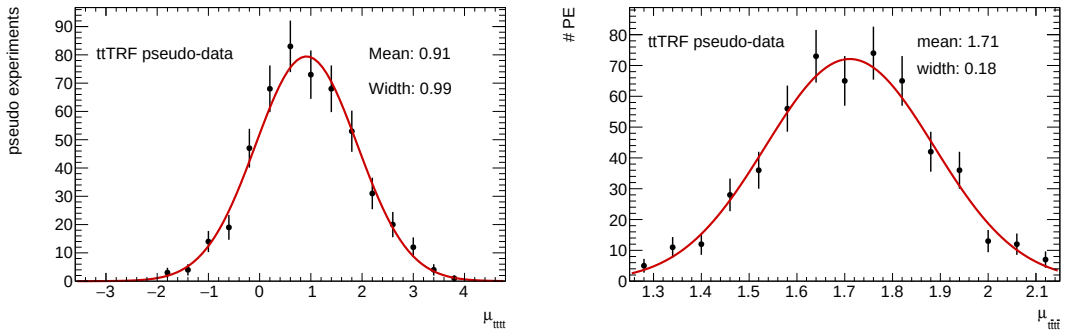


Figure 6.9.: Distribution of the measured signal strength for pseudo experiments generated by fluctuating each bin within the statistical uncertainties (Poissonian) (left) and by fluctuating the $t\bar{t}$ prediction of the pseudodataset within the uncertainties of the ttTRF prediction (right) 500 times each. The uncertainties are statistical only.

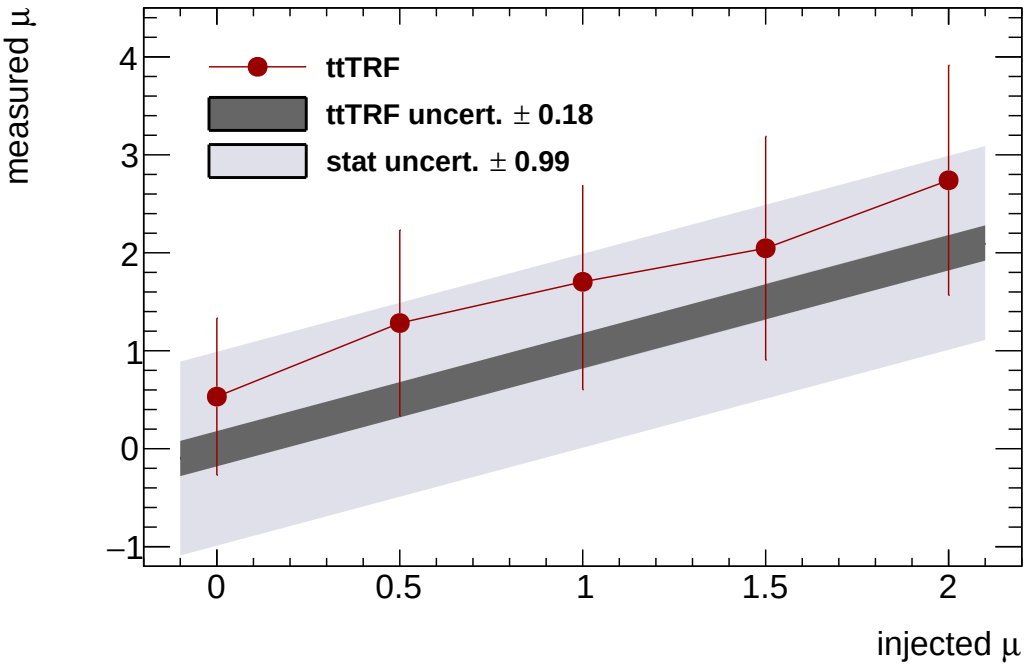


Figure 6.10.: Linearity test for the measured signal strength in a fit to pseudodata generated by the ttTRF method. For the exact same fit setup, the signal strength is measured for different injected signal strengths where 1 corresponds to the SM prediction.

6.2. Cross-Section Measurement in the Single Lepton and Opposite-Sign Dilepton Channel

In the previous chapter it was shown that in three scenarios with different pseudodata no clear bias of the fit is observed while the pulls and constraints of the profile likelihood fit are reasonable and within the assumptions. Furthermore, large correlations between signal strength and parton-shower uncertainties are observed. In this chapter, the final results of the fit to data in the 1L+OS channel are presented for which the exact same setup as for the fit to ttTRF pseudodata is used including all systematic uncertainties. In a fit to the blinded regions

$$\mu_{t\bar{t}t\bar{t}}^{\text{blind}} = 4.03 \pm 2.23 \tag{6.5}$$

is measured which, on a first glance, does not agree well with the expected SM prediction or the results from fits to pseudodatasets. The problem in this setup is that in the blinded regions not enough signal events are available to make use of the difference in shapes between signal and background. As consequence, the signal strength, as a single free parameter, is increased to cover background effects which are a combination of many other NPs. In particular, the $t\bar{t}$ + c/b normalisations have to be measured in parallel to the signal strength while for the corresponding ttTRF setup, in which a signal strength of approximately 1.5 was measured, the normalisations are predicted more precisely.
A similar effect in blinded regions was observed by the SS+ML channel and is reflected by the large uncertainty of ± 2.23 which makes the result only unbiased within approximately $1.5\,\sigma$. To check the dependence of the signal strength on the blinding-cut, and therefore on the signal to background ratio, the regions with exactly three b-jets are fully unblinded while the regions with more than four b-jets remained blinded according to the signal to background ratio of less than 5%. Here, $\mu_{t\bar{t}t\bar{t}} = 3.42 \pm 1.9$ is measured indicating that the possible bias decreases when more signal is included. This behaviour is confirmed by fitting only the regions with exactly three b-jets using the ttTRF pseudodata. By observing an increase of the measured signal strength $\mu_{t\bar{t}t\bar{t}} = 2.2 \pm 1.58$ it is reassured that the signal strength decreases when higher b-jet multiplicities are unblinded.
Next, all regions are fully unblinded. First, a fit to each channel (1L and OS) is performed individually. The measured signal strengths are

$$\mu_{t\bar{t}t\bar{t}}^{\text{1L}} = 2.27 \pm 1.39 \tag{6.6}$$

$$\mu_{t\bar{t}t\bar{t}}^{\text{OS}} = 1.32 \pm 1.45, \tag{6.7}$$

the pulls and constraints are shown in Appendix D. The observed (expected) significances are 1.79/0.97 (0.82/0.78) σ for the 1L/OS channel. Both channels agree within the total uncertainty but more signal is fitted in the 1L channel. This is also reflected by the increased significance and the dependence of the likelihood on the measured signal strength as shown in Figure 6.11 (a) and (b). For the 1L channel, the tail towards higher values is flatter than for the OS channel indicating that the uncertainties are asymmetric

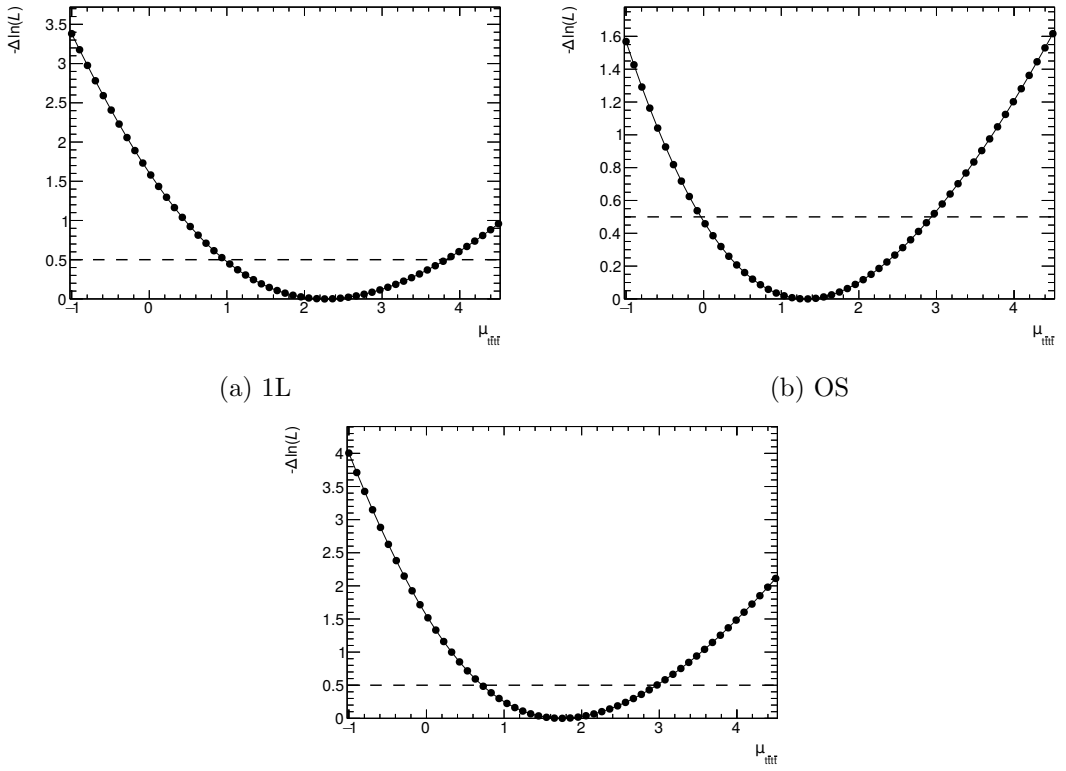
towards higher values.

Next, both channels (1L and OS) are combined. As introduced in the previous Section, Figure 6.12 (a) shows the pulls and constraints for each systematic uncertainties by comparing $\hat{\Theta}$ and Θ while respecting the uncertainty Δ_Θ for each uncertainty before fitting. All other NPs together with the correlation matrix are shown in Appendix D. Figure 6.12 (b) shows the most important systematic uncertainties based on their impact on the measured signal strength. The ranking is extracted by repeating the fit while leaving out each uncertainty at a time and measuring the change of the obtained $\mu_{t\bar{t}t\bar{t}}$. In general, modelling uncertainties are highly ranked, as already observed in the fits to pseudodata. Detector and b-tagging uncertainties only appear in the lower half of the figure. Besides the uncertainties of the scale factors (RW) for the $H_{\mathrm{T}}^{\mathrm{all}}$ component, the $t\bar{t}t\bar{t}$ scale and parton-shower uncertainties are ranked high. In addition, the $t\bar{t}$ $+\geq$ 2b parton-shower and cross-section uncertainties as part of the dominant background must be measured precisely as they are not predicted by the MC-based method. The highest correlations of the signal strength is given by the $t\bar{t}$ $+\geq$ 2b cross-section uncertainty (40.0%) as also observed by the fit to ttTRF pseudodata. Other high correlations exist between the $t\bar{t}$ +B and $t\bar{t}$ +b cross-sections (64%) as well as the $t\bar{t}$ $+\geq 1c$ cross-section and the normalisation effect of the $t\bar{t}$ $+\geq$ 2b parton-shower uncertainty (42%). Similar to the individual fits, a likelihood scan is performed to verify the convergence of the fit. The result is shown in Figure 6.11 (c). It is observed that no second minima or plateaus in the likelihood occur showing that the fit converged well. Furthermore, the observed curve is symmetric close to the minimum and, similar to the 1L fit, a larger tail towards larger signal strengths is observed as expected by the curves from the fits to the individual channels. The kinematic distributions after fitting are shown in Figure 6.13 and Figure 6.14 for the 1L channel and in Figure D.5 and Figure D.6 for the OS channel. The agreement between data and MC is good but some statistical fluctuation in the highest multiplicities could have been reduced by rebinning. The final result for the measured signal strength is

$$\mu_{t\bar{t}t\bar{t}}^{\mathrm{comb}} = 1.72 \pm 1.08 \approx 1.7^{+1.2}_{-1.0} = 1.7^{+0.7}_{-0.7}(\mathrm{stat})\,^{+1.0}_{-0.7}(\mathrm{syst}) \tag{6.8}$$

with a (expected) significance to observe the signal plus background hypothesis compared to the background only hypothesis of 1.76 (1.08) σ.

In summary, as already observed in the fit to blinded data by the ttTRF method and consequently in the fit to the ttTRF post-fit prediction, more signal events than predicted by the SM are observed. The slight increase in the measured signal strength (1.72) and the significance (1.76), compared to the nominal SM prediction (1), originates mostly from the 1L channel. As expected by the previous studies, in particular the $t\bar{t}$ $+\geq$ 2b cross-section and parton-shower uncertainties have the largest impact on the measured signal and are therefore the most constrained NPs in the fit.

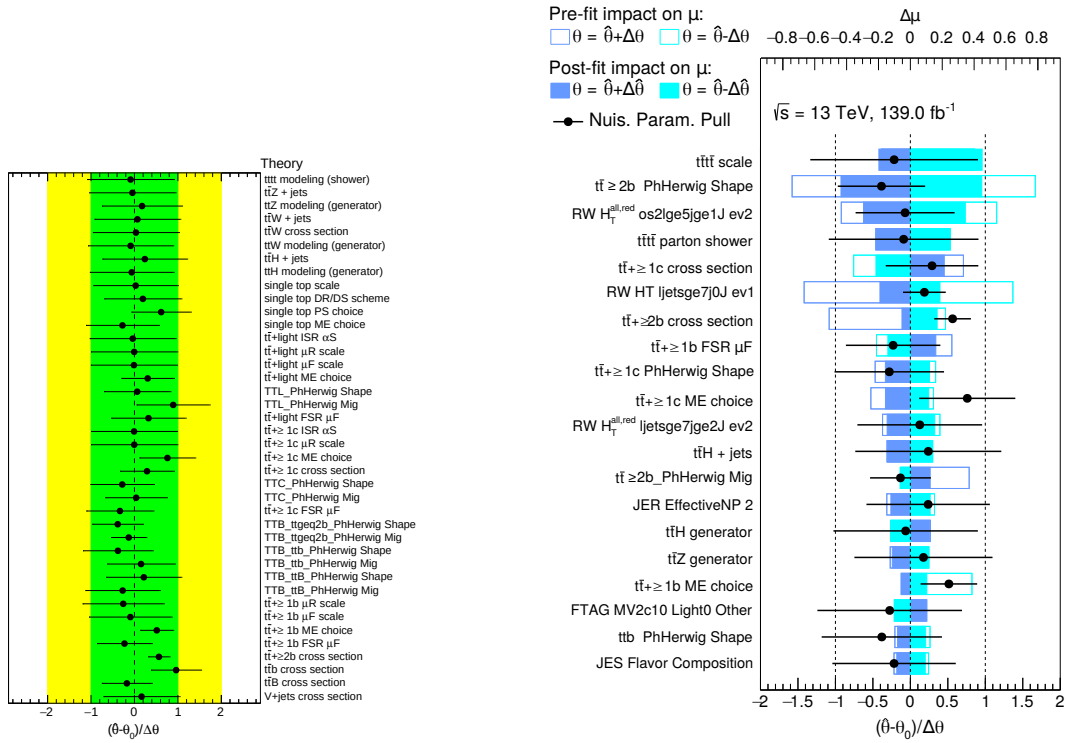


(a) 1L

(b) OS

(c) 1L and OS

Figure 6.11.: Change of the fitted likelihood as function of the fitted signal strength for fits in each channel and their combination.



(a) Modelling uncertainties (b) Ranking

Figure 6.12.: Pulls and constraints for systematic uncertainties affecting the theory prediction in the fit to data for the 1L+OS channel (left) and the ranking of systematic uncertainties, based on the impact to the measured signal strength, for all NPs (right). The plot on the right shows the pre- and post-fit uncertainties (coloured boxes) as well as the fitted NPs (black dots)

6.2.1. Studies on the Fit-setup with the Monte-Carlo Based Reweighting Method

While the result in the previous chapter is the final result of this thesis, further studies are done to understand the origin of the measured signal strength and to allow for a better interpretation. Furthermore, the studies in this chapter aim to provide possible improvements for future analyses.

First, the signal strengths for the 1L and OS channel are measured separately while combining both channels in the fit. While there is no explicit physics motivation in this analysis, as both channels are treated to have the same backgrounds and possible new physics scenarios, this setup provides feedback on the fit behaviour in each channel. The measured signal strengths in the combined fit are

$$\mu_{t\bar{t}t\bar{t}}^{\text{comb,1L}} = 1.88 \pm 1.26 \tag{6.9}$$

$$\mu_{t\bar{t}t\bar{t}}^{\text{comb,OS}} = 1.48 \pm 1.37. \tag{6.10}$$

They agree better with each other than the individually measured signal strengths in separate fits as all other NPs are still fully correlated. Still, a higher signal is fitted in the 1L channel and it is more sensitive, as shown by the smaller uncertainties. One of the possible reasons is the lower statistics in the OS channel which might require a more dedicated binning in the future.

For the same fit setup, the signal strengths in each b-jet multiplicity are decorrelated and measured to be

$$\mu_{t\bar{t}t\bar{t}}^{\text{comb},3b} = 1.60 \pm 1.47 \tag{6.11}$$

$$\mu_{t\bar{t}t\bar{t}}^{\text{comb},\geq 4b} = 1.77 \pm 1.18. \tag{6.12}$$

As expected and confirmed by the smaller uncertainty, the signal is measured most precisely in the $\geq 4b$ regions. Including the regions with three b-jets further reduced the uncertainty by 0.09 compared to the final result form in the previous chapter while not altering the central value significantly. Therefore, the splitting in b-jet multiplicities should be adapted in future analyses. In addition, the study shows that the splitting into 3bL and 3bH regions still allows for measuring the $t\bar{t}t\bar{t}$ signal in the 3b regions and therefore justifies that the splitting can be applied to be more sensitive to the $t\bar{t}$ $+\geq 1b$ sub-components.

Similarly, the signal strengths are measured for each jet multiplicity. The results of a combined fit when decorrelating $\mu_{t\bar{t}t\bar{t}}$ per region are summarised in Table 6.3. While negative values have no physical interpretation, they show that in these regions the overall background prediction might be overestimated and has been compensated by a decreased signal. The smaller uncertainties show that the higher jet multiplicities are more sensitive to the signal strength, as expected. The measured signal strength in the

lowest multiplicities (7j and 5j) have very large uncertainties (±11.95/18.61) and it is therefore possible that they are not required by the fit or might even disturb the final result by introducing a bias. To study this hypothesis, the fit is repeated but without these regions. The signal strength is measured to be

$$\mu_{t\bar{t}t\bar{t}}^{\text{comb},\geq 8/6j} = 1.96 \pm 1.16 \tag{6.13}$$

corresponding to an increased uncertainty of 0.07 compared to the final result. In addition, a shift of the central value by +0.25 is observed. Figure 6.15 shows again the comparison of $\hat{\Theta}$ and Θ but this time for the two cases when the lowest jet/multiplicities have been included (black) in the fit. The largest change is the absence of the $t\bar{t}$ +B normalisation effect of the parton-shower when excluding the lower n-jet multiplicities. After investigating, this is explained by the *pruning* of the nuisance paramaters which excludes small uncertainties (below 1%) to improve fit stability. While no final conclusion can be made if the low jet multiplicities are required, the observations suggest that the increased signal-strength covers effects on the background prediction that are not sufficiently derived in the control regions. As the setup was optimised and studied when all regions have been included, future analyses should investigate the impact of these regions prior to unblinding. In particular the absence of the $t\bar{t}$ +B normalisation parameter indicates that the missing normalisation effect might be covered by an increased signal contribution. However, the effect should be small as the pruning threshold is only 1%. By applying pruning, uncertainties with an effect on the normalisation or the shape of less then the chosen threshold are excluded. Typically this cut is required for numerical stability as otherwise too manz degrees of freedom are allowed in the fit. Therefore, the change in the signal strength is most likely a cumulative effect as other NPs change slightly as well.

Another possibility to alter the fit setup is to reconsider the correlation scheme. By default, only the parton-shower uncertainty, as most important uncertainty, is decorrelated in normalisation and shape effects to introduce additional degrees of freedom. Besides the decorrelation in normalisation and shape effects, assumed to be mostly independently for a given systematic, other correlation schemes are possible but have to be motivated first.
While not in agreement with the general strategy of this analysis to use the lower jet multiplicities to predict the background distributions in higher multiplicities, it is feasible that each jet multiplicity could be treated as uncorrelated. The physics motivation to do so is that each additional jet, to first order, is produced independently of the previously produced jets. This is not the default in this analysis as, on the contrary, all additional jets are produced by the same mechanism which, in MC samples, corresponds to a certain set of choices in the given generator and, in data, to the properties of the strong force. For example, the parton-shower uncertainty is derived by comparing different sets of generators which differ in the way additional radiation is produced. As the choice of the generator is independent of the jet multiplicity, they are treated as correlated in all regions. Nevertheless, the alternative correlation scheme is studied by

Table 6.3.: Signal strength $\mu_{t\bar{t}t\bar{t}}$ and its uncertainty Δ after combined fit to data in the 1L+OS channels. The signal strength is measured separately in each individual jet region while all other NPs remained correlated among the regions.

	1L			OS	
Region	$\mu_{t\bar{t}t\bar{t}}$	$\Delta\mu_{t\bar{t}t\bar{t}}$	Region	$\mu_{t\bar{t}t\bar{t}}$	$\Delta\mu_{t\bar{t}t\bar{t}}$
7j	1.26	11.95	5j	-10.00	18.61
8j	-2.62	5.14	6j	-2.76	6.95
9j	1.02	2.11	7j	5.14	3.15
$\geq 10j$	1.91	1.33	$\geq 8j$	1.09	1.27

decorrelating the parton-shower uncertainties among regions. In parallel, they are kept correlated in shape and normalisation to not introduce too many degrees of freedom at the same time. As a result

$$\mu_{t\bar{t}t\bar{t}}^{\text{comb,alt.}} = 1.38 \pm 0.95, \tag{6.14}$$

with a (expected) significance of 1.63 (1.27) σ is obtained. This is in a better agreement with the nominal SM prediction but might also hint that the increased signal that is measured in data is not fit correctly. While the decision should not be made on the fitted result, this shows that it is worth to reconsider the correlation schemes in the future.

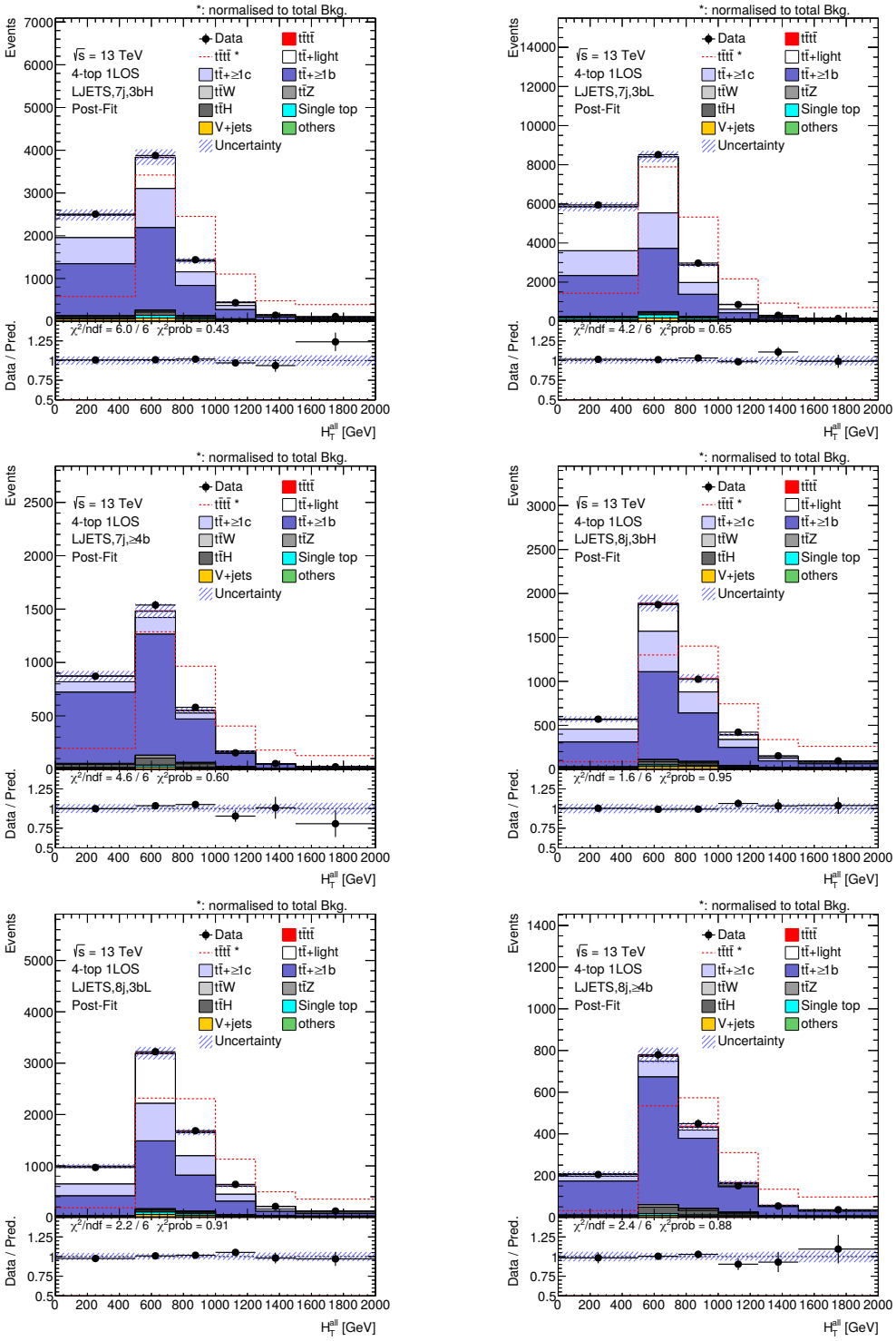


Figure 6.13.: Post-fit distribution for the seven and eight jet regions of the 1L channel after a combined fit to the 1L+OS channels.

