

Master's Thesis

Studien zur Top Modellierung für Run II

Top modelling studies towards Run II

prepared by

Jannik Geisen

from Nordhorn

at the II. Physikalischen Institut

Thesis number: II.Physik-UniGö-MSc-2015/01

Thesis period: 1st October 2014 until 30th September 2015

First referee: Prof. Dr. Arnulf Quadt

Second referee: Prof. Dr. Ariane Frey

Abstract

Mithilfe der Run I Daten des LHC konnte die $t\bar{t}$ Modellierung überprüft werden. Dies zeigte, dass einige Monte Carlo Simulationen die gemessenen Daten nicht präzise beschreiben können, insbesondere da sie härtere Top Quark p_T vorhersagten als in den Daten gemessen wurde. Neue “Next-to-leading Order” (NLO) Monte Carlo Generatoren sind zum Testen verfügbar, welche eine präzisere Übereinstimmung mit den gemessenen Daten versprechen. Diese neuen NLO Monte Carlo Generatoren werden untersucht und “getuned” um bessere Vorhersagen für Run II zu geben. Studien zur Modellierung von $t\bar{t}$, Z -Boson und Dijet Ereignissen bei NLO werden präsentiert.

Abstract

The large dataset from Run I of the LHC allowed to validate the modelling of $t\bar{t}$ events and showed that certain Monte Carlo predictions do not describe the measured data accurately. In particular, Monte Carlo generators predicted a harder top quark p_T than what was measured in data. New next-to-leading order (NLO) Monte Carlo generators became available which are expected to give a more accurate description of data. These new NLO Monte Carlo generators are being tested and tuned to give better predictions for Run II. Studies of modelling of $t\bar{t}$, Z boson and dijet events at NLO are presented.

Contents

1. Introduction	1
2. Theoretical background	3
2.1. The Standard Model of particle physics	3
2.2. Local gauge invariance	4
2.3. The Higgs mechanism	6
2.4. QCD	10
2.5. Top quark physics	12
2.6. The Large Hadron Collider and the ATLAS experiment	14
2.7. Monte Carlo simulations and tuning	16
2.7.1. Matrix element generators	17
2.7.2. Parton shower	18
2.7.3. Hadronisation models	20
2.7.4. Parameter tuning	22
3. Monte Carlo generator studies	23
3.1. A14 tune studies	23
3.2. Studies with MadGraph5_aMC@NLO	30
4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8	33
4.1. What is Monte Carlo generator tuning?	33
4.2. Tuning with Professor	35
4.3. Observables chosen for the tuning	37
4.4. Event selection and ATLAS measurements	38
4.4.1. $t\bar{t}$ events	39
4.4.2. Z events	39
4.4.3. Dijet events	41
4.5. Tuning results	45
4.6. Discussion	49

Contents

4.7. Further studies	54
4.7.1. Sensitivity studies	54
4.7.2. Parameter correlations	57
4.7.3. Comparing global recoil to local recoil	58
4.7.4. Dedicated tunes	62
5. Conclusion and outlook	73
A. Rivet analyses	75
B. Other tuning result plots	77
C. Other sensitivity plots	81
D. Other local recoil comparison plots	85
E. Other dedicated tuning plots	89
F. Parameter studies with POWHEG+Pythia8	93

1. Introduction

In experimental particle physics, Monte Carlo (MC) generators are an indispensable analysis tool. They are able to give theoretical predictions for many distributions of observables based on the input physics model, for example the Standard Model (SM). When comparing such theoretical predictions from the SM to real collider data, one can interpret significant deviations as hints for new physics. On the other hand, if the input model is an exotic model such as supersymmetry, then significant deviations would lead to its exclusion. A MC generator uses different models with parameters of which some are not predicted by theory. To ensure that MC generators give precise predictions those parameters have to be tuned to real data.

In this thesis, studies of the tuning of next-to-leading order (NLO) MC generators are shown, namely MADGRAPH5_aMC@NLO interfaced with PYTHIA8 [1–3]. For the parton showering with PYTHIA8, the A14 tune [4] is tested for different distributions in dileptonic $t\bar{t}$ events at 7 TeV and compared to ATLAS data using the Rivet analysis toolkit [5]. Additionally, studies concerning the interface of MADGRAPH5_aMC@NLO with different parton showers and tunes are shown. The main goal presented in this thesis is the development of a new tune for MADGRAPH5_aMC@NLO+PYTHIA8, the current choice in ATLAS being A14.

This thesis is structured as follows: section 2 provides the theoretical background for the analysis. It precludes a short description of the Standard Model, the Higgs mechanism, Quantum Chromodynamics (QCD) and top quark physics, followed by an overview of the LHC, ATLAS and MC simulation. Section 3 discusses MC generator studies related to NLO generators and the A14 tune. Section 4 describes the tuning procedure and presents the results of the tuning. Conclusions and a short outlook of possible future steps are presented in section 5.

2. Theoretical background

2.1. The Standard Model of particle physics

From the discovery of the electron in 1897 up to the discovery of the Higgs boson by the ATLAS and CMS experiments at the LHC in 2012 [6, 7], the number of different particles that are believed to constitute our universe has grown a lot. The SM provides an accurate description of the matter and its interactions (excluding gravity) as we observe it in nature. It models the interactions via the exchange of so-called gauge bosons, the mediators of the known forces [8].

Aside from the bosons, which possess an integer spin, all other known elementary particles are fermions with a spin of $1/2$. The fermions can be further split into leptons, which have an electric charge of either 1 or 0, and quarks, which possess an electric charge of $1/3$ or $2/3$ in addition to a colour charge which can be red, green or blue. The fermions can be further categorized into three generations. For each fermion of the first generation, which make up most of our known matter, there is another particle with the same quantum numbers but a much higher mass in the second and an even higher mass in the third generation. Lastly, every particle has an antiparticle which has the sign of all quantum numbers inverted, while having the same mass [8]. An overview of these elementary particles and the different gauge bosons, including the recently discovered Higgs boson, can be seen in fig. 2.1.

The interactions of particles involve either the annihilation and creation, radiation or scattering of particles. They realise the electromagnetic, weak and strong force observed in nature. The gravitation can be neglected completely at the available energy scale of current particle collider experiments. The electromagnetic force is described by the exchange of a photon between particles with non-zero electric charge resulting in an energy and momentum change of the involved particles. In the same way, the weak force is realized by the exchange of the charged W^\pm or the neutral Z^0 bosons between quarks and leptons, while the W boson also couples to photons and itself. The strong force describes the exchange of gluons between particles with colour charge i.e. quarks and other gluons. Finally, all massive particles in the SM gain their mass by interacting with

2. Theoretical background

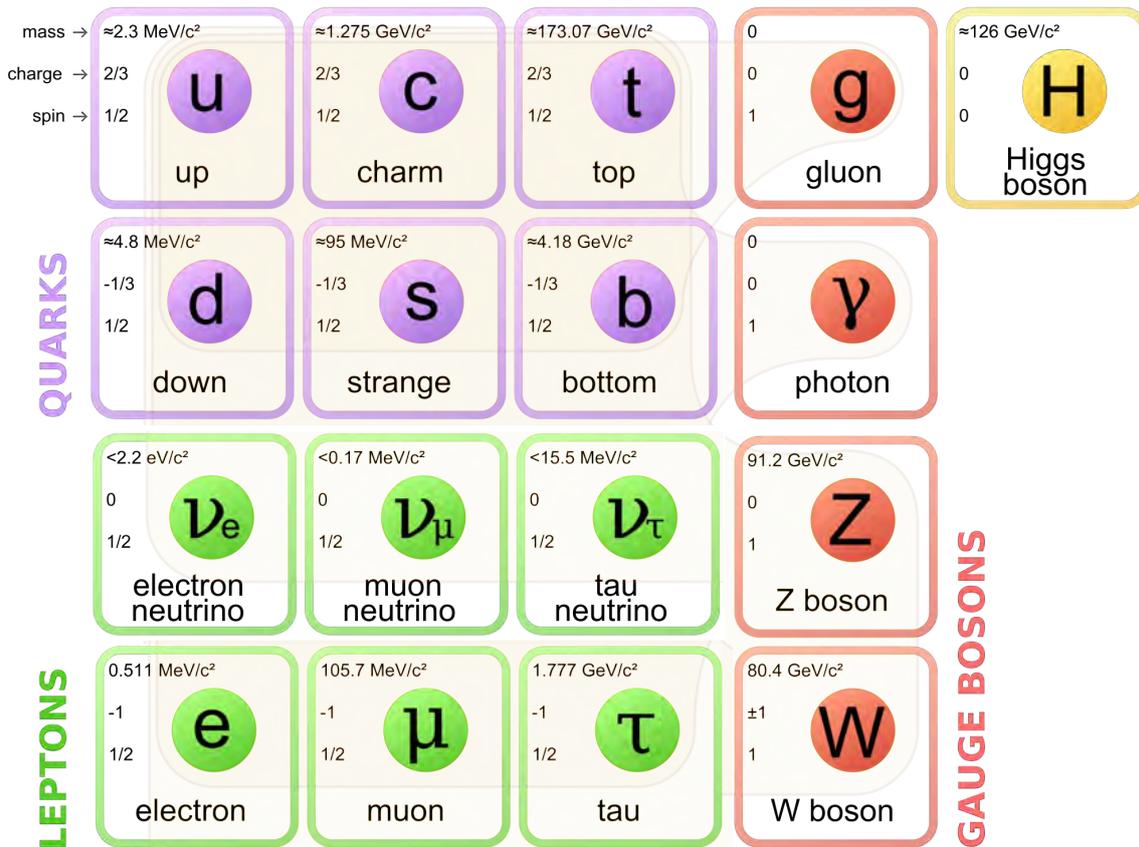


Figure 2.1.: All known elementary particles of the Standard Model.

the Higgs field which is called the Higgs mechanism, explained in section 2.3 [8].

2.2. Local gauge invariance

To calculate predictions of the SM, a mathematical approach using quantum field theory and local gauge invariance is needed. In this way, the SM can be tested and searches for new physics can be performed.

Local gauge invariance is required as a general principle [8]. Most physical situations have an excess of degrees of freedom, i.e. contain some kind of symmetry. As there is no preferred coordinate system, a change of the reference frame according to the symmetry and depending on spacetime can be performed without changing the physical situation. In other words, it is locally invariant under this transformation. Additionally, requiring local gauge invariance allows to conveniently introduce gauge bosons and corresponding interactions to the SM Lagrangian.

Using the example of the free Dirac fermion field and the $U(1)$ phase transformation, the way local gauge invariance can be achieved by introducing a new massless gauge boson

will be shown [8]. The Lagrangian of a free spin-1/2 fermion field ψ can be written in the following way:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi. \quad (2.1)$$

Here, m is the mass of the particle. This equation is invariant under the global U(1) phase transformation

$$\psi(x) \rightarrow \psi'(x) = e^{iq\alpha}\psi(x), \quad (2.2)$$

with $q\alpha$ being the phase shift. However, eq. 2.1 is not invariant under a local phase transformation, that is if $\alpha = \alpha(x)$, because of the derivatives acting on the field. Local gauge invariance can be achieved by replacing the derivative ∂_μ in eq. 2.1 with the so-called covariant derivative D_μ , which is defined as

$$D_\mu = \partial_\mu + iqA_\mu. \quad (2.3)$$

A_μ is a new field and will cancel all unwanted terms violating the local gauge invariance, if this field transforms in the following way:

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\alpha. \quad (2.4)$$

With this, the gauge-invariant Lagrangian for a spin-1/2 fermion can be written as

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - q\bar{\psi}\gamma^\mu\psi A_\mu, \quad (2.5)$$

which now has an interaction term of the fermion field with the new field A_μ , the photon in our example. Thus, the full QED Lagrangian for the electron with elementary charge $q = -e$, the massless photon, their kinetic terms and interactions is given by

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m_e)\psi + e\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.6)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the kinetic term for the photon. This way to achieve local gauge invariance works only for massless gauge bosons, however. For example, if the photon were massive, the QED Lagrangian would need an additional term of the form $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$. Applying a local U(1) gauge transformation would now result in additional

2. Theoretical background

terms, because

$$\frac{1}{2}m_\gamma^2 A_\mu A^\mu \rightarrow \frac{1}{2}m_\gamma^2 (A_\mu - \partial_\mu \alpha) (A^\mu - \partial^\mu \alpha) \neq \frac{1}{2}m_\gamma^2 A_\mu A^\mu. \quad (2.7)$$

This means that the U(1) local gauge symmetry can only be satisfied if the gauge boson of the interaction is massless.

In the same way as for U(1), the SU(2) group of the weak interaction and the SU(3) group of QCD can achieve local gauge invariance, but only if the interacting gauge bosons are massless. This greatly contradicts our observations of the W and Z bosons that have large masses of about 80 GeV and 91 GeV, respectively. The Higgs mechanism is an elegant way to give masses to the three weak bosons, while keeping the Lagrangian invariant under local gauge transformations [8].

2.3. The Higgs mechanism

The Higgs boson is a scalar particle with spin zero, no electromagnetic charge and no colour charge [9]. It was discovered in 2012 by the CMS and ATLAS experiments at the LHC with a mass of about 125 GeV [6, 7]. This discovery was a major success of the SM. The cross sections of the most important production mechanisms of the Higgs boson at the LHC are shown in fig. 2.2 for a centre of mass energy range of $\sqrt{s} = 7 \text{ TeV} \dots 14 \text{ TeV}$ and their corresponding Feynman diagrams are illustrated in fig. 2.3. Its most significant decay channels are shown in fig. 2.4.

The Higgs mechanism exploits the interactions of scalar fields [8]. A complex scalar field

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \quad (2.8)$$

with a potential

$$V(\phi) = \mu^2(\phi^* \phi) + \lambda(\phi^* \phi)^2, \quad (2.9)$$

can be described by the following Lagrangian:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi_1)(\partial^\mu \phi_1) + \frac{1}{2}(\partial_\mu \phi_2)(\partial^\mu \phi_2) - \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) - \frac{1}{4}\lambda(\phi_1^2 + \phi_2^2)^2. \quad (2.10)$$

Here, the first two terms represent the kinetic energy of the scalar particle, the third term can be associated to its mass and the fourth term describes self-interactions of the scalar

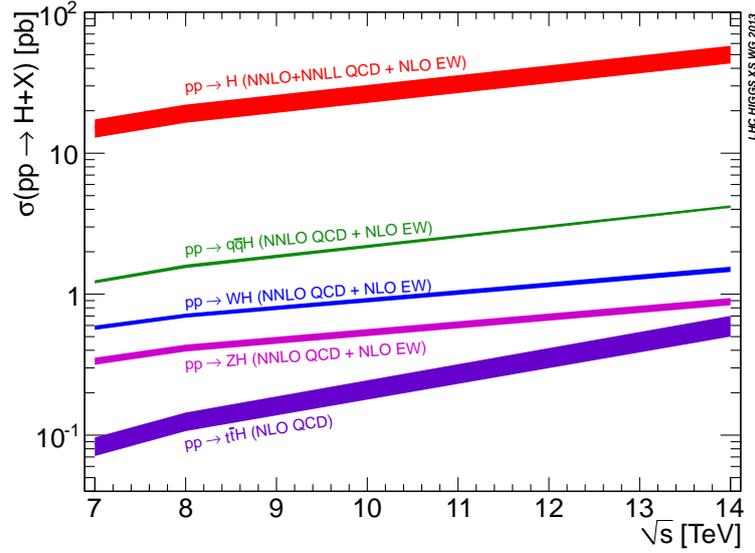


Figure 2.2.: The cross sections of the most important Higgs production mechanisms at the LHC for $\sqrt{s} = 7 \text{ TeV} \dots 14 \text{ TeV}$ [10].

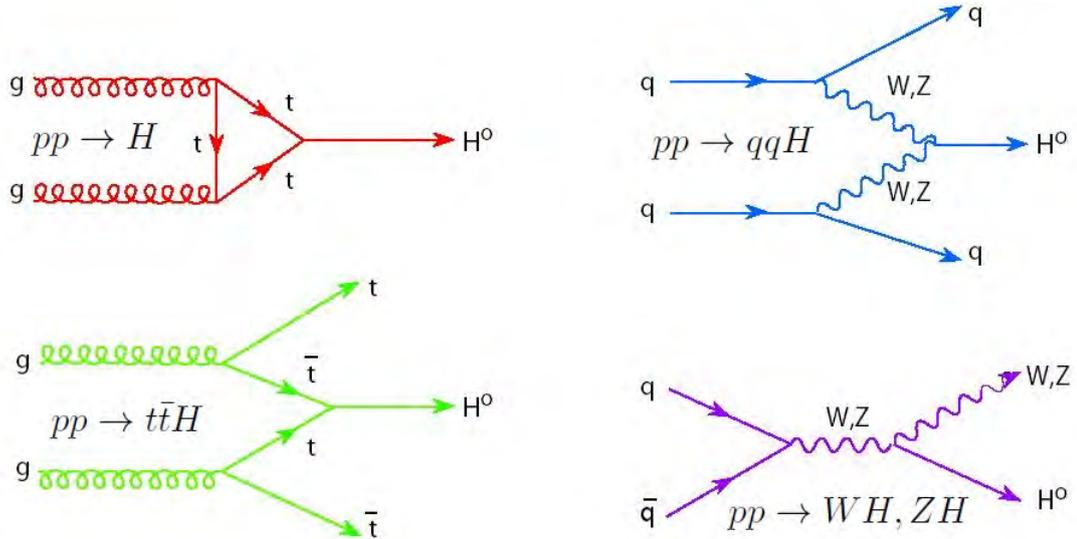


Figure 2.3.: Feynman diagrams of the four most important Higgs production mechanisms at the LHC.

field. The minimum of the potential (eq. 2.9) is only finite, if $\lambda > 0$, but μ^2 can be either greater or less than zero. In the case that $\mu^2 > 0$ the minimum of the potential is given by $\phi_1 = \phi_2 = 0$, but more interestingly, for $\mu^2 < 0$ there is an infinite set of minima given by

$$\phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} \equiv v^2. \quad (2.11)$$

2. Theoretical background

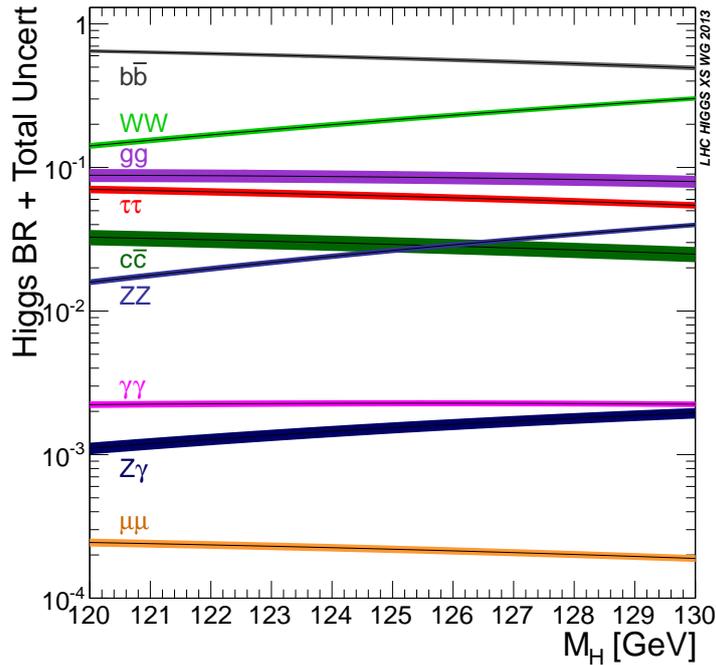


Figure 2.4.: Decay branching ratios of the Higgs boson as a function of its mass within 120 GeV ... 130 GeV [10].

Without loss of generality one can choose the simple case $\phi_1 = v$, $\phi_2 = 0$. With this, the global U(1) gauge symmetry of the Lagrangian is broken. If one now expands the field around its minimum, i.e. $\phi_1(x) = v + \eta(x)$, $\phi_2(x) = \xi(x)$, one can write the field in the following way:

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta + i\xi) \approx \frac{1}{\sqrt{2}} [v + \eta(x)] e^{i\xi(x)/v}, \quad (2.12)$$

where the last approximation results from only considering terms up to the first order in the fields. By introducing a new gauge field and adding its interaction terms, then replacing the derivatives in the Lagrangian with the corresponding covariant derivatives as in the previous section 2.2, one can achieve local gauge invariance by choosing the gauge such that $\alpha(x) = -\xi(x)/gv$, i.e. $\phi(x) \rightarrow \phi'(x) = e^{-ig\xi(x)/gv} \phi(x)$. The effect of the U(1) gauge transformation on the complex scalar field is now given by

$$\phi(x) \rightarrow \phi'(x) = \frac{1}{\sqrt{2}} e^{-i\xi(x)/v} [v + \eta(x)] e^{i\xi(x)/v} = \frac{1}{\sqrt{2}}(v + \eta(x)) \equiv \frac{1}{\sqrt{2}}(v + h(x)). \quad (2.13)$$

This gauge is the so-called *Unitary gauge* and corresponds to choosing the complex scalar field $\phi(x)$ to be entirely real, while the field $\eta(x)$ is now written as the Higgs field $h(x)$. If one writes out the full Lagrangian, one finds a massive scalar h and also a massive gauge

boson that was introduced to restore the local gauge invariance. For the U(1) gauge symmetry, however, the gauge boson is the massless photon, so the above example of giving mass to a gauge boson has to be done for the SU(2) group of the weak interaction instead. In the minimal Higgs model there are two complex scalar fields, placed in a weak isospin doublet as follows:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.14)$$

Here ϕ^0 is the neutral scalar field and ϕ^+ is the charged scalar field, used for the massless photon and the massive W^\pm and Z^0 bosons. The calculation of the minima of the Lagrangian is the same as before, but after the symmetry breaking the photon has to remain massless, so the following simple minimum and gauge can be chosen:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ v + \eta(x) + i\phi_4(x) \end{pmatrix} \xrightarrow{\text{unitary gauge}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (2.15)$$

The resulting Lagrangian is known as the Salam-Weinberg model. It can be written in such a way that it respects the $SU(2)_L \times U(1)_Y$ local gauge symmetry of the electroweak model, namely by replacing the derivatives with the appropriate covariant derivatives

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_W \frac{\vec{\sigma}}{2} \cdot \vec{W}_\mu + ig' \frac{Y}{2} B_\mu, \quad (2.16)$$

where $Y = 2(Q - I_W^3)$ is the hypercharge, I_W^3 is the third component of the weak isospin and B_μ and \vec{W}_μ are new gauge fields. After calculating all terms of the resulting local gauge invariant Lagrangian and rewriting the terms such that the masses of the physical particles are visible, one will find a massless gauge boson field A_μ , three massive spin-1 gauge bosons Z_μ and W_μ^\pm , and a massive scalar Higgs boson. The interaction terms and thus the coupling strength of the Higgs to the W and Z bosons are proportional to the W and Z boson mass, respectively [8].

The Higgs mechanism can also be used to generate masses for the fermions. The so-called *Yukawa coupling* between the Higgs boson and fermions, however, is not predicted by theory, but assumed to be proportional to the mass of the fermions i.e. $g_f = \sqrt{2}m_f/v$, similar to the coupling of the Higgs to the massive gauge bosons [8].

In the SM, the Higgs has neither electromagnetic nor colour charge and therefore does not directly couple to photons or gluons [9]. It may, however, decay into two photons via a triangular top quark or W loop. There is another interesting decay channel of the

2. Theoretical background

Higgs boson, namely the four lepton channel where the Higgs decays into one real and one virtual W or Z which in turn decay into two leptons ($ee, \mu\mu, e\nu$ and $\mu\nu$). Although these decay channels have a small branching ratio, as shown in fig. 2.4, they have very low background as well. The high significance of these channels eventually lead to the discovery of the Higgs boson.

To determine whether the discovered new particle is indeed the SM Higgs boson, all its properties have to be measured. One of the most important ones is the Yukawa-coupling to the top quark, $g_t = \sqrt{2}m_t/v \approx 1$, because it is the highest Higgs to fermion coupling in the SM. It can be measured directly in the $t\bar{t}H$ process which is the production of a Higgs boson in association with a top-antitop pair. Fig. 2.3 shows an example Feynman diagram of this $t\bar{t}H$ production on the bottom left. This process is also important for another reason. As can be seen in fig. 2.4 the decay channel of a 125 GeV Higgs boson with the highest branching ratio is the decay into a $b\bar{b}$ pair. But the simple production of two b quarks at the LHC has a cross section that is many orders of magnitude higher than the expected $pp \rightarrow H \rightarrow b\bar{b}$ process [9, 11]. However, one way to measure the $H \rightarrow b\bar{b}$ decay is in association with a $t\bar{t}$ pair. The main background process, the $t\bar{t}b\bar{b}$ production, has a cross section that is only roughly three orders of magnitude higher [11], which makes it possible to observe this decay with the help of multivariate analysis techniques.

2.4. QCD

QCD is the quantum field theory that describes the SU(3) flavour symmetry of the strong interaction between coloured particles, namely quarks and gluons [12]. The additional degree of freedom of the quarks and gluons, which is the colour, can have the three possible values red, green and blue. The gluons, as the mediators of the strong force, have eight colour degrees of freedom coming from the structure of the SU(N) group, which for $N = 3$ has $N^2 - 1 = 8$ degrees of freedom. The colour was introduced as an additional degree of freedom to explain how spin-3/2 baryons in a symmetrical state of space, spin and SU(3)_f could still have a total antisymmetric wave function and thus obey Fermi-Dirac statistics [12].

As was the case for the Dirac Lagrangian of a free spin-1/2 fermion and the Lagrangian for the weak isospin doublet of two complex scalar fields in earlier sections, QCD also has a classic Lagrangian to describe the interaction of spin-1/2 quarks of mass m and massless

spin-1 gluons, namely:

$$\mathcal{L} = -\frac{1}{4}F_{\alpha\beta}^A F_A^{\alpha\beta} + \sum_{\text{flavours}} \bar{q}_a (i\gamma_\mu D^\mu - m)_{ab} q_b. \quad (2.17)$$

Here, $F_{\alpha\beta}^A$ is the field strength tensor derived from the gluon field \mathcal{A}_α^A ,

$$F_{\alpha\beta}^A = [\partial_\alpha \mathcal{A}_\beta^A - \partial_\beta \mathcal{A}_\alpha^A - gf^{ABC} \mathcal{A}_\alpha^B \mathcal{A}_\beta^C] \quad (2.18)$$

and the indices A, B, C run over the eight colour degrees of freedom of the gluon field. g is the coupling strength of the interaction and f^{ABC} are the structure constants of the SU(3) colour group. The third term on the right-hand-side of eq. 2.18 is the one that distinguishes QCD from QED, because it gives rise to triplet and quartic gluon self-interactions and ultimately to the properties of asymptotic freedom and confinement [12].

Both properties and the difference between QCD and QED can be explained by the so-called *running* of the strong coupling constant α_S . At very low energies, that is large distances, the strong coupling constant α_S increases in contrast to the electromagnetic coupling strength α_{EM} which decreases. The consequence is that quarks cannot be separated from each other over a large distance and be observed as bare quarks. Instead they are confined to bound states of three (anti)quarks into (anti)baryons or a quark-antiquark pair into mesons. The energy needed to separate two quarks from each other is high enough to produce a new quark-antiquark pair out of the vacuum that binds with both separated quarks into new bound states. On the other hand, the strong coupling constant is so large at energies in the order of 1 GeV that it is approximately 1 and thus, perturbative calculations in α_S break down. This problem is known as the *infra-red divergence* of QCD, because the probability of strong interactions, in this case the radiation of soft (low-energetic) and collinear (small angle) gluons, increases over all boundaries. Therefore, quarks or gluons produced in particle collisions at the LHC do not simply form bound states in the detector, but rather radiate many soft gluons which in turn radiate off more gluons or split into a quark-antiquark pair and so on. This signature of the many quarks and gluons inside a narrow cone in the detector is called a *jet* and the forming of the many bound states after such a particle cascade is called *hadronisation*. As no perturbative calculation is possible for these phenomena, different models are necessary. MC generators are able to simulate all these phenomena using so-called *parton shower* algorithms. These rely on different shower models (angular-ordered or dipole shower) as well as different fragmentation models (string or cluster fragmentation) to simulate the forming of jets and the hadronisation process [12]. This will be further described in section 2.7.

2. Theoretical background

The other property of the strong interaction, namely the asymptotic freedom, describes the other extreme of the running of the strong coupling constant. At very high energies, i.e. short distances, the strong coupling constant becomes very small such that the quarks behave as if they were free particles. This particular discovery from deep inelastic proton scattering experiments is very important, because it supports the idea of the parton model in which hadrons and mesons consist of point-like particles, namely the coloured quarks and gluons [12].

2.5. Top quark physics

The top quark was discovered in 1995 [13, 14] and, as all up-type quarks, possesses a colour charge, an electric charge of $2/3$ and a spin of $1/2$ [9]. In addition to this, the top quark is the heaviest particle in the SM with a mass of 173.34 ± 0.27 (stat) ± 0.71 (syst) GeV [15] which results in two unique properties.