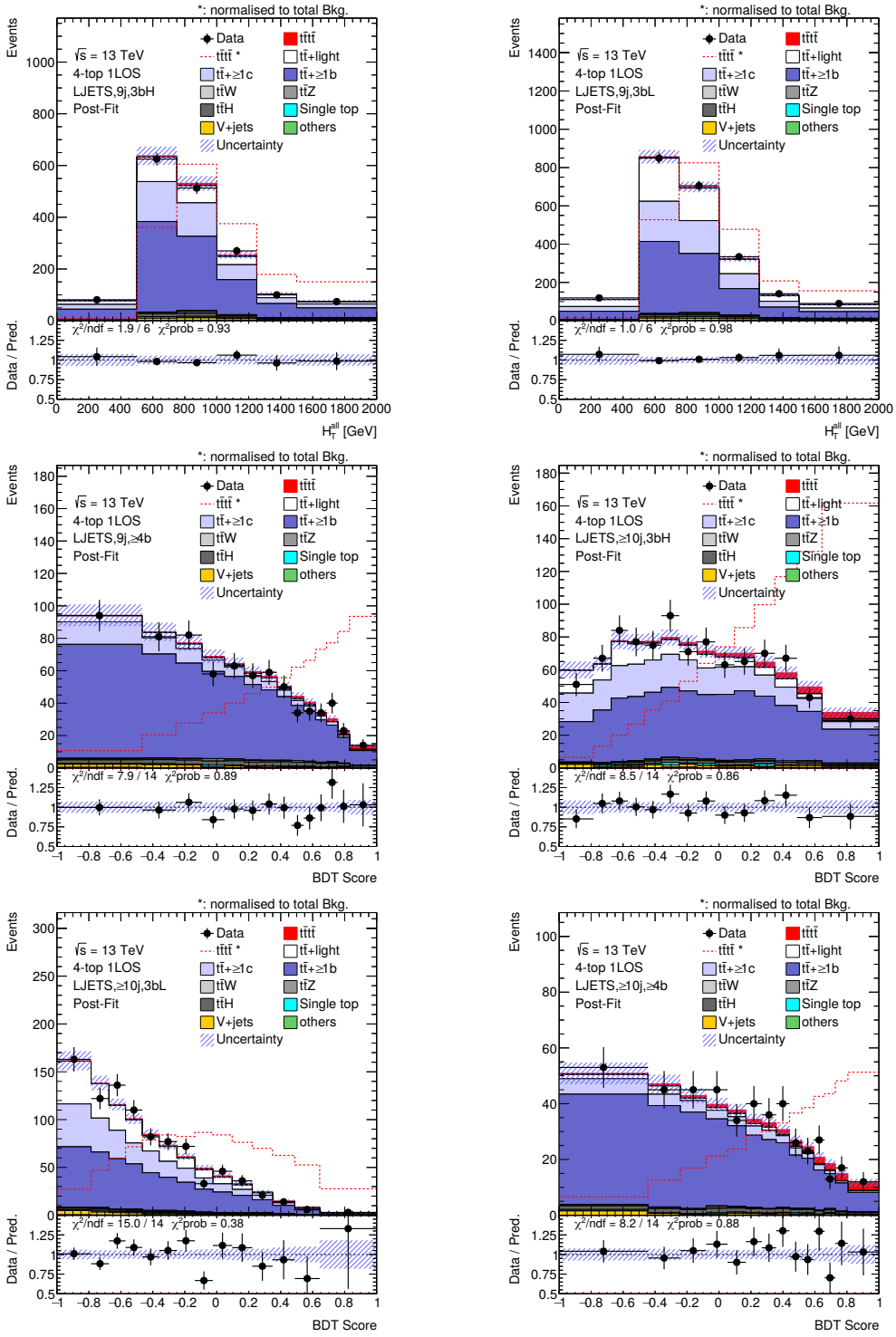


Figure 6.14.: Post-fit distribution for the nine and more than ten jet regions of the OS channel after a combined fit to the 1L+OS channels.

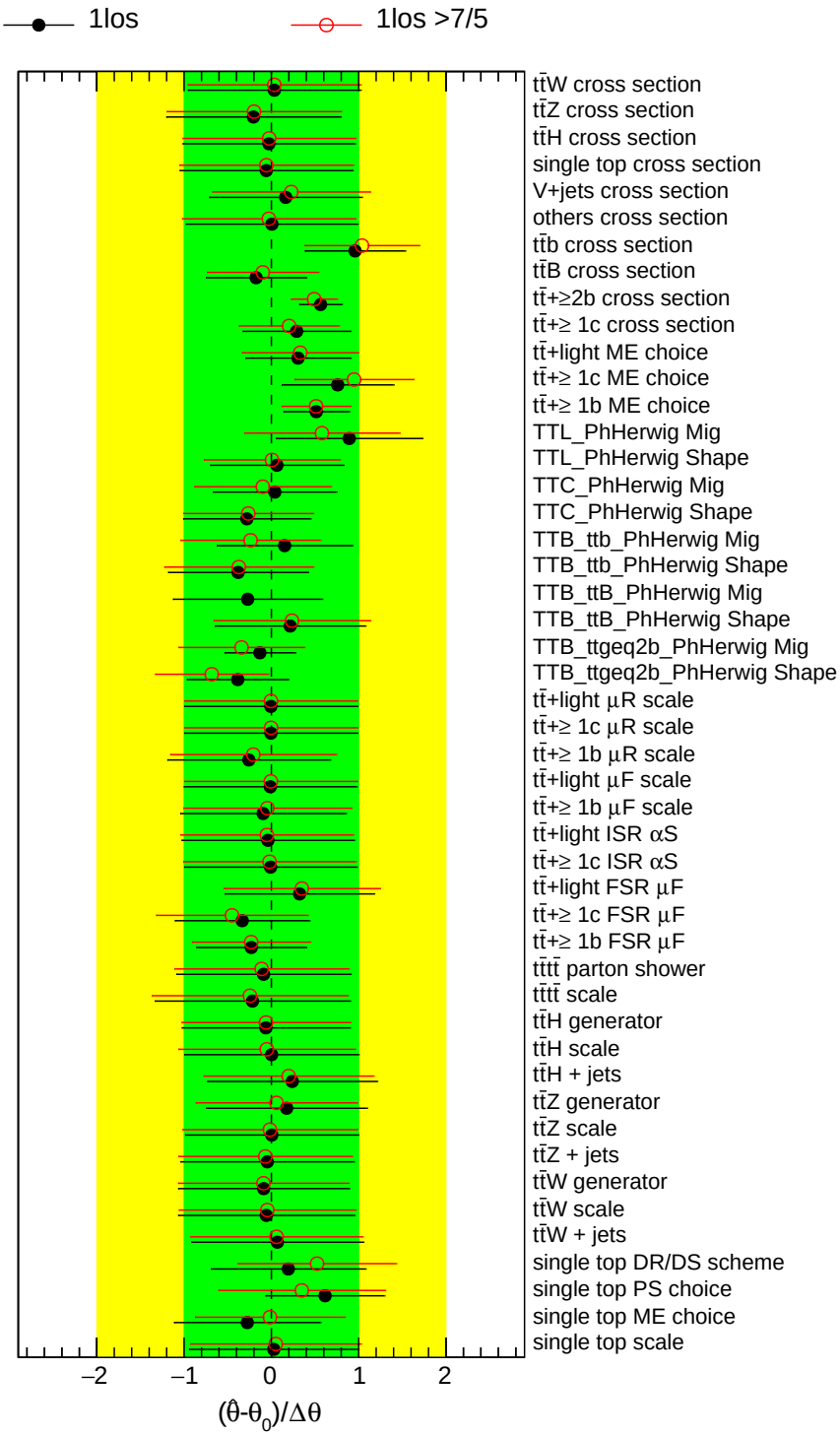


Figure 6.15.: Comparison of the nuisance parameters when including the 7j and 5j multiplicities (black) and when excluding them (red).

6.3. Combination with the Same-Sign Dilepton and Multilepton Channel

After summarising the latest results from the SS+ML channel [52], the combination with the final result from the previous section is presented in this section.

6.3.1. Results in the Same-Sign Dilepton and Multilepton Channel

The fit setup for the SS+ML channel was introduced in Section 5.7. In total, five regions are fit of which four are used to constrain the backgrounds. The remaining region uses the BDT-score to measure the $t\bar{t}t\bar{t}$ signal strength. The post-fit distributions are shown in Figure 6.16, the fit results are included in Appendix D. The good agreement between data and simulation for regions in which the BDT-score is not fit shows that the background contributions are measured accurately. The measured normalisation factors for background and signal are shown in Figure 6.17 (a) and the ranking of the NPs in Figure 6.17 (b). It is observed that the $t\bar{t} + W$ background is the most important uncertainty but the $t\bar{t}t\bar{t}$ cross-section ranks even higher. The uncertainty on material conversion has a similar impact on the normalisation and the uncertainty on the cross-section of three top-quarks, which has a very similar signature to the $t\bar{t}t\bar{t}$ signal, is ranked high. The signal strength of

$$\mu_{t\bar{t}t\bar{t}}^{\text{SSML}} = 2.0^{+0.8}_{-0.6} = 2.0^{+0.4}_{-0.4}(\text{stat})^{+0.7}_{-0.5}(\text{syst}) \tag{6.15}$$

is in agreement with the 1L+OS results of this thesis but the (expected) significance of 4.3 (2.4) σ is much higher. The upper limit at 95% CL is 3.1 times the SM cross-section.

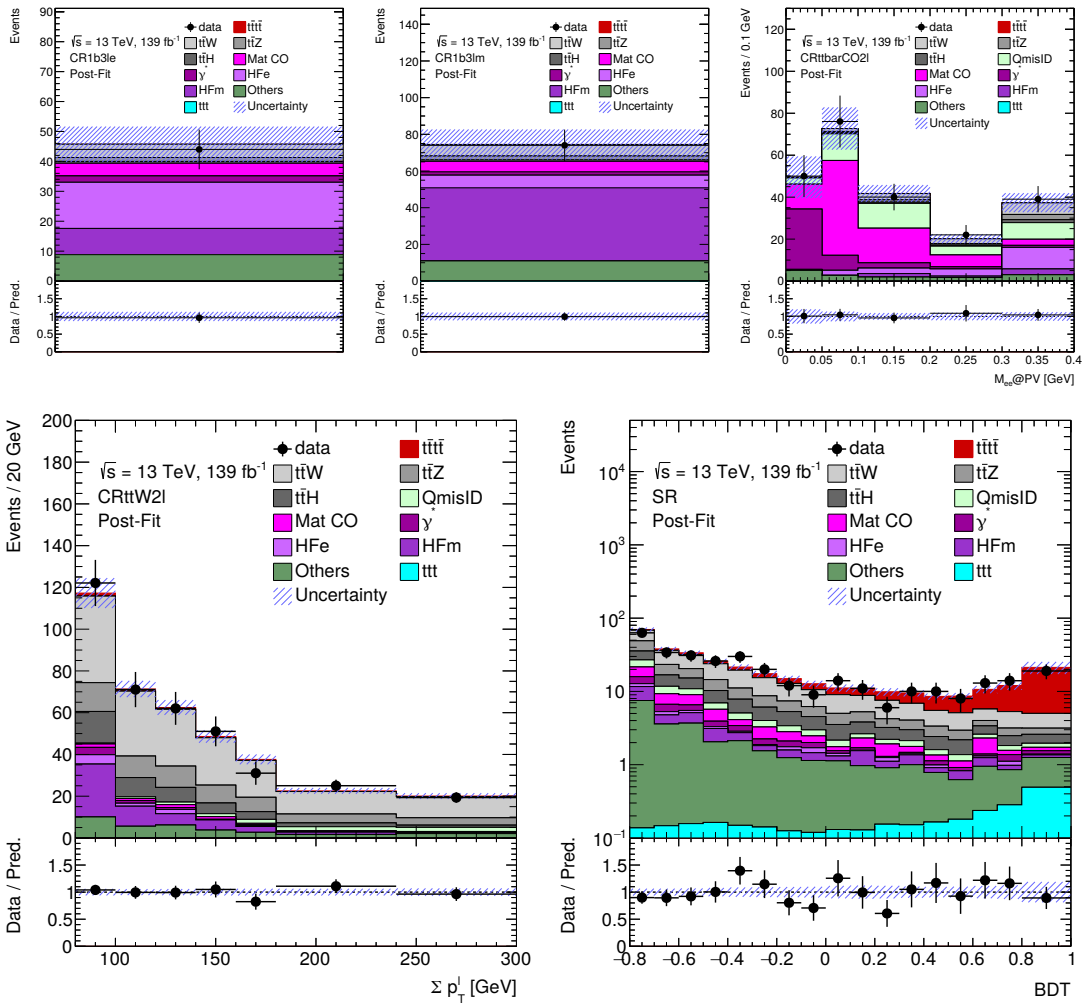
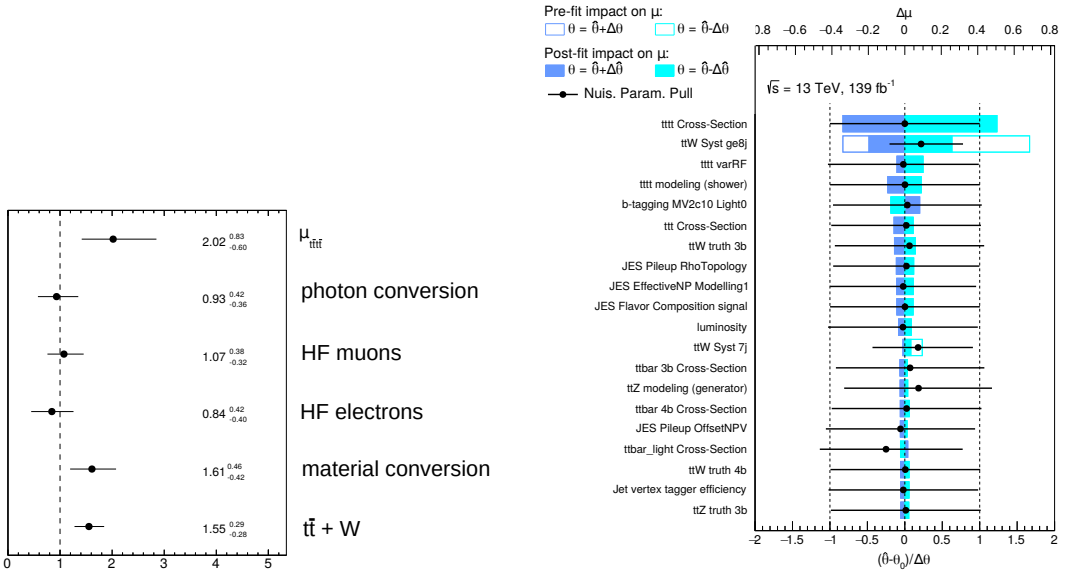


Figure 6.16.: Fitted distributions after a simultaneous fit in the SS+ML channel. The uncertainty band includes statistical and systematic uncertainties.



(a) Normalisation factors (b) Ranking

Figure 6.17.: Normalisation factors of the backgrounds and signal and the ranking of the NPs for the fit in the SS+ML channel. The ranking shows the pre- and post-fit uncertainties as well as the measured NPs.

6.3.2. Combined Results

The combination of the results in both channels is done as follows: The NPs are combined according to their definition in Section 5.5 such that common detector and modelling systematics are fully correlated among all channels. Channel specific uncertainties, such as the uncertainties of the scale factors or additional normalisation factors for photon and material conversion, are excluded in the complementary channel. Afterwards, a profile likelihood fit with a single free parameter to measure the signal strength is performed in all regions simultaneously. As result,

$$\mu_{t\bar{t}t\bar{t}}^{\text{1L+OS+SS+ML}} = 2.0^{+0.5}_{-0.4} = 2.0^{+0.4}_{-0.4}(\text{stat})\,^{+0.3}_{-0.3}(\text{syst}) \tag{6.16}$$

is measured, with a signal significance of $4.71\,\sigma$ ($3.14\,\sigma$). One caveat worth mentioning is that some NPs are under-constrained in the default setup. This means that the pre-fit uncertainties had to be increased to allow the fit to converge. While there are many possible and often technical reasons, such as an unidentified mistake in the correlation scheme or an inconsistency in the exact implementation of the systematic uncertainties, it usually indicates an unsatisfying convergence of the fit. To check how the result changes, the under-constraints are reduced by increasing the pruning threshold to 9%, which is rather large compared to the nominal value of 1%. This approach suppresses many small uncertainties that exhibited the unwanted behaviour and thereby reduces the degrees of freedom in the fit. By studying the excluded systematics it could not be

Table 6.4.: Observed and expected significances as well as measured signal strengths for two different pruning thresholds. In the combination marked with * under-constraints are observed indicating that the fit did not converge properly.

Channel/Pruning	1%			9%
	obs.	exp.	$\mu_{t\bar{t}t\bar{t}}$	$\mu_{t\bar{t}t\bar{t}}$
1L+OS	1.76	1.08	1.72 ± 1.09	1.87 ± 0.85
SS+ML	4.3	2.4	2.02 ± 0.60	2.00 ± 0.39
Comb	4.71*	3.14	1.98 ± 0.46*	1.98 ± 0.45

confirmed that a single systematic uncertainty causes problems in the fit. Since the fit was performed based on the configurations of the individual fits, technical difficulties can be excluded. The pulls and constraints for theory NPs are shown in Figure 6.18, the full list is included in Figures D.8 and D.9. In general, no tension between the channels or the combination is observed except for the $t\bar{t}$ +≥ 2b parton-shower normalisation which is pulled by more than $1\,\sigma$ in the combination. The measured signal strength is $\mu_{t\bar{t}t\bar{t}} = 1.98 \pm 0.45$ when increasing the pruning threshold. To not overestimate the significance, a 10% uncertainty, roughly corresponding to the pruning threshold, on all samples is introduced and a observed (expected) signal significance of $5.03\,\sigma$ ($3.25\,\sigma$) is obtained. The results are summarised in Table 6.4 including a comparison of the measured signal strength for both pruning thresholds. The central values of the measured signal strength agree but, as expected, the uncertainties are reduced in all channels. Even after including an additional uncertainty, the observed (expected) significance is increased compared to the default pruning by approximately 0.32 (0.11) in this setup. Therefore, as final result of the combination Equation 6.16 is quoted and the result with an increased pruning is seen as confirmation.
While the caveat of under-constrained uncertainties does not provide a fully satisfying conclusion, this result is the first full combination with the full Run-2 dataset by ATLAS or CMS. The measured signal strength is consistent in two pruning setups and the significance is well above $3\,\sigma$. Therefore, evidence of the process is obtained.
Furthermore, the combined result shows that the SS+ML channel is dominating the sensitivity and the excess over the expected significance is driven by this channel. Depending on the pruning, the 1L+OS channel improves the significance by around 0.3 (0.7) σ and, given the larger uncertainty of $\Delta\mu_{t\bar{t}t\bar{t}}^{\text{1L+OS}} = \pm\ 1.1\ (0.8)$, only slightly improves the precision on the measured signal strength from $\Delta\mu_{t\bar{t}t\bar{t}}^{\text{SS+ML}} = \pm 0.6$ to approximately $\Delta\mu_{t\bar{t}t\bar{t}}^{\text{Comb}} = \pm 0.5$. The 1L+OS channel is therefore at this stage mostly important to study MC modelling in a new and challenging phase space while the search for the process is driven by the SS+ML channel.

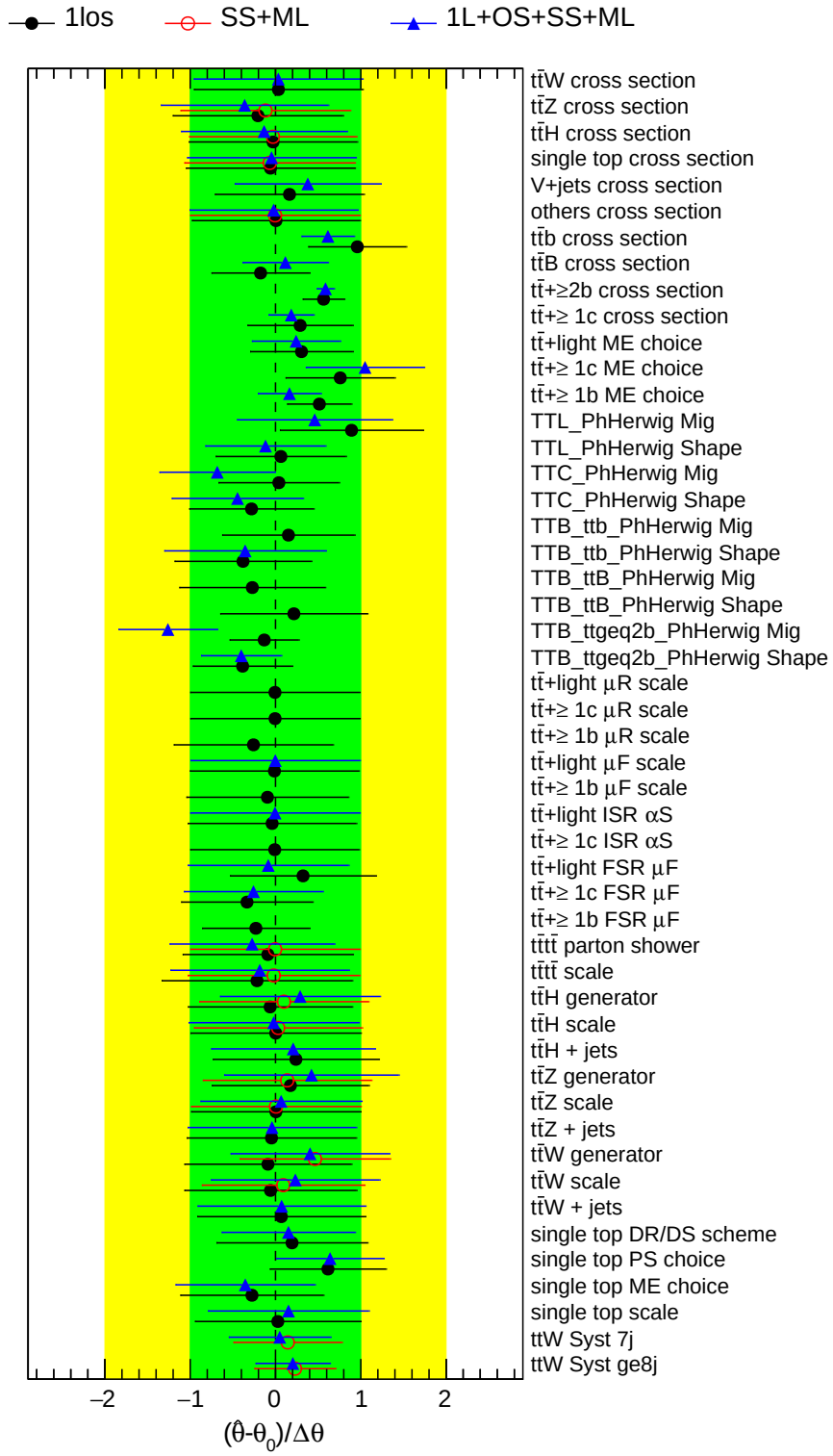


Figure 6.18.: Comparison of the modelling and theory NPs when fitting the 1L+OS and the SS+ML channel separated and combined. To obtain a better fit stability, the pruning threshold is increased to 9% for the combination.

6.4. Comparison with latest CMS Results

No 1L+OS result of the CMS experiment and therefore no combined result in both channels with the full Run II dataset is available at the moment. Therefore, the 1L+OS result with a dataset of 36fb^{-1} is used for the comparison. As the significance of the measured results in the SS+ML is quite different, a short comparison is presented as well even though this is not the focus of this thesis.

6.4.1. Comparison in the 1L+OS Channel

CMS measured the $t\bar{t}t\bar{t}$ cross-section with 35.8fb^{-1} to be

$$\sigma_{t\bar{t}t\bar{t}}^{\text{CMS}} = 13^{+11}_{-9}\,\text{fb}, \tag{6.17}$$

corresponding to s signal strength of $\mu_{t\bar{t}t\bar{t}} = 1.09^{+0.92}_{-0.75}$, with a significance of $1.4\,\sigma$ in a combination with the SS+ML channel. For the same dataset, the signal strength in the 1L+OS channels was measured (expected) to be $\mu_{t\bar{t}t\bar{t}} = 0.0^{+2.2}_{-0.0}$ (0.4). CMS used a multivariate discriminant to exploit the difference between $t\bar{t}t\bar{t}$ and $t\bar{t}$ [144]. First, a BDT is used to reconstruct top-quarks and a second BDT, similar to the analysis presented here, is used to generate a discriminant. The exact choice of input variables differs. In particular angular and mass distributions are more exploited in this analysis while in the case of CMS the focus is mostly on the kinematics of jets and leptons. In the dilepton channel, the region definition differs as the 4-5 and 6-7 jet multiplicities are merged and only ≥3b events, not split = 3 and ≥ 4b, are considered due to the lower data statistics. The choice of the POWHEG at NLO generator for the $t\bar{t}$+jets background is similar while for smaller backgrounds MADGRAPH5_aMC@NLO is used instead of SHERPA. In both cases, the $t\bar{t}t\bar{t}$ signal is simulated by MADGRAPH5_aMC@NLO but in this analysis a newer version (2.6.2 instead of 2.2.2) is used. Due to the different detectors, the event selection and the trigger criteria are necessarily different. CMS benefits from a particle-flow algorithm for the jet reconstruction which is not yet available in ATLAS.
Despite the technical differences discussed above, the general strategy of this analysis is based on correcting the MC prediction while the CMS analysis relies on the nominal MC prediction which, in particular, contains uncorrected shape effects. The splitting of $t\bar{t}$+jets into flavour components of the nominal sample is similar but the flavour sub-categories ($t\bar{t}$ + $b/B/ \geq 2b$) are used for the uncertainties of the parton-shower and cross-sections in this analysis. Furthermore, shape and migration effects of modelling uncertainties are considered while CMS focuses mostly on normalisation differences. In summary, it is not feasible to compare the two analyses in detail as the datasets differ substantially and, due to the difference in time, other techniques and results are available at the time of this analysis. However, in the overall picture, this analysis introduces a new technique which can be compared to the results of CMS as, for example, generator choices and therefore the underlying background modells are similar. While CMS observed no signal, this analysis measured an increase over the nominal SM prediction.