Firstly, it is a very unstable particle with an average lifetime of about $5 \cdot 10^{-25}$ s [9]. It decays into a b quark and a W boson in over 99.9% of the cases [9]. Its lifetime is shorter than the hadronisation time scale ($\propto 10^{-23}$ s [8]), therefore it decays before it hadronises which means it is the only quark in the SM that can be measured before having formed a bound state. This also means that the top quark can only be studied by analysing its decay products. Thus, when studying top-antitop pair production at the LHC, one looks at the different decay channels of the top-antitop pair which are given by the decays of the two W bosons. W bosons either decay into two jets or leptonically.

The second important property is the high top quark mass itself and its contribution to couplings and gauge boson widths. As the Yukawa coupling strength of the Higgs boson is proportional to the mass, it couples strongly to the top quark as explained in the previous section. The Higgs cannot decay into a real top-antitop pair, but instead it may produce a virtual top-antitop pair which then annihilates again to form a Higgs boson. This $t\bar{t}$ loop contributes to the Higgs width and also to the Z and W width, because the Z also has a virtual $t\bar{t}$ loop and the W has a virtual $t\bar{b}$ loop [8]. This is one of the main reasons that the top quark mass is a very important parameter of the SM and needs to be measured with very high precision.

The main production mechanism of the top quark is in pairs via strong interactions. The production cross section of the top-antitop quark pair was measured at the TEVATRON and the LHC [16–18]. Example diagrams of the $t\bar{t}$ process at leading order are shown in fig. 2.5.

The analyses in chapter 3 will focus on $t\bar{t}$ events, specifically on the case when the $t\bar{t}$

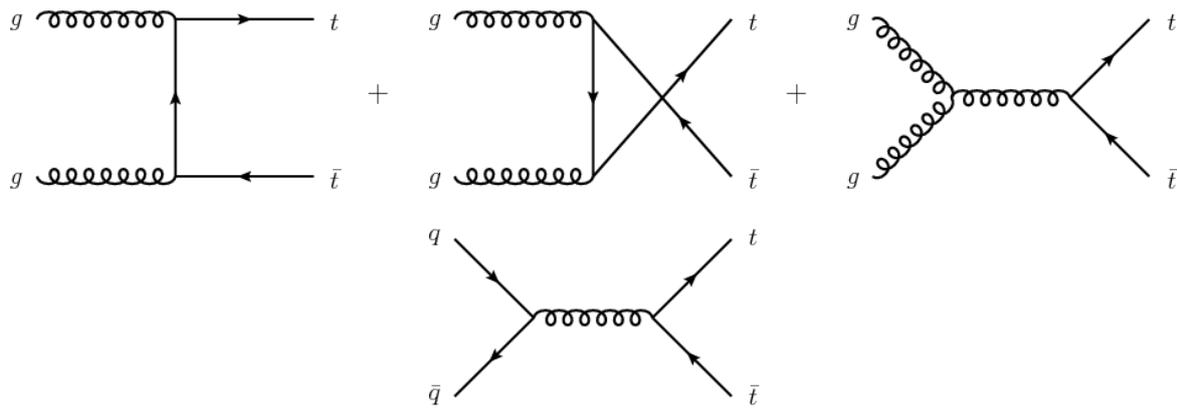


Figure 2.5.: The leading order Feynman diagrams for $t\bar{t}$ production at a hadron collider. The production initiated by quarks was the most dominant process at TEVATRON, while at the LHC it is the gluon-initiated production.

pair decays into two b quarks and both W bosons decay into an electron or muon and corresponding neutrino. The dileptonic channel has the smallest branching ratio of all $t\bar{t}$ decays, as can be seen in fig. 2.6, but it also has the highest purity allowing to distinguish signal from background events easier than in the other decay channels.

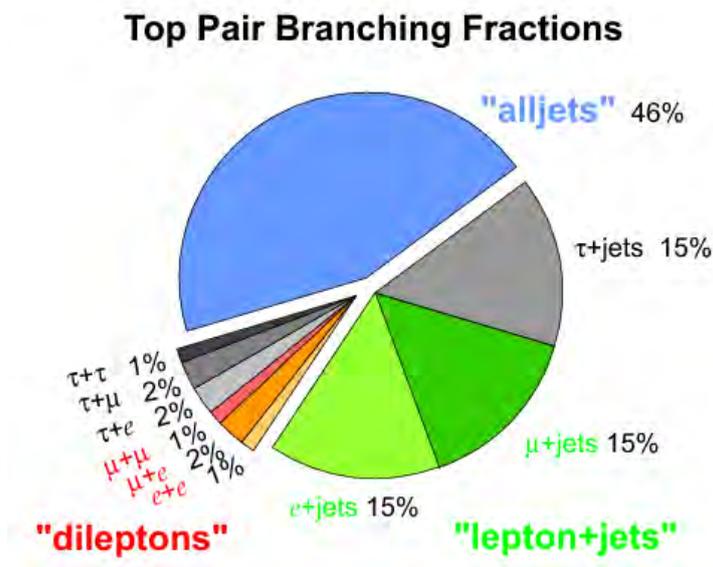


Figure 2.6.: The branching fractions of a top-antitop pair decay [9].

2.6. The Large Hadron Collider and the ATLAS experiment

In 2010 the Large Hadron Collider at CERN in Geneva, Switzerland was turned on and ran for three years. It is a large ring structure with a circumference of approx. 27 km in which protons are accelerated with huge cavities and forced onto a circular trajectory with magnets and then collide with equally accelerated protons to initiate a particle interaction with a centre of mass energy of 7 TeV in 2010 and 2011, and 8 TeV in 2012. This way, every 50 ns two high energetic protons collide inside the particle detectors at the LHC [19]. With such a high centre of mass energy and collision rate, the two largest experiments at the LHC, namely CMS and ATLAS, both succeeded in discovering a Higgs-like particle in 2012 with a mass of about 125 GeV. However, the discovery of a possible Higgs boson is not the end of the potential of the LHC as physicists continue to search for new physics beyond the SM now that Run II of the LHC has started.

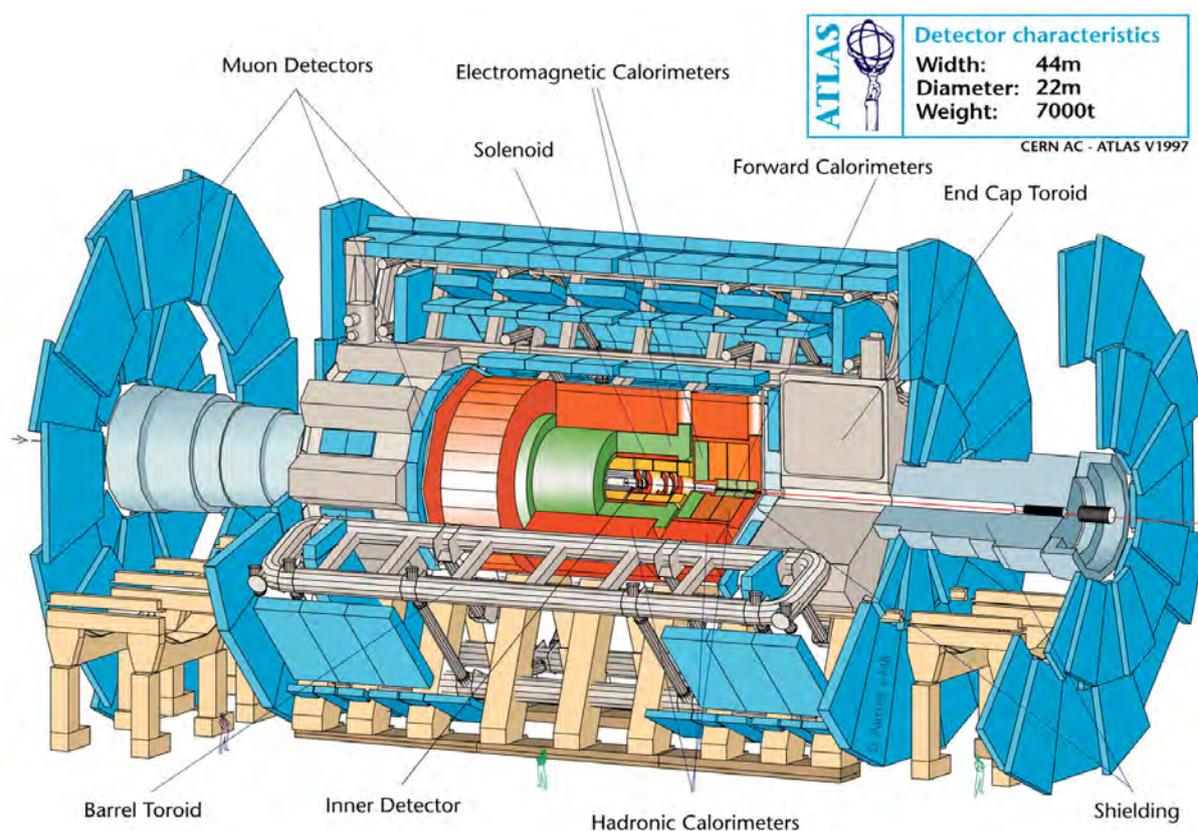


Figure 2.7.: The different parts of the ATLAS detector at the LHC.

A sketch of the ATLAS detector is shown in fig. 2.7. It consists of different parts which provide the identification and measurement of the particles originating from the proton-

proton collisions. They can be categorized as follows [20]:

a) The inner detector consists of a pixel detector, a semiconductor tracker and a transition radiation tracker. It is as close as possible to the collision point to measure the vertices of the particle interactions. In the inner detector charged particles leave tracks through ionisation. The transverse momentum of charged particles is measured through a magnet system, namely a large central solenoid designed for a magnetic field of 2 T, that bends the trajectory of charged particles. The detectors have multiple layers so that the track trajectory can be reconstructed from multiple points to have a higher precision measurement.

b) The electromagnetic and hadronic calorimeters are able to measure the energies of charged particles and photons, and hadrons, respectively. The calorimeters consist of absorber and sensing elements. Interactions in the absorbers transform the incident energy into a shower of particles detected by the sensing elements. In the inner sections of the calorimeter, the sensing element is liquid argon, in the outer sections, the sensors are tiles of scintillating plastic. Electrons, positrons and photons produce showers in the electromagnetic calorimeter and get stopped completely. Muons and charged hadrons leave only a tiny amount of energy in the electromagnetic calorimeter and are not stopped. The hadrons like protons, neutrons or pions produce a wide parton shower inside the hadronic calorimeter from which its energy can be reconstructed.

c) The muon spectrometer is another tracking detector surrounding the calorimeter and measuring the muon tracks to determine their transverse momenta. It consists of thousands of charged particle sensors that are placed in the magnetic field produced by the large superconducting toroidal coils of the magnet system. High precision is achieved by matching the tracks in the muon spectrometer to the tracks from the inner detector.

With these detector parts, electrons, muons, photons and all hadrons can be detected. The only SM particles that will leave the detector undetected are neutrinos. Their transverse momenta, however, can be reconstructed from the missing momentum in the transverse plane.

The focus of this thesis is on $t\bar{t}$, Z and dijet events at $\sqrt{s} = 7$ TeV. Such final states can consist of electrons and/or muons, missing transverse momentum from neutrinos, and jets. Thus, every part of the detector is necessary to reconstruct the events.

2. Theoretical background

The first long shut-down of the LHC (LS1) was used to increase the centre of mass energy to 13 TeV and the instantaneous luminosity to $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, and to upgrade the ATLAS detector. While the increase of the instantaneous luminosity results in more particle collisions per second, it simultaneously increases the radiation damage done to the detector during Run II that the parts have to withstand. As the detector parts cannot be repaired or replaced during operation, the detector has to be prepared in such a way that it does not lose too much detection efficiency due to radiation damage for the full Run II. To achieve this, the *insertable B-Layer* (IBL) was developed [21]. Instead of replacing the old B-Layer, the new IBL is inserted as a fourth layer of the pixel detector even closer to the collision point than the old B-layer. It serves as an additional tracking point such that the precision of the tracking is even higher than before. This allows to improve b -tagging algorithms which rely on the tracks in the inner detectors to identify jets originating from b quarks with a high probability. These algorithms use multiple detector signatures, such as secondary vertices, because of the long lifetime of the b quark. However, with the increased energy in Run II, the b jets will on average have a higher boost which means they are more collimated and thus harder to identify. Therefore, efforts on improving the b -tagging are necessary.

2.7. Monte Carlo simulations and tuning

The MC method is a statistical method to simulate a large number of random experiments through an efficient random number generator in order to numerically solve a problem that is otherwise very difficult or even impossible to solve analytically. This idea of repeating a large number of random experiments relies on the law of large numbers [12]. The MC method is used for simulating particle collisions. The effort that needs to be invested in perturbative calculations increases roughly factorially with the order of α_S . So, instead of going for a precise prediction to some fixed order in perturbation theory, an approximate result which takes into account all important terms to all orders is sufficient. This will later be identified as the parton shower which can be implemented conveniently in computer simulations [12].

In this section, the simulation of a proton-proton collision at the LHC, depicted in fig. 2.8, is briefly discussed and the concept of matrix element generators and parton showers is explained.

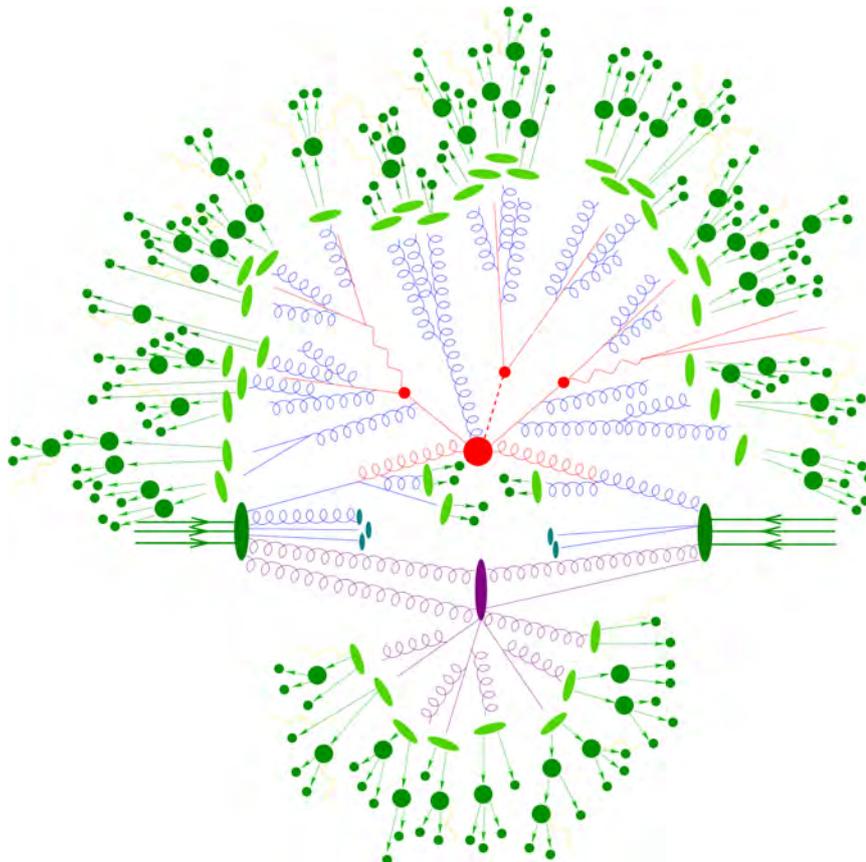


Figure 2.8.: Sketch of a proton-proton collision at the LHC, with the hard process in red, underlying event in purple, radiation and splitting in blue and hadronisation in green [22].

2.7.1. Matrix element generators

To compute the matrix element of a chosen input process, MC generators calculate all of the process' Feynman diagrams up to a certain order. Current generators are either LO or NLO in QCD, so they calculate the hard process, for example $t\bar{t}$ production, with up to one additional emission. Fig. 2.8 shows an example of a possible proton-proton collision and illustrates the hard process, initial state and final state radiation, hadronisation and the underlying event. The matrix element generator calculates all possible Feynman diagrams for the hard process, produces the corresponding final state particles and decays them according to the predictions of the SM. For a $t\bar{t}$ event, for example, the branching fractions of fig. 2.6 are taken into account to select the W boson's decay mode. All energies and momenta of the final state particles are calculated by the matrix element generators. This information is necessary, as it is used in the parton showering afterwards, because the hard process sets the energy scale Q^2 up to which partons can branch into new particles during the showering [12].

2. Theoretical background

2.7.2. Parton shower

The parton shower is able to give an approximate perturbative treatment of QCD dynamics at scales of momentum transfer-squared t greater than some infra-red cut-off value t_0 , usually of the order of 1 GeV². The MC method allows this perturbative treatment to be combined with a non-perturbative model that is necessary for the hadronisation process and assumed to take place at scales $t < t_0$ [12].

The most important part of the parton shower is how to correctly treat the infra-red region of the parton branching, that is if a soft gluon is emitted or when a gluon or light quark splits into two almost collinear partons. To compute cross sections for the various branchings, one can use the fact that the cross section for an $n + 1$ -parton final state can be factorised into the cross section for an n -parton final state multiplied by the probability that an additional parton with momentum fraction z is emitted. The differential cross section for an n -parton final state can be written in the following way:

$$d\sigma_n = \mathcal{F} |\mathcal{M}_n|^2 d\Phi_n, \quad (2.19)$$

where \mathcal{F} is the initial-state flux factor, $|\mathcal{M}_n|^2$ is the absolute squared matrix element for an n -parton process and $d\Phi_n$ is the final-state n -parton phase space. After computing $d\Phi_{n+1} \propto d\Phi_n dt dz d\phi$ and integrating over the azimuthal angle ϕ , one finds:

$$d\sigma_{n+1} = d\sigma_n \frac{dt}{t} dz \frac{\alpha_S}{2\pi} \hat{P}_{ba}(z), \quad (2.20)$$

with $\hat{P}_{ba}(z)$ being the appropriate splitting function which describes the probability of emitting an additional parton with momentum fraction z . This equation can only be integrated analytically within certain limits, with the lower bound set to the cut-off limit t_0 and the upper bound set by the energy scale Q^2 from the hard subprocess. Therefore, a new function can be defined as

$$\Delta(t) \equiv \exp \left[- \int_{t_0}^t \frac{dt'}{t'} \int dz \frac{\alpha_S}{\pi} \hat{P}(z) \right], \quad (2.21)$$

which is called the *Sudakov form factor* [12]. It describes the probability of the parton shower evolving from the infra-red cut-off t_0 to the momentum transfer-squared t without branching. Additionally, $\Delta(t)/\Delta(t')$ represents the probability of evolving from t' to t without branching which is consistent with $\Delta(t_0) = 1$.

The Sudakov form factor allows to handle the basic problem that the MC branching algorithm has to solve, namely: given the virtual mass scale and momentum fraction (t_1, x_1)

after some step of the evolution, or as initial conditions, generate the values (t_2, x_2) after the next step. As shown in fig. 2.9, the first quantity to be generated by the algorithm is the value of t_2 .

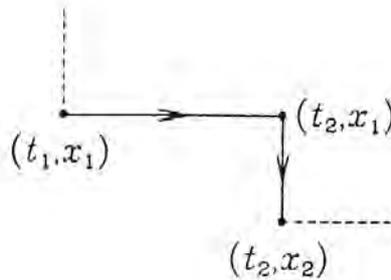


Figure 2.9.: In the parton shower evolution, the value of t_2 is generated first and then x_2 [12].

Using the Sudakov form factor, t_2 can be generated with the correct probability distribution by solving the equation

$$\frac{\Delta(t_2)}{\Delta(t_1)} = \mathcal{R}, \quad (2.22)$$

where \mathcal{R} is a random number distributed uniformly in the interval $[0,1]$. If the value of t_2 is larger than the hard subprocess scale Q^2 , no further branching can occur. Otherwise, the momentum fraction $z = x_2/x_1$ has to be generated for the next branching, which has a probability distribution proportional to $(\alpha_S/2\pi)P(z)$. The generation can be done by solving the following equation:

$$\int_{\epsilon}^{x_2/x_1} dz \frac{\alpha_S}{2\pi} P(z) = \mathcal{R}' \int_{\epsilon}^{1-\epsilon} dz \frac{\alpha_S}{2\pi} P(z), \quad (2.23)$$

with \mathcal{R}' being another random number in $[0,1]$ and $\epsilon = \epsilon(t)$ being another infra-red cut-off, but now for the momentum fraction z instead of momentum transfer-squared t . This cut-off ϵ is chosen in such a way that if $z < \epsilon$ ($z > 1 - \epsilon$) the emitted (emitting) parton is so soft that it cannot be resolved properly and thus would be undetectable.

The above description of branching and evolving the parton shower from (t_1, x_1) to (t_2, x_2) assumes spacelike parton branching. This means that the values of (t_i, x_i) generated by successive applications of the algorithm define the virtual masses and momentum fractions of the exchanged quark, from which the momenta of the emitted gluons can be computed. There is also a timelike parton branching scenario. In contrast to the spacelike case, each parton with timelike momentum in a parton shower can itself undergo further branching,

2. Theoretical background

because here the momentum transfer-squared t evolves downwards towards the cut-off value t_0 rather than upwards towards the hard process scale Q^2 . The evolution can be done mostly in the same way, using eq 2.22 and 2.23, but switching t_1 with t_2 .

Out of the many timelike branchings that trigger new timelike branchings, a parton cascade develops, where each outgoing parton becomes the source of a new cascade, until the MC algorithm generates a step without any further branchings and the parton cascade stops. At this point, the outgoing partons have to be converted into hadrons via a hadronisation model, two of which are discussed in the next subsection [12].

2.7.3. Hadronisation models

Two different hadronisation models are presented in this subsection, namely the *string model* [12] used by PYTHIA [2, 3], and the *cluster model* [12] used by HERWIG [23].

The **string model** is easiest to describe using the example of the $e^+e^- \rightarrow q\bar{q}$ process, which is depicted in fig. 2.10. In the centre of mass frame both quarks move in opposite directions. In section 2.4 the confinement was explained with the running of the strong coupling constant and the gluon self-interactions. It can also be explained with the strong potential having an additional term that increases linearly with the distance. In this model, both quarks move in opposite directions and their colour connection, represented by a string, stretches with distance until the potential energy becomes large enough to create a $q\bar{q}$ pair out of the vacuum. The string breaks and the new quark connects to the antiquark string segment. Similarly the new antiquark connects to the quark string segment. Thus, two new strings are formed that again stretch as the partons continue to move away from each other. This mechanism continues until all the energy has been converted into quark-antiquark pairs connected by short string segments, which can be identified with hadrons.

The **cluster model**, instead of colour-connecting quarks and gluons via strings, forms colourless groups, called clusters, out of partons after the perturbative phase of the parton shower. This is motivated by the observation that colour-connected neighbouring partons have an asymptotic mass distribution that falls rapidly at high masses and is asymptotically Q^2 -independent and universal. The simplest way to form these clusters is through non-perturbative splitting of gluons into $q\bar{q}$ pairs. Neighbouring quarks and antiquarks can then combine into colour-singlets. The resulting mass spectrum is again universal and steeply falling at high masses, while this spectrum is also dependent on the cut-off scale t_0 . With a typical value of $t_0 \approx 1 \text{ GeV}^2$, the clusters have average masses of

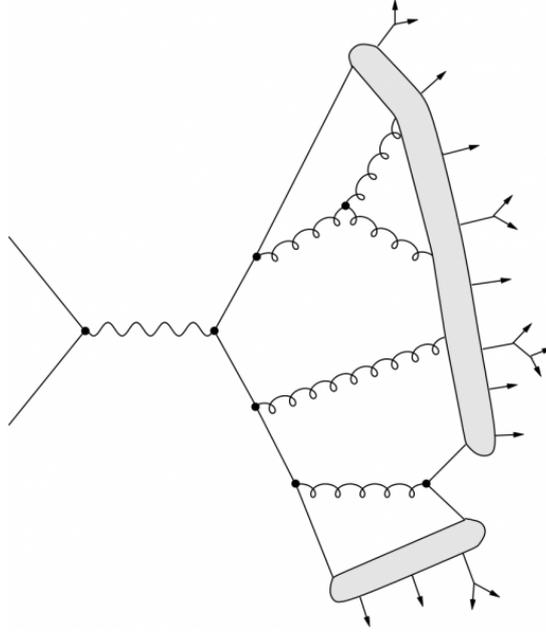


Figure 2.10.: The string model for the process $e^+e^- \rightarrow q\bar{q}$ [12].

about 3 GeV and decay into hadrons using a simple isotropic quasi-two-body phase space model. Fig. 2.11 shows the $e^+e^- \rightarrow q\bar{q}$ process in which colour-singlets are formed using the cluster model.

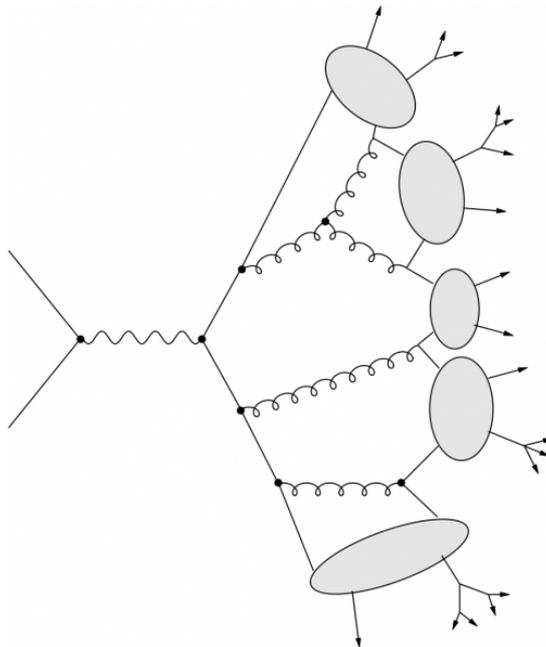


Figure 2.11.: The cluster model for the process $e^+e^- \rightarrow q\bar{q}$ [12].

2.7.4. Parameter tuning

Parton shower models use many different parameters to describe the non-perturbative aspects of particle collisions, for example the radiation of soft partons, the underlying event or hadronisation. However, not all of these parameters are predicted by theory and instead can be set in the MC generators. Meaningful values for these parameters are obtained through proper tuning. In this thesis, the tuning is done with the PROFESSOR tool [24], namely by varying the model parameters a large number of times, each time comparing the respective output of the MC simulation to real data and then performing a fit to determine the best values for these parameters. For this study, the MC simulation is compared to ATLAS $\sqrt{s} = 7$ TeV data, recorded 2011 and which corresponds to an integrated luminosity of 4.6 fb^{-1} . The tuning of MC generators is essential to obtain accurate predictions and uncertainty estimations for Run II of the LHC.

3. Monte Carlo generator studies

In this chapter, comparisons of different MC generators and tunes are presented as motivation for the MC generator tuning studies discussed in chapter 4.

3.1. A14 tune studies

The first study is related to the A14 tune obtained for the PYTHIA8 MC generator. This tune was developed using observables listed in the following and comparing the MC generator output to ATLAS $\sqrt{s} = 7$ TeV data [4]:

- Underlying event (evolution of transverse activity with leading track and calorimeter jets [25, 26])
- Jet structure (track-jet properties [27], jet masses and other substructure variables [28], and jet shapes in inclusive jet and $t\bar{t}$ events [29, 30])
- observables sensitive to additional jet emissions above the lowest-order process (dijet azimuthal decorrelation [31], $t\bar{t}$ gap fraction [32], the three to two jet ratio [33], and Z -boson p_T [34, 35]).

These three classes of observables are sensitive to the modelling of multiple partonic interactions, final state parton showering, and initial state showering [4]. The resulting tuned PYTHIA8 parameters and their values are shown in tab. 3.1.

PYTHIA8 is a LO matrix element generator and therefore does not take into account NLO QCD effects in the hard processes that we observe in the experiments at the LHC. On the other hand, new NLO MC generators such as POWHEG and MADGRAPH5_aMC@NLO are now available and have to be tested and tuned for Run II of the LHC, because deviations from data were observed in Run I. As an example, the Run I baseline MC generator used for $t\bar{t}$ events in ATLAS, namely POWHEG interfaced with PYTHIA6 using the Perugia2011c tune [38], shows a mismodelling of the top quark p_T which is shown in fig. 3.1. This is an important reason to test the A14 tune also for the NLO MC generators and see

3. Monte Carlo generator studies

Parameter	Value
Tune:ee	7
Tune:pp	14
PDF:useLHAPDF [36]	on
PDF:LHAPDFset	NNPDF23_lo_as_0130_qed [37]
SigmaProcess:alphaSvalue	0.1399
SpaceShower:pTmaxMatch	2
SpaceShower:pTdampMatch	1
SpaceShower:rapidityOrder	on
SpaceShower:pT0Ref	1.556
SpaceShower:pTmaxFudge	0.9091
SpaceShower:pTdampFudge	1.054
SpaceShower:alphaSvalue	0.1273
TimeShower:alphaSvalue	0.1273
MultipartonInteractions:pT0Ref	2.086
MultiPartonInteractions:pT0ref	0.1261
BeamRemnants:primordialKThard	0.1879
BeamRemnants:reconnectRange	1.71

Table 3.1.: Settings of the A14 tune for PYTHIA8.

whether this tune is sufficient such that the NLO MC predictions can correctly describe the observed data. However, if the tune is not suitable for the NLO generators, meaning significant deviations from data are observed, then these MC generators have to be tuned further and in a proper way.