6.4.2. Comparison in the SS+ML Channel

CMS performed two independent measurements of the SM cross-section [51]. The first one is based on kinematic cuts while the second one uses a BDT. The cut-based analysis uses 15 regions, based on lepton and (b-)jet multiplicities, for the signal extraction and one region which is enriched in the $t\bar{t}+W$ background. The BDT analysis divides the phase space into 17 regions based on the BDT-score to separate the $t\bar{t}t\bar{t}$ signal from the SM background. In addition, one additional region to control the $t\bar{t}+Z$ background is used. Both analyses use a profile-likelihood fit to measure the signal. The measured cross-section is

$$\sigma_{t\bar{t}t\bar{t}}^{\text{BDT}} = 12.6^{+5.8}_{-5.2}\,\text{fb} \tag{6.18}$$

$$\sigma_{t\bar{t}t\bar{t}}^{\text{Cut-based}} = 9.4^{+6.2}_{-5.6}\,\text{fb} \tag{6.19}$$

which is compared to an expected SM cross-section of $12.0^{+2.2}_{-2.5}$. This corresponds to a signal strength of $\mu_{t\bar{t}t\bar{t}} = 1.05$ (0.78) for the BDT (cut-based) analysis. The observed (expected) significance is 2.6 (2.7) for the BDT and 1.7 (2.5) for the cut-based analysis. While the expected significance is compatible to the expected significance by ATLAS (2.4), the observed significance is considerably lower by 1.7 (2.6) σ compared to the ATLAS result of 4.3 σ. Despite smaller differences in the event and trigger selection, the first striking difference are the choices of the generators.
For CMS, most samples are generated by the MADGRAPH5_aMC@NLO 2.2.2 generator at NLO with a varying number of additional partons in the ME calculation. Only the WZ, ZZ and $t\bar{t}$ +H samples are generated using POWHEG. For ATLAS, the $t\bar{t}$ +W, tW, diboson and triboson samples are generated by SHERPA, $t\bar{t}$ samples are generated by POWHEG and only the nominal signal $t\bar{t}t\bar{t}$ sample is generated by the newer version of MADGRAPH5_aMC@NLO 2.6.2. Since each generator gives a slightly different prediction the different choices in each generator lead to a different prediction of the SM background.
Next, the procedure to generate the fit regions differs substantially. Both analyses use a BDT to separate signal and background which is split into 17 bins or regions. ATLAS uses four control regions to consider different backgrounds while CMS only uses the BDT-score. For CMS, the $t\bar{t}+W$ background is measured in the lower part of the BDT-score where it is enriched. Other backgrounds, which are assigned dedicated regions to be measured in the ATLAS measurement, are treated by corresponding uncertainties in the signal regions. Another difference, originating from the fit itself, is the measured normalisation factor of $t\bar{t}+W/Z$ which, for ATLAS, was measured to be $1.6^{+0.3}_{-0.3}$ and $1.3^{+0.2}_{-0.2}$ for CMS. On a first glance, a smaller normalisation factor could correspond to a higher signal strength as, when assuming the same data, part of the background normalisation could be covered by the signal. However, CMS does not provide the normalisation factors for HF leptons and conversions separately and therefore other backgrounds could be increased as well, effectively reducing the signal strength.
In summary, ATLAS measures an increased signal which is still compatible with the SM signal while the result from CMS is good agreement with the nominal SM prediction. The analyses differ in the choice of the generators and ATLAS focuses more on the

measurement of each background by including more control regions which could explain the difference in the measured signal.

CHAPTER 7

Conclusions and Outlook

In this thesis, the $t\bar{t}t\bar{t}$ signal strength was measured in the 1L+OS channel with the 13 TeV dataset of the ATLAS collaboration. Due to assumptions and limitations in the theoretical modelling, the default background prediction by MC generators is not reliable in the phase-space of high jet and b-jet multiplicities of this analysis. The mismodelling is addressed by dedicated scale factors which have been derived sequentially in the distributions of jet-multiplicity, a variable related to $H_{\mathrm{T}}^{\mathrm{all}}$ and the average distance between two jets. In particular the treatment of parton-shower systematics was studied in detail leading to a dedicated correlation scheme in the profile likelihood fit. For this systematic uncertainty and the cross-section uncertainty, the $t\bar{t} + \geq 1b$ component is split into three sub-categories as well as migration and shape effects. Together with the additional scale factors, this setup provides a reliable measurement of the $t\bar{t}$+jets background. In control regions, $H_{\mathrm{T}}^{\mathrm{all}}$ is used as discriminant to constrain the background modelling. For higher (b-)jet multiplicities, a BDT is trained to separate $t\bar{t}t\bar{t}$ events from $t\bar{t}$+jets events to measure $\mu_{t\bar{t}t\bar{t}}$.

The final result with an observed (expected) signal significance of $1.76\,\sigma$ ($1.08\,\sigma$) is

$$\mu_{t\bar{t}t\bar{t}} = 1.7^{+1.2}_{-1.0} \;=\; 1.7^{+0.7}_{-0.7}(\mathrm{stat})\,^{+1.0}_{-0.7}(\mathrm{syst}) \tag{7.1}$$

which, for the theoretical SM prediction of $11.97^{+18\%}_{-21\%}$ fb, corresponds to a measured cross-section of

$$\sigma_{t\bar{t}t\bar{t}} \;=\; 20.6^{+8.4}_{-8.1}(\mathrm{stat})^{+11.84}_{-9.0}(\mathrm{syst})\,^{+3.7}_{-4.3}(\mathrm{theory})\,\mathrm{fb} \tag{7.2}$$

and an (expected) upper limit of 3.86 (1.97) times the SM cross-section @ 95% CL. This is in agreement with the previous ATLAS result [49] of

$$\sigma_{t\bar{t}t\bar{t}} \;=\; 20.3^{+22.7}_{-20.3} = 20.3^{+13.2}_{-12.0}(\mathrm{stat})^{+19.2}_{-16.8}(\mathrm{syst})\,\mathrm{fb} \tag{7.3}$$

with an upper limit of 5.3 times the SM cross-section (63 fb) in this channel. No CMS result in the 1L+OS channel with the full Run-2 is available for a direct comparison. However, the result is an improvement to the latest result with a dataset of 36fb^{-1} where

$$\sigma_{t\bar{t}t\bar{t}} = 0.0^{+26.3} \tag{7.4}$$

was measured with an upper limit of 4.01 times the SM cross-section (48 fb) at 95%CL. The result agrees with the latest measurement the SS+ML channel from ATLAS and CMS

$$\sigma_{t\bar{t}t\bar{t}}^{\text{ATLAS}} = 24^{+7}_{-6}\,\text{fb} \tag{7.5}$$

$$\sigma_{t\bar{t}t\bar{t}}^{\text{CMS}} = 12.6^{+8.8}_{-5.2}\,\text{fb} \tag{7.6}$$

which, for ATLAS, corresponds to an upper limit of $\sigma_{t\bar{t}t\bar{t}} = 37.1\,\text{fb}$. Furthermore, the first combined result with the full Run-2 dataset of

$$\sigma_{t\bar{t}t\bar{t}}^{\text{1L+OS+SS+ML}} = 23.7^{+5.6}_{-5.2}\,\text{fb} \tag{7.7}$$

and a signal significance of 4.71 σ was presented. This shows that evidence for the production of $t\bar{t}t\bar{t}$ is found by ATLAS with a consistent result among all channels. The corresponding limit at 95% is 2.76 times the SM cross-section (33.0 fb).

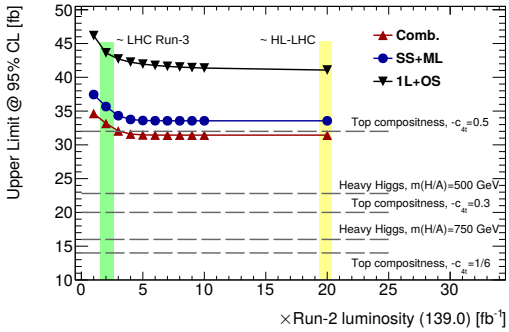
While this analysis aimed at a measurement of the SM $t\bar{t}t\bar{t}$ signal, the excess over the nominal SM prediction motivates the interpretation of the result in scenarios beyond the SM. Within ATLAS, a follow up analysis with interpretations in the context of the scenarios introduced in Section 2.1.2, in particular to study effective field theories, is planned. However, for such a study dedicated MC samples are required as shapes and behaviour in different regions might change. These samples are not available at this stage and therefore the interpretation is limited. Figure 7.1a shows the projected limits, obtained by scaling the statistical component of the uncertainty by the square-root of the multiple of the Run II luminosity, as a function multiples of the Run II luminosity. Furthermore, two theories beyond the standard model for specific parameter setting are compared.
The first theory (Top compositeness) describes the scenario when the SM is a manifestation at the currently available energy scales of a more fundamental theory that cannot be yet observed. In such a theory, the top quark is not an elementary particle but rather a composition of the fundamental fields. Hence, the production of four top quarks can be described by an effective field theory with additional coefficients that reflect new physics. In the given case, an additional parameter c_{4t} is introduced which gives the interaction strength of the $t\bar{t}t\bar{t}$ production in the underlying theory [46]. For example, $c_{4t} = -0.5$ corresponds to the scenario where the four top interaction originates from a singlet resonance but other parameter dependent interpretations are possible.
The second theory (Heavy Higgs) predicts the existence of a second but heavier Higgs boson (H/A) in the Two Higgs Doublet Model [40] (2HDM) at low tanβ, where tanβ is the ratio between the vacuum expectation values of the doublets. In the 2HDM, due

Table 7.1.: Estimated limits and previously set limits on selected BSM theories based on previous ATLAS results [48]. The limits are estimated by comparing the measured cross-section limits in the 1L+OS (combined) result with theory predictions. The limits are estimated for masses of vector-like quarks ($m_{B/T/T_{5/3}}$), the lightest resonance of a Kaluza-Klein tower (m_{KK}) and additional neutral Higgs bosons (m_H) depending on the ratio between the vacuum expectation values of the two Higgs doublets tanβ.

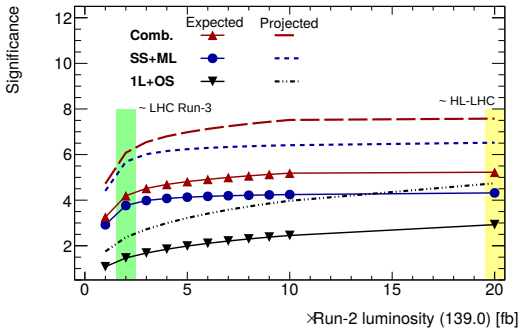
Theory	Theory parameter	Previous limit	Estim. limit 1L+OS (comb)
Vector-like quarks	m_B/m_T	>1 TeV />0.98 TeV	>900 GeV (>1.1 TeV)
	$m_{T_{5/3}}$	>1.19 TeV	>900 GeV (>1.0 TeV)
Extra dimensions	m_{KK}	>1.45 TeV	>1.25 TeV (>1.31 TeV)
2HDM	m_H (tan$\beta = 0.3$)	>700 GeV	>700 GeV (>800 GeV)
	m_H (tan$\beta = 0.5$)	>400 GeV	>500 GeV (>650 GeV)

to the large Yukawa coupling with the top quark, one of the preferred decay modes is, for $m_H > 2 \times m_t$, $H \to t\bar{t}$. A direct resonance search is often not feasible due to the large $t\bar{t}$ background, similar to the 1L+OS channel in this analysis. Therefore, the $t\bar{t}t\bar{t}$ production is a good opportunity to search for such a new particle, in particular in the SS+ML channel due to the smaller background. By comparing the observed limit with the mass-dependent cross-section, a limit on the mass itself can be set. However, in an actual search, more differences in the angular and kinematic distributions between the SM $t\bar{t}t\bar{t}$ signal and the $t\bar{t}$ + H/A production can be exploited.
Similar to the mass of the new boson in the previous case, the cross-section of other BSM theories depend on a variety of exact parameter settings. Therefore, the obtained limits on the cross-section for 139.0fb^{-1} are compared to selected scenarios that were discussed in a previous SS+ML analysis with 36.1 fb^{-1} [48]. By comparing the cross-section limits with theory predictions, provided in the analysis as function of several parameters, limits on the theory parameters can be estimated. This estimation assumes that the response of the final fit is identical for all scenarios and the only difference occurs from the difference in the cross-section. The estimated limits are summarised in Table 7.1. In general, the obtained values agree with the previous limits which were derived more sophisticated, and can therefore be confirmed. For future analyses, improved limits can be expected as already shown in Figure 7.1a.

In a similar fashion, Figure 7.1b shows the expected significance as function of luminosity which is normalised to the Run II dataset of 139 fb^{-1}. In addition, the observed results are projected by assuming that the relative behaviour between observed and expected significance is unchanged. In this idealized approach, the projected significances reflect the scenario when only the statistical uncertainties are reduced by acquiring more data and all other aspects stay unchanged. As reference, an estimate of the luminosity



(a) Projected upper limit 95% CL for each channel and the combination. The limits are obtained by scaling the statistical component of the uncertainty w.r.t. the luminosity. The dashed lines show scenarios for two BSM scenarios with different parameter settings.



(b) Expected and projected significances as function of luminosity normalised to the collected Run-2 luminosity ($139\,\mathrm{fb}^{-1}$) of ATLAS. The projected significance is obtained by scaling the expected significances by the ratio of observed to expected significances for the current dataset.

that is expected in Run III (green area) and for the high luminosity LHC (HL-LHC) are given. For Run III a significant improvement of the result of approximately $2\,\sigma$ can be expected. However, in both figures the curves flatten towards higher luminosities which shows that more advanced techniques are required to reduce systematic effects as the analysis becomes more limited by systematic uncertainties.

For ATLAS in general, one of these improvements will be the implementation of particle-flow jets to improve the precision of jet related measurements and a corresponding reduction of jet related uncertainties. Some possible improvements to the fit setup were already discussed in Section 6.2.1. However, more fundamental changes to the fit setup are possible such as new developments from theorists to improve the modelling of the $t\bar{t}$ background and their systematic uncertainties. This would be reflected by an improved modelling of the MC generators making it, ideally, unnecessary to derive additional scale factors. However, typically these developments take time to be implemented and are not expected for Run III. Therefore, in the near future, it will likely be still necessary for analysers to derive a reliable background modelling, for example, by deriving additional scale factors as presented here.

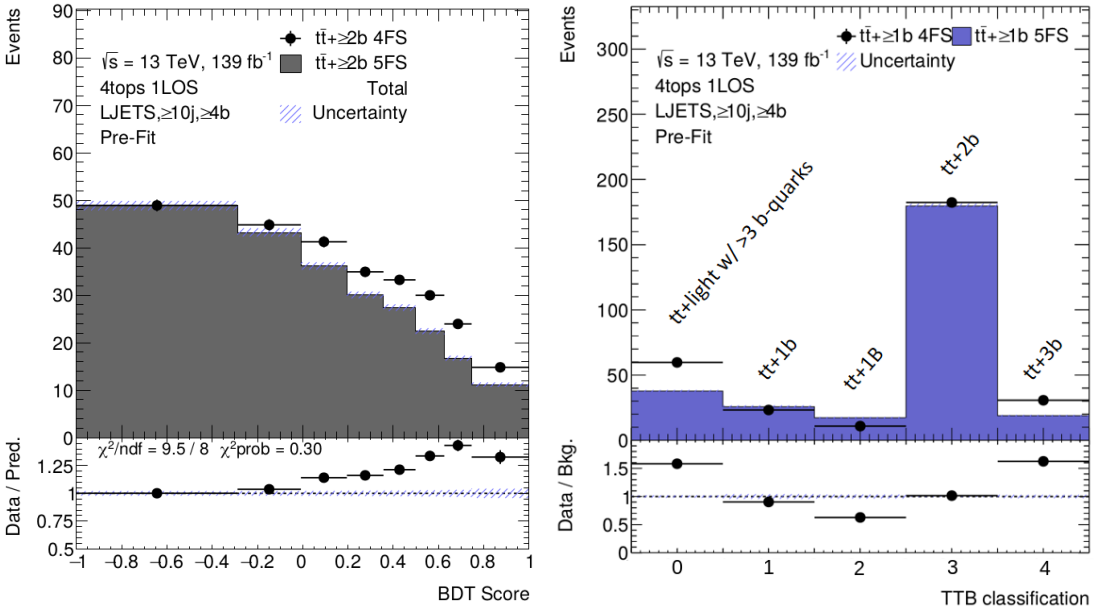
One disadvantage of the method presented is the complexity and the amount of additional scale factors with the corresponding uncertainties. This adds many parameters to the fit and makes the setup slow, and therefore more difficult to study, when all uncertainties are included. As shown in the combination, it is difficult to combine the approach reliably without dedicated additional studies with the SS+ML channel. While this issue is possibly of technical nature, a possibility to simplify the setup is exploited by ATLAS in a search for heavy Higgs-bosons. The analysis studies the usage of a neural network to derive the additional scale factors. This approach uses the same but po-

tentially even more input variables while reducing the output to a single event-weight instead of three sequential weights. Furthermore, the uncertainty is obtained as single variation of the event weight instead of 102 parameters. However, no results are available at the moment as the heavy Higgs analysis is in an early stage.

As discussed, $H_{\mathrm{T}}^{\mathrm{all}}$ has been used in lower (b)-jet multiplicities to exploit shape differences that are not as visible in the BDT even though the separation is slightly worse. This approach can be developed further by, for example, considering the invariant masses of the three leading b-jets in these regions. This variable shows a different behaviour for the $t\bar{t}$ $+\geq 1b$ sub-components and could therefore help to reduce the large constraints and the high correlation with the signal strength of the $t\bar{t}$ $+\geq$ 2b component which shows up in all pseudodata and data fits. This idea can further be exploited by adjusting the splitting of the regions with exactly three b-jets into 3bL and 3bH. For example, more working-points of the b-tagger can be used to define more control regions. On the contrary, at some point, the splitting into regions is limited by statistics. Therefore, in parallel, the binning of these distributions has to be evaluated. At the moment, all regions use the same binning which, due to the difference in statistics, can be improved further.

Another difficulty regarding b-jets is that the analysis has to deal with b-jets from hard-processes and from soft-radiation. The modelling of these processes is based on different steps during the generation of the MC samples but cannot easily be further distinguished by the detector. In this analysis, the five flavour scheme in which only the top-quark is treated as massive quark, is used as this simplifies theoretical calculations and is the default in many Monte-Carlo generators. On the contrary, the four flavour scheme includes the b-quark in the PDFs and is often better suited in resolved events. Figure 7.2 (a) shows a comparison of the two schemes in the Powheg + Pythia sample in the BDT-score in the $t\bar{t}$ $+\geq$ 2b sub-categories for more than ten jets and more than four b-jets. While a clear difference is observed it should be noted that both distributions use the scale factors that have been derived in the five flavour scheme which might not be the correct ones to apply to the four flavour scheme. It is currently investigated how to included this observation in the analysis. Possible solutions are to include an additional uncertainty to cover this effect or to re-train the BDT for the four flavour scheme to have a dedicated distribution to include in the fit. Furthermore, it is possible to define a control region to measure the effect which seems to mostly originate from $t\bar{t}$ +3b events as shown in Figure 7.2(b) where the different sub-categories are shown in the highest (b-)jet multiplicity. In a longer perspective, a combination of the four and five flavour scheme would be advantageous to provide good modelling for resolved and collinear events. While recent developments show promising results it will take time until these approaches are available by default in MC generators.

The result shows that it is possible to study the $t\bar{t}t\bar{t}$ process at the Lhc. Due to limitations of the MC generators new techniques have to be developed to obtain a reliable



(a) (b)

Figure 7.2.: Comparison of the four (4FS) and five flavour scheme (5FS) in the BDT-score (a) and in normalisation for $t\bar{t}$ $+\geq$ $1b$ sub-categories in the highest (b-)jet multiplicity for the 1LOS channel. The uncertainty band includes the statistical component.

MC prediction. The MC-based presented in this theses is one possible approach to do so and can be adapted to other analyses that work in a similar phase-space. The increased signal in the combined result motivates to study the process with more data and future versions of MC generators to limit possible scenarios of physics beyond the SM.

Bibliography

[1] J. J. Thomson, *XL. Cathode Rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1897) 293–316.

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

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

[4] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2008.

[5] D. Griffiths, *Introduction to Elementary Particles*. WILEY-VCH Verlag, 2008.

[6] M. Gell-Mann, *The interpretation of the new particles as displaced charge multiplets*, Il Nuovo Cimento (1955-1965) **4** (1956) 848–866.

[7] T. Nakano and K. Nishijima, *Charge Independence for V-particles*, Progress of Theoretical Physics **10** (1953) 581–582.

[8] Gell-Mann, M., *THE EIGHTFOLD WAY: A THEORY OF STRONG INTERACTION SYMMETRY*, California Inst. of Tech,. Pasadena. Synchtrotron Lab (1961).

[9] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533.

[10] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, PThPh **49** (1973) 652–657.

[11] E. Fermi, *Versuch einer Theorie der β-Strahlen. I*, Z.Phys **88** (1934) 161–177.

[12] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[13] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509.

[14] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, Phys. Rev. Lett. **13** (1964) 585–587.

[15] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22** (1961) 579 – 588.

[16] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[17] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (1998) 1562–1567.

[18] SNO Collaboration, *Measurement of the rate of $\nu_e + d \to p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301.

[19] SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (2002) 011301.

[20] Planck Collaboration, *Planck 2013 results*, Astron. Astrophys. **A16** (2014).

[21] J. H. Christenson, et al., *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13** (1964) 138–140.

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

[23] 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.

[24] 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.

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

[26] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001, and 2019 updates.

[27] G. Aad, et al., *Measurement of the $t\bar{t}t$ production cross-section in the lepton+jets channel at $\sqrt{s}$=13 TeV with the ATLAS experiment*, Phys. Lett. B **810** (2020) 135797.

[28] ATLAS collaboration, *Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment*, Tech. Rep. ATLAS-CONF-2019-044, CERN, Geneva, Sep, 2019. https://cds.cern.ch/record/2690717.

[29] 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.

[30] 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.

[31] 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.

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

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

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

[35] 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.

[36] 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.

[37] ATLAS Collaboration, *ATLAS Top Working Group Summary Plots*, 2019, https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/TOP/, Accessed: 20/05/2020.

[38] 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.

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

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

[41] T. Plehn and T. M. P. Tait, *Seeking sgluons*, J. Phys. G: Nucl. Part. Phys **36** (2009) 075001.

[42] M. Guchait, F. Mahmoudi, and K. Sridhar, *Associated production of a KaluzaKlein excitation of a gluon with a tt pair at the LHC*, Phys. Lett. B**666** (2008) 347 – 351.

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

[44] A. De Simone, et al., *A first top partner hunters guide*, JHEP **2013** (2013) 4.

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

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

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

[48] 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.

[49] 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.

[50] ATLAS Collaboration, *Combined search for four-top quarks production in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, techreport ATL-COM-PHYS-2018-269, CERN, Geneva, Mar, 2018. `https://cds.cern.ch/record/2310185`.

[51] 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** (2020) 75.

[52] ATLAS Collaboration, *Evidence for* $t\bar{t}t\bar{t}$ *production in the multilepton final state in proton-proton collisions with the ATLAS detector*, ATLAS-CONF-2020-013 (2020), in process of being published (26.05.2020).

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

[54] K. Johnsen, *CERN Intersecting Storage Rings (ISR)*, in *Proc. Natl. Acad. Sci. U.S.A.*, p. , 619626, National Academy of Sciences Autumn Meeting. 1973.

[55] CERN, *Beam-Beam Effects in the SPS Proton-Anti Proton Collider*. CERN, Geneva, 2014. `https://cds.cern.ch/record/1550757`. published as CERN Yellow Report CERN-2014-004.

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

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

[58] C. Beckmann, *Proceedings of the 1990 linear accelerator conference*, Linear accelerator conference, Albuquerque, NM (USA), 9-14 (1991).

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

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

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

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

[63] ATLAS Collaboration Collaboration, ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1*. Technical Design Report ATLAS. CERN, Geneva, 1999. `https://cds.cern.ch/record/391176`.

[64] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Tech. Rep. CERN-LHCC-2010-013. ATLAS-TDR-19, Sep, 2010. `https://cds.cern.ch/record/1291633`.

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

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

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

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

[69] 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.

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

[71] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015.*, Eur. Phys. J. C **77** (2017) 317.

[72] J. C. Collins, D. E. Soperand , G. Sterman, *Factorization of hard processes in QCD*, Advanced Series on Directions in High Energy Physics **5** (1989) 1–91.

[73] A. Buckley, et al., *General-purpose event generators for LHC physics*, Phys. Rep. **504** (2011) 145 – 233.

[74] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1996.

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

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

[77] T. Sjöstrand, et al., *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159 – 177.

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

[79] S. Plätzer and S. Gieseke, *Dipole showers and automated NLO matching in Herwig++*, Eur. Phys. J. C **72** (2012) 2187.

[80] S. Agostinelli, et al., *Geant4a simulation toolkit*, Nucl. Instrum. Methods. Phys. Res. B **506** (2003) 250 – 303.

[81] 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`.

[82] 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`.

[83] 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.

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

[85] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV datas*, Tech. Rep. ATL-PHYS-PUB-2014-021, CERN, Geneva, Nov, 2014. `https://cds.cern.ch/record/1966419`.

[86] 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.

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

[88] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Methods. Phys. Res. B **462** (2001) 152 – 155, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

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

[90] 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.

[91] 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.

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

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

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

[95] 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.

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

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

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

[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*, tech. rep., CERN Yellow Reports: Monographs CERN-2017-002, 2017.

[102] S. Höche, et al., *A critical appraisal of NLO+PS matching methods*, JHEP **2012** (2012) 49.

[103] S. Höche, et al., *QCD matrix elements + parton showers. The NLO case*, JHEP **2013** (2013) 27.

[104] S. Catani, et al., *QCD Matrix Elements + Parton Showers*, JHEP **2001** (2001) 063–063.

[105] S. Höche, et al., *QCD matrix elements and truncated showers*, JHEP **2009** (2009) 053–053.

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

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

[108] 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.

[109] 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`.

[110] T. G. Cornelissen, et al., *The global χ^2 track fitter in ATLAS*, JHEP **119** (2008) 032013.

[111] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Methods. Phys. Res. B **262** (1987) 444 – 450.

[112] 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.

[113] 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.

[114] ATLAS Collboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490.

[115] ATLAS Collboration, *Jet energy scale and resolution measured in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2019-281, CERN, Geneva, Apr, 2019. https://cds.cern.ch/record/2669655.

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

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

[118] ATLAS Collboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at* $\sqrt{s} = 13$ TeV *with the ATLAS detector*, Phys. Rev. D **96** (2017) 072002.

[119] ATLAS Collboration, *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.

[120] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C**77** (2017) 466.

[121] B. Nachman, et al., *Jets from jets: re-clustering as a tool for large radius jet reconstruction and grooming at the LHC*, JHEP **2015** (2015) 75.

[122] D. Krohn, J. Thaler, and L.-T. Wang, *Jet trimming*, JHEP **2010** (2010) 84.

[123] ATLAS Collaboration, *Jet reclustering and close-by effects in ATLAS run II*, Tech. Rep. ATLAS-CONF-2017-062, CERN, Geneva, Jul, 2017. https://cds.cern.ch/record/2275649.

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

[125] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, Tech. Rep. ATL-PHYS-PUB-2017-013, CERN, Geneva, Jul, 2017. https://cds.cern.ch/record/2273281.

[126] A. Hoecker, et al., *TMVA - Toolkit for Multivariate Data Analysis*, 2007, arXiv:physics/0703039 [physics.data-an].

[127] 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.

[128] 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.

[129] 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.

[130] 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.

[131] D. Adams, et al., *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*, Tech. Rep. ATL-PHYS-INT-2014-018, CERN, Geneva, Jul, 2014. https://cds.cern.ch/record/1743654.

[132] ATLAS Collaboration, *Trigger menu in 2018*, Tech. Rep. ATL-DAQ-PUB-2019-001, CERN, Geneva, Oct, 2019. https://cds.cern.ch/record/2693402.

[133] 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.

[134] S. Catani and M. Seymour, *A general algorithm for calculating jet cross sections in NLO QCD*, Nucl. Phys. B **485** (1997) 291 – 419.

[135] S. Catani, et al., *The dipole formalism for next-to-leading order QCD calculations with massive partons*, Nucl. Phys. B **627** (2002) 189 – 265.

[136] 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 **2016** (2016) 160.

[137] 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.

[138] ATLAS Collaboration, *TRexFitter Software Package*, https://trexfitter-docs.web.cern.ch/trexfitter-docs/.

[139] 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 $s\sqrt{s}$= 13 TeV with the ATLAS detector*, JHEP **2019** (2019) 46.

[140] F. Demartin, et al., *tWH associated production at the LHC*, Eur. Phys. J. **77** (2017) 34.

[141] M. Aaboud, et al., *Measurement of the production cross-section of a single top quark in association with a Z boson in protonproton collisions at 13 TeV with the ATLAS detector*, Physics Letters B **780** (2018) 557 – 577.

[142] Z. Marshall, *Simulation of Pile-up in the ATLAS Experiment*, Tech. Rep. ATL-SOFT-PROC-2013-030, CERN, Geneva, Oct, 2013. `https://cds.cern.ch/record/1616394`.

[143] ATLAS Colaboration, *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. `http://cds.cern.ch/record/2693930`.

[144] 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*, JHEP **1911** (2019) 082. 49 p.

Appendices

APPENDIX A

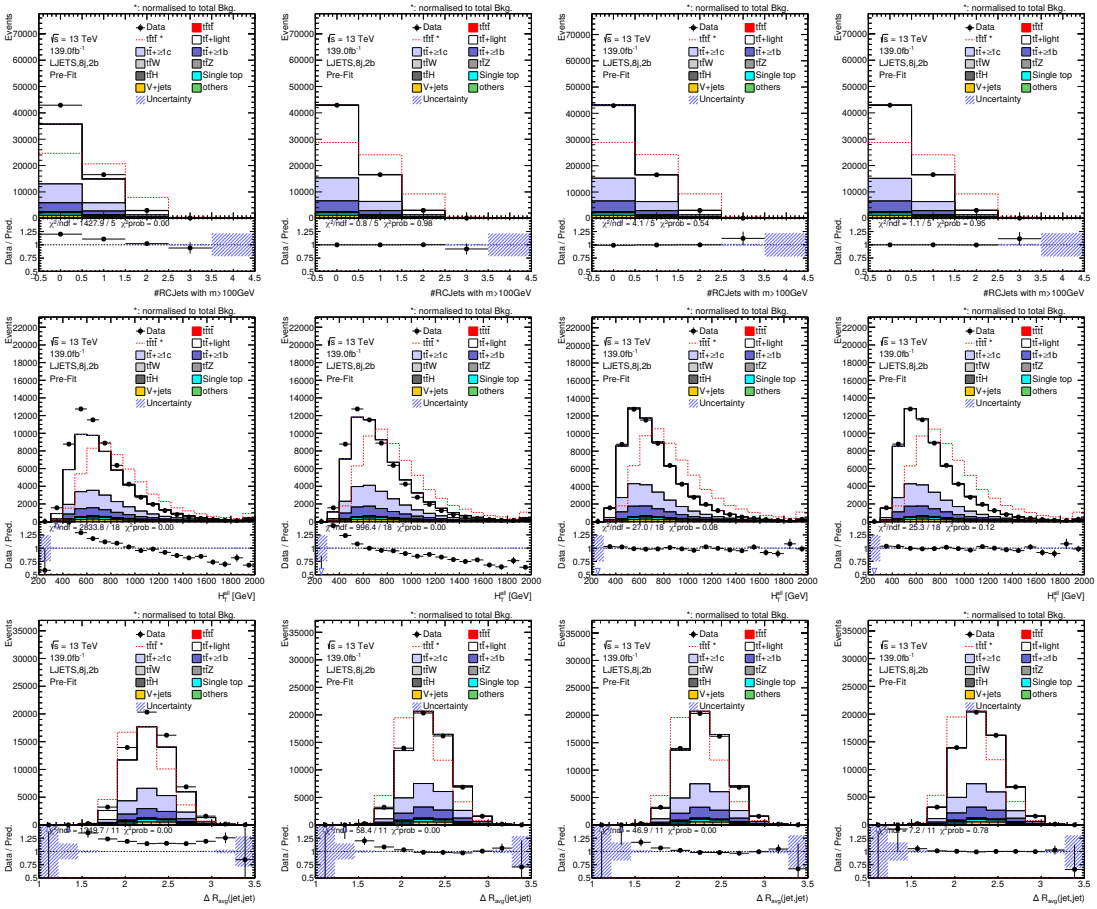
Validation Reweighting Factors

In this appendix, additional plots for the validation of the reweighting factors are shown.

A.1. Validation Sequential Reweighting

Selected variables after each reweighting step to verify that later steps do not significantly impact already corrected variables.

2 b-jet regions



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

Figure A.1.: Distributions of reweighting variables in the in the 1L channel for 8j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.

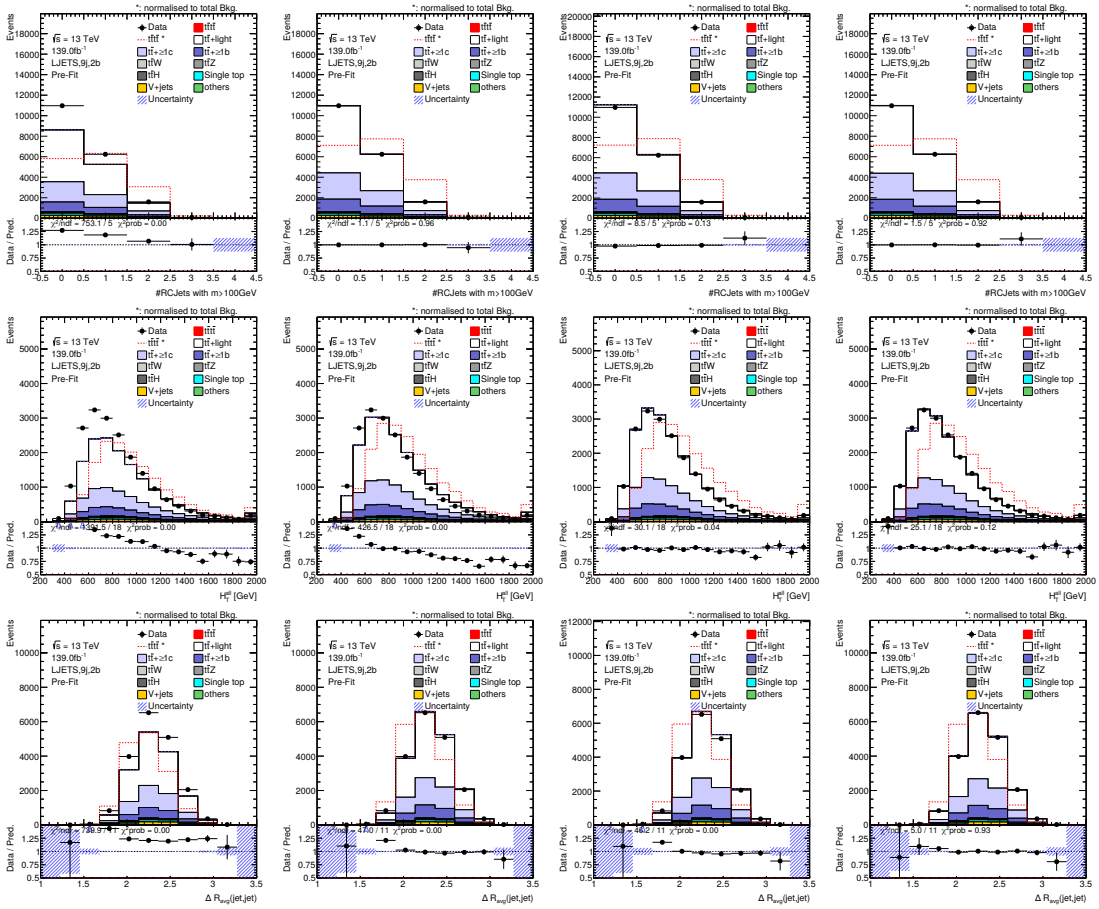
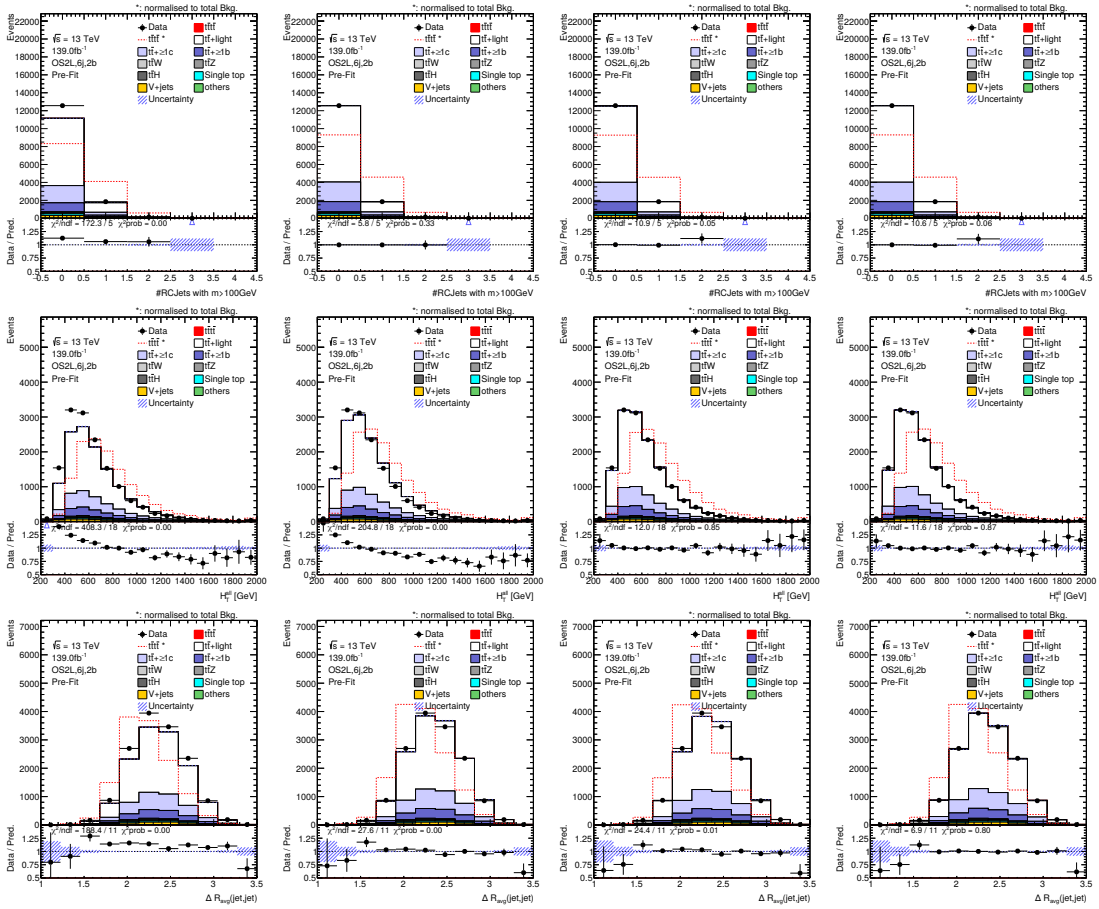
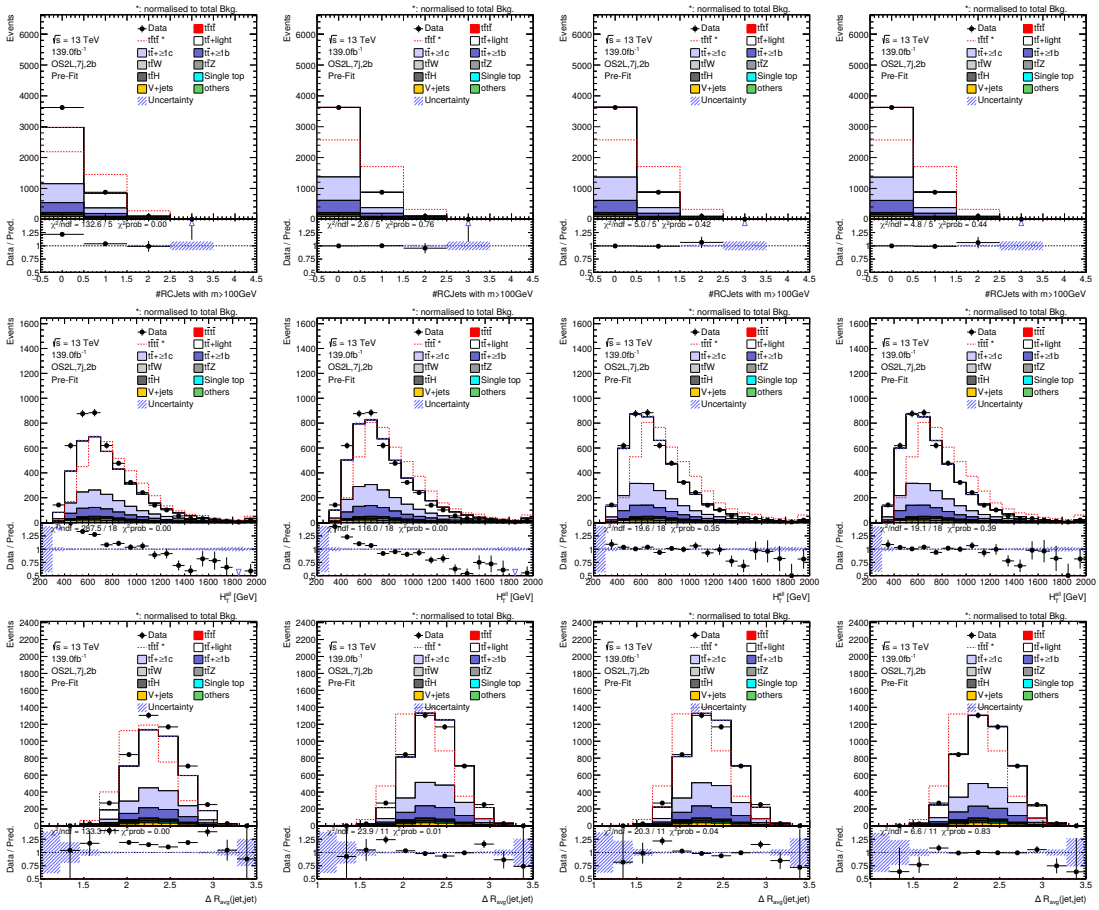


Figure A.2.: Distributions of reweighting variables in the 1L channel for 9j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



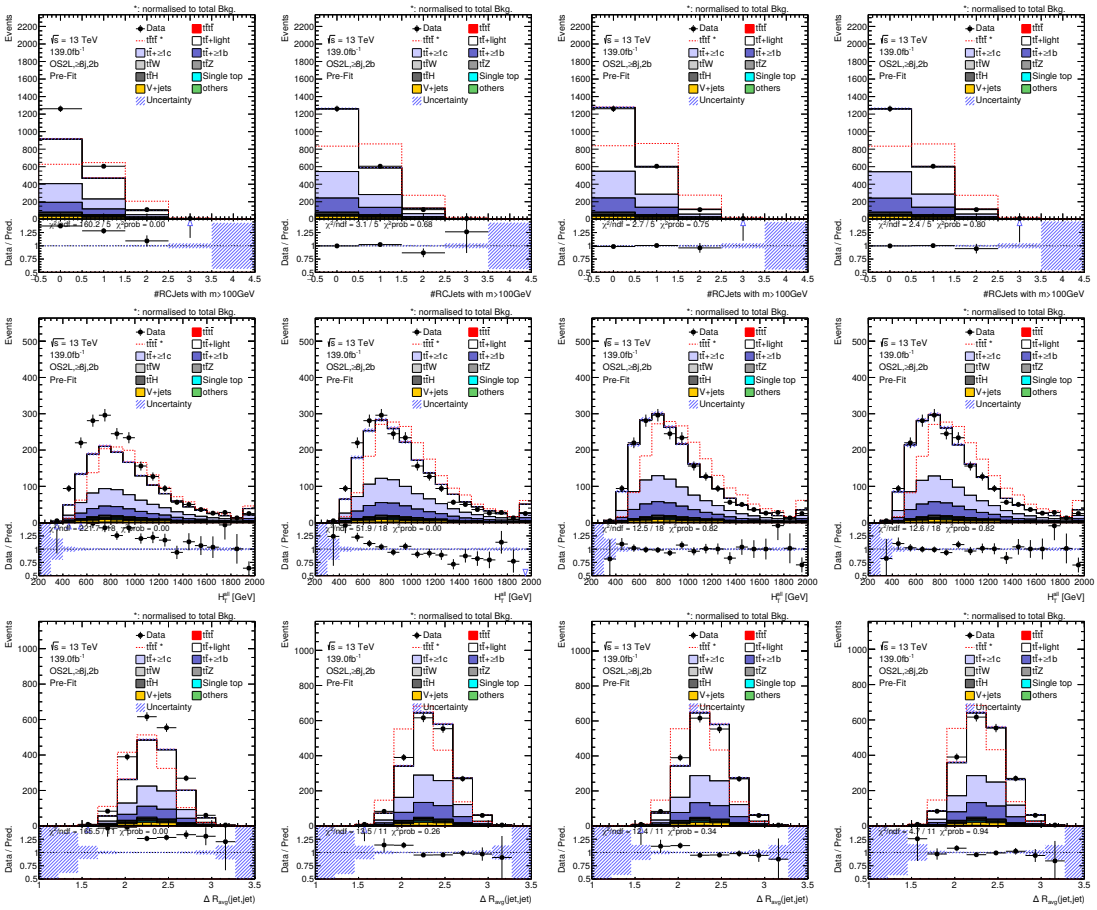
(a) (b) (c) (d)

Figure A.3.: Distributions of reweighting variables in the OS channel for 6j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



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

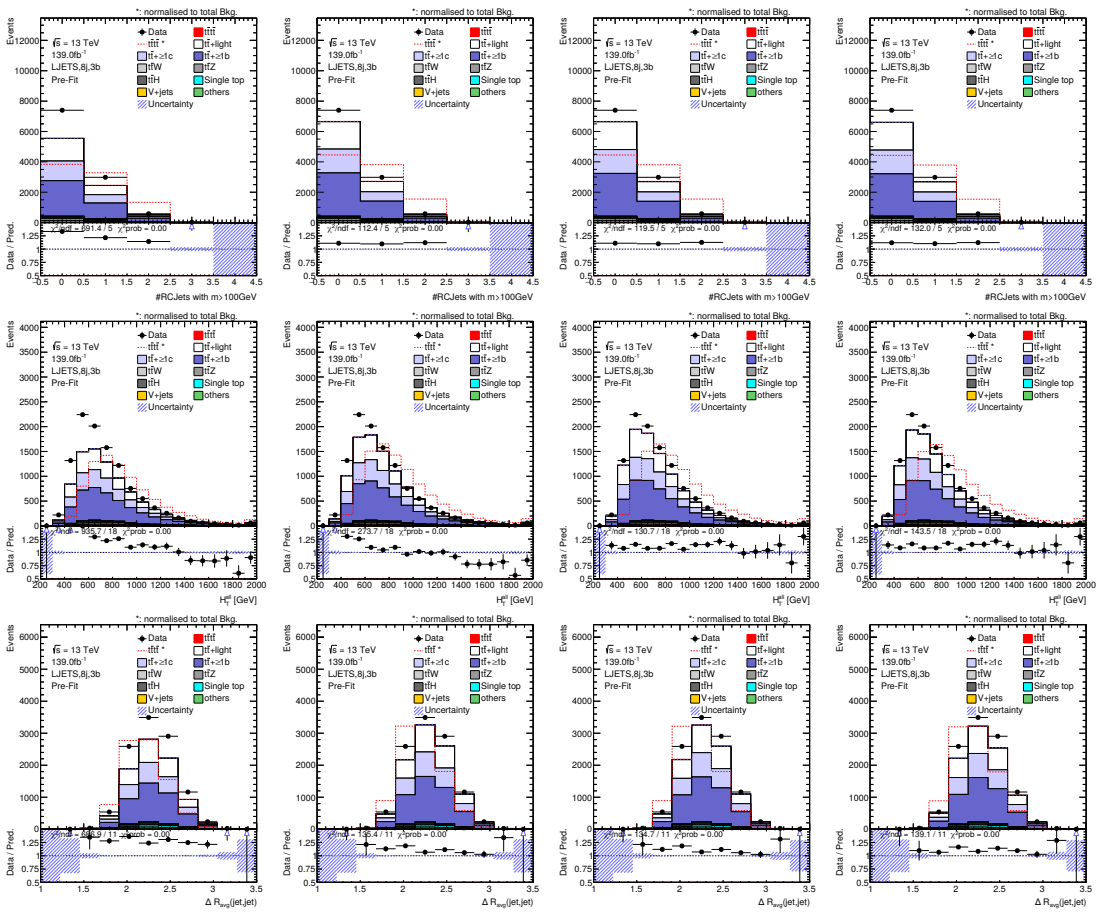
Figure A.4.: Distributions of reweighting variables in the OS channel for 7j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



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

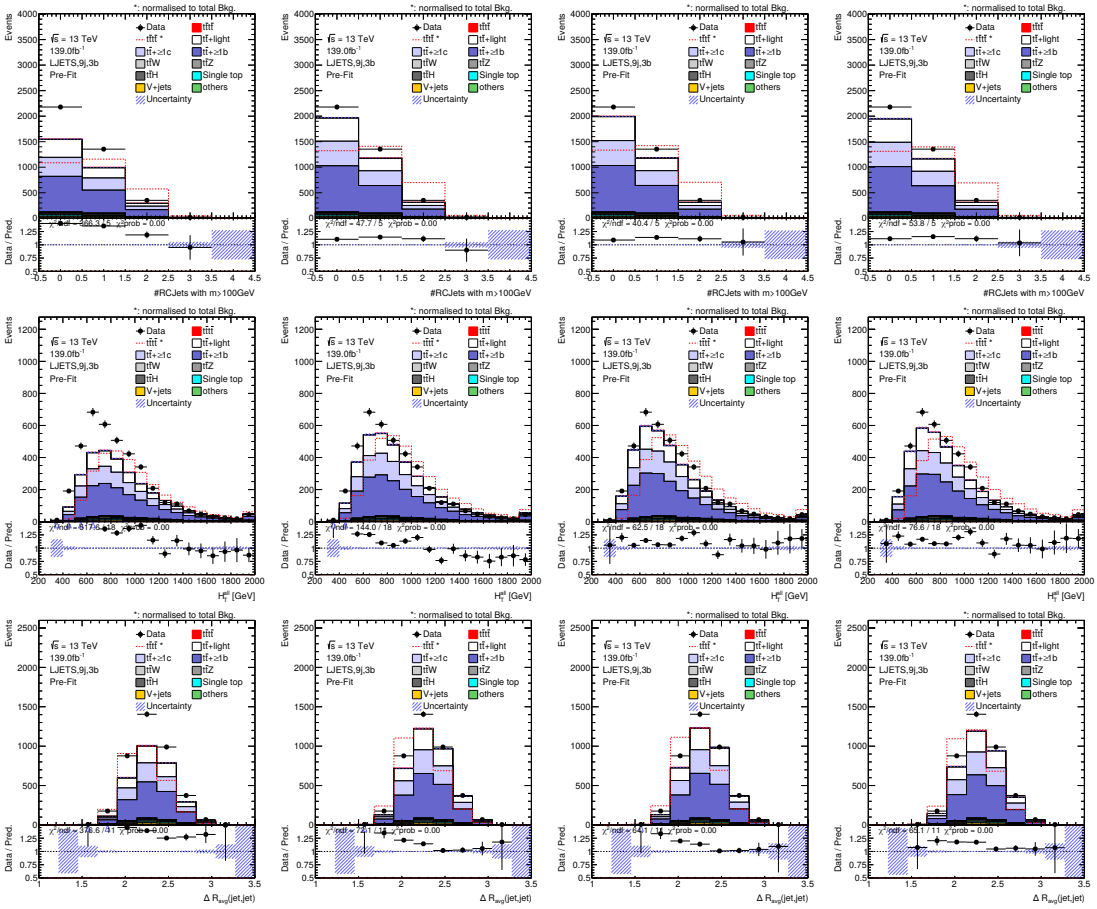
Figure A.5.: Distributions of reweighting variables in the OS channel for ge8j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.