In this first study, the PYTHIA8 A14 tune will be interfaced with the matrix elements generated at NLO with POWHEG [40] and MADGRAPH5_amc@NLO as well as the leading order generator PYTHIA8 stand-alone. These three MC predictions are compared to ATLAS $\sqrt{s} = 7$ TeV data as well as POWHEG+PYTHIA6 with the P2011c tune. The following observables are used for this comparison and are implemented in Rivet analyses:

- $t\bar{t}$ gap fractions [32]
- $t\bar{t}$ jet shapes [29, 30]
- top quark p_T and $t\bar{t}$ system p_T
- Jet multiplicity in $t\bar{t}$ events

The first two are official ATLAS Rivet routines performing $t\bar{t}$ measurements in the dilepton (and semi-leptonic) channel. Other official $t\bar{t}$ Rivet routines are not considered in this chapter. The last three observables listed above are part of a private, unvalidated ATLAS

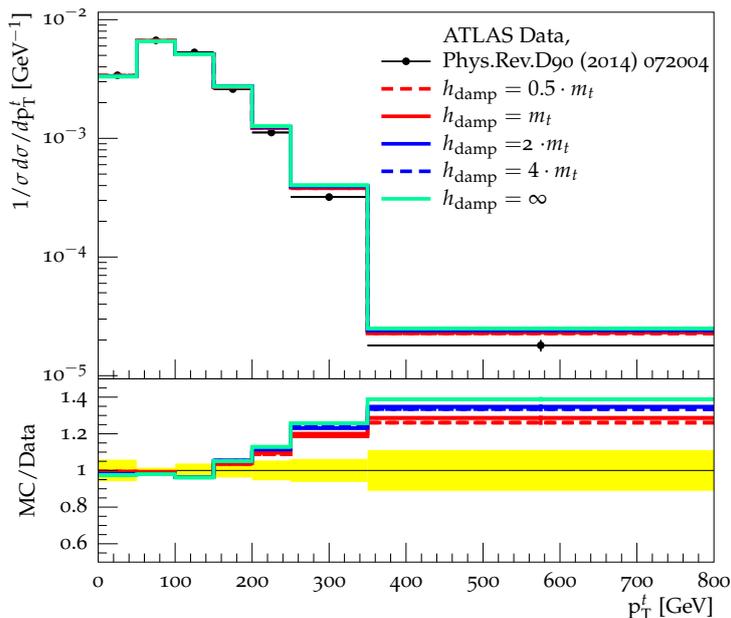


Figure 3.1.: Top quark p_T in $t\bar{t}$ events as predicted by POWHEG+PYTHIA6 with the Perugia2011c tune. The POWHEG parameter h_{damp} is varied from ∞ to $0.5 \cdot m_t$. Large deviations from ATLAS data in the high p_T bins are observed for all h_{damp} parameter values [39].

routine. Therefore, these observables cannot be compared directly to data, but instead can be compared to the $t\bar{t}$ simulation from POWHEG+PYTHIA6. There is a corresponding official ATLAS Rivet analysis to measure these three observables [41], but only in semi-leptonic $t\bar{t}$ events which are not included in this study.

Event selection

In all three Rivet analyses, the dileptonic $t\bar{t}$ events are selected differently. A brief overview of the p_T and η requirements for the charged leptons and jets are given in tab. 3.2.

Cut/Analysis	$t\bar{t}$ gap fraction	$t\bar{t}$ jet shapes	top p_T , $t\bar{t}$ p_T , N_{jets}
Electron p_T & η	$p_T > 25 \text{ GeV}$, $ \eta < 2.47$	$p_T > 25 \text{ GeV}$, $ \eta < 2.47$	$p_T > 30 \text{ GeV}$, $ \eta < 2.5$
Muon p_T & η	$p_T > 20 \text{ GeV}$, $ \eta < 2.5$	$p_T > 20 \text{ GeV}$, $ \eta < 2.5$	$p_T > 30 \text{ GeV}$, $ \eta < 2.5$
Jet p_T & η	$p_T > 25 \text{ GeV}$, $ \eta < 2.4$	$p_T > 25 \text{ GeV}$, $ \eta < 2.5$	$p_T > 25 \text{ GeV}$, $ \eta < 2.5$
Jet overlap removal	$\Delta R(\text{jet}, \text{lepton}) < 0.4$	$\Delta R(\text{jet}, \text{electron}) < 0.2$	$\Delta R(\text{jet}, \text{lepton}) < 0.4$

Table 3.2.: Selection of dileptonic $t\bar{t}$ events in the three used analyses.

The jets are reconstructed using the anti- k_t algorithm with a cone size of $R = 0.4$ [42] and the “jet overlap removal” criterion in tab. 3.2 excludes any jets within $\Delta R < 0.4$ (0.2) of a charged lepton (electron) from the analysis.

The $t\bar{t}$ gap fraction analysis applies additional selection criteria: in events where both

3. Monte Carlo generator studies

charged leptons have the same flavour, i.e. e^+e^- or $\mu^+\mu^-$, charged leptons are required to have opposite charges and their dilepton mass m_{ll} must be above 15 GeV and outside a 10 GeV window around the Z boson mass m_Z . Additionally, the missing transverse momentum from the neutrinos has to be greater than 40 GeV. If the charged leptons have different flavour, that is $e\mu$ events, then, in addition to opposite charges, the scalar sum of the transverse momenta of both charged leptons and all jets, denoted as H_T , is required to be greater than 130 GeV. All events must have at least two b -tagged jets.

Observables defined in $t\bar{t}$ events

The gap fraction analysis measures any additional central jet activity in a dileptonic $t\bar{t}$ event. This is illustrated in fig. 3.2. Two important observables, namely Q_0 and Q_{sum} , serve as p_T thresholds and are used to define the gap fraction. If there is an additional jet within a certain pseudorapidity interval with a transverse momentum greater than Q_0 , then the event will be vetoed. In the same way, if there are multiple additional jets and the sum of their transverse momenta is greater than Q_{sum} , then the event will be vetoed as well. The gap fraction f_{gap} is defined as the ratio of the number of events that are not vetoed and the total number of events. This means for a high p_T threshold the gap fraction is close to 1.

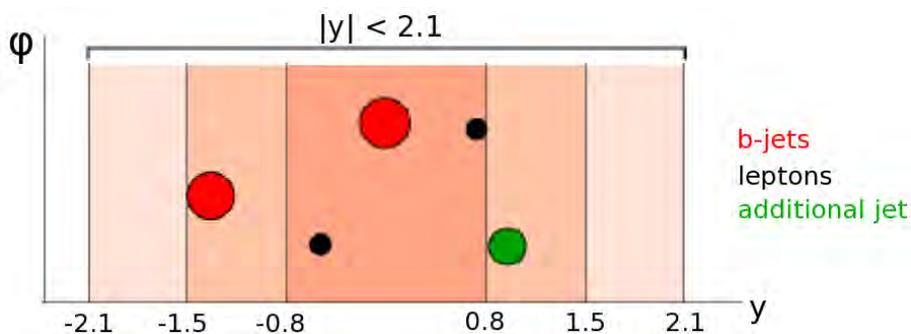


Figure 3.2.: A sketch of a $t\bar{t}$ event with an additional jet in the central rapidity region.

The differential jet shape $\rho(r, \Delta r)$ measures the fraction of the total jet momentum that is inside an annulus of radius Δr around a cone of radius r around the jet. This can be expressed by the following formula:

$$\rho(r, \Delta r) = \frac{1}{\Delta r} \frac{p_T(r - \Delta r/2, r + \Delta r/2)}{p_T(0, R)}. \quad (3.1)$$

The integrated jet shape $\Psi(r)$ is the integral over the differential jet shape from zero to r . For $r = R$ its value is $\Psi(r = R) = 1$. Fig. 3.3 shows a sketch of the integrated jet shape

for some $0 < r < R$ and for $r = R$.

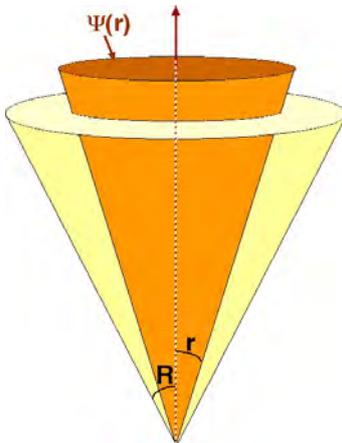


Figure 3.3.: A schematic view of the integrated jet shape for a variable r with respect to the fixed R .

Special Pythia8 settings

In addition to the A14 tune of PYTHIA8, both matrix element generators POWHEG and MADGRAPH5_aMC@NLO need special settings to match the matrix elements of the hard process correctly to the parton shower in PYTHIA8. The MADGRAPH5_aMC@NLO settings are listed in tab. 3.3. These settings are recommended by the MADGRAPH5_aMC@NLO authors. For POWHEG, a special PYTHIA8 routine is used, called main31, which is part of the PYTHIA8 distribution. Using main31, additional PYTHIA8 parameters listed in tab. 3.4 can be set that are only available for the interface with POWHEG. For both NLO matrix element generators, the parameter *SpaceShower:pTdampMatch* is not set to 1 as is proposed for the A14 tune, instead the default value 0 is chosen to not include any special damping in the showering.

Finally, POWHEG has an additional, very important parameter, called h_{damp} , which is set during the generation of the hard process [43]. This parameter describes the damping of radiation with a high transverse momentum in POWHEG [43]. Setting $h_{\text{damp}} = \infty$ corresponds to no damping. In this case, the dependence on the hard scattering scale of the matrix element generation is underestimated at high transverse momentum. The current default value in ATLAS for $t\bar{t}$ events is $h_{\text{damp}} = m_{\text{top}}$.

The results of this A14 tune analysis are shown in fig. 3.4 to 3.7. The MC prediction of the three generators PYTHIA8, POWHEG+PYTHIA8 and MADGRAPH5_aMC@NLO+PYTHIA8 are compared to POWHEG+PYTHIA6 with the Perugia2011c tune as well as ATLAS Run I data for $\sqrt{s} = 7$ TeV. Aside from the tune, the parameter h_{damp} is different

3. Monte Carlo generator studies

Parameter	Value
SpaceShower:pTmaxMatch	1
SpaceShower:pTmaxFudge	1
SpaceShower:MEcorrections	off
TimeShower:pTmaxMatch	1
TimeShower:pTmaxFudge	1
TimeShower:MEcorrections	off
TimeShower:globalRecoil	on
TimeShower:limitPTmaxGlobal	on
TimeShower:nMaxGlobalRecoil	1
TimeShower:globalRecoilMode	2
TimeShower:nMaxGlobalBranch	1

Table 3.3.: Special settings needed for the interface of MADGRAPH5_aMC@NLO with PYTHIA8. These settings are recommended by the MADGRAPH5_aMC@NLO authors.

Parameter	Value
POWHEG:nFinal	2
POWHEG:veto	1
POWHEG:vetoCount	3
POWHEG:pThard	0
POWHEG:pTemt	0
POWHEG:emitted	0
POWHEG:pTdef	1
POWHEG:MPIveto	0

Table 3.4.: New PYTHIA8 parameters that have to be used for the interface of POWHEG with PYTHIA8 using main31.

for POWHEG+PYTHIA6 and POWHEG+PYTHIA8. While in POWHEG+PYTHIA6 h_{damp} is set to m_{top} , POWHEG+PYTHIA8 is using $h_{\text{damp}} = \infty$ which is already known to result in large deviations from data.

From the gap fraction plots in fig. 3.4 one can see multiple aspects: First of all, POWHEG+PYTHIA6 consistently predicts larger gap fractions than the measured data, while the other three generators predict smaller gap fractions than observed in data. Both NLO MC generators interfaced with PYTHIA8 with the A14 tune show significant deviations from data, while the LO generator PYTHIA8 stand-alone describes the data best. This is not surprising, as the A14 tune was developed for PYTHIA8 stand-alone using the gap fraction analysis among others.

The b -jet shapes shown in fig. 3.5 demonstrate that all four MC generators roughly describe data in the low p_T region, while for high p_T b -jets they show large deviations

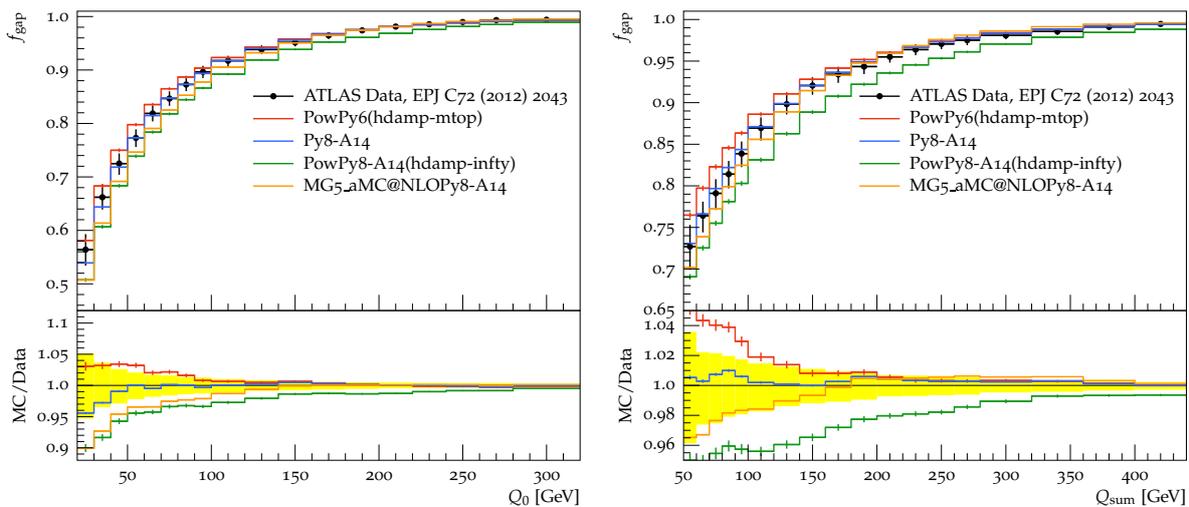


Figure 3.4.: Gap fraction as a function of Q_0 (left) and Q_{sum} (right) for a rapidity region of $|y| < 2.1$.

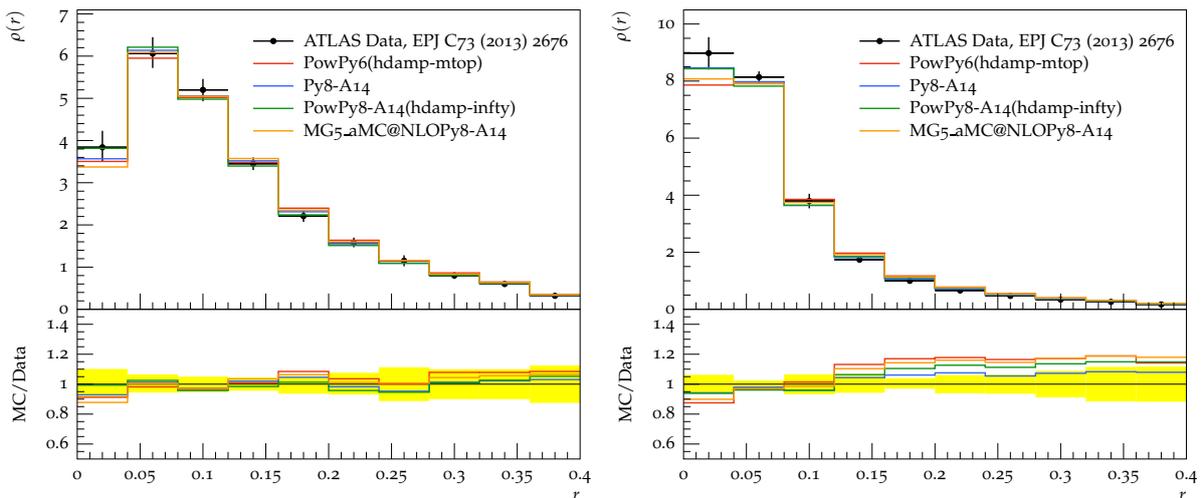


Figure 3.5.: Differential jet shape for b -jets in the low p_T region, that is $30 \text{ GeV} < p_T < 40 \text{ GeV}$ (left), and for the high p_T region $70 \text{ GeV} < p_T < 100 \text{ GeV}$ (right).

from data. This means that the A14 tune is not fully optimal and the NLO generators POWHEG+PYTHIA8 and MADGRAPH5_aMC@NLO+PYTHIA8 need further tuning.

The plots in the figures 3.6 and 3.7 show no comparison to data as these are from an unvalidated Rivet analysis. However, by comparing to POWHEG+PYTHIA6 one can still see significant differences between the four generators. POWHEG+PYTHIA6 with $h_{\text{damp}} = m_{\text{top}}$ generates a softer $t\bar{t}$ system p_T than the other three generators. It also generates a softer top quark p_T than PYTHIA8 and POWHEG+PYTHIA8, but a significantly harder top

3. Monte Carlo generator studies

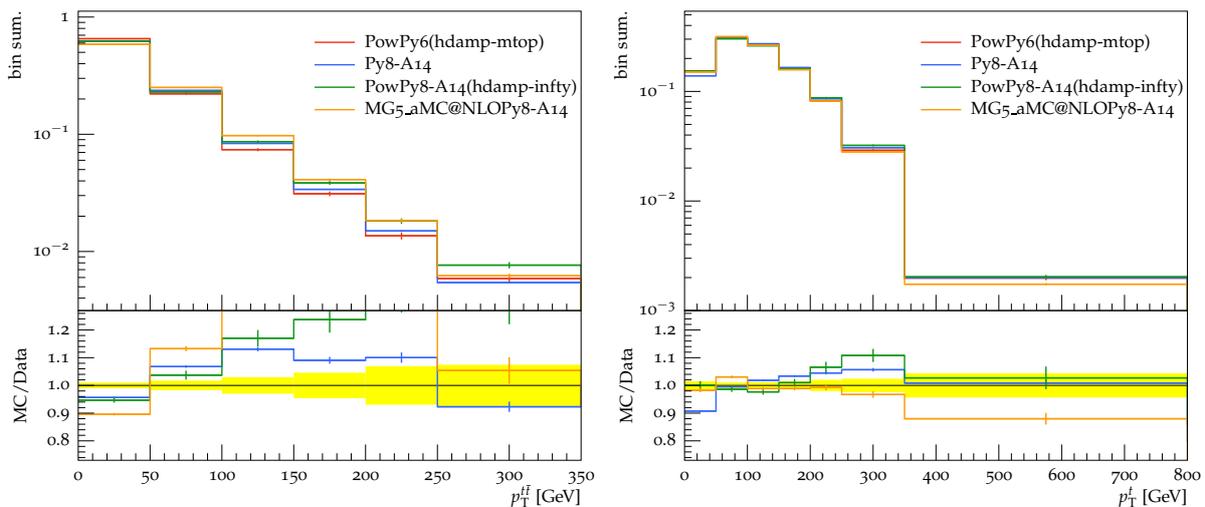


Figure 3.6.: $t\bar{t}$ system p_T (left) and top quark p_T (right).

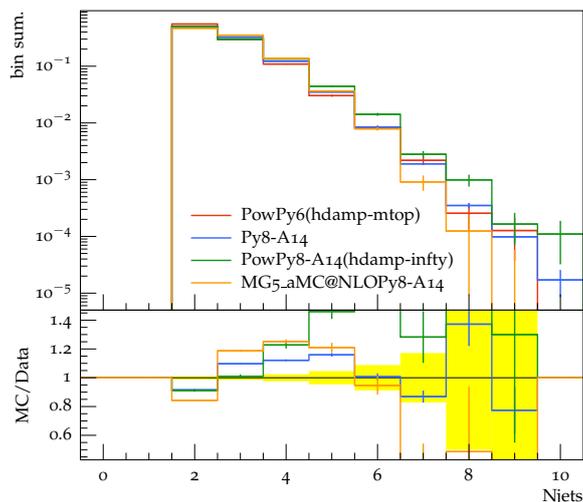


Figure 3.7.: Jet multiplicity for up to 10 jets in the final state.

quark p_T than MADGRAPH5_aMC@NLO+PYTHIA8. Additionally, both NLO MC generators with the PYTHIA8 A14 tune show significant deviations from POWHEG+PYTHIA6 for the number of jets in the $t\bar{t}$ events, especially in the high jet multiplicity bins.

3.2. Studies with MadGraph5_aMC@NLO

In this section, a short comparison of MADGRAPH5_aMC@NLO interfaced with different parton showers and different tunes is given. Fig. 3.8 shows the $t\bar{t}$ gap fractions for MADGRAPH5_aMC@NLO interfaced with HERWIG++, and PYTHIA8 with two different

tunes, compared to POWHEG+PYTHIA6 with $h_{\text{damp}} = m_{\text{top}}$ and data.

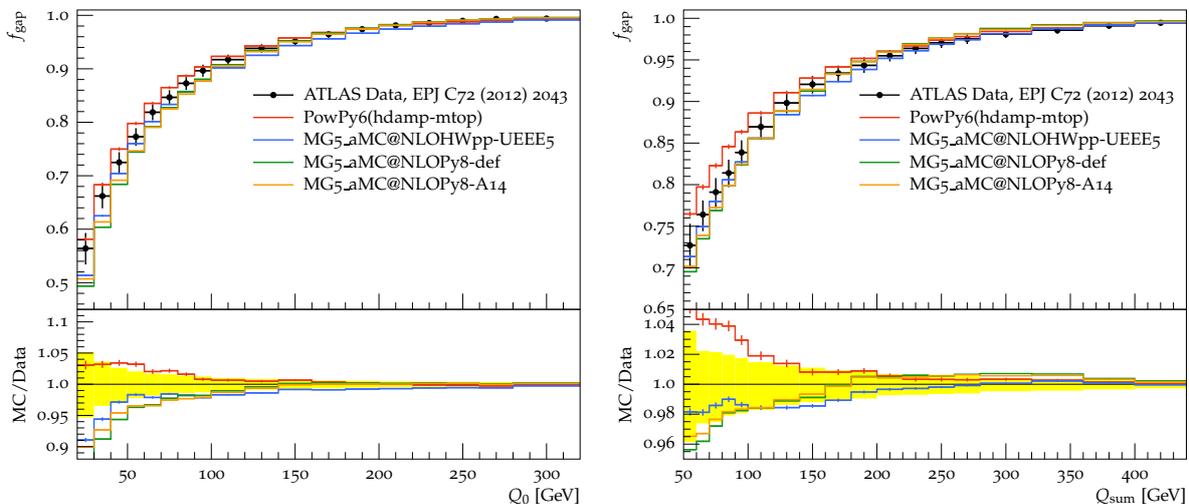


Figure 3.8.: Gap fraction as a function of Q_0 (left) and vs. Q_{sum} (right) for an inclusive rapidity region of $|y| < 2.1$.

One can see that MADGRAPH5_aMC@NLO+HERWIG++ with the Ueee5 tune shows a good agreement with data within uncertainties. It does not seem to require further dedicated tuning. This generator has been considered as a baseline MC generator for $t\bar{t}$ modelling in Run II in ATLAS. The comparison of MADGRAPH5_aMC@NLO with different PYTHIA8 tunes, where “def” corresponds to the default generator settings, shows that the A14 tune only has a small impact on the $t\bar{t}$ gap fraction distributions.

These first studies were performed to motivate that MADGRAPH5_aMC@NLO+PYTHIA8 needs dedicated tuning to better describe data, before it can be used in Run II. The main topic of this thesis is the development of this tune. The procedure and results of this tuning study are presented and discussed in the next chapter.

4. Developing a new tune for Mad-Graph5_aMC@NLO+Pythia8

In this chapter the tuning procedure and strategy as well as the chosen parameters and observables will be described in detail. The tuning results will be presented and compared to other available tunes. In addition to this, sensitivity studies, parameter correlations and uncertainties will be discussed and a short study regarding the PYTHIA8 shower model settings for the interface with MADGRAPH5_aMC@NLO will be shown.

4.1. What is Monte Carlo generator tuning?

The theory of QCD is well known where perturbation theory applies, that is for high energies and a low value of α_S , as explained in section 2.4. However, soft effects like the underlying event and hadronisation are not perturbative and have to be modelled using phenomenological MC models. This means there are model parameters not predicted by theory and therefore unknown beforehand, thus they need to be tuned. A properly tuned MC generator is essential to produce simulated events as similar as possible to real collider data [24].

As discussed in section 2.7, the parton shower MC generators simulate multiple aspects of an event, namely initial and final state radiation (ISR and FSR), the underlying event (UE) and hadronisation [24].

The FSR from the parton showering is assumed to be universal and only the scale Q^2 of the hard process varies depending on the input process. Thus, the relevant parameters in PYTHIA8 that determine FSR are α_S^{FSR} , its cut-off scale $p_{\text{T,min}}^{\text{FSR}}$ and starting scale fudge factors [24]. The first two parameters are tuned in the study presented in this thesis. α_S^{FSR} corresponds to the α_S value at the scale M_Z^2 used for final state emissions, where M_Z is the mass of the Z boson. The actual value is then regulated by the running of the scale p_{T}^2 at which the shower evaluates α_S . The other parameter, $p_{\text{T,min}}^{\text{FSR}}$, is the parton shower cut-off p_{T} in GeV for QCD emissions [2, 3]. Although FSR has been preferably tuned to

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

e^+e^- data from lepton colliders, because there are no initial state hadron complexities, using hadron collider data is also possible.

ISR is very similar to FSR, but preferably tuned to hadron collider data. The two ISR parameters chosen for this tuning are α_S^{ISR} , which is analogous to α_S^{FSR} , and $p_{\text{T0,Ref}}^{\text{ISR}}$ which regularises the divergence of the QCD emission probability in the low p_{T} limit and thus serves as another p_{T} cut-off, similar to the FSR case. This regularisation is done by multiplying the probability with a factor

$$\frac{p_{\text{T}}^2}{p_{\text{T}}^2 + (p_{\text{T0}}^{\text{ISR}})^2}, \quad (4.1)$$

and using $\alpha_S(p_{\text{T}}^2 + (p_{\text{T0}}^{\text{ISR}})^2)$, where $p_{\text{T0}}^{\text{ISR}}$ is given by

$$p_{\text{T0}}^{\text{ISR}} \equiv p_{\text{T0}}^{\text{ISR}}(\sqrt{s}) := p_{\text{T0,Ref}}^{\text{ISR}} \cdot \left(\frac{\sqrt{s}}{E_{\text{CM}}^{\text{Ref}}} \right)^{E_{\text{CM}}^{\text{Pow}}} [2, 3]. \quad (4.2)$$

$E_{\text{CM}}^{\text{Ref}}$ and $E_{\text{CM}}^{\text{Pow}}$ are two PYTHIA8 parameters to introduce the energy dependence for the ISR p_{T} cut-off, but only $p_{\text{T0,Ref}}^{\text{ISR}}$ is tuned here [2, 3].

The UE is also tuned to hadron collider data and sensitive to the PDF choice. The UE parameters that are typically tuned are the beam particle matter distribution, and the α_S and cut-off variables for Multiple Parton Interactions (MPI) [24]. Similar to FSR and ISR, two parameters are included in this tuning study, namely α_S^{MPI} and $p_{\text{T0,Ref}}^{\text{MPI}}$, which have the same meaning as the corresponding ISR parameters [2, 3].

The hadronisation has many parameters, for example string or cluster model parameters. However, these parameters are already tuned to high precision from e^+e^- data, more specifically from identified particle spectra [24], and therefore no hadronisation parameters are included here.

The last parameters of importance are more complex and related to the beam remnants and the primordial k_{T} . The latter determines the p_{T} of the initiators of hard-scattering subsystems according to Gaussian distributions in p_x and p_y separately. The widths $\sigma_{k_{\text{T}}}$ of these distributions are chosen to be dependent on the hard scale of the central process and on the mass of the whole subsystem, and are given by

$$\sigma_{k_{\text{T}}} = \left(\frac{\sigma_{\text{soft}} \cdot Q_{\text{half}} + \sigma_{\text{hard}} \cdot Q}{Q_{\text{half}} + Q} \right) \times \left(\frac{m}{m + m_{\text{half}} \cdot y_{\text{damp}}} \right) [2, 3]. \quad (4.3)$$

In this equation, Q is the hard-process renormalisation scale for the hardest process and the p_{T} scale for subsequent multiparton interactions and m the mass of the system. σ_{soft} , σ_{hard} , Q_{half} , m_{half} and y_{damp} are additional PYTHIA8 parameters [2, 3]. Only one of these

Parameter	Default	Minimum	Maximum	Tuning range
α_S^{FSR}	0.1383	0.06	0.25	0.115 - 0.150
$p_{\text{T,min}}^{\text{FSR}}$	0.4	0.1	2.0	0.5 - 2.0
α_S^{ISR}	0.137	0.06	0.25	0.115 - 0.140
$p_{\text{T0,Ref}}^{\text{ISR}}$	2.0	0.5	10.0	0.75 - 2.5
α_S^{MPI}	0.127	0.06	0.25	0.115 - 0.140
$p_{\text{T0,Ref}}^{\text{MPI}}$	2.15	0.5	10.0	1.5 - 3.0
Primordial $k_{\text{T,hard}}$	2.0	0.0	-	1.5 - 2.0
Reconnect range	10.0	0.0	10.0	1.0 - 5.0

Table 4.1.: List of parameters included in this tuning study.

parameters is tuned in these studies, namely the *primordial* $k_{\text{T,hard}}$, which is σ_{hard} in eq. 4.3 and assigned to initiators in the hard-interaction limit [2, 3]. Finally, the strength of the phenomenological colour-string reconnection mechanism, denoted as *reconnect range*, is the last parameter tuned in this study. A system with a hard scale p_{T} can be merged with one of a harder scale with a probability that is

$$\frac{p_{\text{T0,Rec}}^2}{p_{\text{T0,Rec}}^2 + p_{\text{T}}^2}, \quad (4.4)$$

where $p_{\text{T0,Rec}}$ is the reconnect range multiplied by $p_{\text{T0}}(\sqrt{s})$, the latter being the same energy-dependent dampening parameter as used for MPI. Thus it is easy to merge a low- p_{T} system with any other, but difficult to merge two high- p_{T} ones with each other [2, 3]. Tab. 4.1 shows all chosen parameters, their default, minimum, and maximum possible value in PYTHIA8, as well as the range it is tuned in.

4.2. Tuning with Professor

The tuning procedure pursued in this chapter is described in the following.

1. Choose n parameters and their sampling ranges. For this study, $n = 8$ is used and the parameter list is shown in tab. 4.1.
2. The PROFESSOR tuning toolkit (version 1.4.0) [24] is used to randomly sample N parameter points in the n -dimensional hyperspace. This is illustrated in fig. 4.1. N is given by the number of parameters n and the polynomial order of the interpolation one chooses to fit. The tuning is done with the default interpolation method, which is a 3rd order polynomial. Thus, $N_{\text{min}}(n) = 1+n+n(n+1)/2+n(n+1)(n+2)/6 = 165$

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

for $n = 8$ [24]. An oversampling of a factor of 3 is recommended [24] and therefore, $N = 500$ is chosen here.

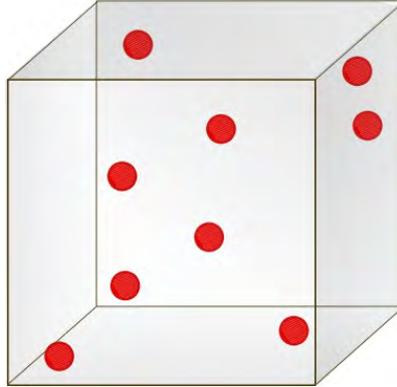


Figure 4.1.: Example of a random sampling of 9 points in a 3-dimensional space.

3. Generate MC events according to the observables one wants to tune to and simulate a high number of events such that the statistical uncertainty is small compared to the data uncertainty. This is explained in more detail in section 4.3.
4. Run the parton shower using the randomly sampled parameters, analyse the events with Rivet (version 2.2.0) and fill the histograms for the chosen observables. In this study, the A14 tune is taken as the basis before the additional tuning parameter variations are applied in PYTHIA8.
5. For each histogram bin use N points to fit the 3rd order polynomial interpolation. This interpolation calculates the sensitivity of each bin to the respective parameters, i.e. determines how the bin content changes when changing the tuning parameters.
6. Construct the overall χ^2 , which is given by

$$\chi^2 \approx \sum_{\text{bins}} \frac{(\text{interpolation} - \text{data})^2}{\text{error}^2} [24]. \quad (4.5)$$

Here, the error corresponds to the data uncertainty.

7. Numerically minimise this χ^2 , namely by varying the parameters according to the interpolation. This is shown in fig. 4.2 for 1 bin in 1 dimension. The resulting output of PROFESSOR is the parameter values that minimise this χ^2 .
8. Compare the tuning result with other tunes to determine whether the result is an improvement to already existing tunes.

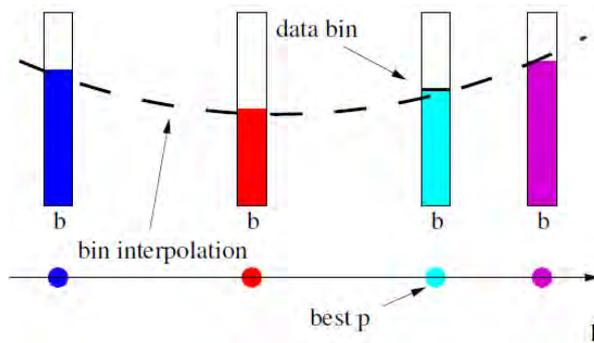


Figure 4.2.: Minimising the χ^2 by varying the parameter until the MC prediction matches the data. b labels the bin and p the tuned parameter.

4.3. Observables chosen for the tuning

After the parameters for the tuning have been chosen, one needs to find observables which are sensitive to these parameters. Sensitivity studies allow to find such observables, but in this thesis the observables used for developing the A14 tune, described in section 3.1, are taken as baseline. Additionally, some observables are included in this study that are known to be sensitive to these parameters, but for which measurements were not accessible at the time of the development of the A14 tune. All chosen observables are related to three main processes, namely inclusive dijet, $t\bar{t}$ and Z boson production, which are generated with MADGRAPH5_AMC@NLO (version 2.2.1) at next-to-leading order and showered with PYTHIA8 (version 8.186). The PDF set used for the hard process is CT10 [44], while NNPDF2.3LO [37] is used in PYTHIA8, with the PDFs taken from LHAPDF (version 6.1.3) [36]. A MC generator that is able to describe all data measured at the LHC with one general tune is desirable and thus, by tuning the parameters to more than one individual process, the resulting tune is more universal, because of the different contributions to the sensitivities. For this tuning, the $t\bar{t}$ events are analysed either in the dileptonic or semi-leptonic channel. Z events are studied in the e^+e^- or $\mu^+\mu^-$ channel, due to their purity. Finally, the dijet final state consists of two hard jets of arbitrary flavour. Tab. 4.2 gives a brief overview of all tuning observables, the number of events generated for each process and the respective decay channel. The corresponding Rivet analyses are listed in tab. A.1 in the appendix and correspond to ATLAS measurements using 2010 or 2011 data at $\sqrt{s} = 7$ TeV.

The $t\bar{t}$ gap fraction analysis has a filter to only accept dileptonic events, while the other $t\bar{t}$ analyses require semi-leptonic events. This is the reason that two different $t\bar{t}$ samples of different sizes are taken for this study, namely a dilepton sample with three million,

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

Process	$t\bar{t}$	Z	dijet
Observables	gap fractions (dilepton) [32] jet shapes (inclusive) [30] N_{jets} & jet p_{T} (inclusive) [41]	$Z p_{\text{T}}$ ($Z \rightarrow \mu^+\mu^-$) [35] $Z \phi_{\eta}^*$ ($Z \rightarrow e^+e^-$) [45] Z UE ($Z \rightarrow e^+e^-$) [46]	jet shapes [29] angular decorrelation [31] gap fractions [47] multi-jet cross sections [33] track-jet properties [27] jet substructure [28] track-jet UE [25] η dependence on E_{T} [48] leading jet UE [26]
Events generated	3M dilepton 6M inclusive	2M $Z \rightarrow \mu^+\mu^-$ 2M $Z \rightarrow e^+e^-$	6M low p_{T} 3M high p_{T}

Table 4.2.: The observables used for the tuning with respective decay channels in brackets. The last row lists the number of events generated for each sample.

and an inclusive sample with six million events.

In a similar way, one cannot tune the $Z p_{\text{T}}$ and $Z \phi_{\eta}^*$ using the same events due to strong correlation. Thus, the $Z p_{\text{T}}$ is analysed in the $\mu^+\mu^-$ channel and $Z \phi_{\eta}^*$ in $Z \rightarrow e^+e^-$ events.

These observables are very sensitive to ISR, but only weakly sensitive to FSR and MPI, as will be shown in section 4.7.1. To increase the sensitivity to FSR, dijet observables are included in the tuning. The sensitivity to MPI is mainly driven by the Z UE and dijet UE analyses.

Two different dijet samples are used, but this is purely a technical strategy. The default event generation with MADGRAPH5_aMC@NLO sets a minimum transverse momentum for final state particles from the hard process. Due to the fast decrease of the dijet production cross section with increasing jet transverse momentum [43], two samples with different p_{T} slices have been generated: six million events with a low minimum p_{T} of 10 GeV and three million events with a higher minimum p_{T} , namely 100 GeV. By adding this second sample, one is able to describe dijet observables for jets with transverse momenta of up to 1000 GeV.

4.4. Event selection and ATLAS measurements

This section describes the event selection applied in each of the Rivet analyses used in the tuning.

4.4.1. $t\bar{t}$ events

For the dileptonic $t\bar{t}$ event selection of the gap fraction analysis [32], refer to section 3.1.

The $t\bar{t}$ jet shape analysis requires semi-leptonic events, since, instead of measuring the b -jet shapes as in chapter 3, the light-jet shapes are taken for the tuning. The lepton selection remains identical, additionally two light-jets are selected as the pair of non- b -tagged jets whose invariant mass is closest to the W mass [30].

The $t\bar{t}$ N_{jets} & jet p_{T} analysis also selects semi-leptonic $t\bar{t}$ events by requiring the charged lepton and all jets to have $p_{\text{T}} > 25 \text{ GeV}$ and $|\eta| < 2.5$. If a jet is within $\Delta R < 0.4$ of a charged lepton or $\Delta R < 0.5$ of another jet, the event is vetoed. The neutrino is required to have $p_{\text{T}}^{\nu} > 30 \text{ GeV}$ and the transverse mass of the leptonically decaying W has to be greater than 35 GeV [41]. This transverse W mass is defined as

$$M_{\text{T}}^W = \sqrt{2 \cdot p_{\text{T}}^l \cdot p_{\text{T}}^{\nu} \cdot (1 - \cos(\phi^l - \phi^{\nu}))} \quad [8]. \quad (4.6)$$

The three measurements are highly sensitive to the production of additional jets in $t\bar{t}$ events.

4.4.2. Z events

The $Z \rightarrow l^+l^-$ events are all selected in the same way: the two leptons are required to have a $p_{\text{T}} > 20 \text{ GeV}$, $|\eta| < 2.4$ and the invariant mass of the dilepton pair has to be within 25 GeV of the Z mass, meaning $66 \text{ GeV} < m_{ll} < 116 \text{ GeV}$. Only dressed leptons are taken, which means they are reconstructed taking into account all radiated photons within a $\Delta R = 0.1$ cone around the lepton [35, 45, 46].

The optimal experimental observable to probe the low Z p_{T} domain of Z/γ^* production was found to be ϕ_{η}^* which is defined as

$$\phi_{\eta}^* = \tan(\phi_{\text{acop}}/2) \cdot \sin(\theta_{\eta}^*), \quad (4.7)$$

with $\phi_{\text{acop}} = \pi - \Delta\phi$ and $\Delta\phi$ being the azimuthal opening angle between the two leptons. The angle θ_{η}^* is defined by $\cos(\theta_{\eta}^*) = \tanh[(\eta^- - \eta^+)/2]$ where η^- and η^+ are the pseudorapidities of the negatively and positively charged lepton, respectively [45]. The ϕ_{η}^* variable is strongly correlated to the Z p_{T} [45]. While the high Z p_{T} regime is described by perturbative QCD, the low Z p_{T} range is governed by ISR [35].

The observables of the Z underlying event ATLAS measurements are very sensitive to numerous properties of the UE, such as the $\sum p_T$ of final state particles, number of charged particles and their mean p_T . These properties are measured in different regions relative to the moving direction of the Z boson and each region is sensitive to different effects [46]. This is illustrated in fig. 4.3 and explained in the following. The “away” region is dominated by particles balancing the momentum of the Z boson. The “transverse” region is sensitive to the underlying event, since it is by construction perpendicular to the direction of the Z boson and hence it is expected to have less activity from the hard scattering process compared to the away region. The two opposite transverse regions may be distinguished on an event-by-event basis through their amount of activity, as measured by the $\sum p_T$ of the charged particles in each of them. The more or less active regions are then referred to as “trans-max” and “trans-min”, respectively. The trans-max side is more likely to be affected by wide-angle emissions associated with the hard process and correspondingly the trans-min observables have the potential to be more sensitive to soft MPI and beam-remnant activity. The “toward” region is similarly unaffected by additional activity from the hard scatter and thus included in this measurement.

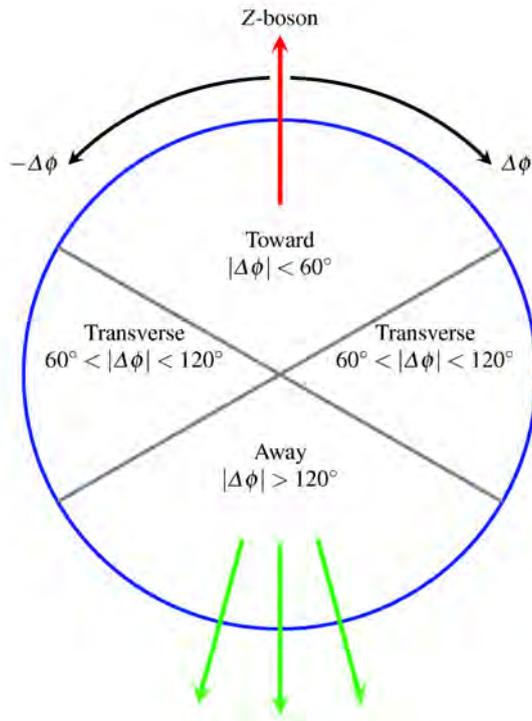


Figure 4.3.: The Z UE analysis defines three regions relative to the moving direction of the Z boson, namely the toward, away and transverse region.

4.4.3. Dijet events

The event selection applied in the dijet measurements is summarised in tab. 4.3. Depending on the analysis and observable, the jets are required to have a different minimum p_T , maximum η , or they are reconstructed with a different cone size R . Except for the dijet substructure analysis, the jets are always reconstructed with the anti- k_t algorithm [42].

Analysis/Cut	Jet p_T	Jet η	Jet cone size R
Jet shapes	$30 \text{ GeV} < p_T < 600 \text{ GeV}$	$ \eta < 2.8$	$R = 0.6$
Angular decorrelation	$p_T > 100 \text{ GeV}$	$ \eta < 2.8$	$R = 0.6$
Gap fractions	$p_T > 20 \text{ GeV}$, 2 lead. jets with $\sum p_T > 100 \text{ GeV}$, 2 Fwd/Bkwd jets with $\sum p_T > 100 \text{ GeV}$	$ \eta < 4.4$	$R = 0.6$
Multi-jets	$p_T > 60 \text{ GeV}$, but at least 1 jet with $p_T > 80 \text{ GeV}$	$ \eta < 2.8$	$R = 0.4, 0.6$
Track-jet properties	$p_T > 0.3 \text{ GeV}$	$ \eta < 2.5$	$R = 0.4, 0.6$
Substructure	$p_T > 200 \text{ GeV}$	$ \eta < 2.0$	anti- k_t : $R = 1.0$ C/A [49]: $R = 1.2$
Track-jet UE	$p_T > 4 \text{ GeV}$	$ \eta < 1.5$	$R = 0.2, 0.4, 0.6, 0.8, 1.0$
E_T flow (η dependence of E_T)	$p_T > 20 \text{ GeV}$, 2 lead. jets with $E_T > 20 \text{ GeV}$, $E_T^{2\text{nd}}/E_T^{1\text{st}} > 0.5$	$ \eta < 2.5$, back-to-back	$R = 0.4$
Leading jet UE	$p_T > 20 \text{ GeV}$	$ \eta < 2.8$	$R = 0.4$

Table 4.3.: Event selection for dijet events in the different analyses used for the tuning.

The jet shape analysis for dijet events and thus the definition of the jet shape is analogous to the $t\bar{t}$ case [29]. Additionally, the jet shapes are measured in different rapidity regions.

The dijet angular decorrelation routine measures the azimuthal angle $\Delta\phi$ between the two leading jets. Distributions in $\Delta\phi$ are sensitive to the presence of additional high p_T jets. Events with only two high p_T jets have small azimuthal decorrelations, i.e. $\Delta\phi \sim \pi$, while $\Delta\phi \ll \phi$ is evidence of events with several high p_T jets. Thus, distributions in $\Delta\phi$ allow to test perturbative QCD (pQCD) calculations for multiple jet production without requiring the measurement of additional jets. Smaller values of $\Delta\phi$ require additional activity such as soft radiation, and therefore are sensitive to soft QCD effects [31].

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

The dijet gap fraction analysis is similar to the $t\bar{t}$ analogue, but provides additional observables such as measuring the gap fraction and the jet multiplicity as functions of the rapidity separation $|\Delta y|$ for a fixed threshold Q_0 . Therefore, this analysis allows to study a diverse range of pQCD phenomena and the effects of wide-angle soft-gluon radiation. The fact that at large values of $|\Delta y|$ the gap fraction predicted by MC generators deviates from the data is expected, because the NLO matrix element plus parton shower approximation does not contain full QCD calculation contributions that become important as $|\Delta y|$ increases [47].

The multi-jet cross section measurement is important to study the background of many searches for new physics and to test LO and NLO pQCD effects. In this study, the 3-to-2 jet ratios are the only observables considered from this routine, which are sensitive to additional jet radiation [33]. The 3-to-2 jet ratio measures the number of events containing at least three high p_T jets with respect to number of events containing at least two.

The track-jet properties analysis measures a wide range of observables with low p_T jets reconstructed from charged particle tracks and is sensitive to perturbative as well as soft QCD [27]. The observable used in this study is the density of charged particles in the $\phi - y$ space, denoted as $\rho_{\text{ch}}(r)$, which is measured as a function of the radial distance r of charged particles from the axis of the jet that contains them. $\rho_{\text{ch}}(r)$ is defined as

$$\rho_{\text{ch}}(r) = \frac{1}{N_{\text{jet}}} \frac{dN_{\text{ch}}}{2\pi r dr}. \quad (4.8)$$

This represents a particle number density, rather than the related energy density variable used for calorimeter-based jet shapes defined before.

Jet substructure studies allow to identify single, highly boosted jets of interest from the overall jet background [28]. Such techniques have been found promising for boosted W decay identification, Higgs searches and boosted top identification amongst others. Two different jet reconstruction algorithms are tested, namely the anti- k_t and the Cambridge-Aachen algorithm [49] and the jets are required to have a high p_T typical from a boosted regime. In addition to the jet mass, two other types of observables are measured in the analysis included in this routine, namely the k_t splitting scales $\sqrt{d_{ij}}$ and the N -subjettiness τ_N .

The k_t splitting scales are defined by reclustering the constituents of the jet with the k_t recombination algorithm. The k_t -distance of the final clustering step can be used to define

a splitting scale variable $\sqrt{d_{12}}$:

$$\sqrt{d_{12}} = \min(p_{Tj_1}, p_{Tj_2}) \times \Delta R_{j_1, j_2}, \quad (4.9)$$

where j_1 and j_2 are the two jets before the final clustering step. The parameter $\sqrt{d_{12}}$ can therefore be used to distinguish heavy particle decays, which tend to be more symmetric, from the largely asymmetric splittings of quarks and gluons. The variable $\sqrt{d_{23}}$ is defined analogously, but for the two objects combined in the penultimate clustering step [28].

The N -subjettiness variables τ_N are a measure of how likely a jet is composed of N different subjects. These N subjects define axes within the jet around which the jet constituents may be concentrated. The τ_N are then defined as the following sum over all constituents k of the jet:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \times \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}) \quad (4.10)$$

$$\text{with } d_0 = \sum_k p_{T,k} R, \quad (4.11)$$

where $\Delta R_{i,k}$ is the distance from the subject i to the constituent k and R is the cone size of the original jet algorithm. With this, one can define $\tau_{21} \equiv \tau_2/\tau_1$ and $\tau_{32} \equiv \tau_3/\tau_2$, which are measures of how much the jet is better described by two(three) subjects than one(two), respectively [28]. All observables of the substructure analyses are therefore sensitive to radiation and additional splittings of quarks and gluons.

The pseudorapidity dependence of E_T measures the sum of the transverse energy of either low p_T particles or two jets with a transverse energy greater than 20 GeV in different rapidity intervals. The latter case is of particular interest since it corresponds to an event with hard parton-parton scattering. Measurements in the region transverse to this hard scatter allow to probe the particle kinematics in the UE [48].

The track-jet UE and leading jet UE analyses for dijet events are very similar to the Z UE one. However, in contrast to the Z UE analysis, the toward region is not measured for dijet topologies, since it is dominated by the leading jet and thus less sensitive to the UE [25, 26].

All discussed dijet analyses use ATLAS $\sqrt{s} = 7$ TeV data recorded in 2010 and thus some of the measured observables show large statistical uncertainties. The $t\bar{t}$ and Z analyses use ATLAS $\sqrt{s} = 7$ TeV data from 2011.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

Finally, a concrete list of the specific observables included in the tuning and their respective bin ranges is shown in tab. 4.4.

Observable
$t\bar{t}$ gap fractions [32] Gap fraction vs Q_0 and Q_{sum} for $ y < 2.1$ (2 distributions)
$t\bar{t}$ jet shapes [30] Differential light-jet shape $\rho(r)$ (5 distributions)
$t\bar{t}$ N_{jets} & jet p_{T} [41] $t\bar{t}$ cross-section vs. jet multiplicity for jets above 25 and 80 GeV (2 distributions) $t\bar{t}$ cross-section vs. 1 st and 5 th jet p_{T} (2 distributions)
Z/γ^* boson angular correlations [45] $Z \phi_{\eta}^*$ profile for $Z \rightarrow ee$ (dressed) and $\phi_{\eta}^* < 0.5$ (1 distribution)
Z/γ^* boson p_{T} [35] $Z p_{\text{T}}$ profile for $Z \rightarrow \mu\mu$ (dressed) and $Z p_{\text{T}} < 50$ GeV (1 distribution)
Z/γ^* boson underlying event [46] $\sum p_{\text{T}}$ vs. $Z p_{\text{T}}$ for $Z \rightarrow ee$ (dressed) and $Z p_{\text{T}} < 5$ GeV (2 distributions) N_{ch} vs. $Z p_{\text{T}}$ for $Z \rightarrow ee$ (dressed) and $Z p_{\text{T}} < 5$ GeV (2 distributions)
Dijet jet shapes [29] Jet shape ρ (49 distributions)
Dijet angular decorrelations [31] Dijet azimuthal decorrelations (3 distributions)
Dijet gap fractions [47] Gap fraction vs $ \Delta y $ with Fwd/Bwd jets for $ \Delta y < 3.0$ (5 distributions) \bar{N}_{jet} vs $ \Delta y $ with Fwd/Bwd jets for $ \Delta y < 3.0$ (4 distributions)
Multi-jets [33] 3-to-2 jet ratios for $p_{\text{T}}^{\text{jets}} > 80, 110$ GeV ($R = 0.6$) (2 distributions)
Track-jet properties [27] Charged jet $\rho_{\text{ch}}(r)$ for $R = 0.4$ (16 distributions)
Substructure [28] Jet mass, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, τ_{21} , τ_{23} (27 distributions)
Track-jet UE [25] Mean N_{ch} vs. $p_{\text{T}}^{\text{lead}}$ in the transverse region for $p_{\text{T}}^{\text{lead}} > 30$ GeV (5 distributions) Mean p_{T} vs. $p_{\text{T}}^{\text{lead}}$ in the transverse region for $p_{\text{T}}^{\text{lead}} > 30$ GeV (5 distributions)
η dependence on E_{T} [48] $\sum E_{\text{T}}$ for the dijet selection (6 distributions)
Leading jet UE [26] Trans-min $\sum p_{\text{T}}^{\text{ch}}$ vs. $p_{\text{T}}^{\text{lead}}$ in excl. dijet events for $p_{\text{T}}^{\text{lead}} > 200$ GeV (1 distribution) Trans-min N_{ch} vs. $p_{\text{T}}^{\text{lead}}$ in excl. dijet events for $p_{\text{T}}^{\text{lead}} > 200$ GeV (1 distribution)

Table 4.4.: The specific observables and bin ranges from the Rivet analyses chosen for the tuning. All dijet observables have a weight factor of 1, while all $t\bar{t}$ and Z observables have a weight factor of 5.

4.5. Tuning results

In this section, the results of the tuning are presented. Fig. 4.4 - 4.6 show a subset of the observables used in the tune. The remaining plots are shown in appendix B. For each observable, the obtained tuning result is compared to ATLAS $\sqrt{s} = 7$ TeV data as well as MADGRAPH5_aMC@NLO+PYTHIA8 with the A14 tune. In the case of $t\bar{t}$ and Z observables, comparisons to MADGRAPH5_aMC@NLO+PYTHIA8 with the 4C tune [50] and default MADGRAPH5_aMC@NLO+PYTHIA8 options are added. For $t\bar{t}$ distributions a comparison to the prediction of POWHEG+PYTHIA6 with the Perugia2011c tune and $h_{\text{damp}} = m_{\text{top}}$ is also shown. The 4C tune is a PYTHIA8 tune (since version 8.150) developed by the PYTHIA8 authors [2, 3], while the default MADGRAPH5_aMC@NLO+PYTHIA8 options are listed in tab. 3.3 in section 3.1. Tab. 4.5 lists the parameter fit result from the PROFESSOR tool including the goodness of the fit and shows comparisons to a selection of other existing tunes. In this table, the Monash tune [51] is a general tune for PYTHIA8 and also the basis of the A14 tune, while ATTBAR [52] and AZ [35] are dedicated tunes for $t\bar{t}$ and Z boson modelling, respectively. Finally, tab. 4.6 compares the performance of the new tune against the A14 tune for the used analyses. Section 4.6 thoroughly discusses the results presented here.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

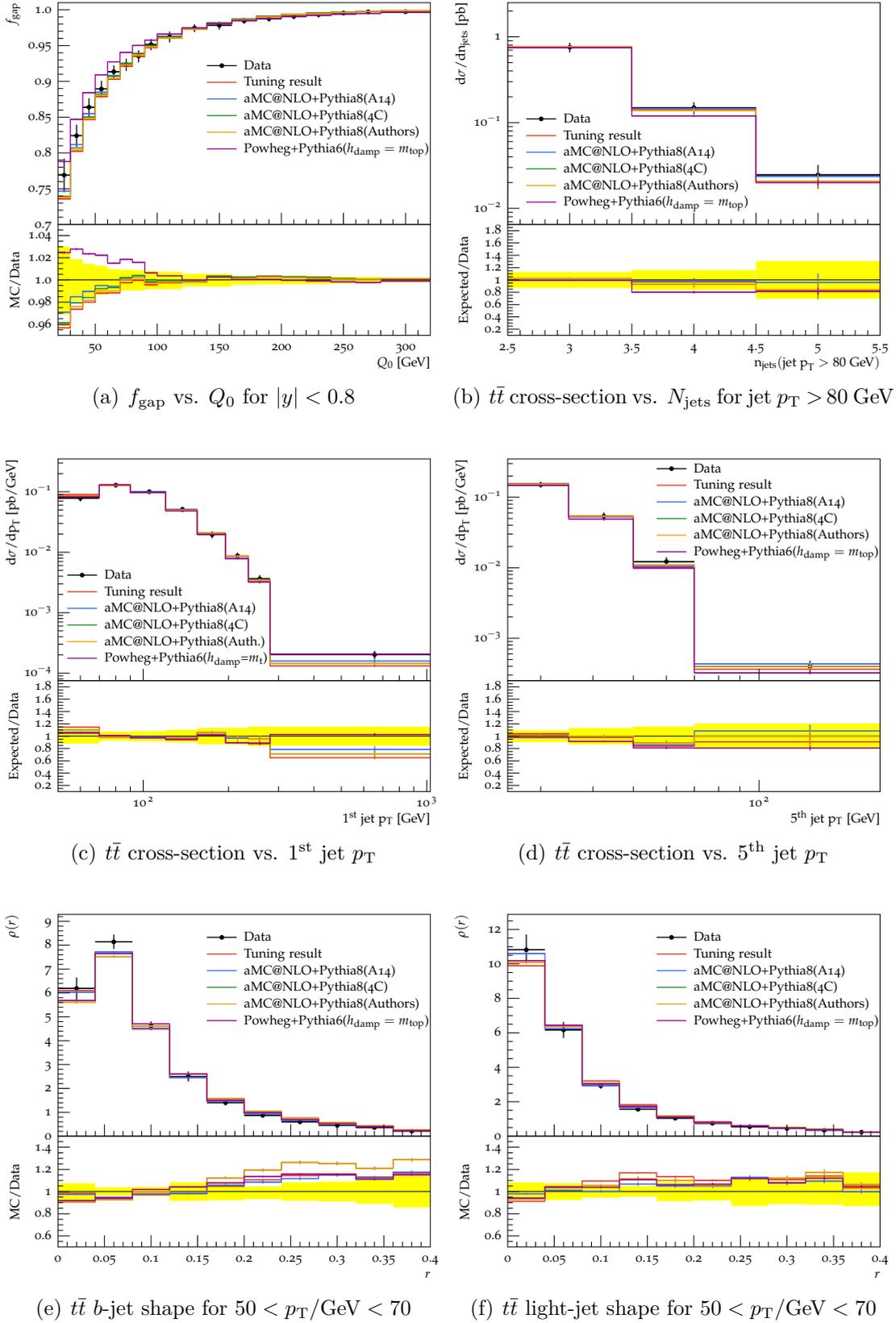


Figure 4.4.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (New tune, A14, 4C, Authors) and POWHEG+PYTHIA6 (P2011c, $h_{\text{damp}} = m_{\text{top}}$) compared with ATLAS $t\bar{t}$ gap fraction, jet multiplicity, jet p_{T} and jet shape distributions. The yellow shaded areas represent data uncertainty.