3 b-jet regions



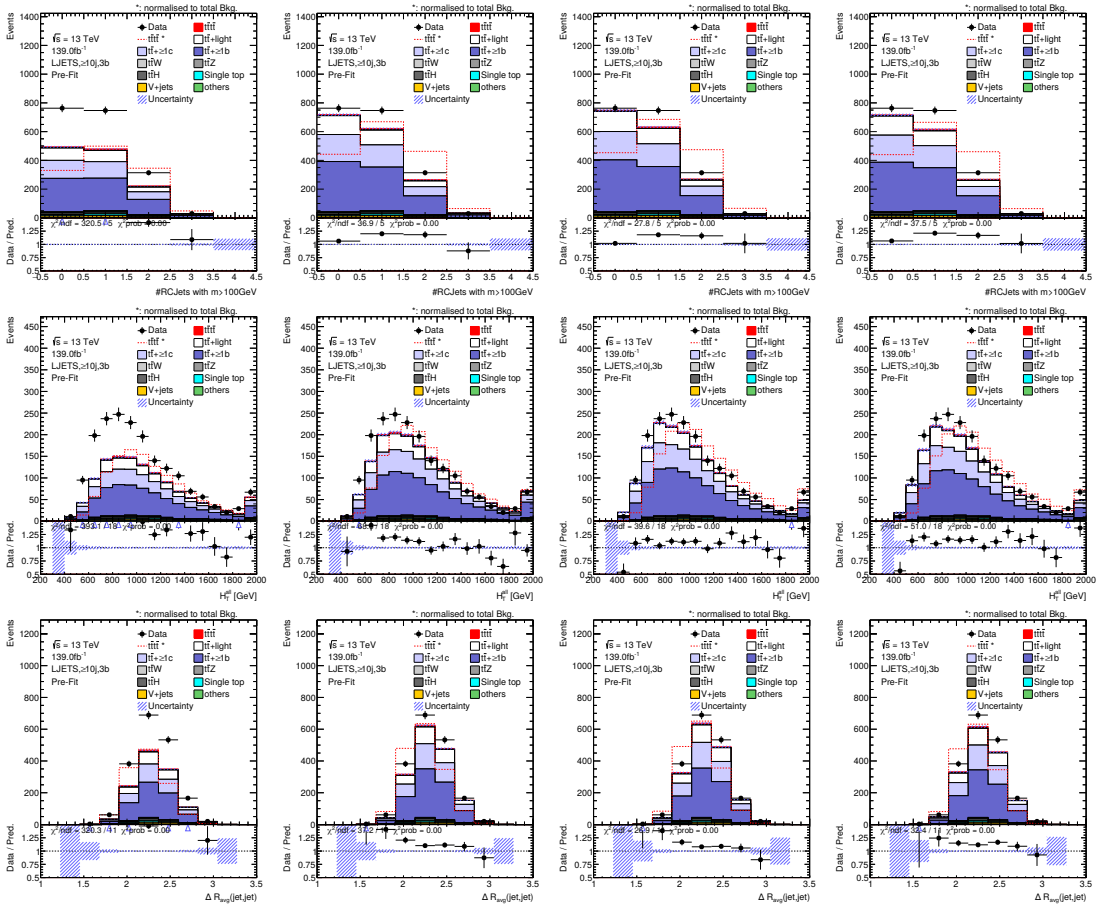
(a) (b) (c) (d)

Figure A.6.: Distributions of reweighting variables in the in the 1L channel for 8j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



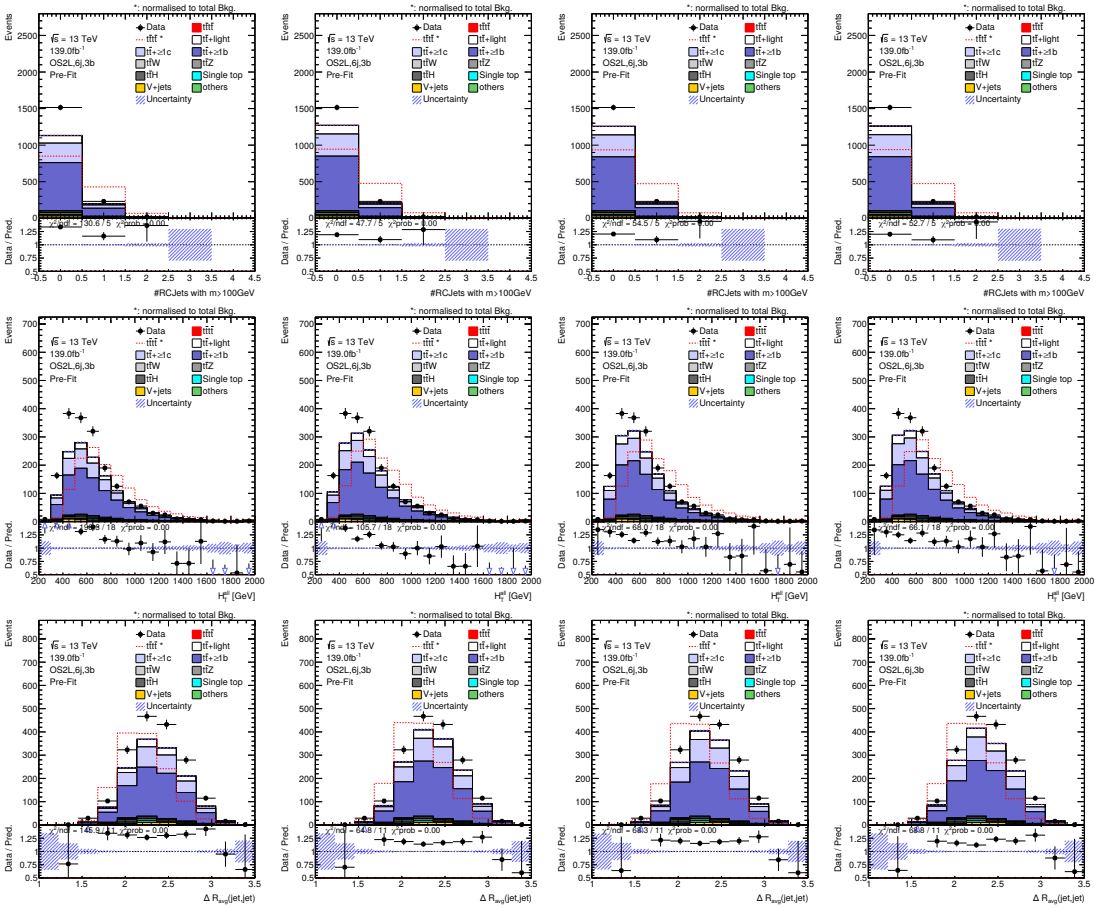
(a) (b) (c) (d)

Figure A.7.: Distributions of reweighting variables in the 1L channel for 9j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



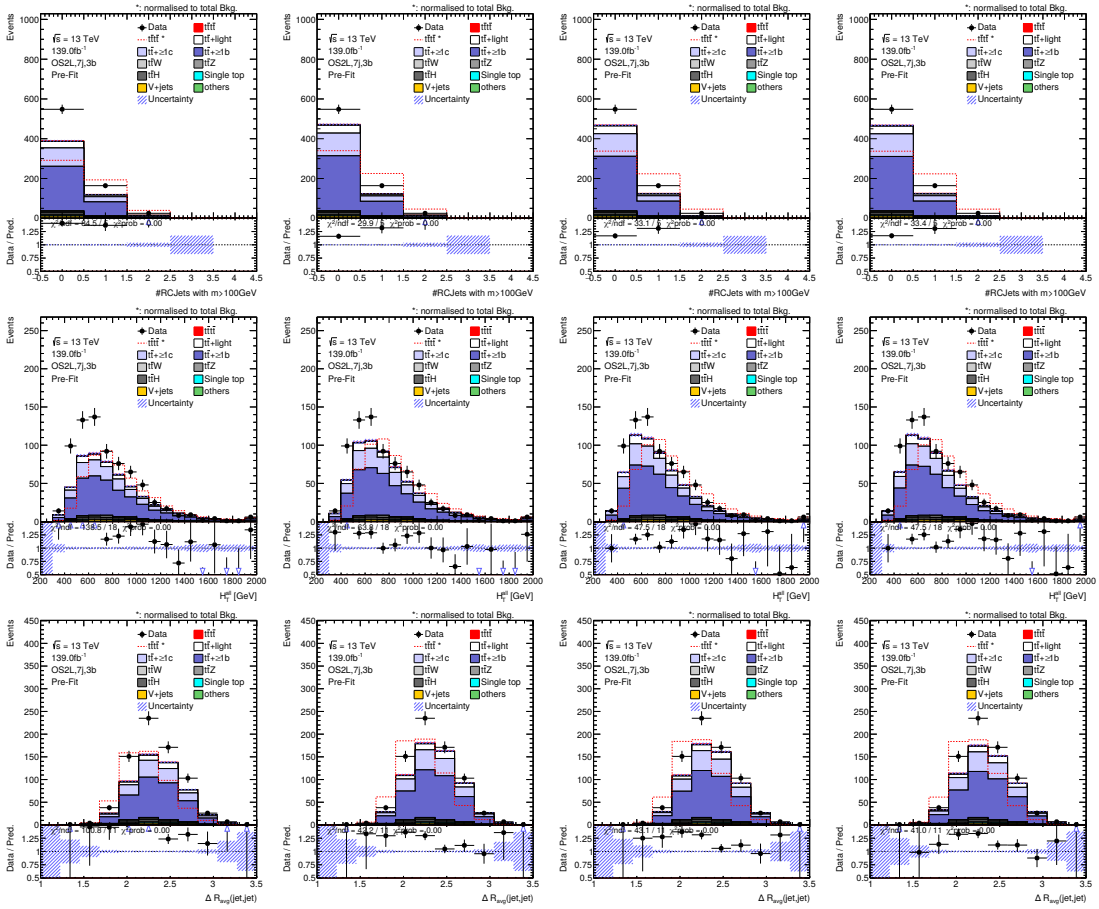
(a) (b) (c) (d)

Figure A.8.: Distributions of reweighting variables in the 10j inclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



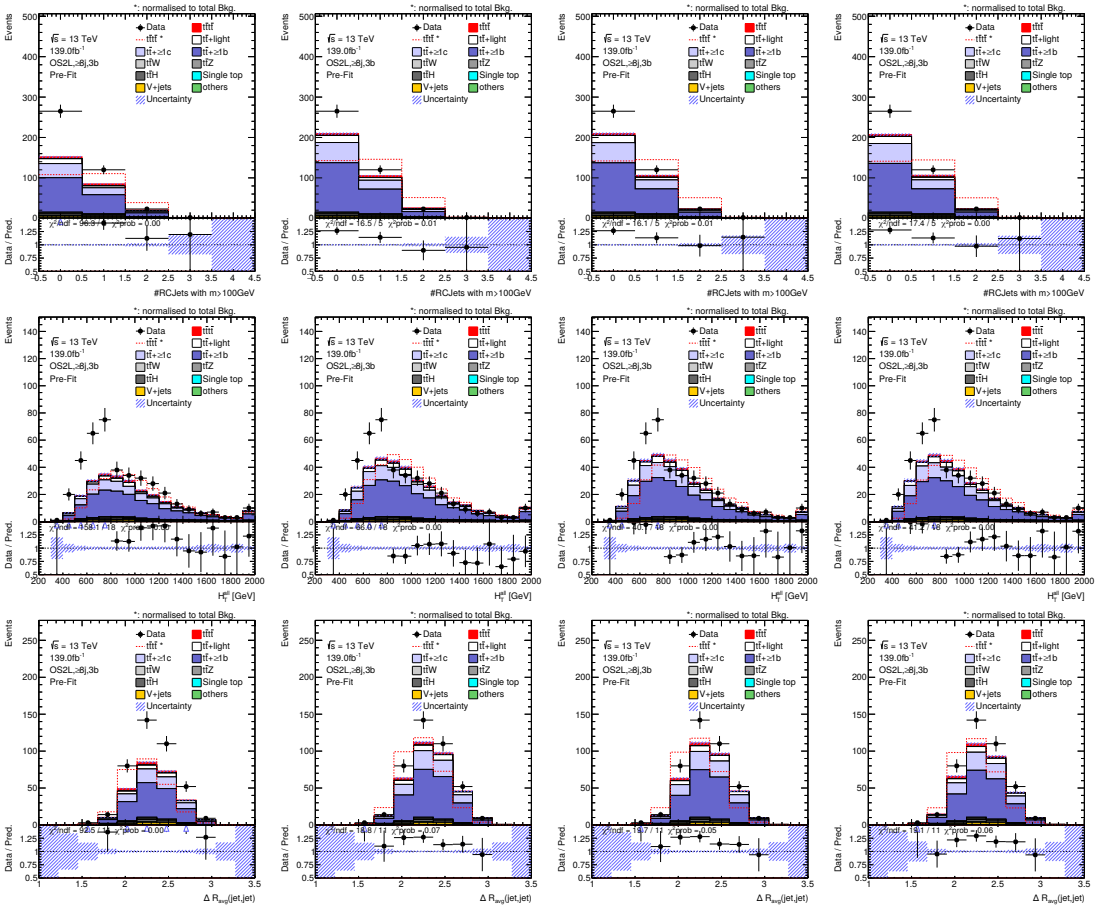
(a) (b) (c) (d)

Figure A.9.: Distributions of reweighting variables in the OS channel for 6j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



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

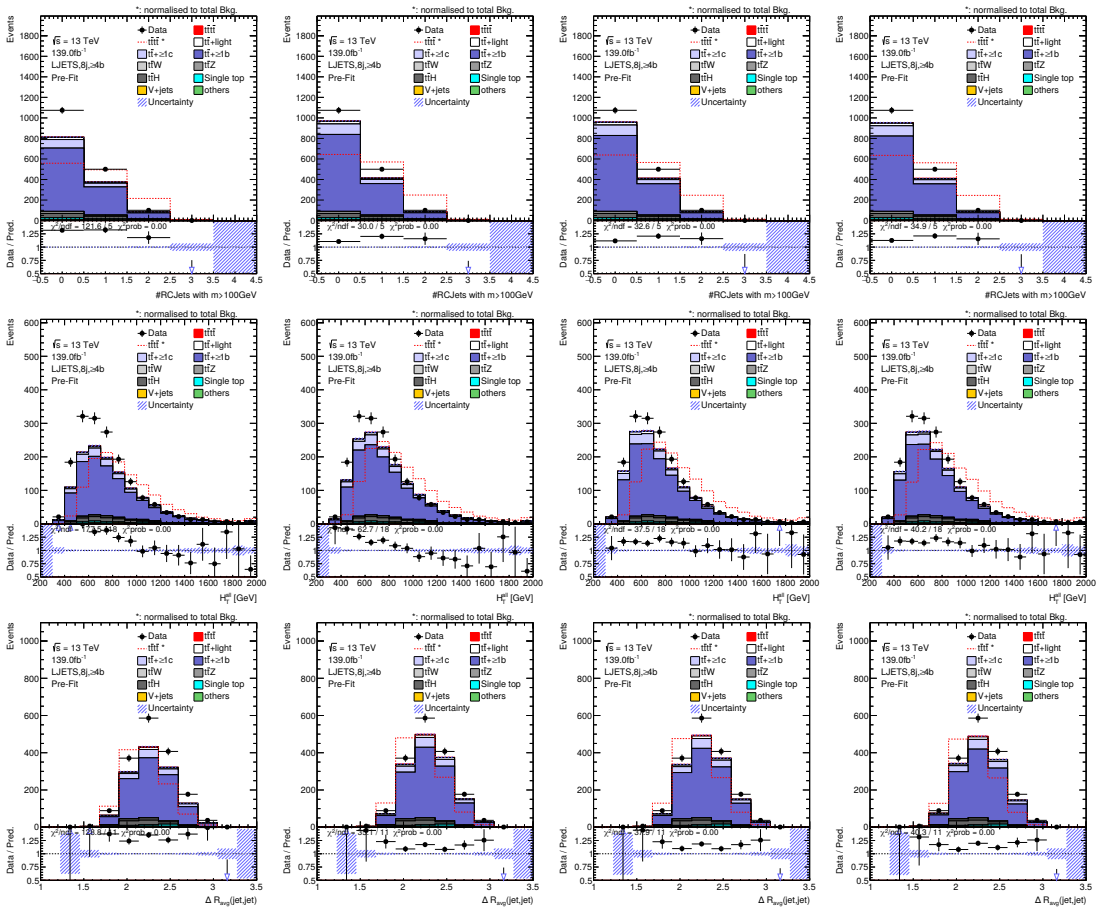
Figure A.10.: Distributions of reweighting variables in the OS channel for 7j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



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

Figure A.11.: Distributions of reweighting variables in the OS channel for ge8j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.

More than 4b regions



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

Figure A.12.: Distributions of reweighting variables in the in the 1L channel for 8j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.

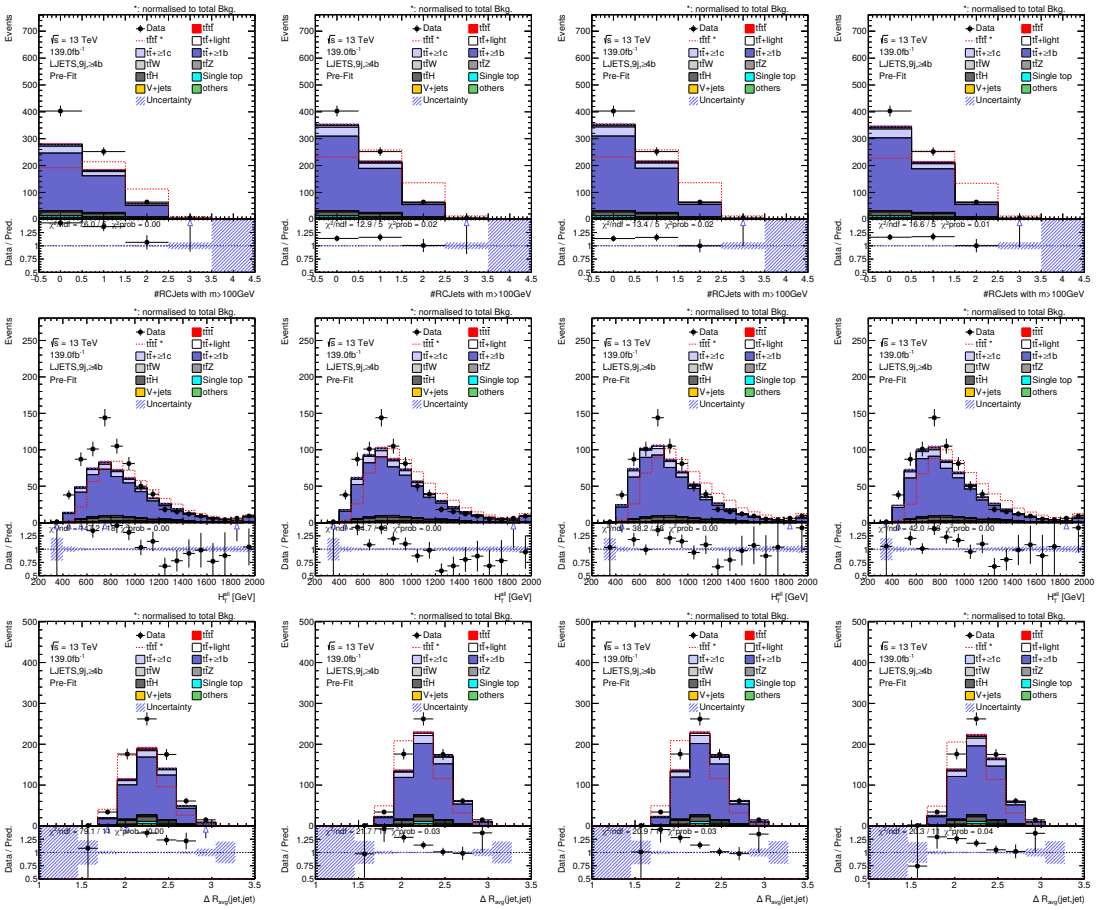
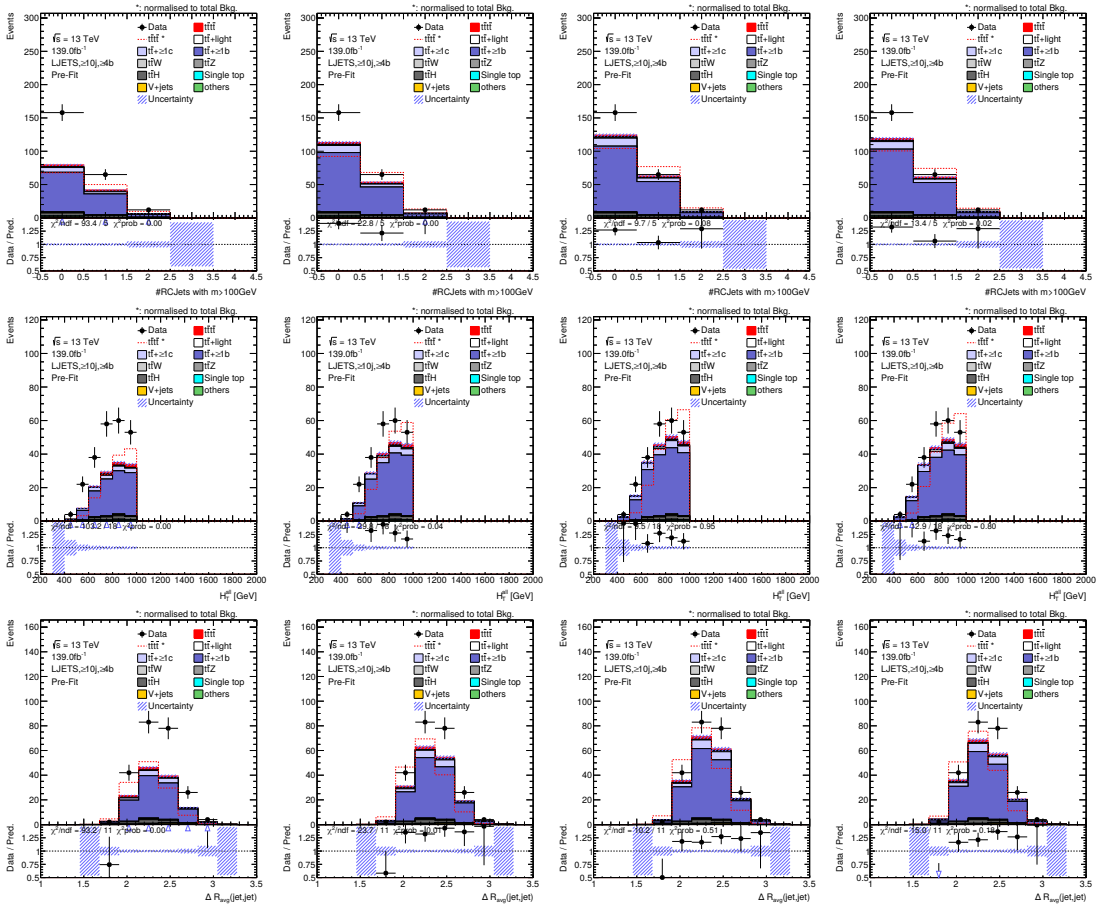
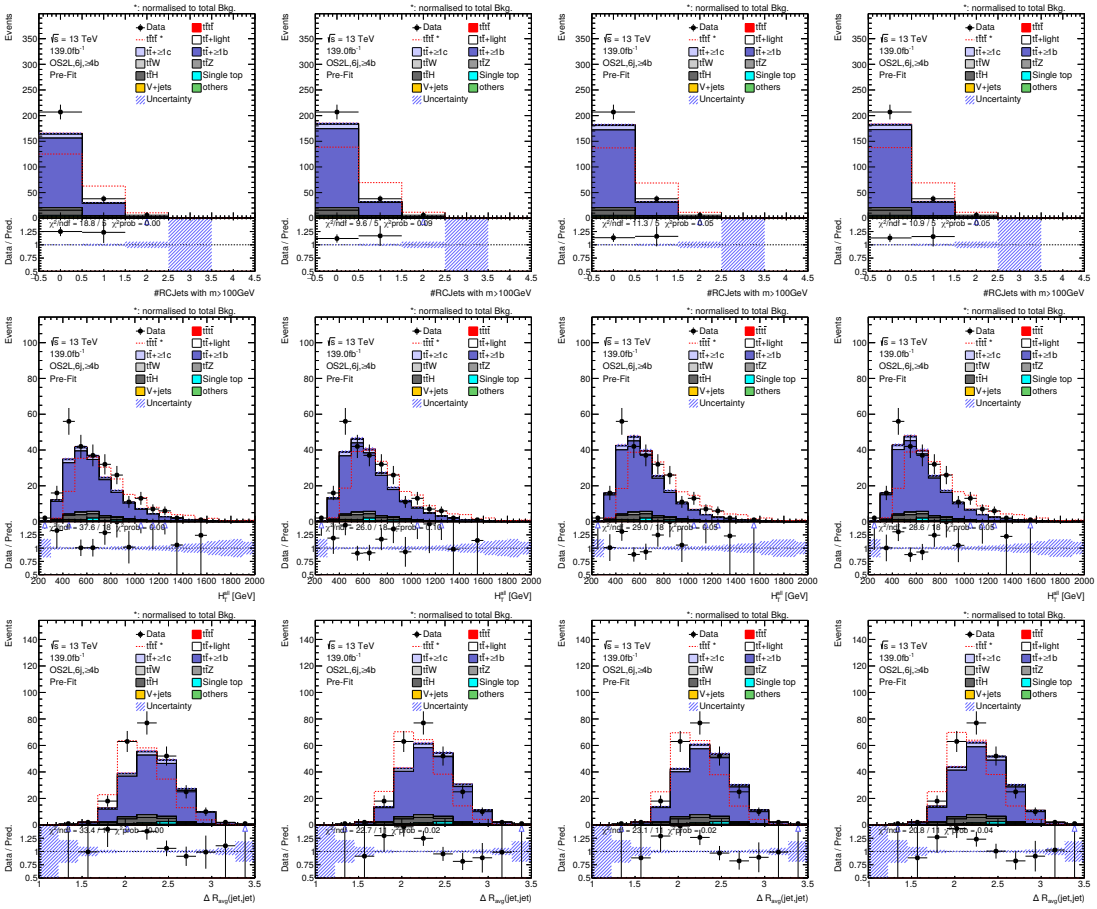