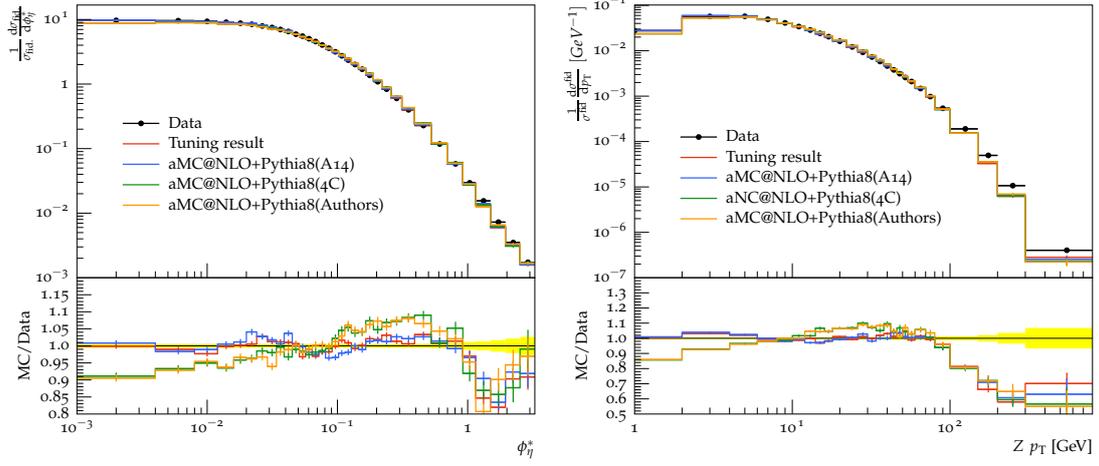
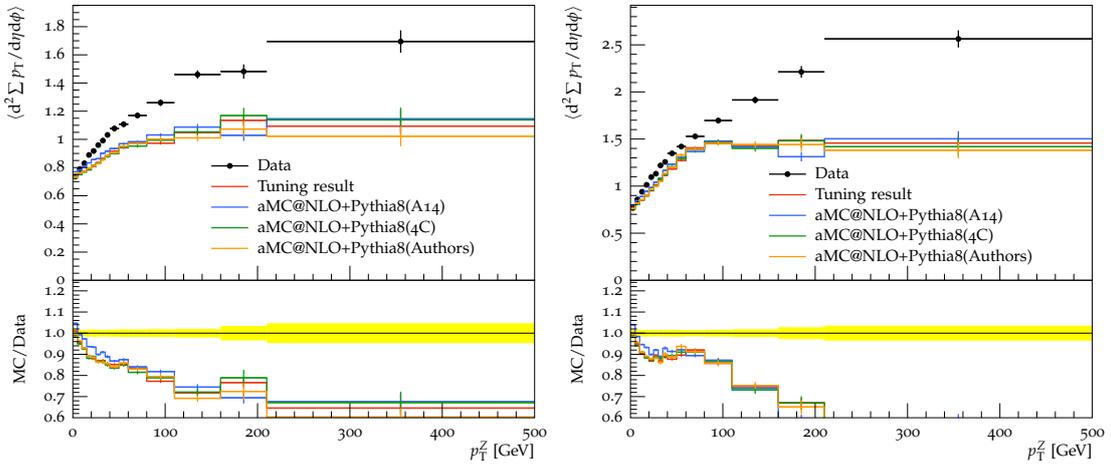
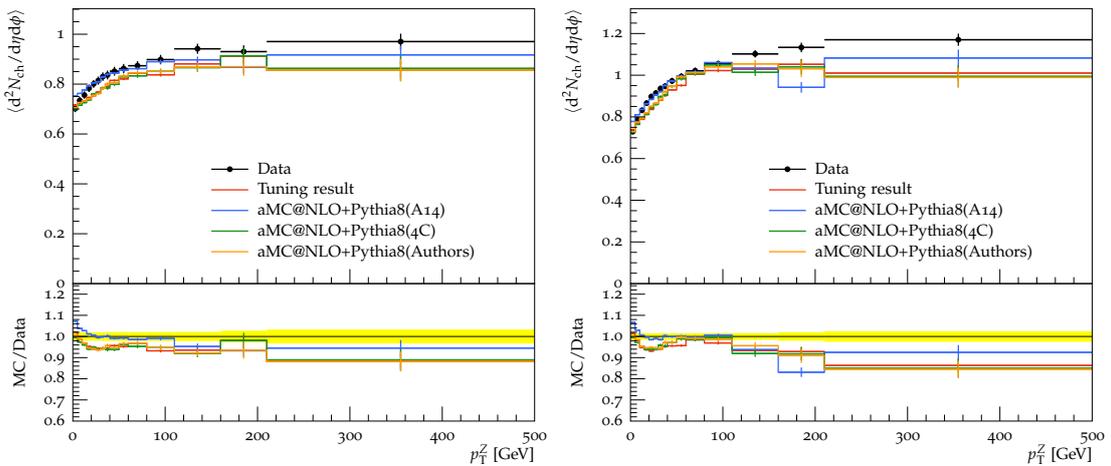
(a) $Z \phi_{\eta}^*$ spectrum(b) $Z p_T$ spectrum(c) $\sum p_T$ vs p_T^Z in toward region(d) $\sum p_T$ vs p_T^Z in transverse region(e) N_{ch} vs p_T^Z in toward region(f) N_{ch} vs p_T^Z in transverse region

Figure 4.5.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (New tune, A14, 4C, Authors) compared with ATLAS $Z \phi_{\eta}^*$, $Z p_T$ and Z UE distributions. The yellow shaded areas represent data uncertainty.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

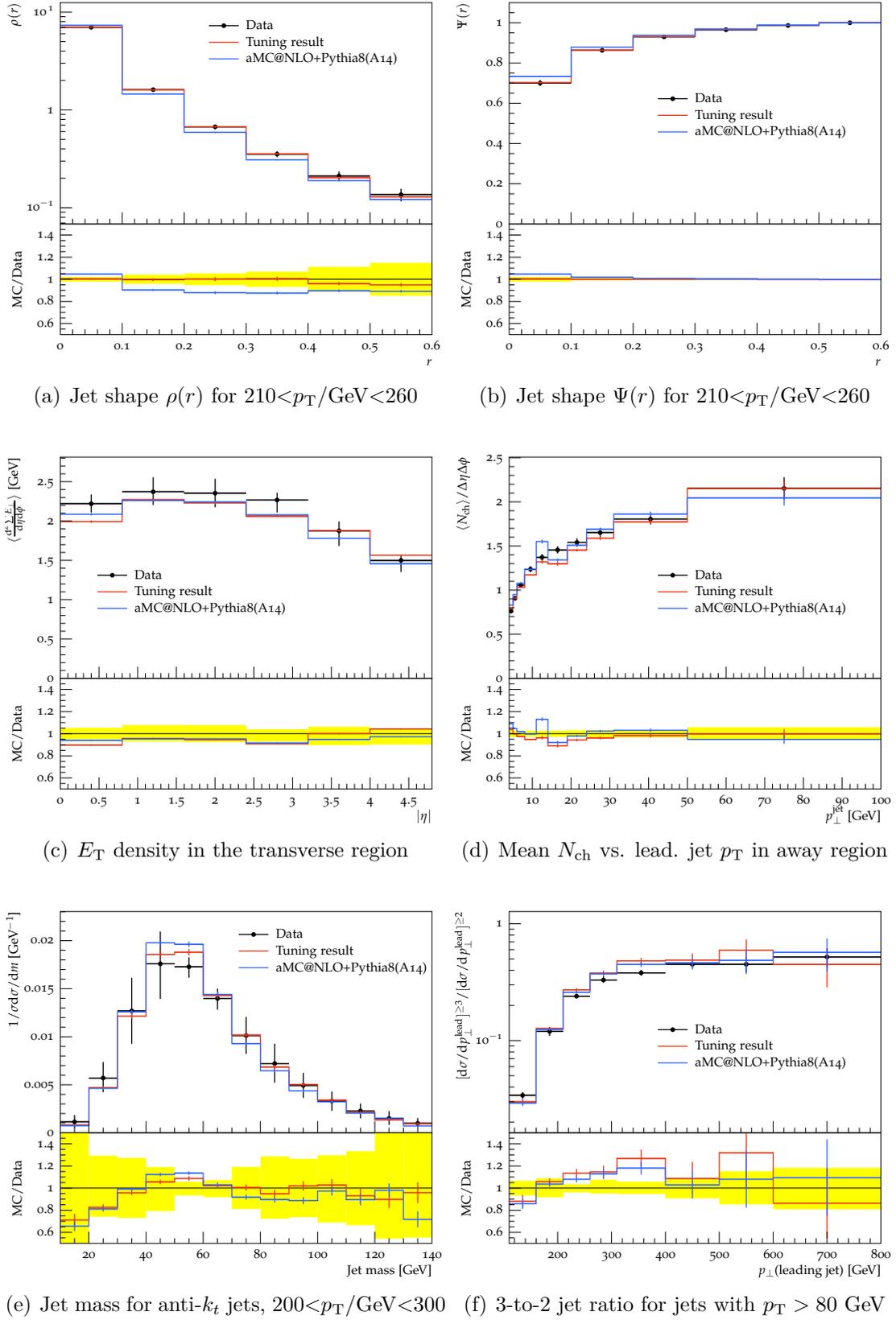


Figure 4.6.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (New tune and A14) compared with ATLAS dijet jet shape, transverse energy flow, track-jet UE, jet substructure and multi-jet 3-to-2 jet ratio distributions. The yellow shaded areas represent data uncertainty.

Parameter	4C	Monash	A14 NNPDF	ATTBAR	AZ	Tuning result
α_S^{FSR}	0.1383	0.1365	0.127	0.137	0.1383	0.1385 ± 0.0007
$p_{\text{T,min}}^{\text{FSR}}$	0.4	0.5	0.5	1.26	0.4	1.18 ± 0.05
α_S^{ISR}	0.137	0.1365	0.127	0.121	0.1237	0.1267 ± 0.0002
$p_{\text{T0,Ref}}^{\text{ISR}}$	2.0	2.0	1.56	2.0	0.59	0.87 ± 0.03
α_S^{MPI}	0.135	0.130	0.126	0.130	0.135	0.124 ± 0.002
$p_{\text{T0,Ref}}^{\text{MPI}}$	2.085	2.28	2.09	2.16	2.18	2.06 ± 0.06
Primordial $k_{\text{T,hard}}$	2.0	1.8	1.88	1.8	1.71	1.74 ± 0.02
Reconnect range	1.5	1.8	1.71	1.8	1.5	2.1 ± 0.1

Table 4.5.: Comparison of the tuned parameters to a selection of previous ATLAS tunes. The goodness of the fit for the tuning result is $\chi^2/\text{Ndf} = 1800/2301 = 0.89$.

	New tune better	A14 better	both equal
$t\bar{t}$		gap fractions jet shapes jet p_{T} N_{jets}	
Z	$Z \phi^*$ $Z p_{\text{T}}$ $Z \text{ UE}$		
dijet	angular decorrelations track-jet UE gap fractions substructure leading jet UE		jet shapes track-jet properties multi-jets E_{T} flow

Table 4.6.: List of observables comparing the new tune with the A14 tune.

4.6. Discussion

After the tuning results have been presented in the last section, it is meaningful to study and discuss them in more detail.

First of all, it has to be noted that the choice of which observables and bin ranges to include in the tuning is considered carefully. Observables sensitive to the tuning parameters are included, while correlated ones are avoided as much as possible. Therefore, out of 1010 available $t\bar{t}$, Z and dijet distributions, only 141 are taken. The aim when choosing the tuning observables is to maximise the sensitivity to each parameter in order to constrain them. The constraint is reflected in the uncertainty of each parameter shown in tab. 4.5.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

However, it is important to note that PROFESSOR computes these uncertainties simply from the χ^2 minimisation. This assumes that the tuned MC models are perfect i.e. able to describe all data after tuning. Thus, these uncertainties are an underestimate of the real error [24]. The common routine in ATLAS to introduce meaningful uncertainties to this tuning procedure is to perform *eigentune* variations for the tuning result. Performing these eigentune studies is beyond the scope of this thesis due to time constraints.

As mentioned in section 4.3, some dijet observables, especially those from the dijet UE analyses, show an unintended p_T threshold problem that arises from the event generation with MADGRAPH5_aMC@NLO, which can be seen in some plots in appendix B. This is suboptimal for the tuning, but accounted for by excluding the problematic p_T regions from the tuning ranges as listed in tab. 4.4.

The weights for all dijet observables of the tune are chosen to be equal, namely 1, while for all $t\bar{t}$ and Z it is 5. The weights for $t\bar{t}$ and Z observables is higher than for dijet observables to compensate the higher number of dijet observables included in the tuning, which would leave the respective contribution from $t\bar{t}$ and Z observables insignificant otherwise. Within the processes, the weights are equal for direct comparison of the impact of each observable on the tuning parameters and also to not introduce any bias or preference to tuned observables.

The description of data by the obtained tune in fig. 4.4-4.6 is discussed in the following.

$t\bar{t}$ tuning plots

For the $t\bar{t}$ gap fraction the MADGRAPH5_aMC@NLO+PYTHIA8 generator shows the same behaviour as in chapter 3. This generator predicts a smaller gap fraction for all shown tunes than what is measured in data for low p_T threshold values Q_0 , while POWHEG+PYTHIA6 consistently predicts a higher gap fraction than data. MADGRAPH5_aMC@NLO+PYTHIA8 with the A14 tune shows the best agreement with data for this observable, while the tuning result agrees with the 4C tune and “Authors” settings.

The jet multiplicity for jets with $p_T > 80$ GeV shows a similar behaviour for all generators. The two differences are that POWHEG+PYTHIA6 predicts less events that contain 4 jets than the other generators and data, while the A14 tune shows the best agreement with data compared to all other generators, especially in the last bin.

For the leading jet p_T distribution, all generators describe data and show an overall similar behaviour in the low p_T region. The most significant difference is visible in the last bin of the leading jet p_T spectrum. POWHEG+PYTHIA6 agrees very well with data here,

while the other generators show significant deviations from data. On the other hand, the A14 and 4C tunes and the “Authors” settings give the best description of the 5th jet p_T distribution. The tuning result predicts much softer jets compared to the other MADGRAPH5_aMC@NLO+PYTHIA8 tunes and data for both observables.

Lastly, the jet shapes are not well described by any of the shown generators. All of them predict broader b -jets in $t\bar{t}$ events than what is measured in data. A similar behaviour can be seen for the light-jet shapes, except the MC predictions are mostly within the data uncertainty. Overall, the A14 tune shows the best agreement with data.

Z tuning plots

The MADGRAPH5_aMC@NLO+PYTHIA8 generator shows the same behaviour for the Z ϕ_η^* and Z p_T observables, because both are strongly correlated. The ϕ_η^* is tuned up to 0.5 and the Z p_T spectrum is tuned up to 50 GeV. The tuning result and the A14 tune agree well with data, while the 4C tune and the “Authors” settings show significant deviations. Overall the MC prediction of the tuning result gives the best description of data here.

The four observables from the Z UE analysis, namely $\sum p_T$ and N_{ch} as functions of p_T^Z in the toward and transverse region, are tuned only up to $p_T^Z < 5$ GeV which is the first bin. This is because modelling limitations cause significant deviations from data for greater Z p_T values. On the other hand, high p_T regions are described by the matrix element and should therefore not be included in the tuning of soft effects.

The 4C tune and “Authors” settings agree best with data, except for the $\sum p_T$ in the transverse region, where the tuning result performs best. The A14 tune shows significant deviations from data here.

Dijet tuning plots

The MC prediction for the differential jet shape $\rho(r)$ for jets with $210 < p_T/\text{GeV} < 260$ and $|y| < 2.8$ shows a very good agreement between the tuning result and data, while the A14 tune predicts narrower jets. The differential jet shape is strongly correlated to the integrated jet shape, which shows a good agreement for both tunes, except for the first bin where the A14 tune is above data.

The two centre plots of fig. 4.6 show observables sensitive to the UE. Both tunes predict a smaller E_T density in the transverse region than observed in data, but the A14 tune performs slightly better than the tuning result. On the other hand, the tuning result shows a better agreement with data for the mean N_{ch} than the A14 tune. The visible excess predicted by the A14 tune at approximately 12 GeV is caused by the p_T threshold problem discussed in section 4.3.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

The MC prediction of the tuning result for the jet mass for anti- k_t jets with $200 < p_T/\text{GeV} < 300$ agrees well with data within the uncertainties and performs better than the A14 tune. Both tunes show significant deviations from data for the 3-to-2 jet ratio with jet $p_T > 80$ GeV, while the A14 tune agrees slightly better with data here.

Parameter results

The best fit tune parameters from PROFESSOR are compared to other currently used tunes in tab. 4.5. The most important parameters are the three α_S values for FSR, ISR and MPI, respectively. The tuning result suggests an α_S^{FSR} which is significantly higher than for the A14 tune, but comparable to the other tunes, especially 4C and AZ, which choose this value based on tuning to LEP data. The constrain on this parameter comes mostly from dijet observables such as jet shapes. The α_S^{ISR} value is close to the A14 result which is very different from the other four tunes listed. The constrain on this parameter is driven mostly by the $Z \phi_\eta^*$ and $Z p_T$ observables. The α_S^{MPI} value is also comparable with the A14 tune, but noticeably smaller than for the other tunes. It is constrained mostly by Z and dijet observables sensitive to the UE, but has a noticeably larger uncertainty than the corresponding FSR and ISR parameters, which may hint at the sensitivity to this parameter being too weak.

Now, from the corresponding p_T cut-off parameters, only the MPI parameter is comparable to the other tunes. As will be shown in section 4.7.1 and 4.7.2, the constraint on this cut-off is driven by the same observables as α_S^{MPI} to which it is strongly correlated. The result for $p_{T,\text{min}}^{\text{FSR}}$ is closest to the ATTBAR tune value, while the $p_{T0,\text{Ref}}^{\text{ISR}}$ parameter is closest to the AZ tune, but otherwise significantly different compared to the other tunes. This may hint at a weak constraint on both parameters in this tuning study.

The primordial $k_{T,\text{hard}}$ parameter agrees well with the AZ tune and is close to the other tunes. The reconnect range for the new tune is slightly different from the other tunes, but still comparable with them within the uncertainties. In the next sections it will be shown that most observables are not or only slightly sensitive to these parameters and that they are correlated to both MPI parameters and to each other. Therefore, the constraint on them may be given mostly through α_S^{MPI} .

Whether the parameters are comparable to other ATLAS tunes is only of secondary interest. Since these parameters are modelling non-perturbative effects, they are not predicted by theory. Therefore any physical parameter value cannot be ruled out by first principles. Lastly, the goodness of the fit $\chi^2/\text{Ndf} = 1800/2301 = 0.89$ indicates that the overall agreement of the tune with all chosen tuning observables is reasonable.

A very important aspect of these results is that some tuning observables show a strong tension between each other. This means that it is not possible to tune those observables at the same time such that the MC prediction matches the data for all of them. This is the case for example for the $t\bar{t}$ gap fraction, $Z \phi_\eta^*$ and $Z p_T$. Both Z observables require an α_S^{ISR} value which agrees with the tuning result and are also responsible for the strong constraint on this parameter, while the $t\bar{t}$ gap fractions require a much lower value for those parameters. This is the reason that the $t\bar{t}$ gap fraction plots in fig. 4.4 show significant deviations between the “Tuning result” and data in the low p_T threshold bins, while there is a good agreement for the mentioned Z observables in their respective tuning ranges. Another such example is the tension between the jet shapes from $t\bar{t}$ and dijet events. Jet shapes are sensitive to α_S^{FSR} and α_S^{MPI} , but the $t\bar{t}$ jet shapes require both values to be considerably lower than the tuning result, which in turn would cause significant deviations of the dijet jet shapes from data.

It is beyond the scope of this thesis to study whether these tensions are a failure of this tuning procedure such as wrong assumptions and approximations used in the PROFESSOR tool, lack of statistics or too few parameter variations - or whether this is a problem of the respective MC generators themselves, of the models they use or the parameters chosen for the tuning. On the other hand, the necessity to limit the tuning ranges for some observables already shows a problem of the MC models that are tuned in this study, which is shown for example in fig. 4.5. Known shortcomings of the MC models limit the tuning ranges of the $Z p_T$ measurement to 50 GeV, $Z \phi_\eta^*$ to 0.5 and the Z UE observables to $Z p_T < 5$ GeV, as listed in tab. 4.4. Because of this, it cannot be ruled out that the tensions may arise from modelling problems of the MC generators.

Since the aforementioned tensions are mainly found between $t\bar{t}$ and Z or $t\bar{t}$ and dijet, but not between Z and dijet observables, the tuning result shows an equal or better agreement with data than A14 for Z and dijet observables, as shown in tab. 4.6. There are no expected conflicts between Z and dijet observables, because the Z observables are mostly sensitive to ISR, while the dijet observables are more sensitive to FSR and MPI. However, the tuning result’s agreement with data for $t\bar{t}$ observables compared to the A14 tune is noticeably worse, precisely because of these tensions. Thus, it is reasonable to pursue a dedicated tune to model $t\bar{t}$ observables as best as possible instead of using a universal tune, while the latter is still meaningful to use as a general tune and will therefore be denoted as “A15_AG”. A dedicated $t\bar{t}$ tune and a dedicated tune for Z and dijet processes, both obtained with the PROFESSOR tool as well, are presented in section 4.7.4.

4.7. Further studies

This section discusses the sensitivity of the observables to the tuning parameters and the correlations between the different tuning parameters. Additionally, a study regarding the local recoil option for parton showering with MADGRAPH5_aMC@NLO+PYTHIA8 is shown which compares the A15_AG tune for global recoil with local recoil. Lastly, two dedicated tunes are presented for $t\bar{t}$ modelling as well as Z and dijet modelling, respectively.

4.7.1. Sensitivity studies

Fig. 4.7 shows the sensitivity for a selection of observables to each tuned parameter. The sensitivity plots for all other tuned observables are shown in appendix C. PROFESSOR calculates the sensitivity S_i for each bin b using the following equation:

$$S_i(p) = \frac{df_i}{\epsilon \cdot f_\sigma + f(p)} \cdot \frac{\epsilon \cdot p_\sigma + p}{dp} \quad (4.12)$$

where f is the interpolation function, p the parameter under study, f_σ and p_σ the respective widths, and ϵ is a constant set to 0.001 to avoid problems in case $f = 0$ or $p = 0$ [24].

The top left plot of fig. 4.7 shows the sensitivity of the $t\bar{t}$ gap fraction as a function of Q_0 in $|y| < 0.8$. A strong negative sensitivity to α_S^{ISR} is given in the low p_T bins which showed the most significant deviation to data in the tuning plots in fig. 4.4. This means by decreasing α_S^{ISR} , the $t\bar{t}$ gap fraction increases and thus the deviation from data decreases in the low p_T bins. Additionally, a small sensitivity to the MPI parameters can be observed. In the top right plot, the sensitivity of the $t\bar{t}$ light-jet shape is shown as a function of r for jets with a p_T between 50 GeV and 70 GeV. A strong negative sensitivity to α_S^{FSR} is shown in the first bin and a strong positive sensitivity in the other bins. As explained in the previous section, the $t\bar{t}$ light-jet shape requires α_S^{FSR} to be smaller than the value obtained for the A15_AG tune, because it would increase the jet shape in the small r bin and decrease it in the large r bins which brings it closer to data, as can be seen in fig. 4.4. Additionally, a noticeable sensitivity to both MPI parameters is also observed and decreasing α_S^{MPI} would have a similar, but smaller effect than decreasing α_S^{FSR} .

The left centre plot in fig. 4.7 shows the sensitivity of the Z p_T distribution. Note that only the region $p_T^Z < 50$ GeV is included in the tune. A very strong sensitivity to α_S^{ISR} is observed being negative in the low p_T region and positive in the high p_T region.

The right centre plot shows the sensitivity of the $\sum p_T$ profile from the Z UE analysis.

As expected, a very strong sensitivity to both MPI parameters is observed in all bins. Although the A15_AG and A14 tunes have a comparable $p_{T0,Ref}^{MPI}$, the A14 tune gives a higher α_S^{MPI} which explains why the A14 tune predicts a higher $\sum p_T$ for smaller values of p_T^Z .

On the bottom left, the sensitivity of jet shapes $\rho(r)$ in dijet events is shown for jets with $260 < p_T/\text{GeV} < 310$ and $|y| < 2.8$. As was the case for the $t\bar{t}$ light-jet shapes, these observables are sensitive to α_S^{FSR} . Here, the sensitivity to α_S^{ISR} is much stronger and the sensitivity to α_S^{MPI} fluctuates. The A15_AG tune has a comparable α_S^{ISR} to the A14 tune, but α_S^{FSR} is much smaller for the A14 tune. This explains why the A14 tune predicts a narrower jet shape than A15_AG, as seen in fig. 4.6.

Lastly, the plot on the bottom right shows the sensitivity of the jet mass for anti- k_t jets with $200 < p_T/\text{GeV} < 300$. This observable shows strong sensitivity to both MPI parameters as well as α_S^{FSR} . To decrease the deviation of both tunes observed in fig. 4.6 in the first bins, they require either a higher $p_{T0,Ref}^{MPI}$ or a smaller α_S^{MPI} or α_S^{FSR} .

Overall, the sensitivity to all tuned parameters is as expected. It is important to note that many parameter requirements mentioned above interfere with each other. Thus, changing one parameter can improve the MC prediction for one particular observable, but worsen it for another. So, as pointed out before, there are some tensions between observables.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

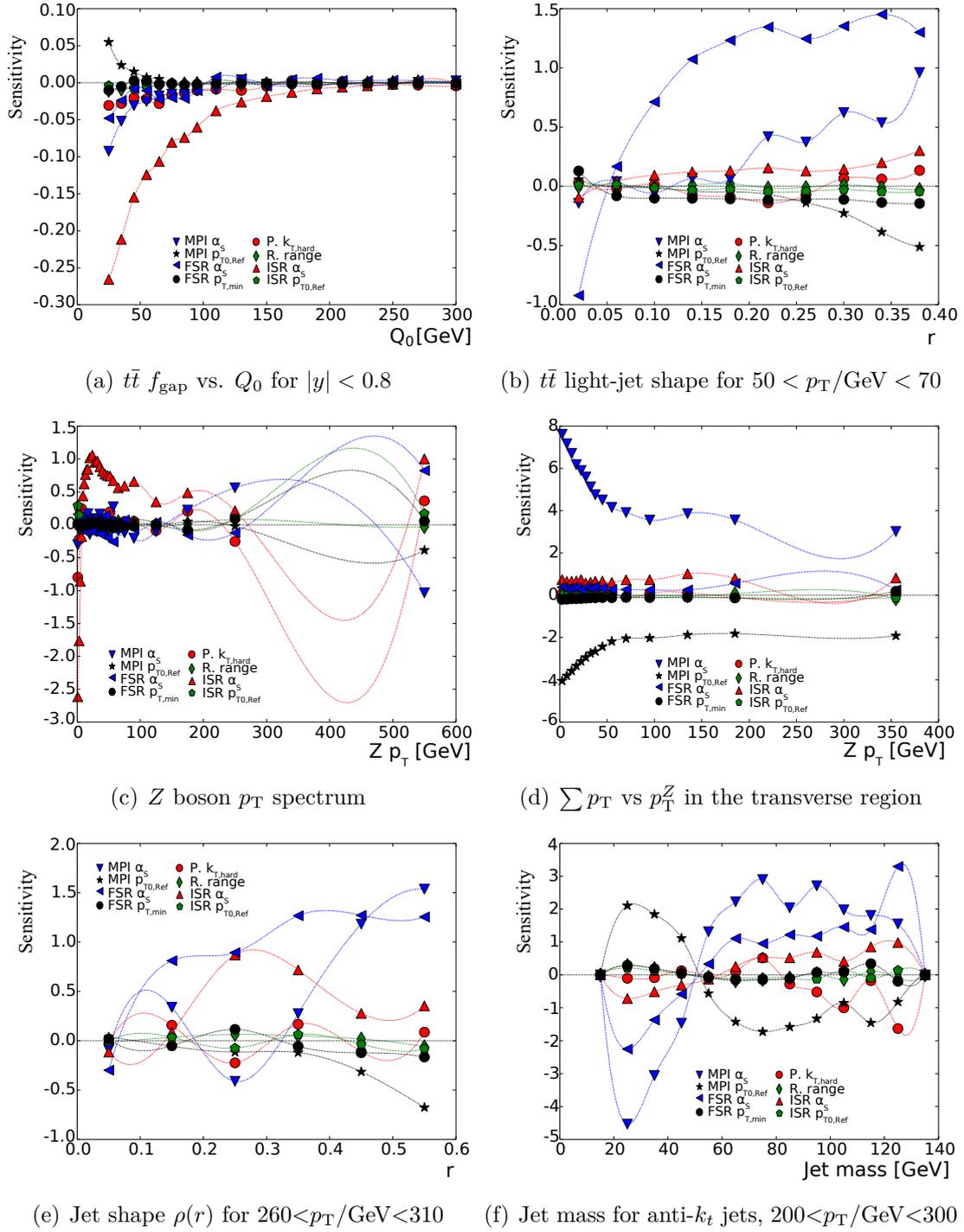


Figure 4.7.: Sensitivity plots for $t\bar{t}$ gap fraction, $t\bar{t}$ light-jet shape, Z p_T , Z UE, dijet jet shape and jet substructure distributions.