Figure A.13.: Distributions of reweighting variables in the 1L channel for 9j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



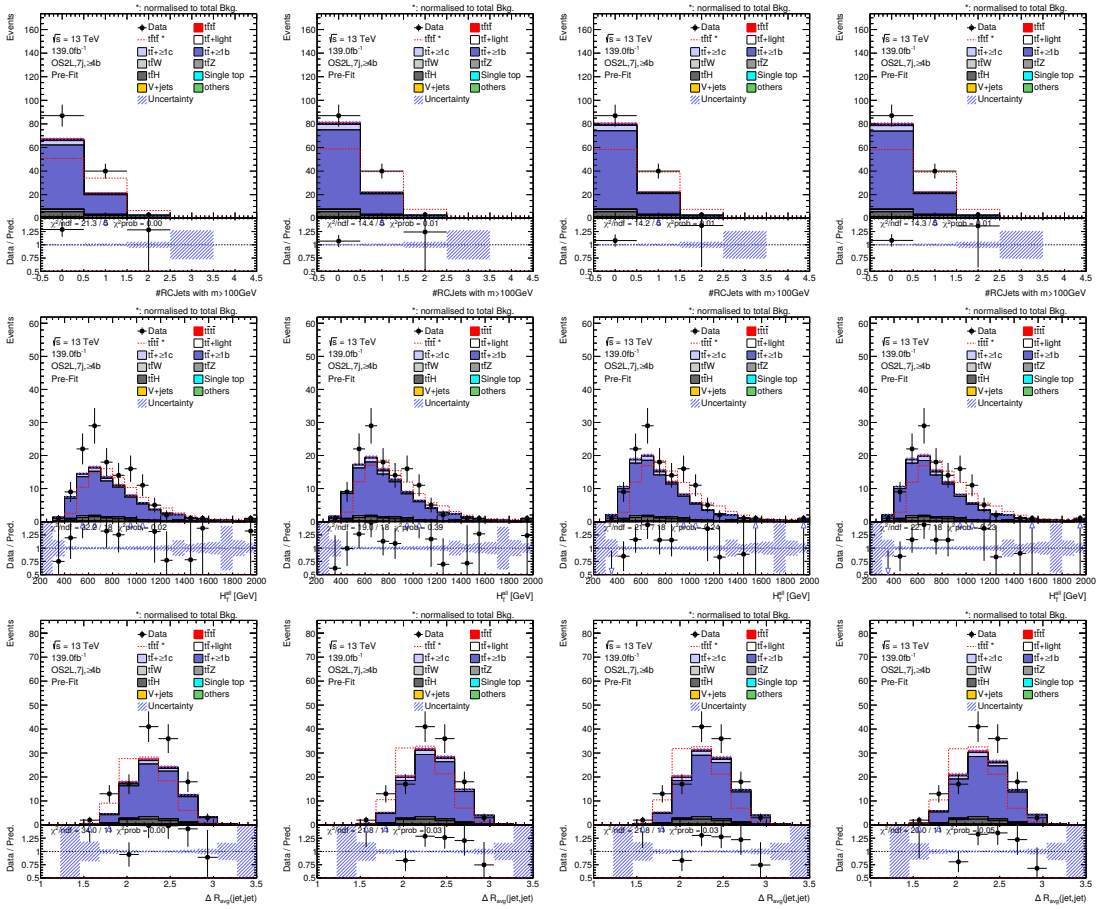
(a) (b) (c) (d)

Figure A.14.: Distributions of reweighting variables in the 10j inclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



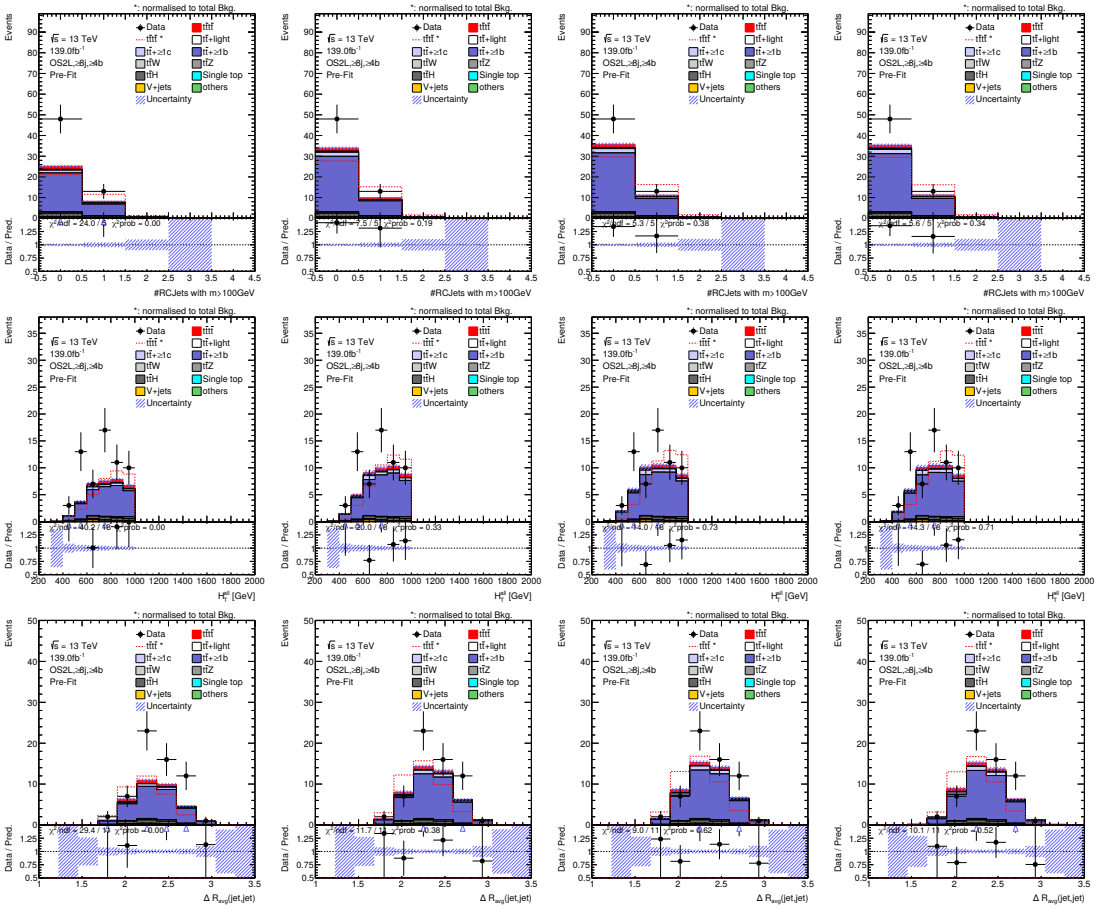
(a) (b) (c) (d)

Figure A.15.: Distributions of reweighting variables in the OS channel for 6j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



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

Figure A.16.: Distributions of reweighting variables in the OS channel for 7j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.



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

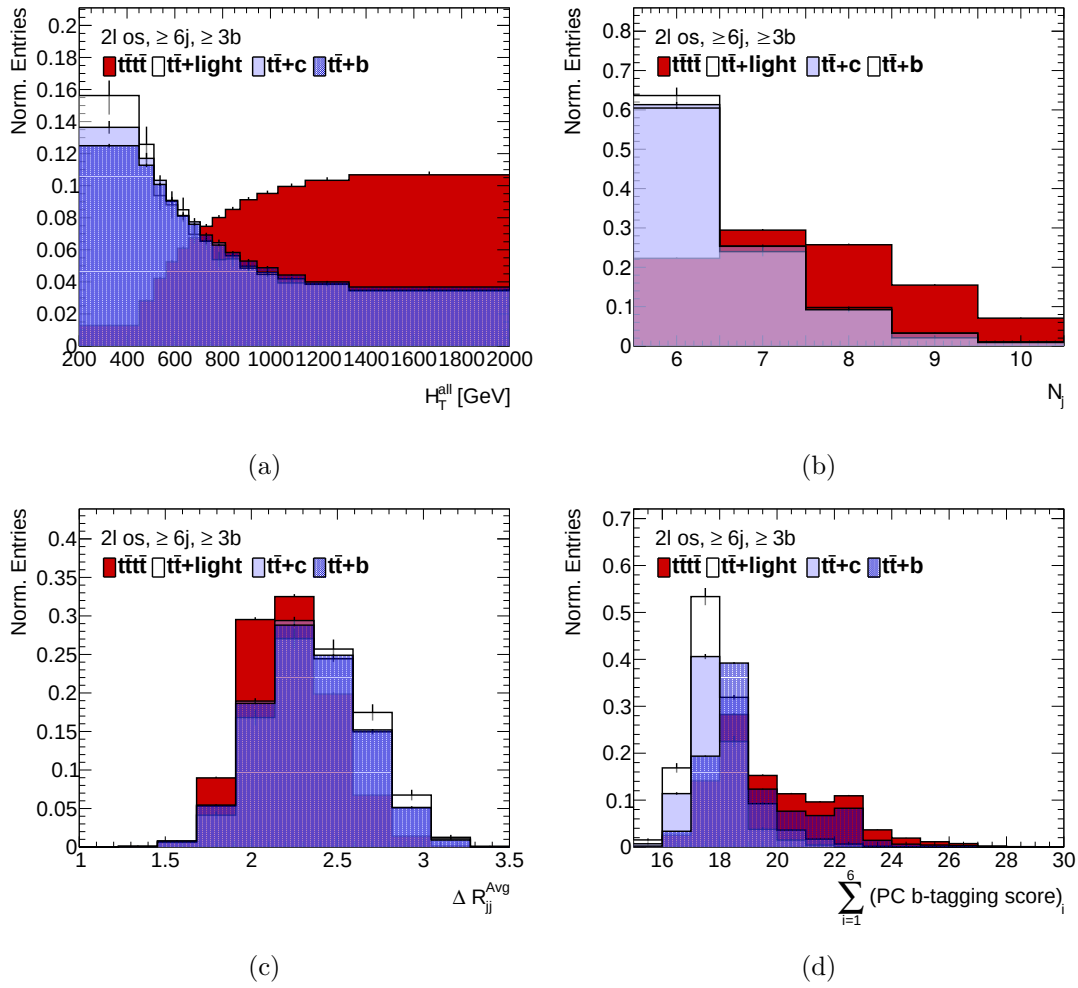
Figure A.17.: Distributions of reweighting variables in the OS channel for ge8j exclusive regions after each reweighting step. (a) Unweighted distributions, (b) after $N_{\text{RC-jets}}$ reweighting, (c) after $H_{\text{T}}^{\text{all,red}}$ reweighting and (d) after applying $\Delta R_{\text{jj}}^{\text{Avg}}$ SFs. The uncertainties are statistical.

APPENDIX B

BDT input variables

This appendix shows additional plots relevant for the BDT. Section B.1 shows the separation for variables highly ranked in the BDT training for the OS channel. Section B.2 shows further plots of the data MC agreement after applying the derived scale factors.

B.1. Separation of BDT Input Variables

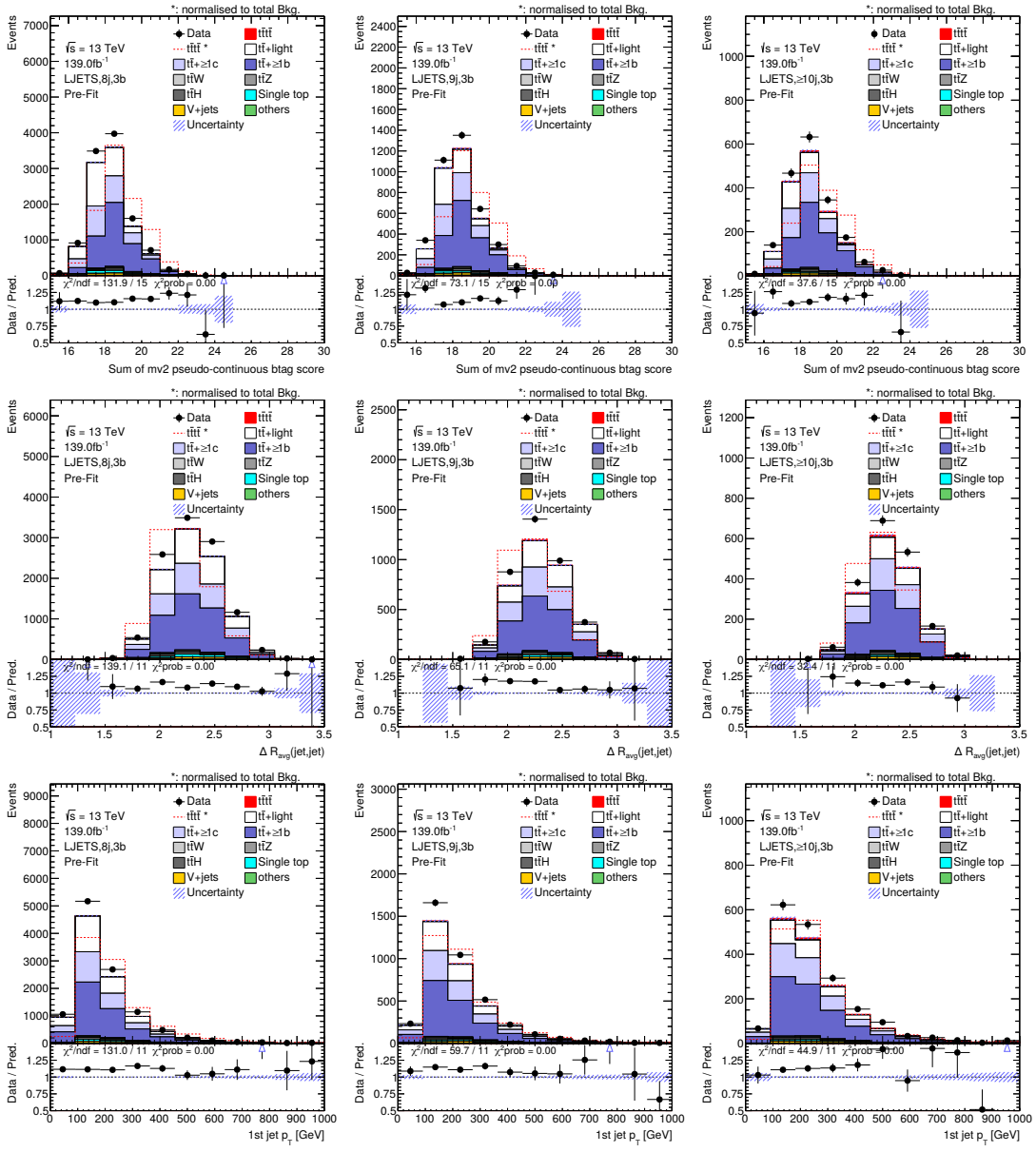


(a) (b) (c) (d)

Figure B.1.: Separation of highest ranked BDT input variables in the OS channel.

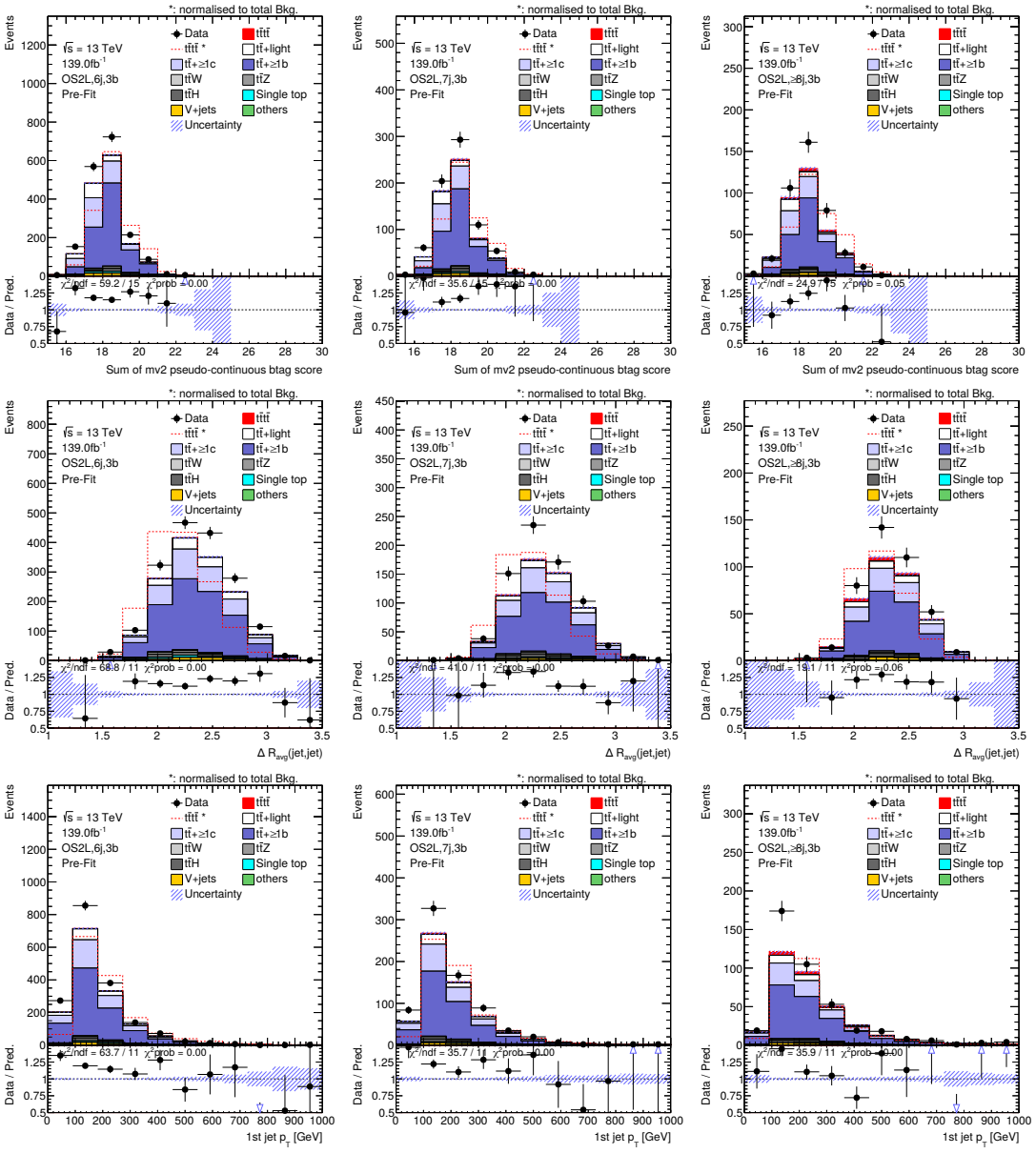
B.2. Data-MC Agreement

Data MC agreement for BDT input variables after applying the additional SFs of the MC-based method.



(a) (b) (c)

Figure B.2.: The agreement between data and MC of the most important BDT-input variables in the 1L channel in region with exactly 3 b-jets in the (a) 8-jet, (b) 9-jet and (c) at least 10 jet regions. The uncertainty band includes statistical uncertainties.



(a) (b) (c)

Figure B.3.: The agreement between data and MC of the most important BDT-input variables in the OS channel in region with exactly 3 b-jets in the (a) 6-jet, (b) 7-jet and (c) at least 8 jet regions. The uncertainty band includes statistical uncertainties.

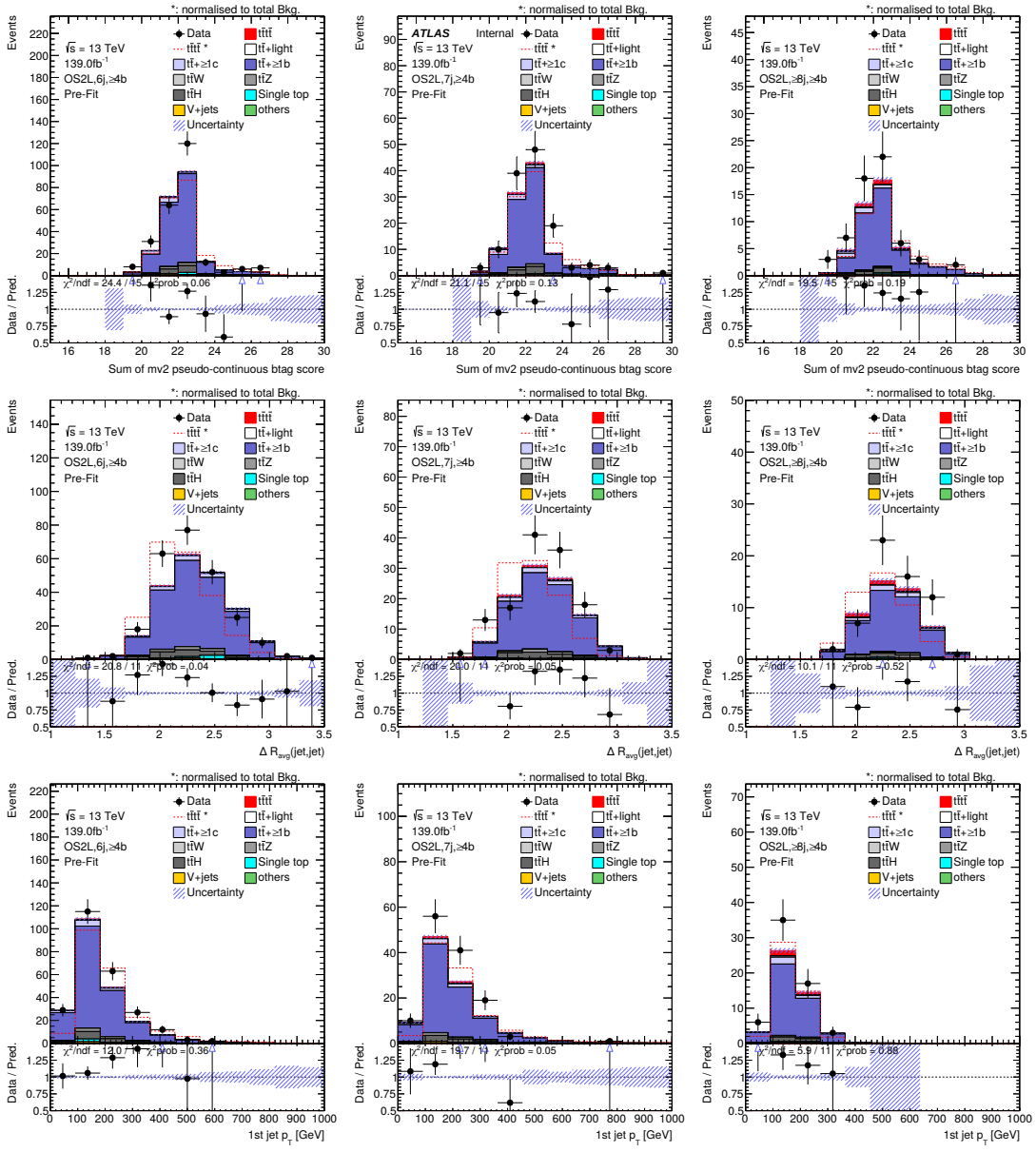


Figure B.4.: The agreement between data and MC of the most important BDT-input variables in the OS channel in region with at least 4 b-jets in the (a) 6-jet, (b) 7-jet and (c) at least 8 jet regions. The uncertainty band includes statistical uncertainties.

APPENDIX C

Fit to Control Regions with ttTRF

This chapter show additional plots from the fit to data in blinded regions using the ttTRF method. The fit-result is used to generate a pseudo-data set in the full phase-space.

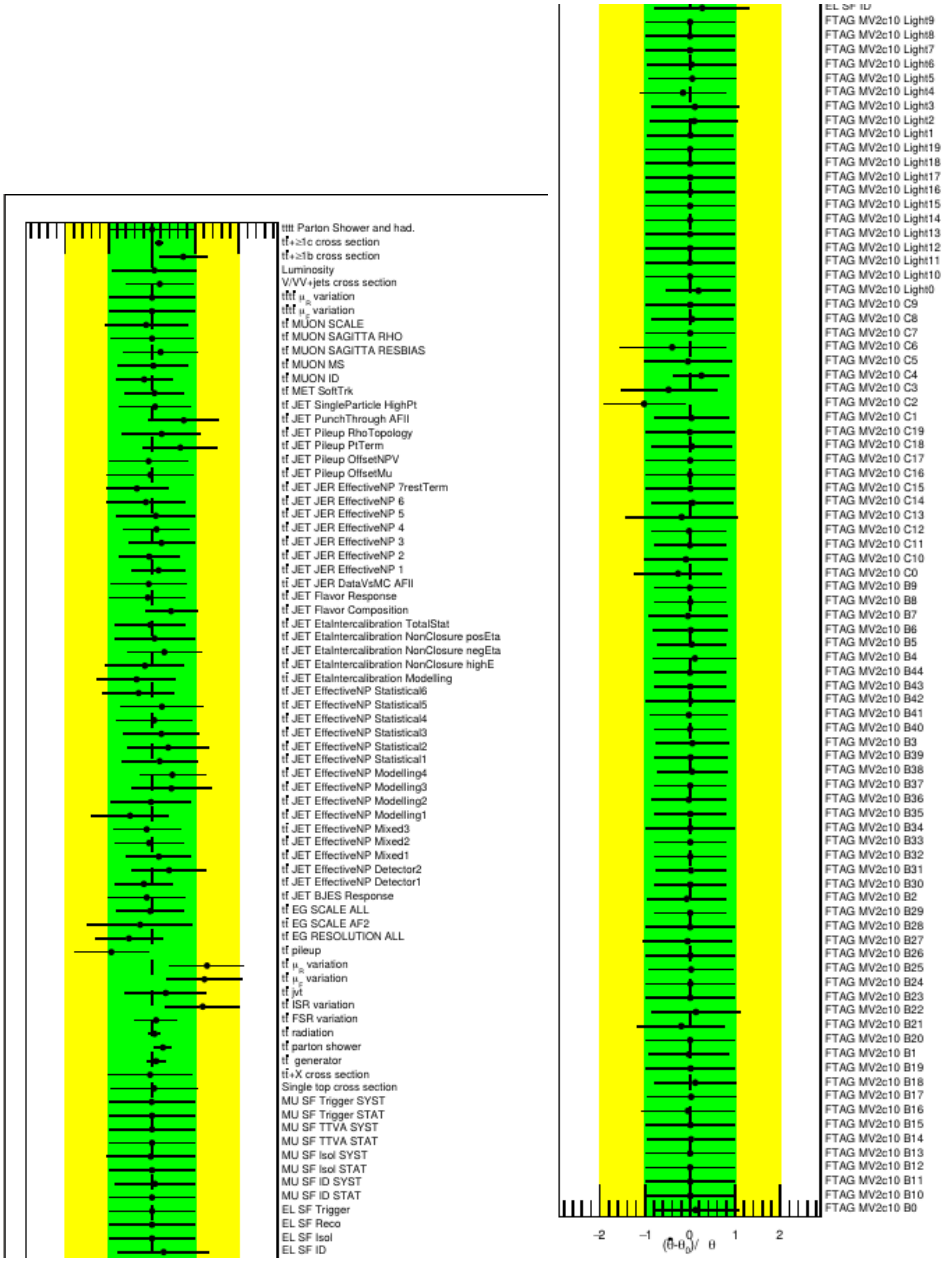
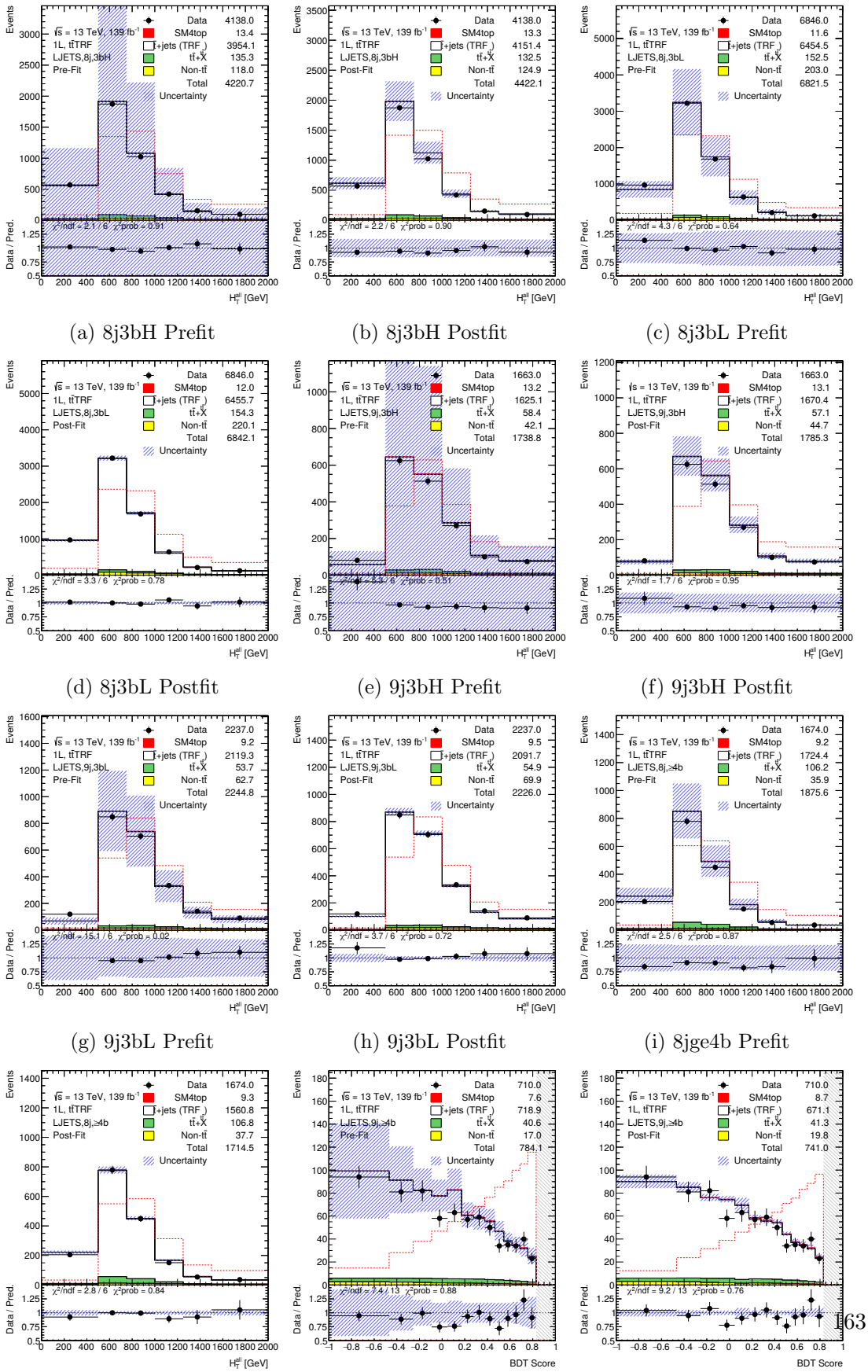


Figure C.1.: Pulls and constraints of the fit to data in blinded regions with the ttTRF method.



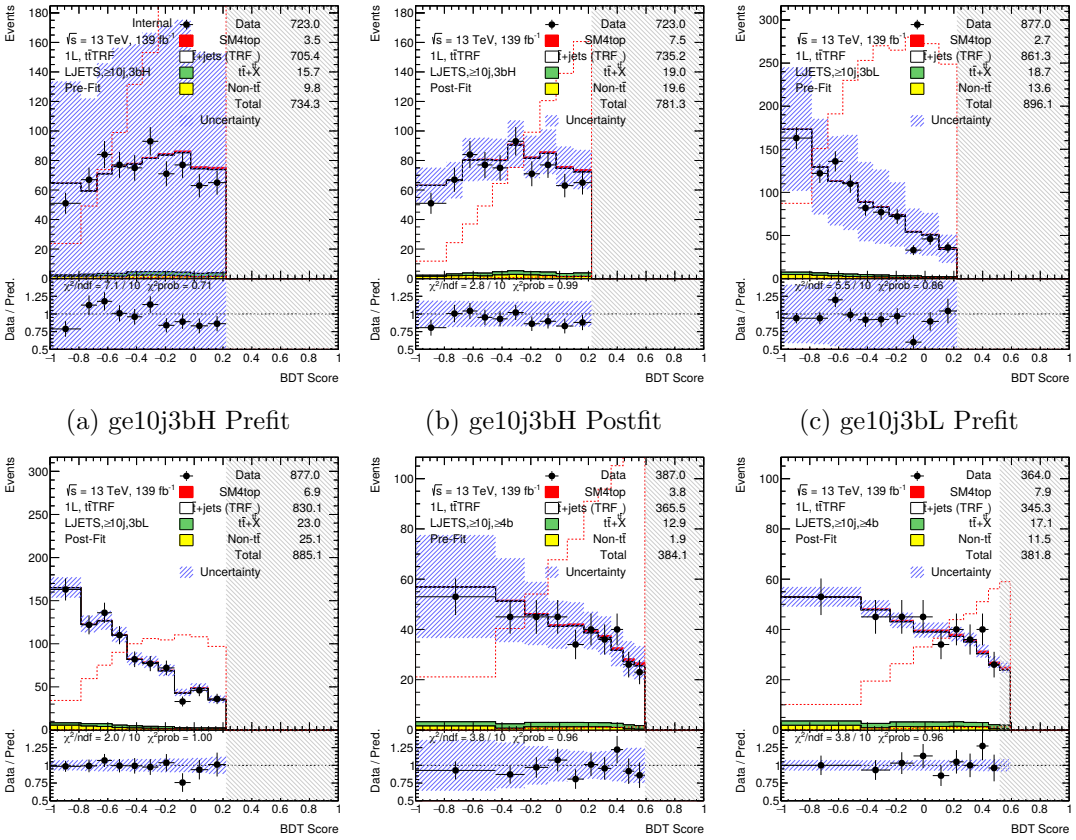
(a) 8j3bH Prefit (b) 8j3bH Postfit (c) 8j3bL Prefit

(d) 8j3bL Postfit (e) 9j3bH Prefit (f) 9j3bH Postfit

(g) 9j3bL Prefit (h) 9j3bL Postfit (i) 8jge4b Prefit

(j) 8jge4b Postfit (k) 9jge4b Prefit (l) 9jge4b Postfit

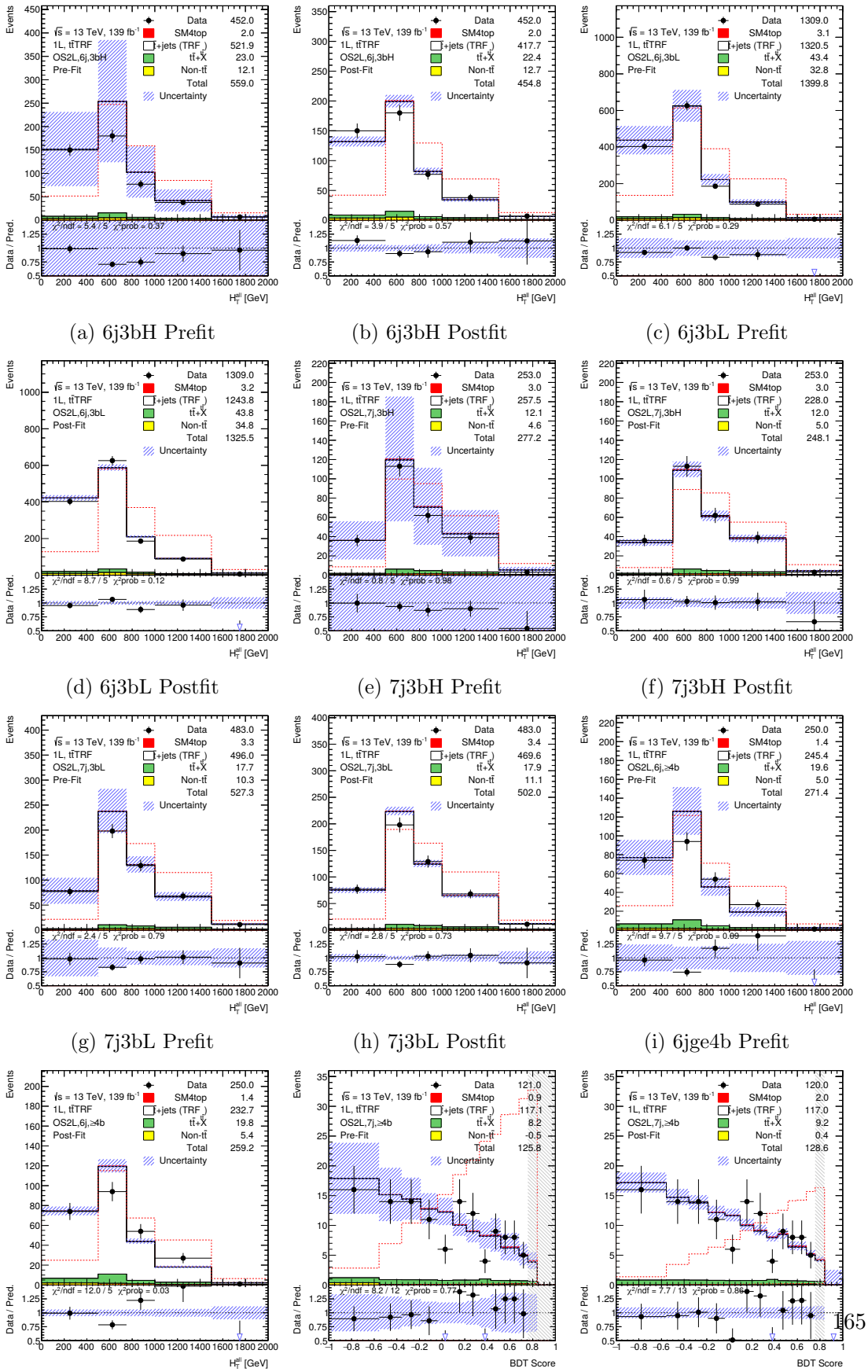
Figure C.2.: Pre- and post-fit comparison of MC to data in the 1L channel of the fit (1L+OS) to data in blinded regions with the ttTRF method. The H (L) stands for $3b^{\geq 4b@85\%}$ ($3b^{<4b@85\%}$). The uncertainty band includes statistical and systematical uncertainties.



(a) ge10j3bH Prefit (b) ge10j3bH Postfit (c) ge10j3bL Prefit

(d) ge10j3bL Postfit (e) ge10jge4b Prefit (f) ge10jge4b Postfit

Figure C.3.: Pre- and post-fit comparison of MC to data in the 1L channel of the fit (1L+OS) to data in blinded regions with the ttTRF method. The H (L) stands for $3b^{\geq 4b@85\%}$ ($3b^{<4b@85\%}$). The uncertainty band includes statistical and systematical uncertainties.



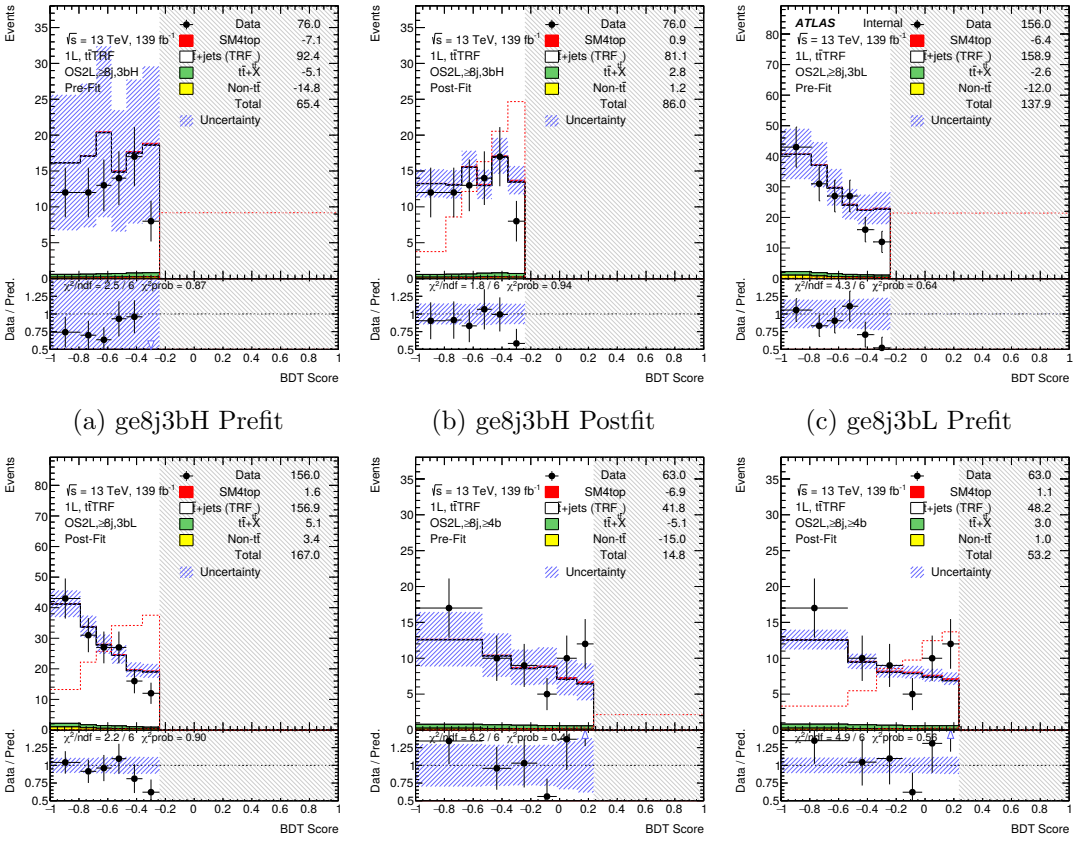
(a) 6j3bH Prefit (b) 6j3bH Postfit (c) 6j3bL Prefit

(d) 6j3bL Postfit (e) 7j3bH Prefit (f) 7j3bH Postfit

(g) 7j3bL Prefit (h) 7j3bL Postfit (i) 6jge4b Prefit

(j) 6jge4b Postfit (k) 7jge4b Prefit (l) 7jge4b Postfit

Figure C.4.: Pre- and post-fit comparison of MC to data in the 1L channel of the fit (1L+OS) to data in blinded regions with the ttTRF method. The H (L) stands for $3b^{\geq 4b@85\%}$ ($3b^{<4b@85\%}$). The uncertainty band includes statistical and systematical uncertainties.



(a) ge8j3bH Prefit (b) ge8j3bH Postfit (c) ge8j3bL Prefit

(d) ge8j3bL Postfit (e) ge8jge4b Prefit (f) ge10jge4b Postfit

Figure C.5.: Pre- and post-fit comparison of MC to data in the 1L channel of the fit (1L+OS) to data in blinded regions with the ttTRF method. The H (L) stands for $3b^{\geq 4b@85\%}$ ($3b^{<4b@85\%}$). The uncertainty band includes statistical and systematical uncertainties.

APPENDIX D

Additional Fit Results

This Appendix includes additional plots for the fits to pseudo data in Section D.1 and for the final results in Section D.2.

D.1. Fit to Pseudodata with the MC-based Method

D.1.1. Fit to Fit to MadGraph5_aMC@NLO +Pythia

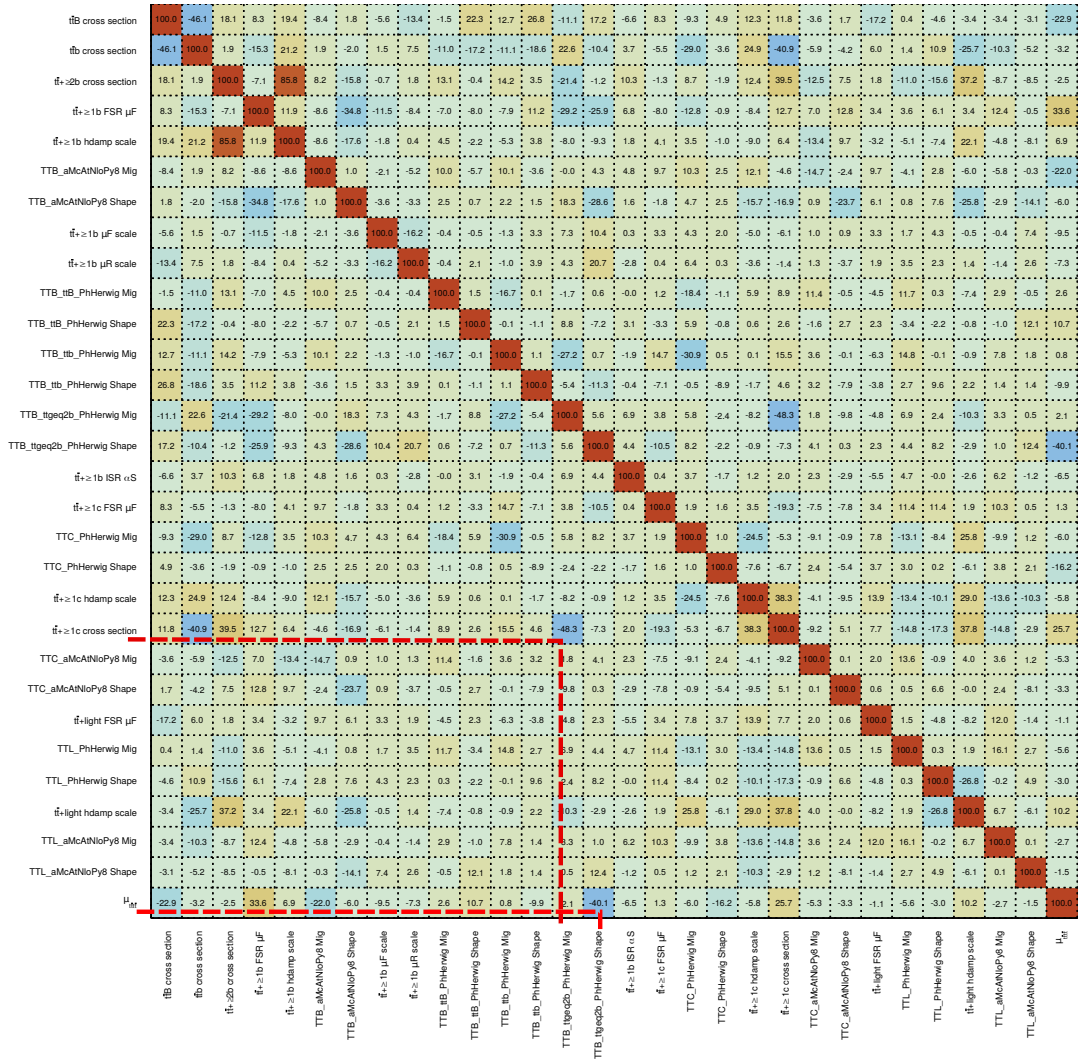


Figure D.1.: Correlation matrix for the fit to the MadGraph5_aMC@NLO +Pythia pseudo-dataset with all modelling systematics.

D.1.2. Fit to ttTRF Post-fit Prediction

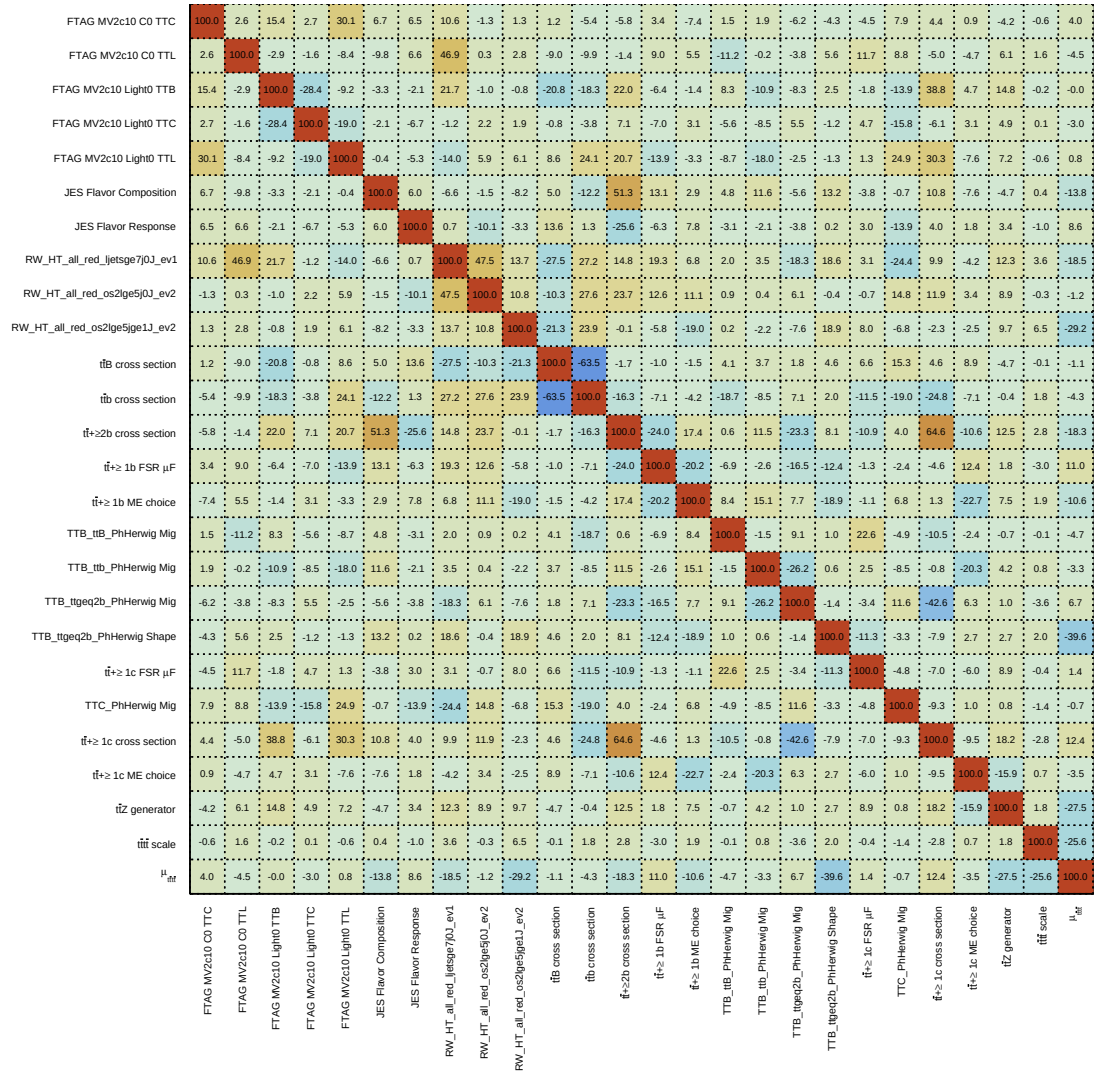


Figure D.2.: Correlation matrix for the fit to ttTRF generated pseudo-data.