4.7.2. Parameter correlations

Fig. 4.8 displays the correlations between the different tuning parameters as a colour map and their corresponding values are listed in tab. 4.7. These correlation coefficients ρ_{ij} for each pair of parameters (i, j) are defined by the n -dimensional covariance matrix C , which PROFESSOR computes when minimising the χ^2 function:

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \quad [24]. \quad (4.13)$$

A strong positive correlation between each α_S and the respective p_T cut-off is observed for ISR, FSR and MPI, which means changing one of the α_S parameters also has a significant effect on the respective p_T cut-off. With this, the strong sensitivity to the α_S parameters, discussed in the last section, translates into a weaker sensitivity to the p_T cut-off parameters. Additionally, the primordial $k_{T,\text{hard}}$ and reconnect range show a strong positive and negative correlation to both MPI parameters respectively, and a significant negative correlation to each other. Thus, changing the MPI parameters also affects both beam remnants parameters.

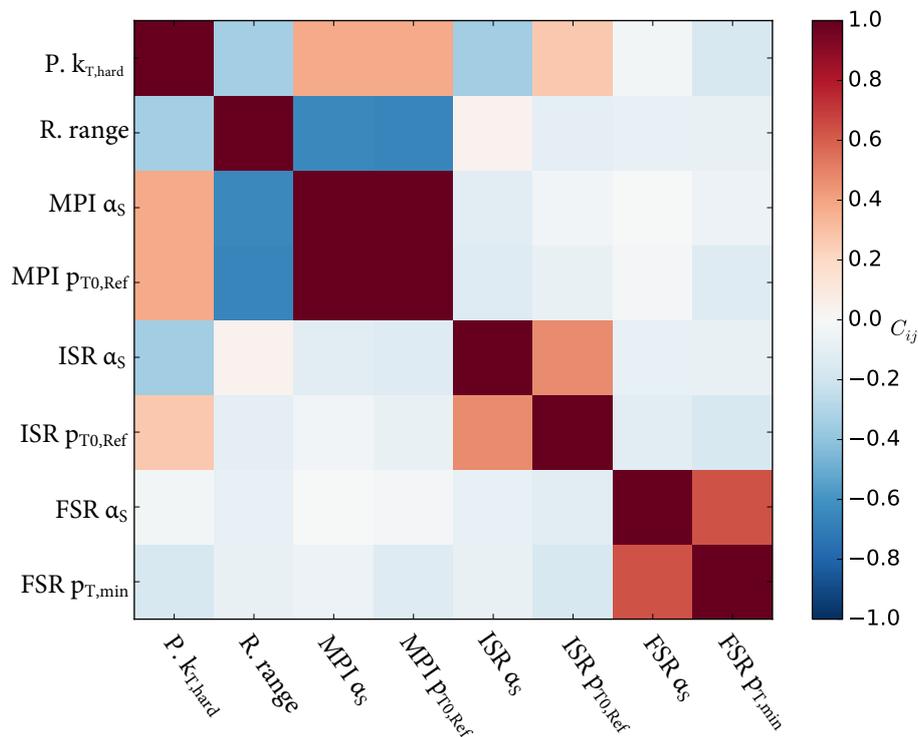


Figure 4.8.: Correlations between the tuning parameters as a result of the interpolation.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

Parameters	α_S^{FSR}	$p_{\text{T,min}}^{\text{FSR}}$	α_S^{ISR}	$p_{\text{T0,Ref}}^{\text{ISR}}$	α_S^{MPI}	$p_{\text{T0,Ref}}^{\text{MPI}}$	P. $k_{\text{T,hard}}$	R. range
α_S^{FSR}	1	0.64	-0.08	-0.12	-0.01	-0.02	-0.03	-0.08
$p_{\text{T,min}}^{\text{FSR}}$	0.64	1	-0.07	-0.17	-0.05	-0.13	-0.16	-0.07
α_S^{ISR}	-0.08	-0.07	1	0.47	-0.12	-0.128	-0.35	0.04
$p_{\text{T0,Ref}}^{\text{ISR}}$	-0.12	-0.17	0.47	1	-0.04	-0.07	0.27	-0.10
α_S^{MPI}	-0.01	-0.05	-0.12	-0.04	1	0.99	0.38	-0.65
$p_{\text{T0,Ref}}^{\text{MPI}}$	-0.02	-0.13	-0.13	-0.07	0.99	1	0.38	-0.66
P. $k_{\text{T,hard}}$	-0.03	-0.17	-0.35	0.27	0.38	0.38	1	-0.34
R. range	-0.08	-0.07	0.04	-0.10	-0.65	-0.66	-0.34	1

Table 4.7.: List of the correlation values for all tuning parameters.

4.7.3. Comparing global recoil to local recoil

The authors of MADGRAPH5_aMC@NLO [1] strongly recommend to use the “global recoil” scheme for the interface with PYTHIA8 to generate emissions in which the recoil of an emission is shared among all final state particles [2, 3]. Thus, the radiation pattern is unrelated to colour correlations. This is especially convenient for some matching algorithms, such as MC@NLO, where a full analytic knowledge of the shower radiation pattern is needed to avoid double-counting. However, this approach will miss out on the colour coherence phenomena. Specifically, the phase space for subsequent emissions is larger than for the normal dipole shower algorithm (“local recoil”). The phase space difference grows as more and more gluons are created, and thus leads to a too steep multiplication of soft gluons. Therefore, the application of the global recoil option is only for the first one or few emissions of the shower, where a potential overestimate of the emission rate is to be corrected for anyway, by matching to the relevant matrix elements. Thereafter, subsequent emissions should be handled as before with the local recoil scheme, i.e. with dipoles spanned between nearby partons. That is why the two settings “TimeShower:globalRecoil on” and “TimeShower:nMaxGlobalRecoil 1” listed in tab. 3.3 are used.

With that being said, recent studies in ATLAS have shown that not using the global recoil option for the MADGRAPH5_aMC@NLO+PYTHIA8 generator can have benefits for some observables [52].

Thus, for completeness, the tuning procedure described in this chapter has been repeated using the local recoil option for the parton showering with PYTHIA8. The same tuning weights as for the global recoil tune have been used and the results are presented in this section. Fig. 4.9 and 4.10 compare a selection of $t\bar{t}$ and dijet observables for the local

recoil and global recoil schemes, while the comparison for other observables can be found in appendix D. Tab. 4.8 and tab. 4.9 compare the parameters and the performance of the global recoil tune with the local recoil tune, respectively.

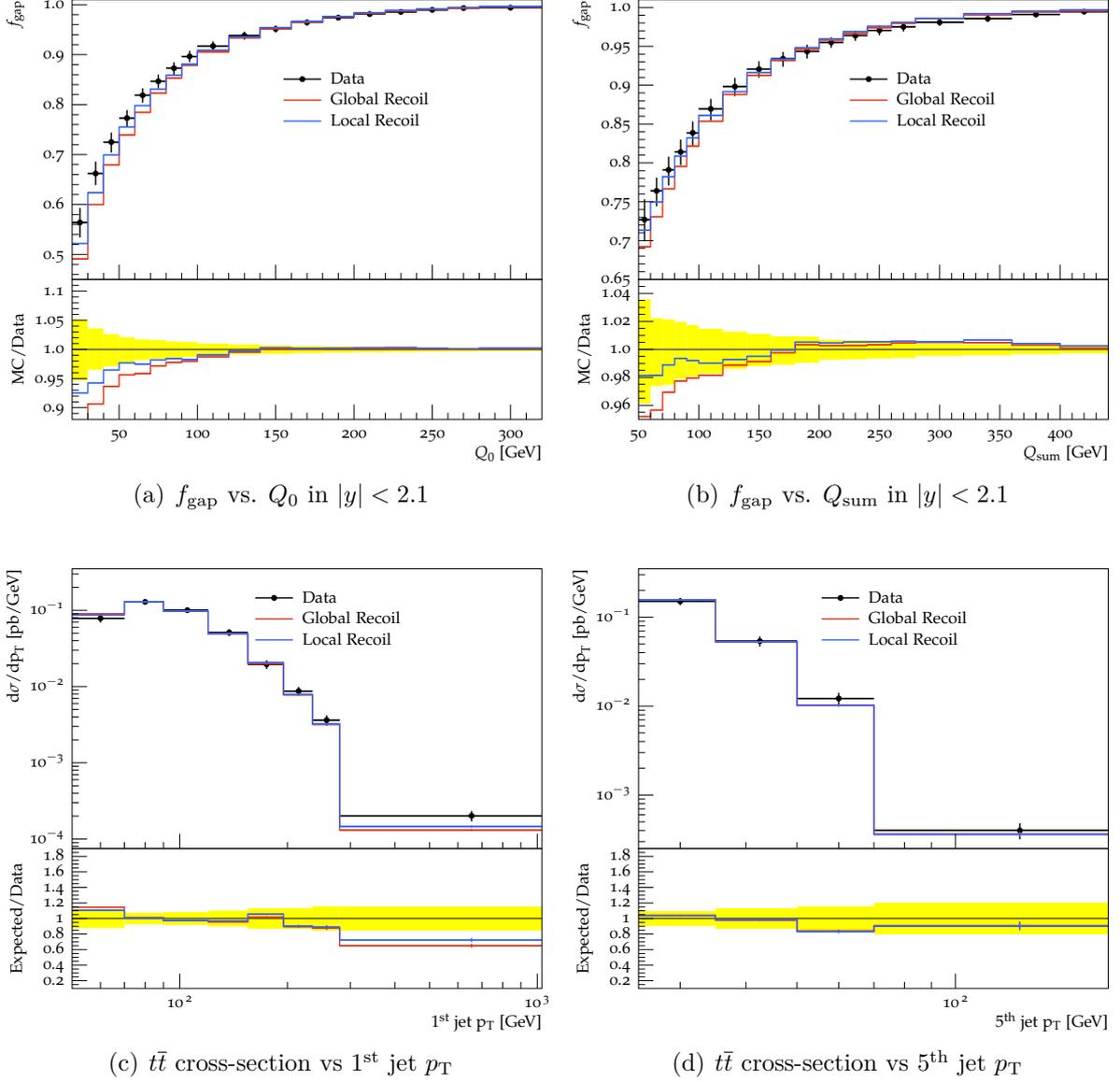


Figure 4.9.: Comparison plots of A15_AG using global recoil or local recoil compared with ATLAS $t\bar{t}$ gap fraction and jet p_T distributions. The yellow shaded areas represent data uncertainty.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

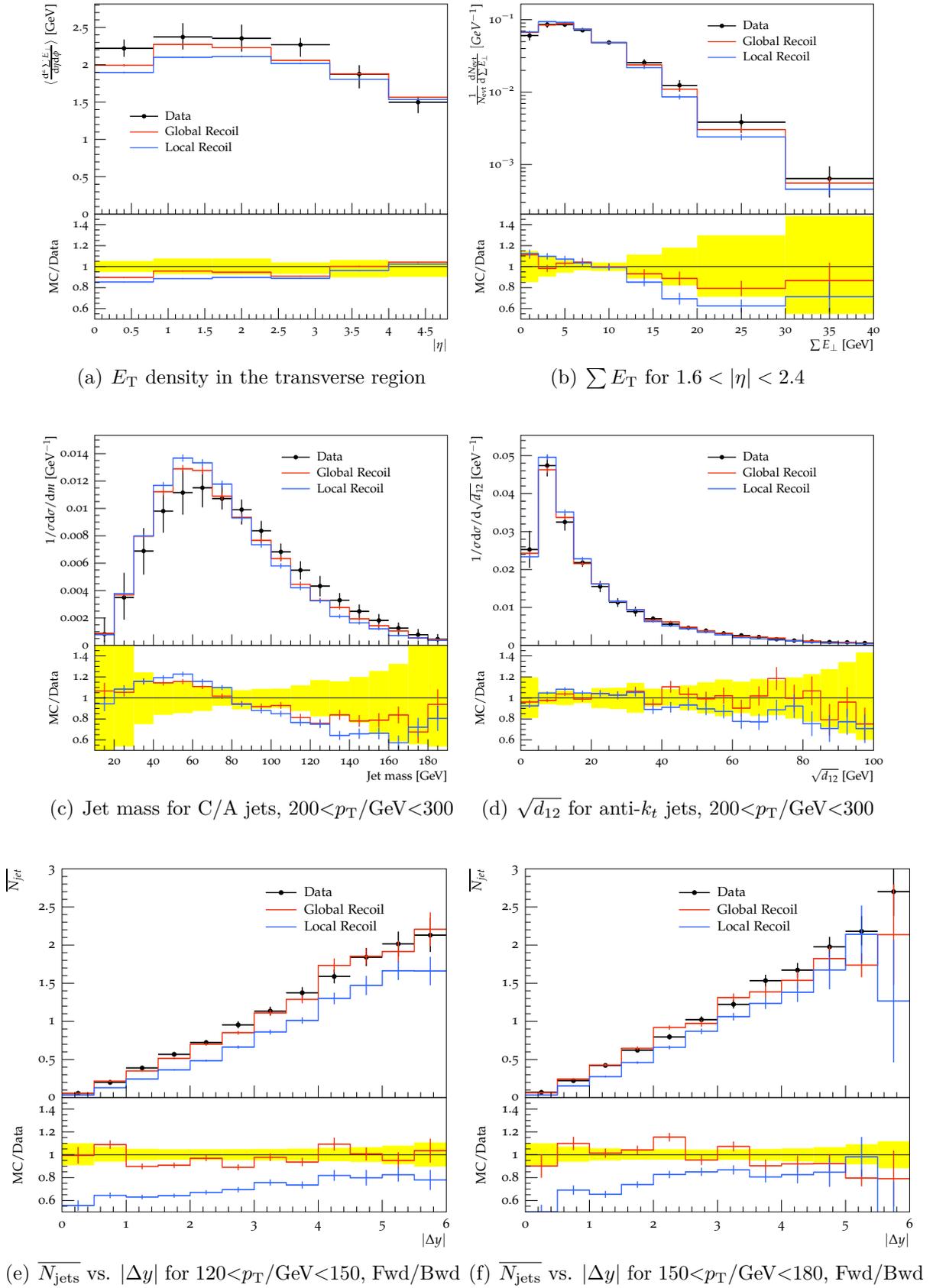


Figure 4.10.: Comparison plots of A15_AG using global recoil or local recoil compared with ATLAS dijet transverse energy flow, jet substructure and gap fraction distributions. The yellow shaded areas represent data uncertainty.

Parameter	4C	A14 NNPDF	ATTBAR	AZ	GR Tune	LR Tune
α_S^{FSR}	0.1383	0.127	0.137	0.1383	0.1385 ± 0.0007	0.1441 ± 0.0006
$p_{\text{T,min}}^{\text{FSR}}$	0.4	0.5	1.26	0.4	1.18 ± 0.05	1.52 ± 0.04
α_S^{ISR}	0.137	0.127	0.121	0.1237	0.1267 ± 0.0002	0.1269 ± 0.0002
$p_{\text{T0,Ref}}^{\text{ISR}}$	2.0	1.56	2.0	0.59	0.87 ± 0.03	0.82 ± 0.03
α_S^{MPI}	0.135	0.126	0.130	0.135	0.124 ± 0.002	0.128 ± 0.001
$p_{\text{T0,Ref}}^{\text{MPI}}$	2.085	2.09	2.16	2.0	2.06 ± 0.06	2.22 ± 0.03
P. $k_{\text{T,hard}}$	2.0	1.88	1.8	1.71	1.74 ± 0.02	1.74 ± 0.02
R. range	1.5	1.71	1.8	1.5	2.1 ± 0.1	1.7 ± 0.1
χ^2/Ndf					0.89	1.17

Table 4.8.: Comparison of the global recoil tune parameters to the local recoil tune as well as the 4C, A14, ATTBAR and AZ tunes.

	Global recoil better	Local recoil better	both equal
$t\bar{t}$		gap fractions jet p_{T}	jet shapes N_{jets}
Z			$Z \phi^*$ $Z p_{\text{T}}$ $Z \text{ UE}$
dijet	angular decorrelations gap fractions substructure E_{T} flow		jet shapes multi-jets track-jet properties track-jet UE leading jet UE

Table 4.9.: List of observables comparing the global recoil option with the global recoil option.

As shown in fig. 4.9, MADGRAPH5_aMC@NLO+PYTHIA8 with the new A15_AG tune predicts a higher gap fraction in the low p_{T} threshold bins when using local recoil instead of the global recoil option. This significantly improves the agreement with data. Additionally, the local recoil scheme predicts slightly harder jets in $t\bar{t}$ events and thus describes data better than the global recoil option.

The top left plot in fig. 4.10 shows the E_{T} density in the region transverse to the direction of the leading jet in the event for $|\eta| < 4.8$, while the top right plot shows the transverse energy sum in $1.6 < |\eta| < 2.4$. In both cases, the MC prediction of the global recoil tune agrees better with data than the local recoil tune. The same can be said for the two centre plots from the jet substructure analysis. The left plot shows the jet mass for jets reconstructed with the Cambridge-Aachen algorithm and the right one shows $\sqrt{d_{12}}$

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

for anti- k_t jets, respectively for jets with $200 < p_T/\text{GeV} < 300$. While both recoil options show a similar trend for the jet mass, the global recoil prediction gives a noticeably better agreement with data for the $\sqrt{d_{12}}$ observable.

The two bottom plots show the mean number of additional jets as a function of $|\Delta y|$ for two different average p_T intervals using the most forward and the most backward jet in the event. The global recoil scheme predicts a jet multiplicity that is compatible with data within the uncertainties, while the local recoil option shows an overall huge deviation from data.

As explained in the beginning of this section, the global recoil scheme affects the final state shower algorithm. Thus, the largest difference between the two recoil options are expected for observables that are sensitive to FSR parameters. This is reflected in the tune parameter results that show significantly higher FSR parameters for the local recoil tune than for the global recoil tune, while all other parameters are comparable.

Indeed, the $Z \phi_\eta^*$ and $Z p_T$ observables are most sensitive to the ISR parameters and therefore invariant under the change of the recoil scheme, as can be seen in appendix D. However, $t\bar{t}$ jet shapes, which are most sensitive to the FSR and MPI parameters, are also invariant under this change, while the $t\bar{t}$ gap fractions show significant differences, although they are mostly sensitive to α_S^{ISR} . It is beyond the scope of this thesis to study the different impact of the local recoil option on the discussed observables.

In conclusion, the local recoil tune is able to describe $t\bar{t}$ observables as well as or even better than the global recoil tune. This is especially visible for the $t\bar{t}$ gap fractions in fig. 4.9. On the other hand, the global recoil prediction describes data better for some dijet observables, as shown in fig. 4.10. Additionally, PROFESSOR computes a worse χ^2 value for the local recoil tune than for the global recoil tune. Thus, it cannot be concluded that using the local recoil option for MADGRAPH5_aMC@NLO+PYTHIA8 is generally preferred over the recommended global recoil option.

4.7.4. Dedicated tunes

MADGRAPH5_aMC@NLO+PYTHIA8 with the general tune A15_AG is not able to describe the measured $t\bar{t}$ observables as well as the A14 tune. Especially the low p_T bins of the $t\bar{t}$ gap fractions and the larger r bins of the $t\bar{t}$ jet shapes show significant deviations from ATLAS data. While the A14 tune gives a better prediction, it also shows deviations from data in the same areas as A15_AG. Therefore, a dedicated tune to optimise $t\bar{t}$ modelling has been developed. Additionally, since Z and dijet observables do not show

significant tensions between each other, a dedicated tune to optimise Z boson and dijet modelling is pursued as well. Both tunes, referred to as “A15_AT” and “A15_AJZ” respectively, are obtained with the PROFESSOR tool and presented in this section.

Tab. 4.10 and 4.11 show the observables and bin ranges included in the A15_AT and A15_AJZ tune, respectively. For both tunes, the tuned observables have the same weight of 1. Fig. 4.11 - 4.13 show a selection of $t\bar{t}$, Z and dijet observables, where the three A15 tunes are compared with the A14 tune. Additionally, for the $t\bar{t}$ observables a comparison to the ATTBAR tune and for Z observables a comparison to the AZ tune is included. Similar to A15_AT, the ATTBAR tune for MADGRAPH5_aMC@NLO+PYTHIA8 was tuned to gap fraction, jet shape, jet multiplicity and jet p_T observables in $t\bar{t}$ events, but it uses the local recoil scheme, while A15_AT is a global recoil tune [52]. The AZ tune for PYTHIA8 is obtained from tuning to the $Z p_T$ and ϕ_η^* distributions [35]. Finally, the tuning parameters of the three A15 tunes are compared to A14, ATTBAR and AZ in tab. 4.12.

Observable
$t\bar{t}$ gap fraction [32] Gap fraction vs Q_0 and Q_{sum} for $ y < 2.1$ (2 distributions)
$t\bar{t}$ jet shapes [30] Differential light-jet shape $\rho(r)$ (5 distributions)
$t\bar{t}$ N_{jets} & jet p_T [41] $t\bar{t}$ cross-section vs. jet multiplicity for jets above 25, 40, 60 and 80 GeV (4 distributions) $t\bar{t}$ cross-section vs. 1 st to 5 th jet p_T (5 distributions)

Table 4.10.: The observables used for the dedicated $t\bar{t}$ tune, A15_AT. Every observable has the same weight 1.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

Observable
$t\bar{t}$ gap fractions [32] Gap fraction vs Q_0 and Q_{sum} for $ y < 2.1$ (2 distributions)
$t\bar{t}$ jet shapes [30] Differential light-jet shapes $\rho(r)$ (5 distributions)
$t\bar{t}$ N_{jets} & jet p_{T} [41] $t\bar{t}$ cross-section vs. jet multiplicity for jets above 25 and 80 GeV (2 distributions) $t\bar{t}$ cross-section vs. 1 st and 5 th jet p_{T} (2 distributions)
Z/γ^* boson angular decorrelations [45] $Z \phi_{\eta}^*$ profile for $Z \rightarrow ee$ (dressed) and $\phi_{\eta}^* < 0.5$ (1 distribution)
Z/γ^* boson p_{T} [35] $Z p_{\text{T}}$ profile for $Z \rightarrow \mu\mu$ (dressed) and $Z p_{\text{T}} < 50$ GeV (1 distribution)
Z/γ^* boson underlying event [46] Σp_{T} vs. $Z p_{\text{T}}$ for $Z \rightarrow ee$ (dressed) and $Z p_{\text{T}} < 5$ GeV (2 distributions) N_{ch} vs. $Z p_{\text{T}}$ for $Z \rightarrow ee$ (dressed) and $Z p_{\text{T}} < 5$ GeV (2 distributions)
Dijet jet shapes [29] Jet shape ρ (49 distributions)
Dijet angular decorrelations [31] Dijet azimuthal decorrelations for $310 < p_{\text{T}}^{\text{max}}/\text{GeV} < 400$ (1 distribution)
Dijet gap fractions [47] Gap fraction vs $ \Delta y $ with Fwd/Bwd jets for $ \Delta y < 3.0$ (5 distributions) \bar{N}_{jet} vs $ \Delta y $ with leading jets and Fwd/Bwd jets for $ \Delta y < 3.0$ (8 distributions)
Multi-jets [33] 3-to-2 jet ratios for $p_{\text{T}}^{\text{jets}} > 80, 110$ GeV (2 distributions)
Track-jet properties [27] Charged jet $\rho_{\text{ch}}(r)$ for $R = 0.4$ (20 distributions) and $R = 0.6$ (20 distributions)
Substructure [28] Jet mass, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$, τ_{21} , τ_{23} (36 distributions)
Track-jet UE [25] Mean N_{ch} vs. $p_{\text{T}}^{\text{lead}}$ in transverse & away region for $p_{\text{T}}^{\text{lead}} > 30$ GeV (10 distributions) Mean p_{T} vs. $p_{\text{T}}^{\text{lead}}$ in transverse & away region for $p_{\text{T}}^{\text{lead}} > 30$ GeV (10 distributions) Mean Σp_{T} vs. $p_{\text{T}}^{\text{lead}}$ in transverse & away region for $p_{\text{T}}^{\text{lead}} > 30$ GeV (10 distributions)
η dependence on E_{T} [48] ΣE_{T} for the dijet selection (6 distributions)
Leading jet UE [26] Transverse, trans-max, trans-min $\Sigma p_{\text{T}}^{\text{ch}}$ vs $p_{\text{T}}^{\text{lead}}$ for $p_{\text{T}}^{\text{lead}} > 200$ GeV (6 distributions) Transverse, trans-max, trans-min N_{ch} vs $p_{\text{T}}^{\text{lead}}$ for $p_{\text{T}}^{\text{lead}} > 200$ GeV (6 distributions) Transverse ΣE_{T} vs $p_{\text{T}}^{\text{lead}}$ for $p_{\text{T}}^{\text{lead}} > 200$ GeV (4 distributions) Transverse $\Sigma p_{\text{T}}^{\text{ch}} / \Sigma E_{\text{T}}$ vs $p_{\text{T}}^{\text{lead}}$ for $p_{\text{T}}^{\text{lead}} > 200$ GeV (2 distributions) Transverse mean p_{T}^{ch} vs $p_{\text{T}}^{\text{lead}}$ for $p_{\text{T}}^{\text{lead}} > 200$ GeV (2 distributions)

Table 4.11.: The observables and bin ranges used for the dedicated dijet and Z boson tune, A15_AJZ. Every observable has the same weight 1.

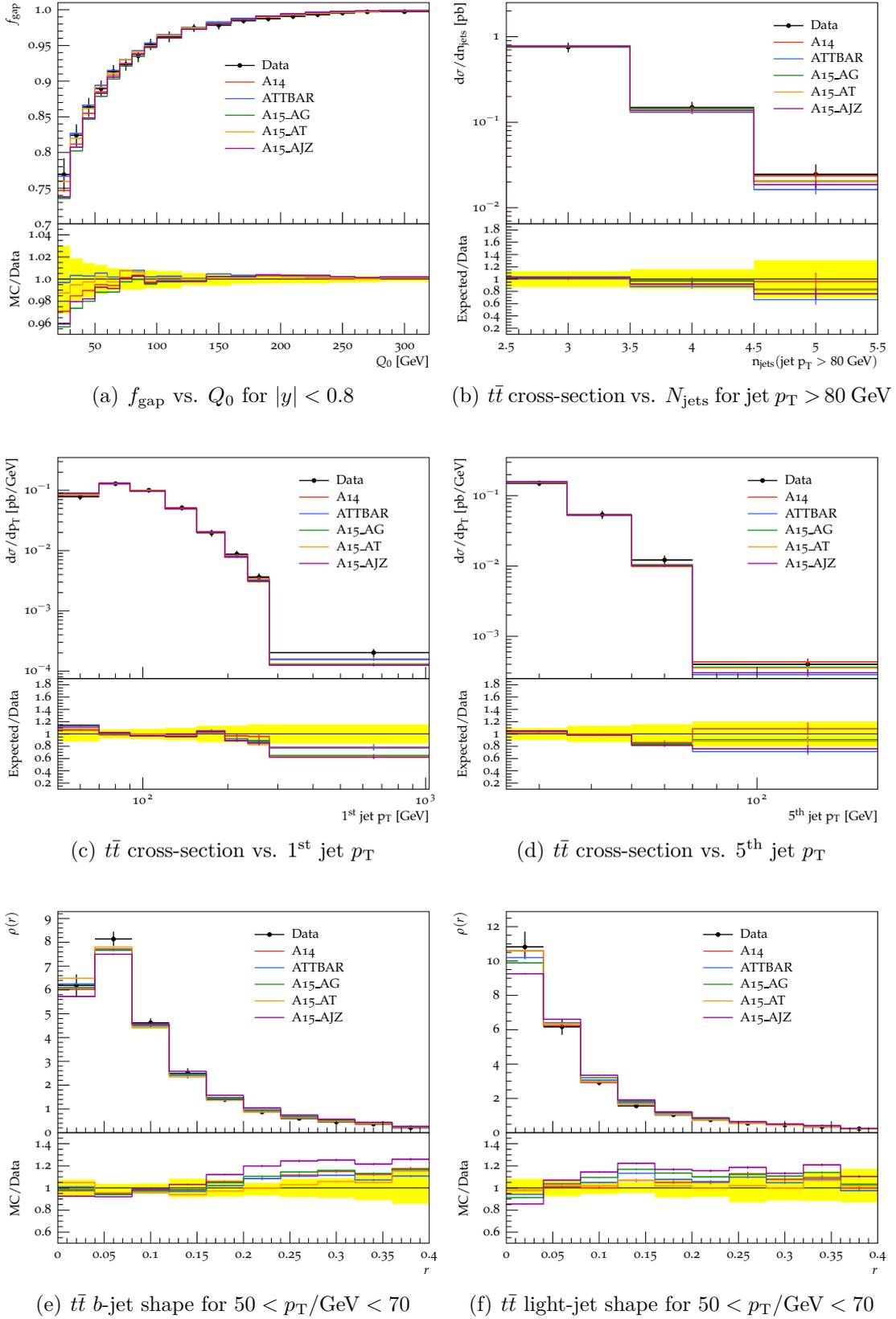


Figure 4.11.: The MC predictions of MADGRAPH5_amc@NLO+PYTHIA8 (A14, ATTBAR, A15_AG, A15_AT, A15_AJZ) compared with ATLAS $t\bar{t}$ gap fraction, jet multiplicity, jet p_{T} and jet shape distributions. The yellow shaded areas represent data uncertainty.

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

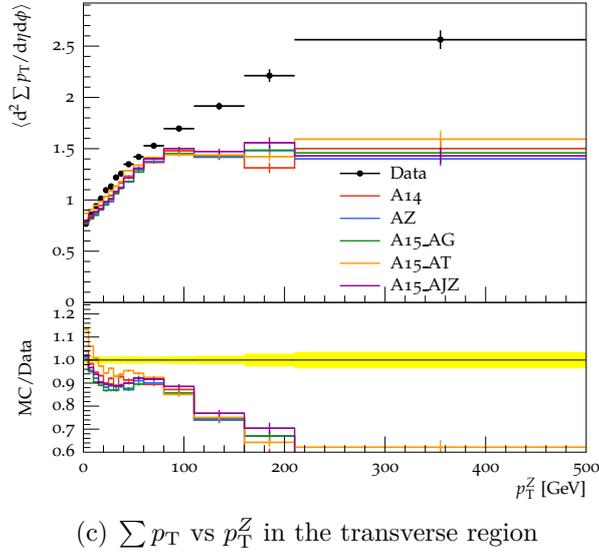
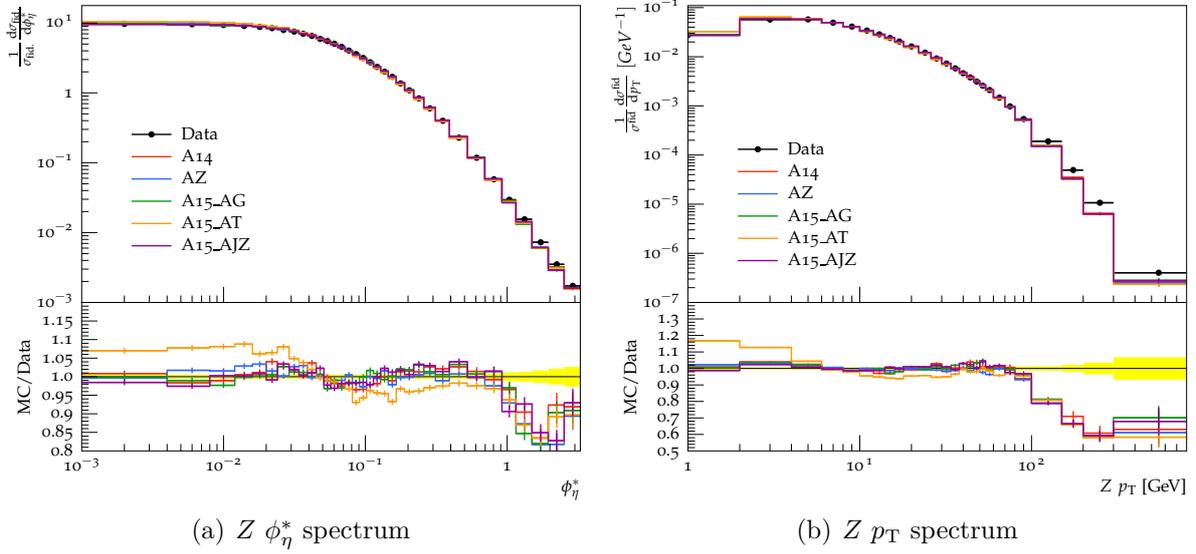


Figure 4.12.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (A14, AZ, A15_AG, A15_AT, A15_AJZ) compared with ATLAS $Z \phi_\eta^*$, $Z p_T$ and Z UE distributions. The yellow shaded areas represent data uncertainty.

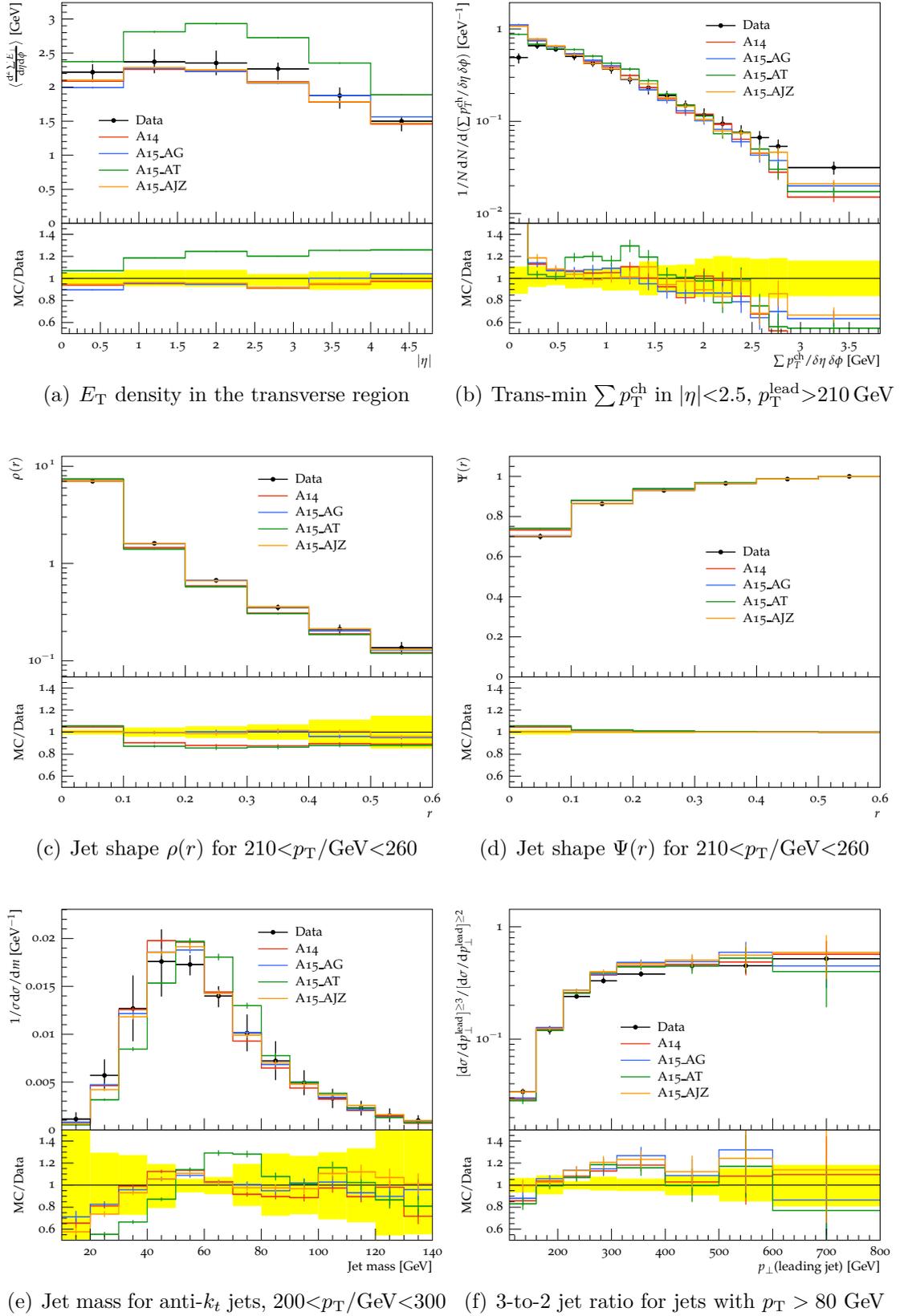


Figure 4.13.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (A14, A15_AG, A15_AT, A15_AJZ) compared with ATLAS dijet E_T density, leading jet UE, jet shape, jet substructure and multi-jet 3-to-2 jet ratio distributions. The yellow shaded areas represent data uncertainty. 67

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

Parameter	A14	ATTBAR	AZ	A15_AG	A15_AT	A15_AJZ
α_S^{FSR}	0.127	0.137	0.1383	0.1385 ± 0.0007	0.122 ± 0.003	0.1409 ± 0.0006
$p_{\text{T,min}}^{\text{FSR}}$	0.5	1.26	0.4	1.18 ± 0.05	1.0 ± 0.3	0.76 ± 0.04
α_S^{ISR}	0.127	0.121	0.1237	0.1267 ± 0.0002	0.119 ± 0.003	0.1275 ± 0.0003
$p_{\text{T0,Ref}}^{\text{ISR}}$	1.56	2.0	0.59	0.87 ± 0.03	0.7 ± 0.3	1.13 ± 0.06
α_S^{MPI}	0.126	0.130	0.135	0.124 ± 0.002	0.115 ± 0.003	0.128 ± 0.001
$p_{\text{T0,Ref}}^{\text{MPI}}$	2.09	2.16	2.18	2.06 ± 0.06	1.7 ± 0.2	2.28 ± 0.05
P. $k_{\text{T,hard}}$	1.88	1.8	1.71	1.74 ± 0.02	1.6 ± 0.2	1.80 ± 0.02
R. range	1.71	1.8	1.5	2.1 ± 0.1	0.1 ± 0.5	2.3 ± 0.1
χ^2/Ndf				0.89	0.41	0.63

Table 4.12.: Comparison of the A15_AG, A15_AT and A15_AJZ tune parameters with the A14, ATTBAR and AZ tunes.

Dedicated tuning plots for $t\bar{t}$ observables

The top left plot in fig. 4.11 shows the $t\bar{t}$ gap fraction as a function of Q_0 for $|y| < 0.8$. The two dedicated $t\bar{t}$ tunes, ATTBAR and A15_AT, perform noticeably better than the other tunes for this observable. Especially in the low p_{T} threshold bins, ATTBAR and A15_AT agree well with data, while A14, A15_AG and A15_AJZ predict significantly smaller gap fractions than what is observed in data.

The top right plot shows the jet multiplicity for jets with $p_{\text{T}} > 80$ GeV. Overall, A14 shows the best agreement with data, while ATTBAR deviates from data the most, although its MC prediction is still within data uncertainty. The dedicated A15_AT tune performs slightly worse than the general A15_AG tune here.

In the two centre plots, the 1st and 5th jet p_{T} spectra are displayed. For the 1st jet p_{T} , the A14 and ATTBAR tunes show the best agreement with data compared to the A15 tunes which predict significantly fewer hard jets than observed in data. A14, A15_AT and A15_AG predict a 5th jet p_{T} spectrum closest to data, while A15_AJZ and ATTBAR significantly deviate from data. Overall, the A14 tune performs best for the jet p_{T} observables.

The two plots at the bottom depict the b -jet and light-jet shape, respectively, for jets with $50 < p_{\text{T}}/\text{GeV} < 70$. The A15_AT tune agrees best with data for both observables, while A15_AJZ deviates from data the most. Additionally, all tunes except for A15_AT predict significantly broader jets than what is measured in data. For both jet shapes, the ATTBAR tune performs second best, while noticeably worse than A15_AT.

Dedicated tuning plots for Z observables

Fig. 4.12 shows the Z ϕ_η^* and Z p_T spectra as well as the $\sum p_T$ of final state particles as a function of p_T^Z . For all three Z boson observables, the MC prediction of the A15_AT tune deviates significantly from data within the respective tuning ranges. This illustrates that a dedicated $t\bar{t}$ tune may not be useful to describe other processes.

Aside from this, the four other tunes predict a Z p_T spectrum close to data for $p_T^Z < 50$ GeV. Additionally, the A14, AZ, A15_AG and A15_AJZ tunes show a similar behaviour for the ϕ_η^* spectrum and perform equally well. Note that AZ is a tune developed for the PYTHIA8 generator and thus not necessarily the optimal Z boson tune for MADGRAPH5_aMC@NLO interfaced with PYTHIA8.

Finally, the A15_AG tune predicts a $\sum p_T$ that agrees best with data for $p_T^Z < 5$ GeV. Both dedicated Z tunes also perform well, while A14 and A15_AT show significant deviations from data.

Dedicated tuning plots for dijet observables

In the top left plot of fig. 4.13 the transverse energy density is shown as a function of $|\eta|$ in the transverse region. Similar to the Z boson observables, the A15_AT tune deviates significantly from data by predicting a much higher energy density than what is observed in data. The A14 and A15_AJZ tunes predict a slightly smaller E_T density than data and perform slightly better than A15_AG here.

The top right plot shows the p_T sum of charged particles in the trans-min region for $|\eta| < 2.5$ and $p_T^{\text{lead}} > 210$ GeV in inclusive dijet events. All tunes predict softer UE activity than what is measured in data, while the A15_AJZ tune agrees reasonably well with data within $0.5 \text{ GeV} < \sum p_T^{\text{ch}} < 2.5 \text{ GeV}$ and performs best overall. The A15_AT shows significant deviations from data here and the A14 and A15_AG tune perform comparably well.

The two centre plots show the differential and integrated jet shape, respectively, for jets with $210 < p_T/\text{GeV} < 260$ and $|y| < 2.8$. The MC predictions of A15_AG and A15_AJZ agree well with data, while A15_AJZ performs slightly better. The A14 and A15_AT tunes both significantly deviate from data here.

The bottom left plot displays the jet mass spectrum for anti- k_t jets with $200 < p_T/\text{GeV} < 300$. The A15_AG and A15_AJZ tunes perform best out of the four tunes and agree reasonably well with data, while A15_AG performs slightly better than A15_AJZ and A14. The A15_AT tune shows overall large deviations from data here.

Lastly, the bottom right plot depicts the 3-to-2 jet ratio as a function of the leading jet p_T for jets with $p_T > 80$ GeV. This observable is not modelled well by any of the four

4. Developing a new tune for MadGraph5_aMC@NLO+Pythia8

tunes, but the MC predictions are mostly limited by lack of statistics. Especially in the $200 < p_T^{\text{lead}}/\text{GeV} < 400$ region, all tunes predict a significantly higher 3-to-2 jet ratio than what is measured in data. Overall, the A14 tune performs best for this observable. Although A15_AT is a dedicated $t\bar{t}$ tune, its MC prediction is comparable to the other three tunes and performs slightly better than A15_AG and A15_AJZ here.

Parameter results

Overall, the parameters of the A15 tunes are very different among each other. The uncertainty and thus the constraint on the parameters is roughly one order of magnitude worse for A15_AT compared to A15_AG and A15_AJZ, because of the much smaller number of tuning observables included in the A15_AT tune. Additionally, the A15_AJZ tune has a stronger constraint on FSR and MPI, but weaker constraint on the ISR parameters than the A15_AG tune, because the A15_AG tune includes fewer dijet observables, but puts a higher weight on Z observables.

The α_S^{FSR} in A15_AG is closest to the ATTBAR and AZ value, the latter coming from a tuning to LEP data. On the other hand, the dedicated $t\bar{t}$ tune has a significantly smaller α_S^{FSR} which is closer to A14. This agrees with the discussions in section 4.6, namely that the $t\bar{t}$ jet shapes require a smaller α_S^{FSR} to describe data. This parameter is slightly greater for A15_AJZ than for A15_AG which comes from the fact that more dijet observables are tuned in A15_AJZ.

The α_S^{ISR} parameters of A15_AG and A15_AJZ are very comparable to each other and close to A14, but slightly greater than the AZ value. This reflects that their respective description of the Z observables are very similar. On the other hand, A15_AT has a significantly smaller α_S^{ISR} than the other tunes, which is necessary to improve the MC prediction for the $t\bar{t}$ gap fractions. In contrast to that, the ATTBAR tune has a higher α_S^{ISR} than A15_AT, but it achieves a comparably good agreement with data through the use of the local recoil option.

The α_S^{MPI} parameters of A15_AG and A15_AJZ are comparable to the A14 tune. Additionally, for A15_AJZ this parameter is also closer to ATTBAR and AZ. Since this parameter is mostly sensitive to the UE, A15_AG and A15_AJZ describe the respective observables reasonably well, while A15_AJZ performs slightly better. On the other hand, this parameter is significantly smaller for the A15_AT tune compared to the five other tunes in order to improve the $t\bar{t}$ jet shape prediction.

Most of the tune observables do not show a significant sensitivity to the p_T cut-off parameters, but they are strongly correlated to the respective α_S parameters. The $p_{T,\text{min}}^{\text{FSR}}$ parameter for A15_AG can be compared to the ATTBAR tune, while the A15_AJZ value

is closer to A14 and AZ which again comes from the higher number of dijet observables considered in this tune. The A15_AT tune parameter agrees with the ATTBAR value within the uncertainties.

The ISR p_T cut-off for A15_AG is significantly different from A14 and ATTBAR, but comparable to the other tunes. In contrast to this, the $p_{T0,Ref}^{MPI}$ parameters of A15_AG and A15_AJZ are comparable to A14, ATTBAR and AZ. The value for A15_AG is closest to A14, while for A15_AJZ it is closest to AZ. This parameter is smaller for A15_AT than for the other tunes to improve $t\bar{t}$ modelling.

The primordial $k_{T,hard}$ parameter of A15_AG agrees with AZ, while the A15_AJZ value agrees with ATTBAR. For A15_AT this parameter is closest to AZ, but agrees with all other tunes as well, due to its large uncertainty.

Lastly, the reconnect range for A15_AG and A15_AJZ is noticeably greater than for the other tunes. Similar to the p_T cut-off parameters, the tuning observables are only weakly sensitive to the reconnect range which may cause this observed discrepancy. On the other hand, the reconnect range of A15_AT is significantly smaller than for all other tunes. Its uncertainty is larger than its mean value and thus it is compatible with zero. However, a reconnect range of zero is not physically meaningful and should be avoided by using the mean value.

In conclusion, the A15_AT tune performs very well for $t\bar{t}$ observables. Especially the modelling of $t\bar{t}$ jet shapes with A15_AT outperforms the other tunes and for $t\bar{t}$ gap fractions its performance is comparable to ATTBAR. However, this tune is not recommended to describe Z boson and dijet observables, because of the significantly smaller FSR, ISR and MPI parameters compared to the other tunes.

Overall, the A15_AJZ tune performs well for Z and dijet measurements. It performs slightly better than A15_AG here, but is worse for $t\bar{t}$.

A15_AG is a universal tune for all processes and compromises in describing all tuned observables reasonably well. However, the observed tensions between some observables in this study make it impossible to find one single tune that is able to perfectly model all processes.

5. Conclusion and outlook

In this thesis, the theoretical background of particle physics was presented with a focus on top quark physics, the Higgs mechanism and QCD. The experimental setup of the LHC and the ATLAS experiment were discussed. The MC event simulation and the concept of parton shower and hadronisation models were explained as well as the strategy of MC generator tuning to LHC data. During Run I of the LHC, POWHEG+PYTHIA6 with the P2011c tune was used as a baseline generator in ATLAS to simulate $t\bar{t}$ events. Given a mismodelling of top quark p_T and $t\bar{t}$ p_T distributions observed in Run I, new MC generators, for example MADGRAPH5_aMC@NLO interfaced with PYTHIA8, are being studied and tuned aiming to give more accurate predictions for Run II of the LHC.

The performance of the latest universal ATLAS tune, called A14 and developed for PYTHIA8 stand-alone, was compared for two different NLO matrix element generators, namely POWHEG and MADGRAPH5_aMC@NLO, interfaced with PYTHIA8. The results showed that this tune is not optimal for the NLO generators and that further tuning is necessary. For the tuning of MADGRAPH5_aMC@NLO+PYTHIA8, ATLAS $\sqrt{s} = 7$ TeV measurements with $t\bar{t}$, Z , and dijet events were used. The tuning was performed with a framework developed within ATLAS which uses the Rivet analysis toolkit [5] for analysing the generated events, and the PROFESSOR MC tuning system [24] which is a powerful tool for parameter variation and fitting to determine the best parameter values for a tune.

The result of this tuning study is a set of three different tunes, denoted as A15_AG, A15_AJZ and A15_AT. Tensions between tuning observables have been observed which prevent the pursue of a single tune that perfectly models every process within the scope of this thesis. Therefore, in addition to the general tune A15_AG, the two dedicated tunes A15_AJZ and A15_AT were developed and presented here. The A15_AT tune is a dedicated tune to optimise $t\bar{t}$ modelling, while A15_AJZ is dedicated to Z boson and dijet measurements. Both tunes give good predictions for the processes they were developed for and describe the corresponding observables as well as or better than the ATLAS tunes A14, ATTBAR and AZ, the latter two being dedicated $t\bar{t}$ and Z tunes, respectively.

5. Conclusion and outlook

In addition to this, different studies were performed with the universal A15_AG tune to study the observables' sensitivity to the tuned parameters, correlations between them and the impact of the local recoil scheme on the MC predictions of the A15_AG tune.

Future studies could pursue the following steps to continue this study. Since the uncertainty estimation of the χ^2 fit is an underestimate for the real error coverage of the presented tuning procedure, eigentune variations should be performed to introduce systematic uncertainties to the tune results. Additionally, other studies with PROFESSOR could be done such as testing whether the tunes are robust against reasonable shifts of the “Goodness of Fit” definition, or the interactive parametrisation explorer which is a graphical interactive tool to show how an observable behaves under slight parameter changes. Unfortunately, these studies were not accessible, due to technical issues with the PROFESSOR tool that could not be resolved within the given time.

Finally, apart from further improving this study, tuning studies should be pursued using other MC generators, other models, other parameters, and other observables - or a different tuning procedure altogether, because this study shows that the used MC models are not perfectly tunable to describe all ATLAS $\sqrt{s} = 7$ TeV data from Run I. This became increasingly more important with the start of Run II of the LHC physics programme.

A. Rivet analyses

$t\bar{t}$	Z	dijet
$t\bar{t}$ gap fraction [32] (ATLAS_2012_I1094568)	$Z p_T$ [35] (ATLAS_2014_I1300647)	jet shapes [29] (ATLAS_2011_S8924791)
$t\bar{t}$ jet shapes [30] (ATLAS_2013_I1243871)	$Z \phi^*$ [45] (ATLAS_2012_I1204784)	angular decorrelation [31] (ATLAS_2011_S8971293)
$t\bar{t} N_{\text{jets}}$, jet p_T [41] (ATLAS_2014_I1304688)	Z UE [46] (ATLAS_2014_I1315949)	gap fractions [47] (ATLAS_2011_S9126244)
		multi-jet cross sections [33] (ATLAS_2011_S9128077)
		track-jets properties [27] (ATLAS_2011_I919017)
		jet mass and substructure [28] (ATLAS_2012_I1094564)
		track-jets UE [25] (ATLAS_2012_I1125575)
		η dependence on E_T [48] (ATLAS_2012_I1183818)
		leading jet UE [26] (ATLAS_2014_I1298811)

Table A.1.: The observables used for the tuning and the corresponding Rivet analyses in brackets.

B. Other tuning result plots

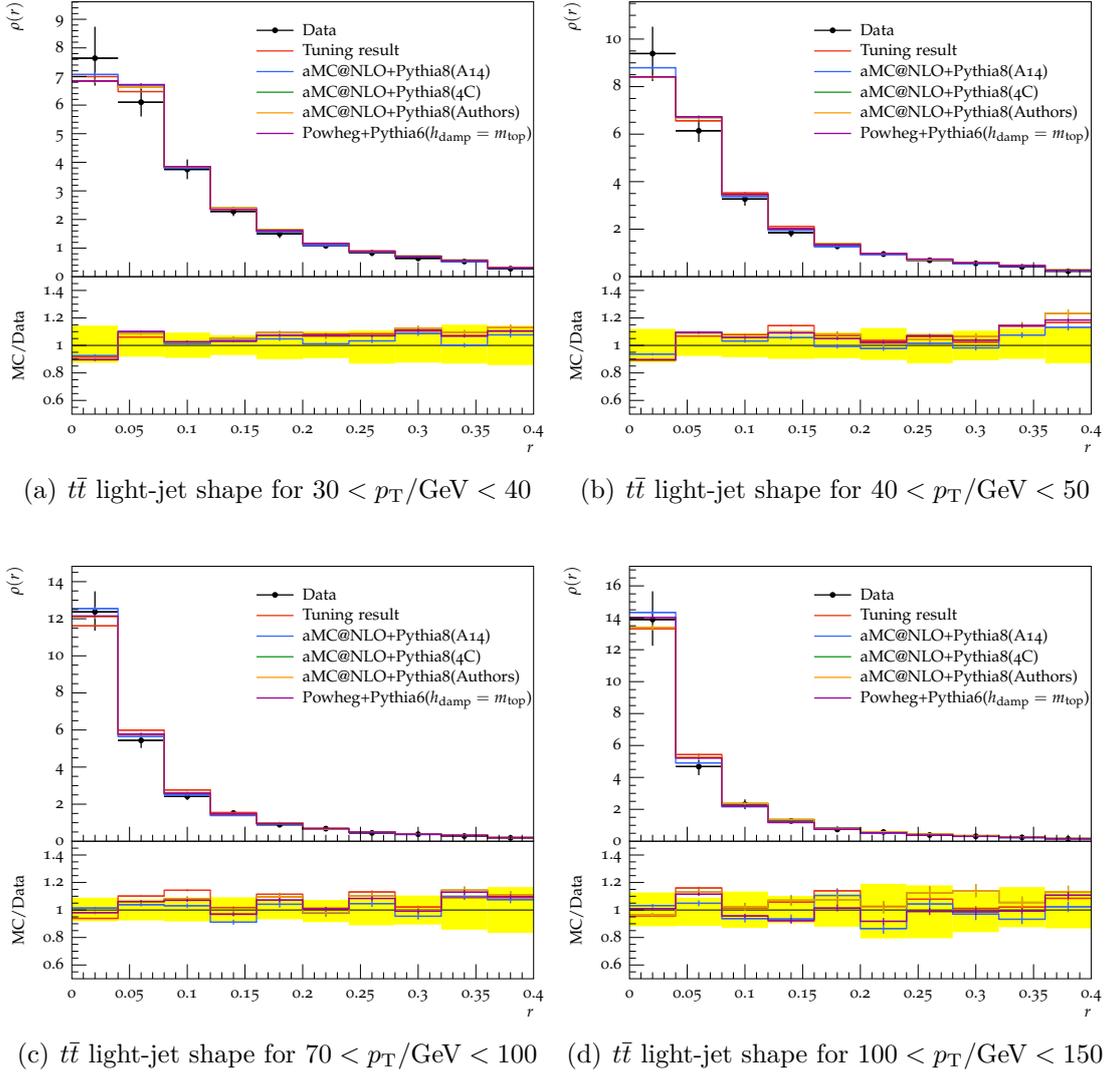
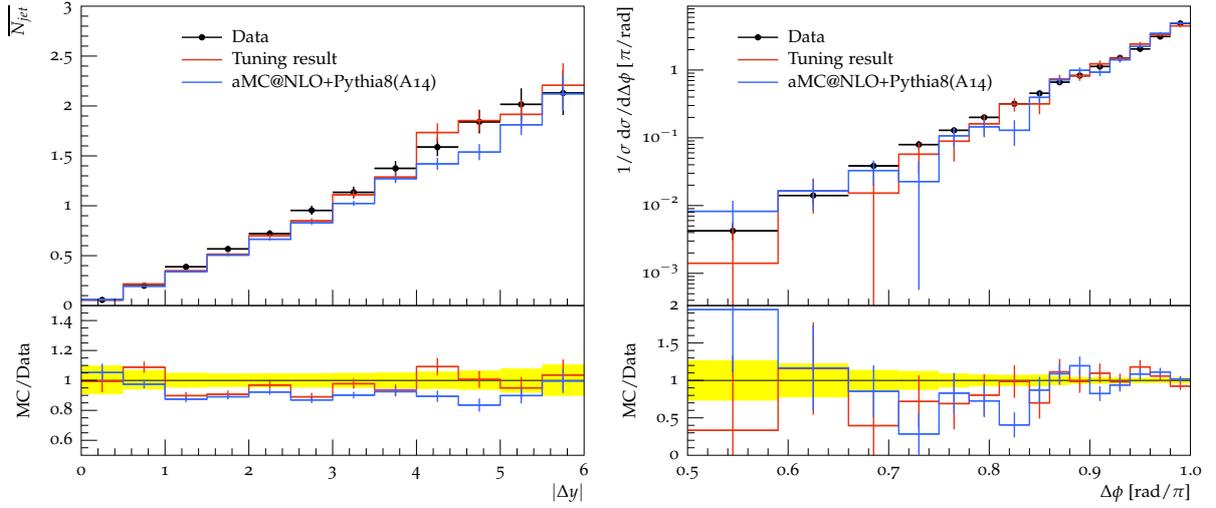
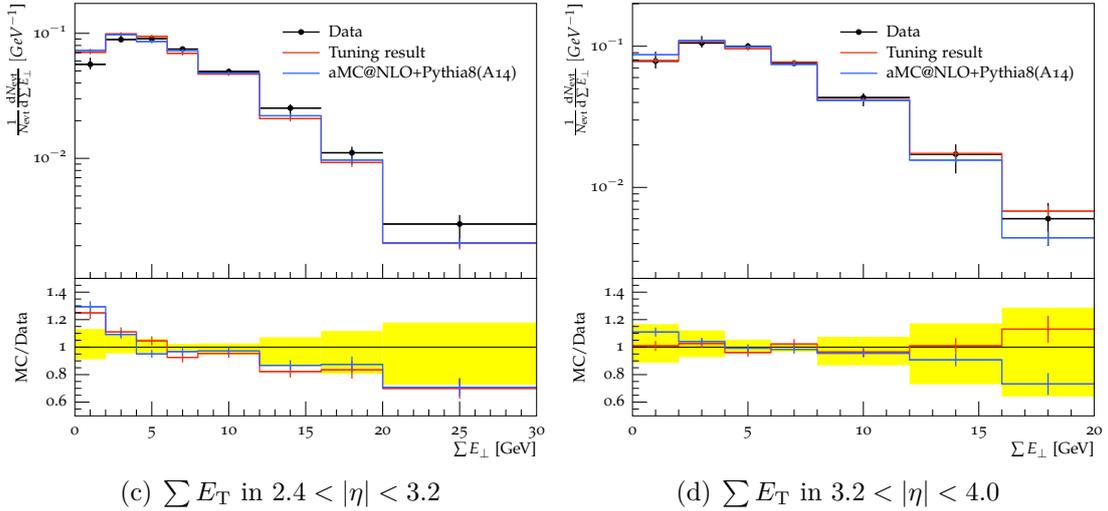


Figure B.1.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (A15_AG, A14, 4C, Authors) and POWHEG+PYTHIA6 (P2011c, $h_{\text{damp}} = m_{\text{top}}$) compared with ATLAS $t\bar{t}$ jet shape distributions. The yellow shaded areas represent data uncertainty.