D.2. Fit to Data with the MC-based Method

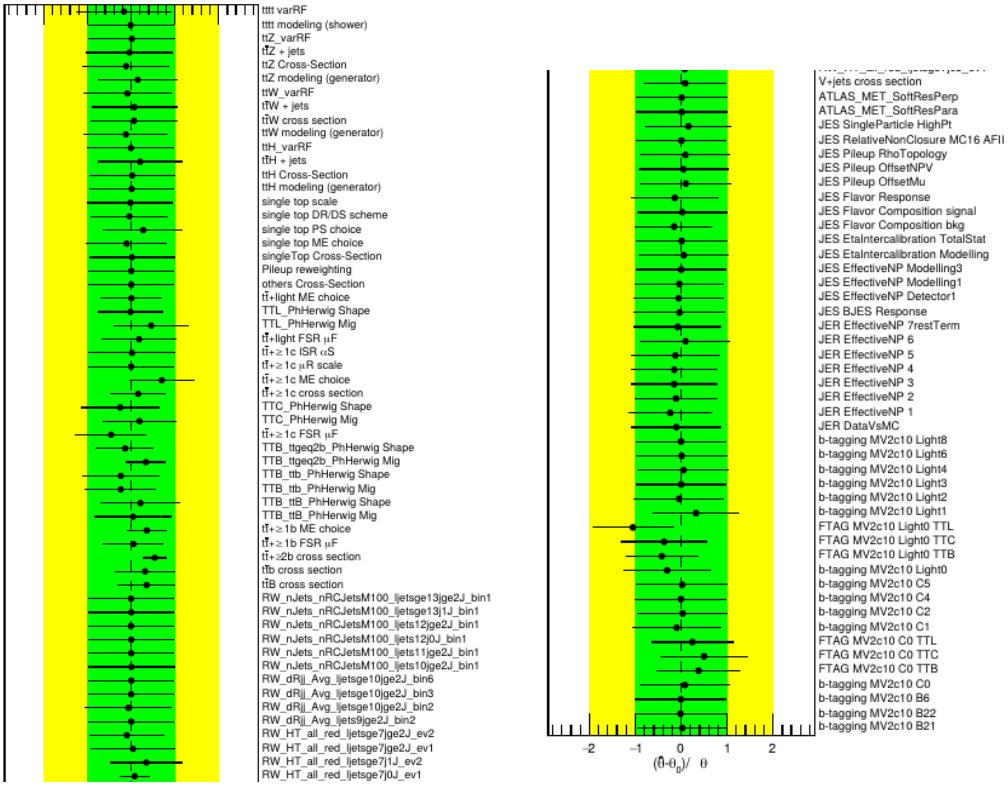


Figure D.3.: Pulls and constraints of the fit to data in unblinded regions with the MC-based method in the 1L channel.

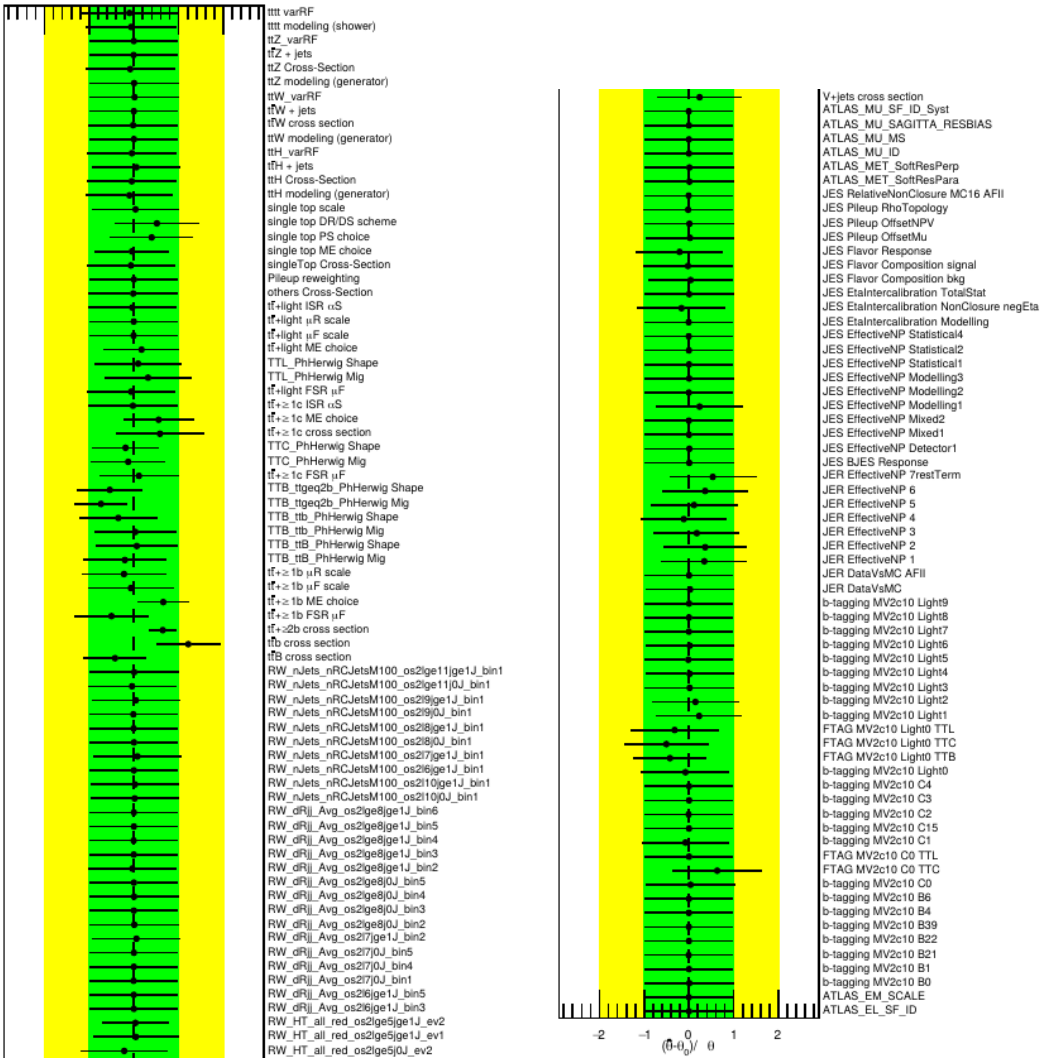


Figure D.4.: Pulls and constraints of the fit to data in unblinded regions with the MC-based method method in the OS channel.

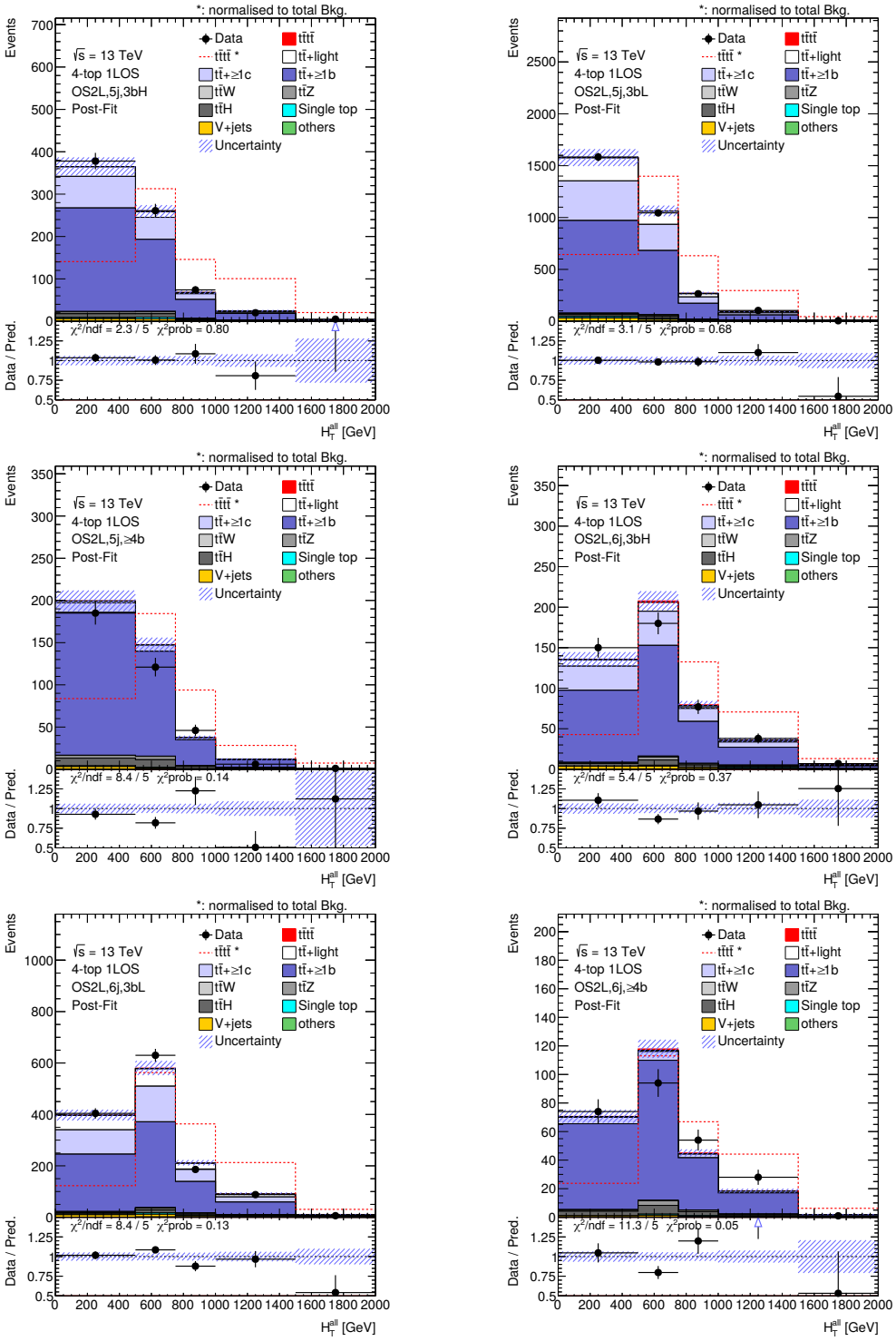


Figure D.5.: Post-fit distribution for the five and six jet regions of the OS channel after a combined fit to the 1L+OS channels. The uncertainty band includes statistical and systematical uncertainties.

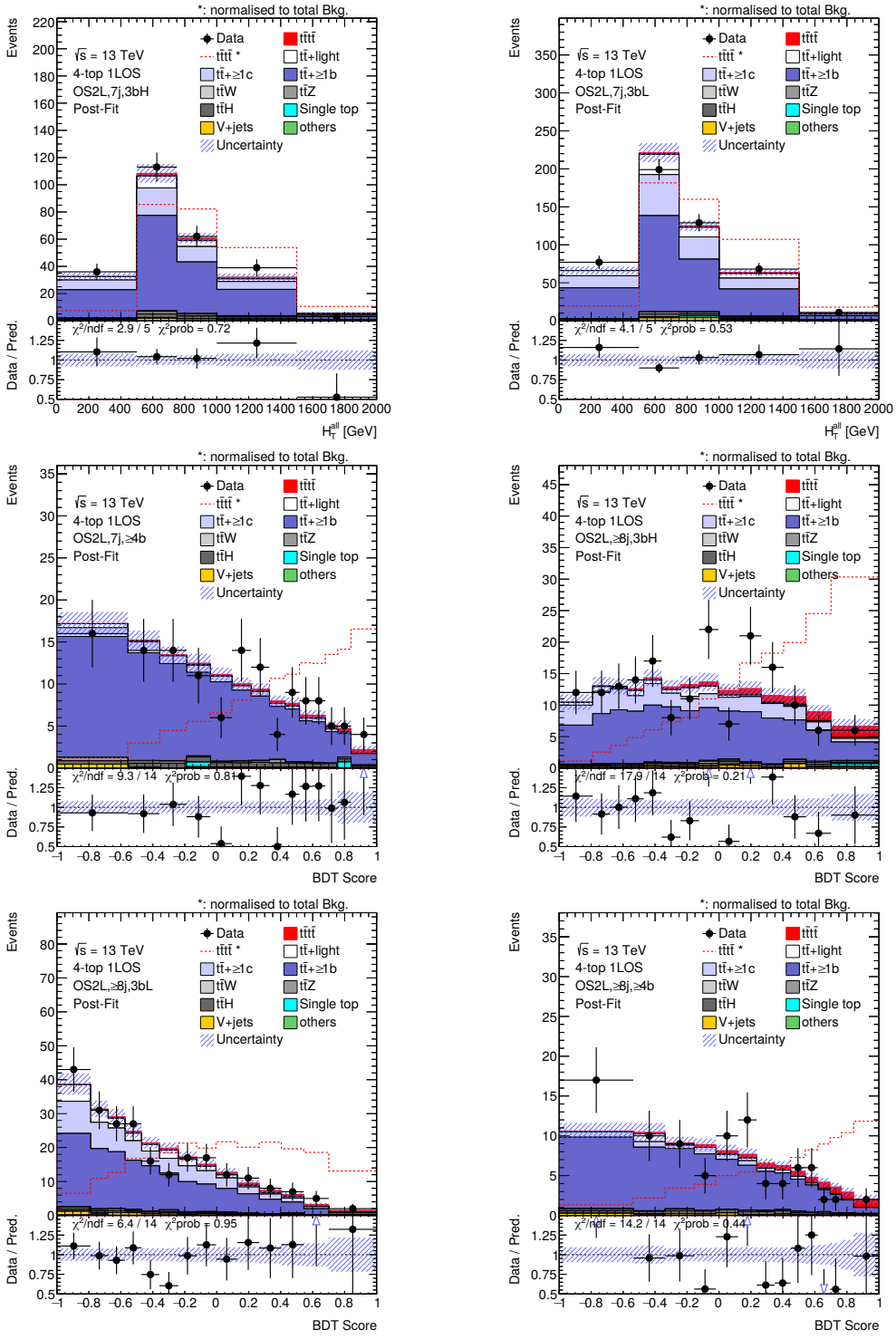


Figure D.6.: Post-fit distribution for the seven and more than eight jet regions of the OS channel after a combined fit to the 1L+OS channels. The uncertainty band includes statistical and systematical uncertainties.

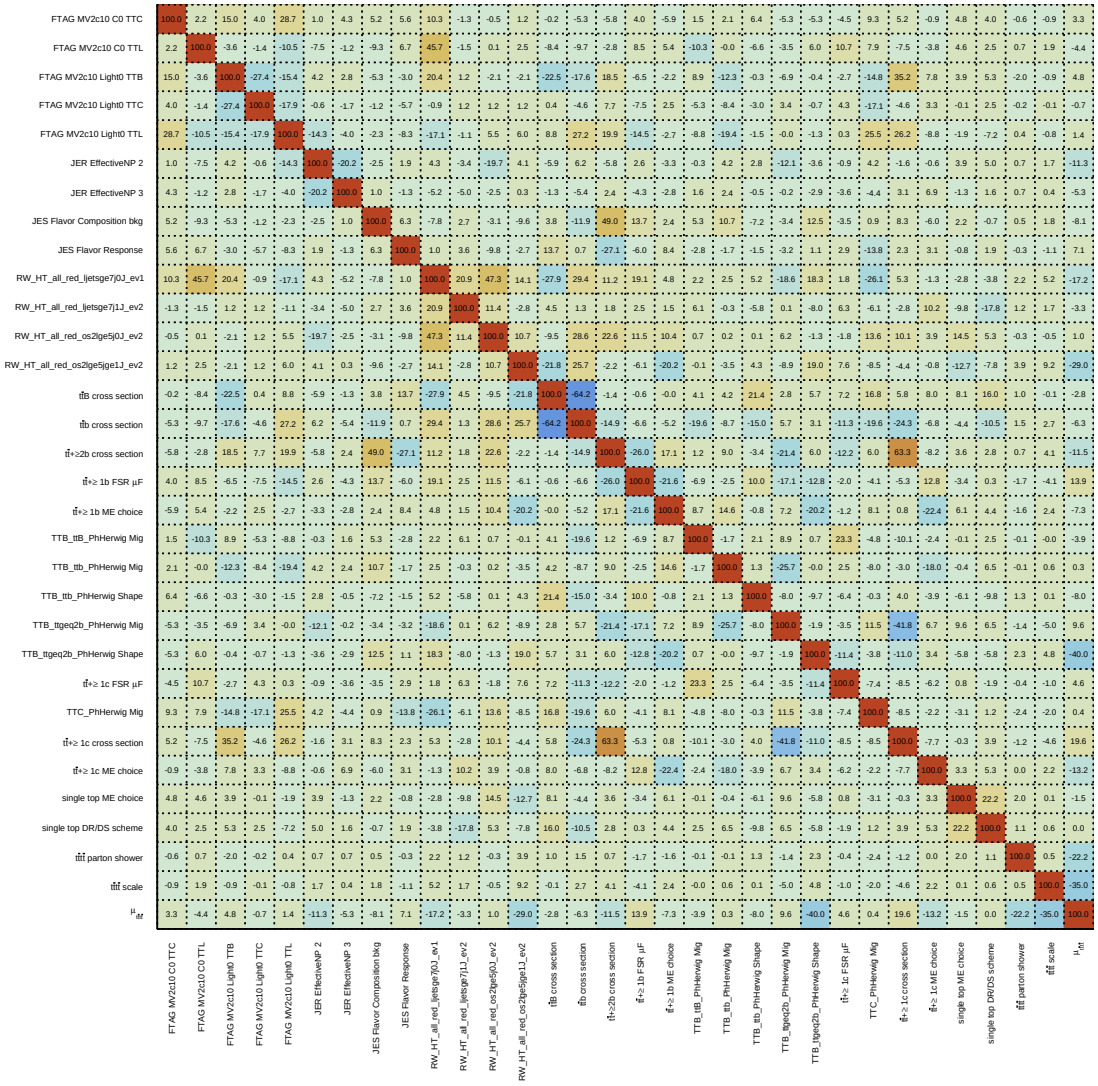


Figure D.7.: Correlation matrix of all NPs with at least 20% correlation for the fit to unblinded data with the MC-based method.

D.3. Combined Fit

All nuisance parameters for the combined fit are compared with the individual pulls and constraints of the 1L+OS and SS+ML channels.

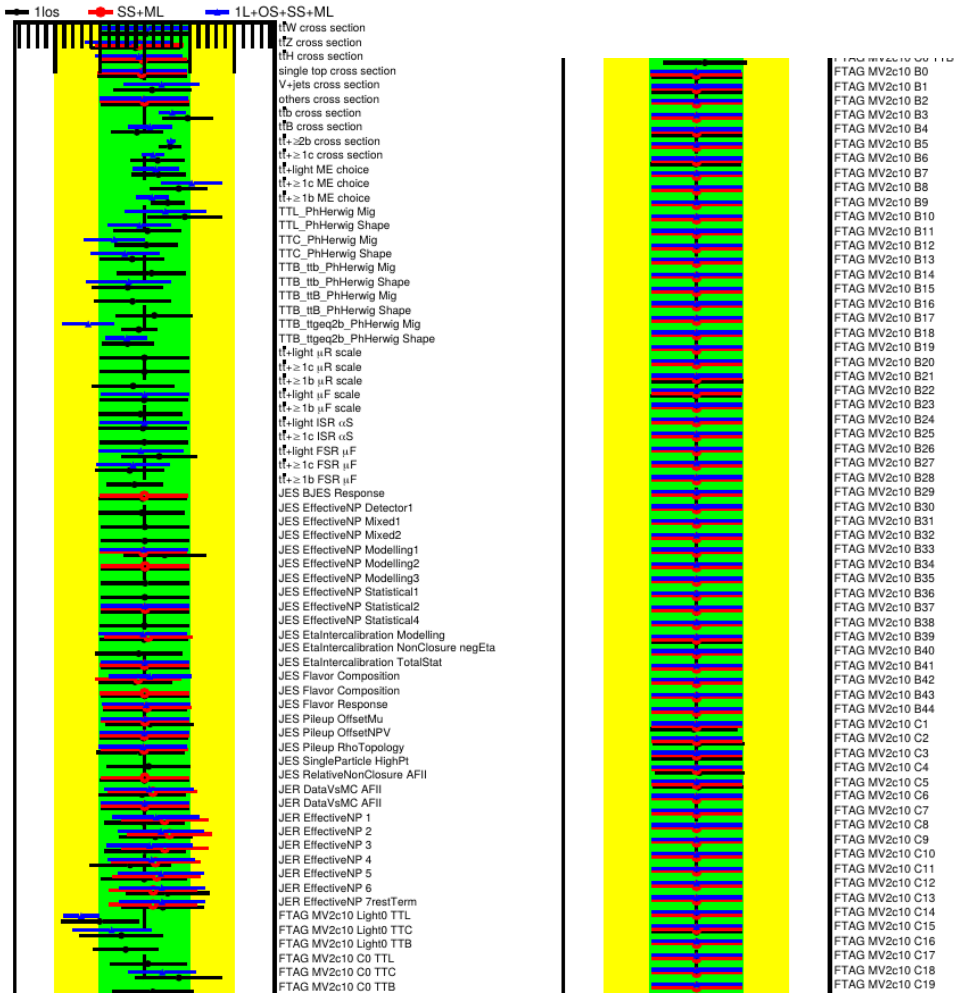


Figure D.8.: Pulls and constraints of the fit to data in unblinded regions in the 1L+OS and SS+ML channel.

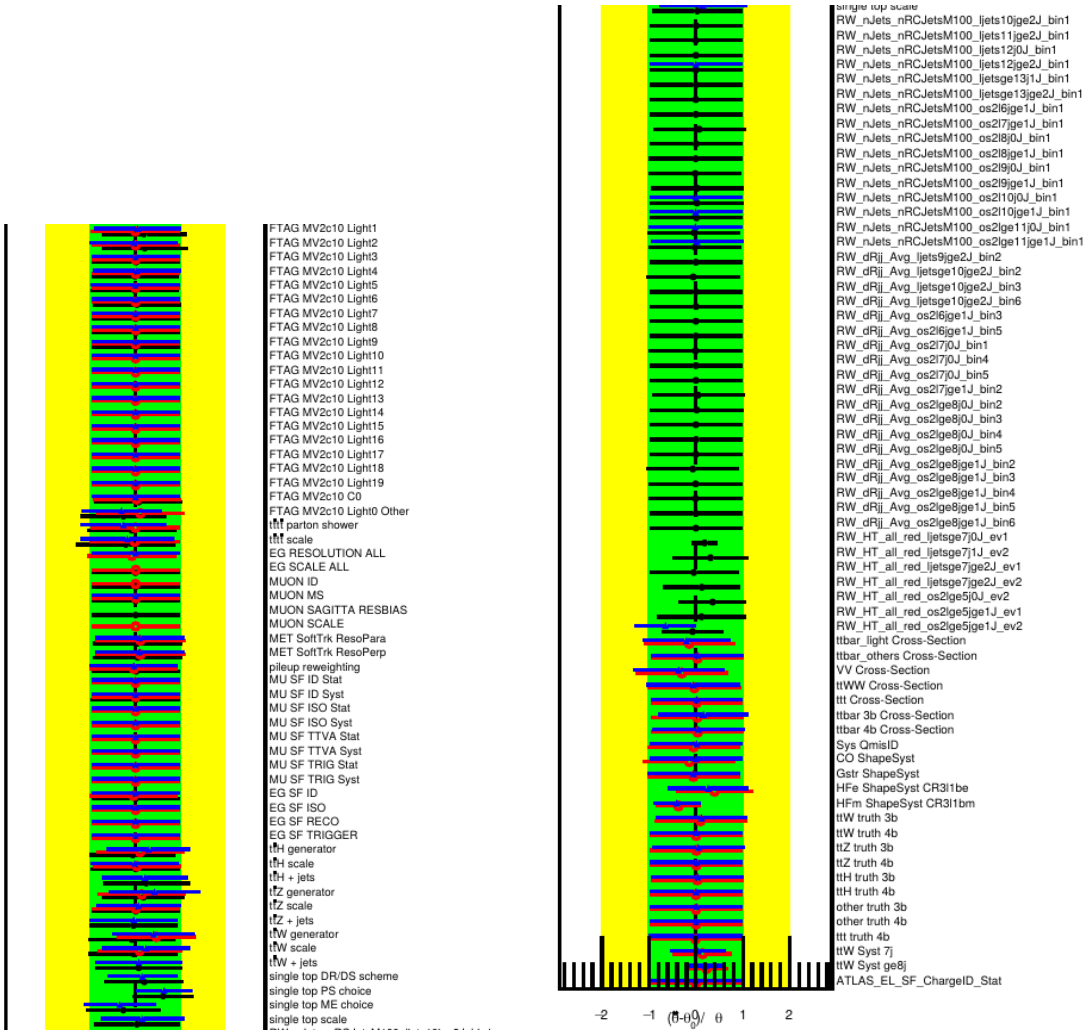


Figure D.9.: Pulls and constraints of the fit to data in unblinded regions in the 1L+OS and SS+ML channel.