B. Other tuning result plots

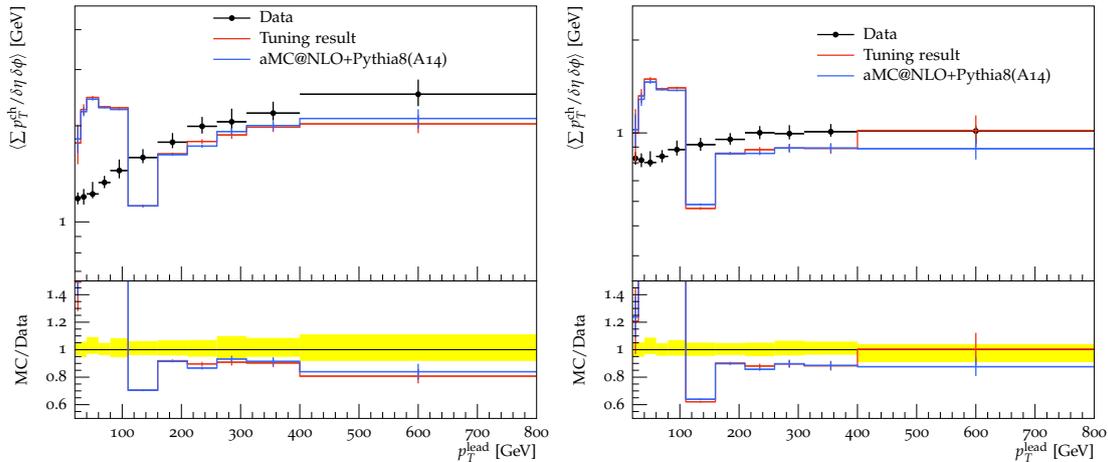


(a) $\overline{N_{jets}}$ vs $|\Delta y|$ for $120 < p_T / \text{GeV} < 150$, Fwd/Bwd (b) Dijet azim. decorr. for $210 < p_T^{\text{max}} / \text{GeV} < 260$



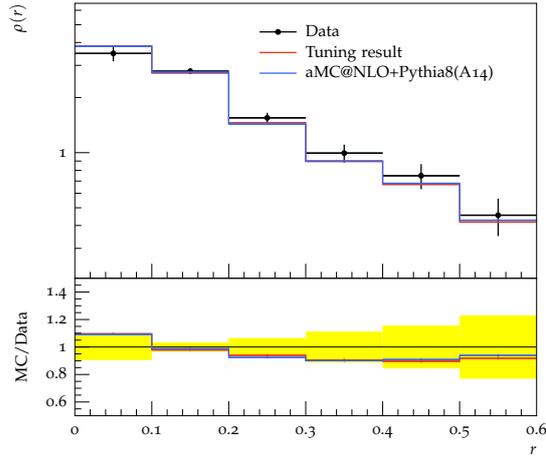
(c) $\sum E_T$ in $2.4 < |\eta| < 3.2$

(d) $\sum E_T$ in $3.2 < |\eta| < 4.0$

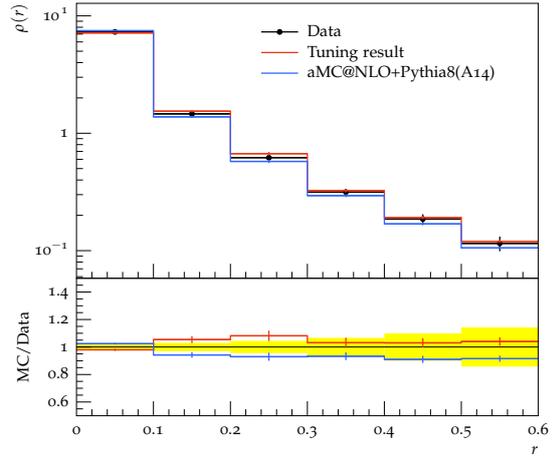


(e) Transverse $\sum p_T^{\text{ch}}$ vs p_T^{lead} , $|y| < 2.5$, incl. (f) Trans-min $\sum p_T^{\text{ch}}$ vs p_T^{lead} , $|y| < 2.5$, incl.

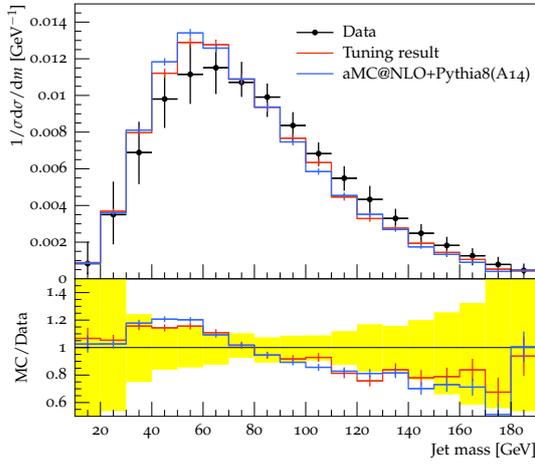
Figure B.2.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (A15_AG and A14) compared with ATLAS dijet gap fraction, dijet angular decorrelation, transverse energy flow and leading jet UE distributions. The yellow shaded areas represent data uncertainty.



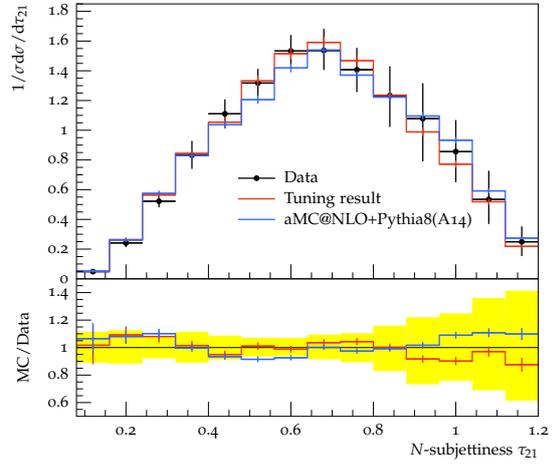
(a) Jet shape $\rho(r)$ for $30 < p_T / \text{GeV} < 40$



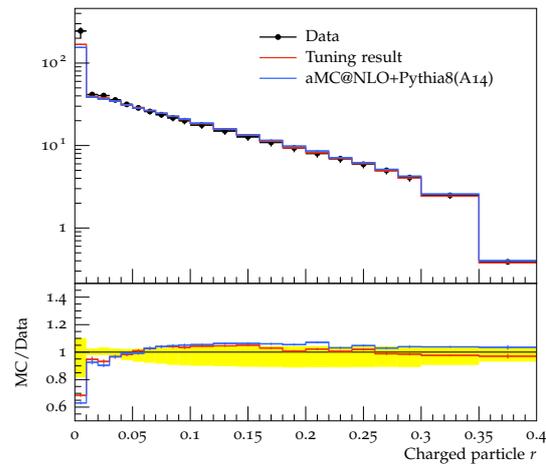
(b) Jet shape $\rho(r)$ for $260 < p_T / \text{GeV} < 310$



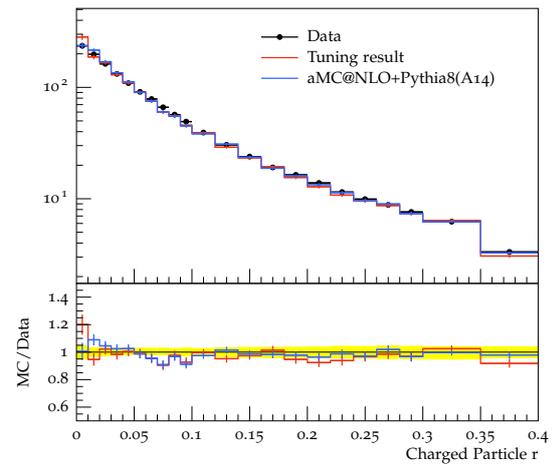
(c) Jet mass for C/A jets, $200 < p_T / \text{GeV} < 300$



(d) τ_{21} for anti- k_t jets, $200 < p_T / \text{GeV} < 300$



(e) Density $\rho_{\text{ch}}(r)$, $4 < p_T / \text{GeV} < 6$, $|y| < 1.9$



(f) Density $\rho_{\text{ch}}(r)$, $24 < p_T / \text{GeV} < 40$, $|y| < 1.9$

Figure B.3.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (A15_AG and A14) compared with ATLAS dijet jet shape, jet substructure and charged particle density distributions. The yellow shaded areas represent data uncertainty.

C. Other sensitivity plots

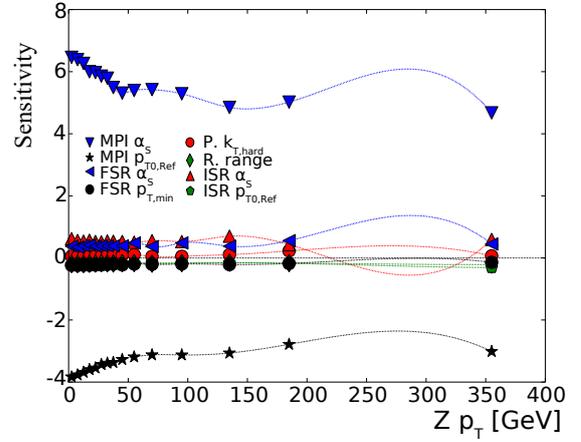
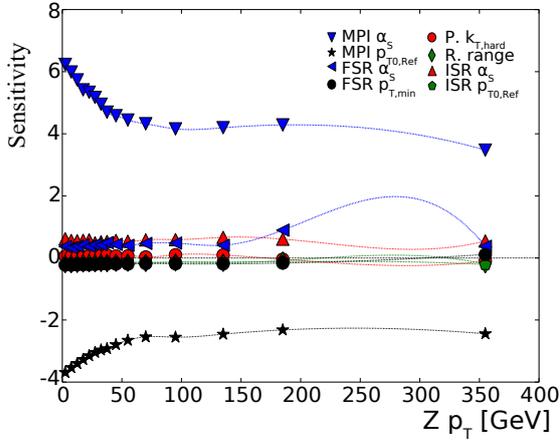
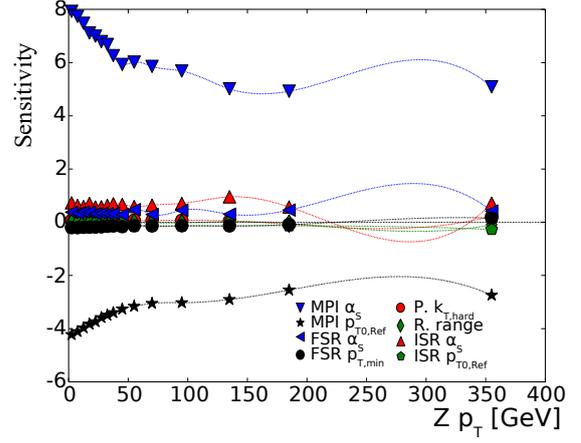
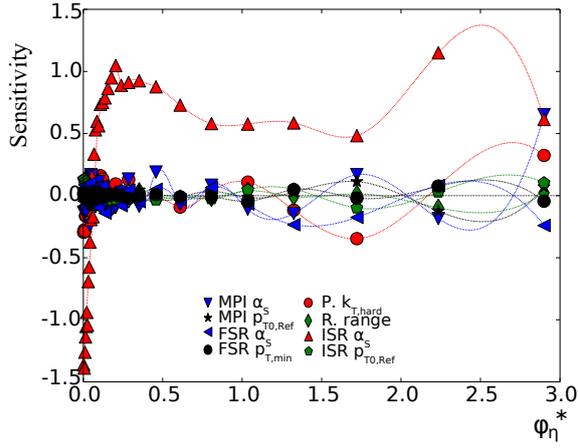
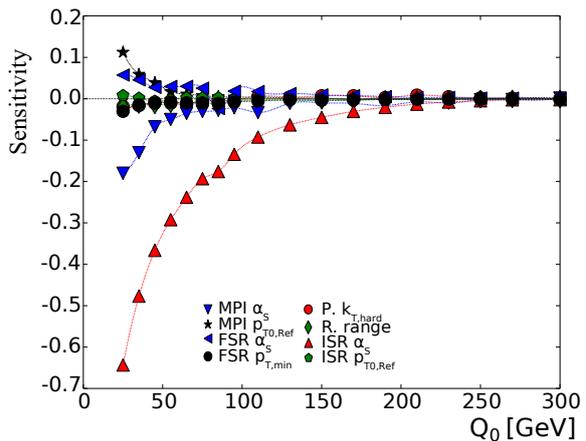
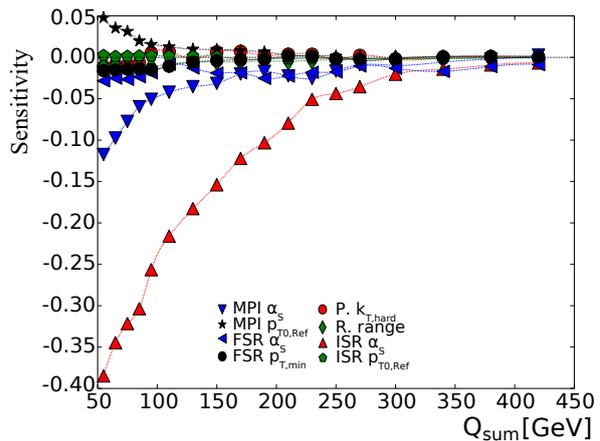


Figure C.1.: Sensitivity plots for $Z \phi_\eta^*$ and remaining Z UE observables.

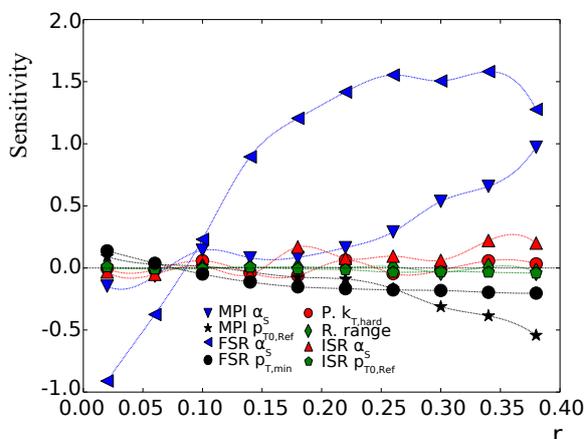
C. Other sensitivity plots



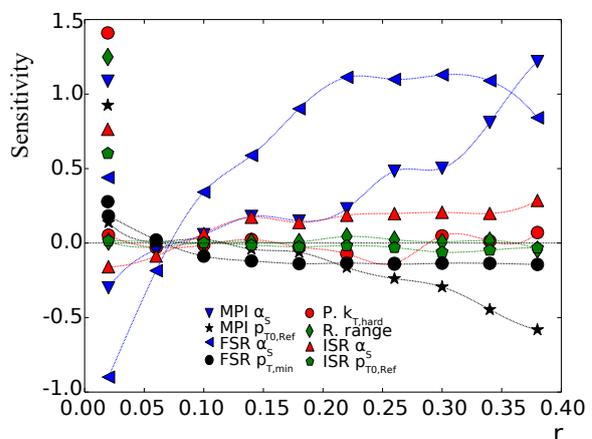
(a) $t\bar{t}$ f_{gap} vs. Q_0 for $|y| < 2.1$



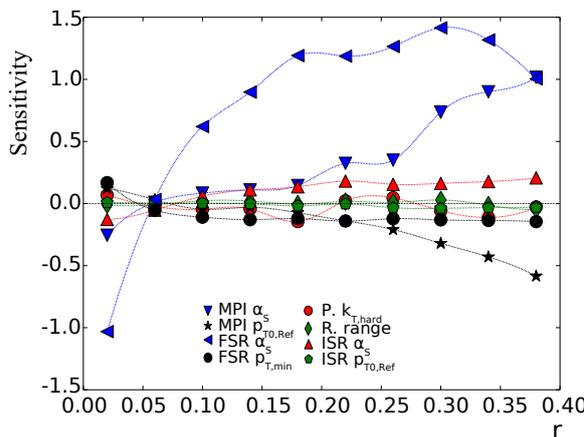
(b) $t\bar{t}$ f_{gap} vs. Q_{sum} for $|y| < 2.1$



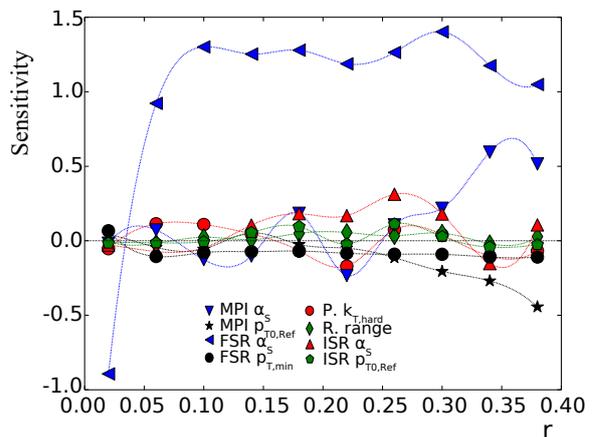
(c) $t\bar{t}$ b -jet shape for $50 < p_T/\text{GeV} < 70$



(d) $t\bar{t}$ light-jet shape for $30 < p_T/\text{GeV} < 40$

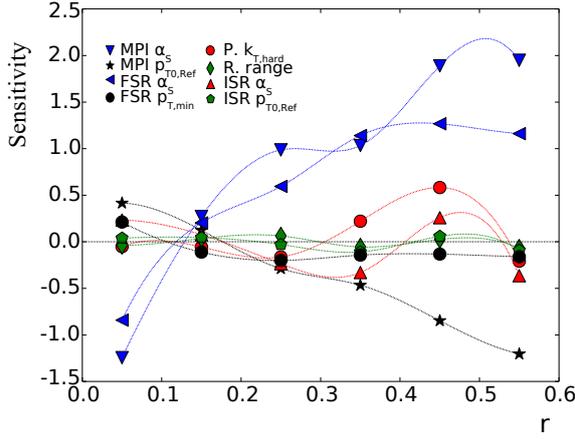


(e) $t\bar{t}$ light-jet shape for $40 < p_T/\text{GeV} < 50$

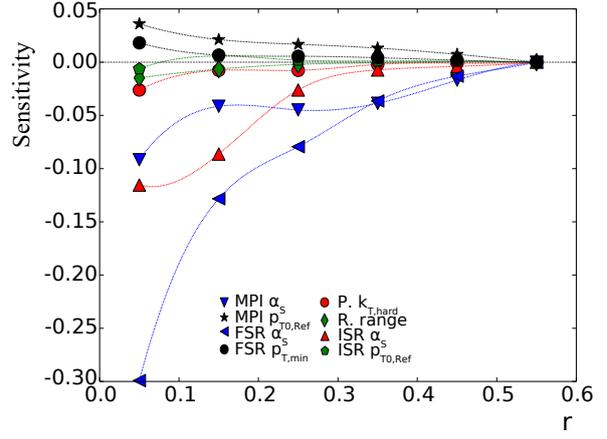


(f) $t\bar{t}$ light-jet shape for $100 < p_T/\text{GeV} < 150$

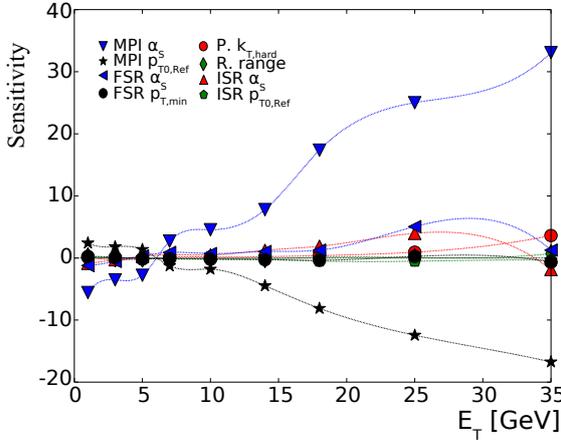
Figure C.2.: Sensitivity plots for $t\bar{t}$ gap fraction vs Q_0 and Q_{sum} , and $t\bar{t}$ b -jet and light-jet shapes.



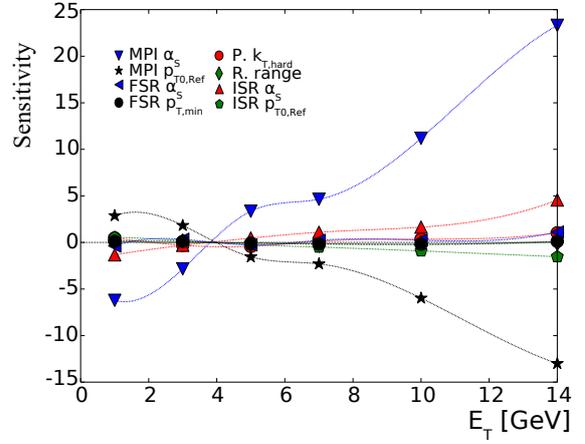
(a) J.shape $\rho(r)$ for $30 < p_T/\text{GeV} < 40, 0.8 < |y| < 1.2$



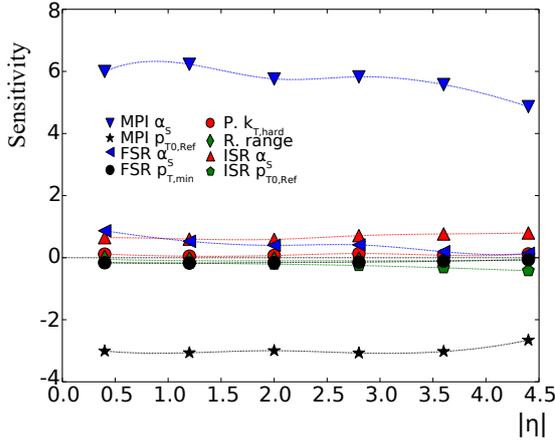
(b) Jet shape $\Psi(r)$ for $260 < p_T/\text{GeV} < 310$



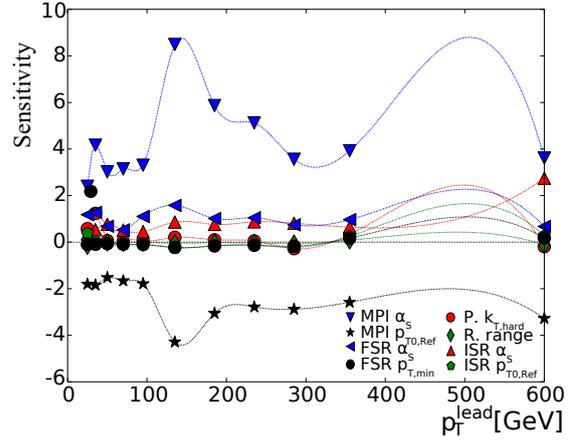
(c) $\sum E_T$ for $|\eta| < 0.8$



(d) $\sum E_T$ for $4.0 < |\eta| < 4.8$



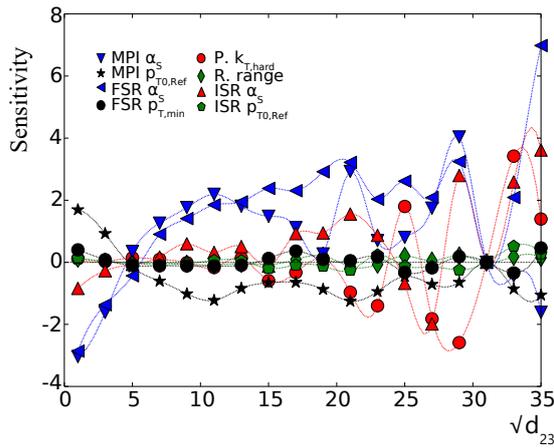
(e) E_T density in the transverse region



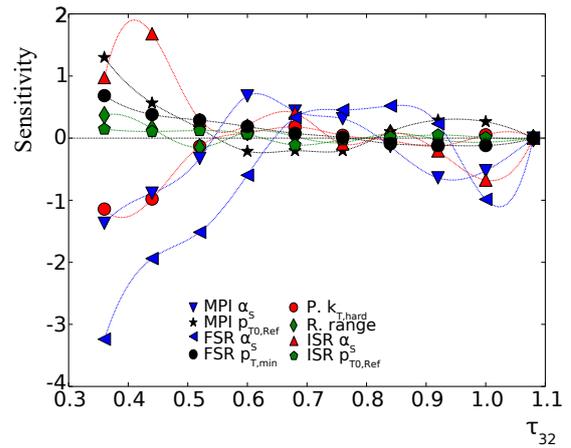
(f) $\sum p_T^{\text{ch}}$ vs p_T^{lead} in trans-min region

Figure C.3.: Sensitivity plots for dijet jet shapes, transverse energy flow and leading jet UE observables.

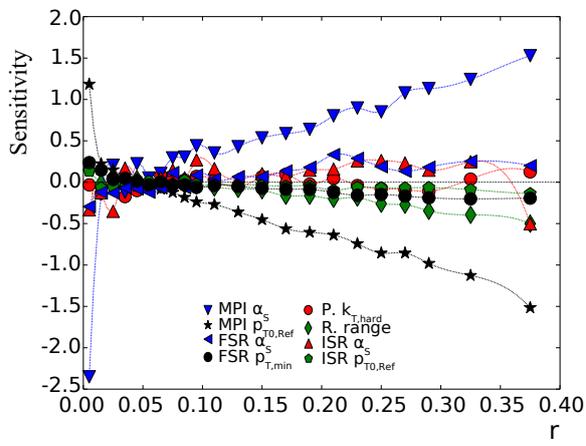
C. Other sensitivity plots



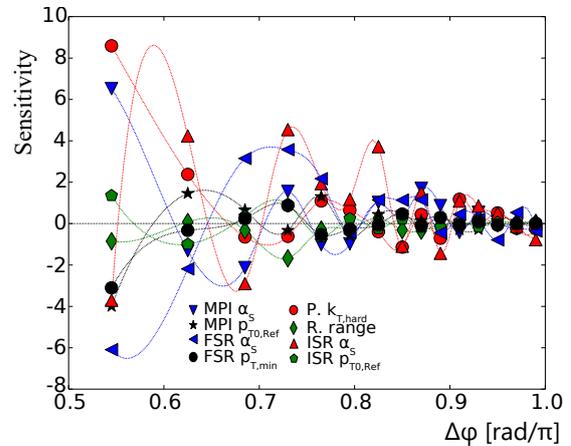
(a) $\sqrt{d_{23}}$ for anti- k_t jets, $200 < p_T/\text{GeV} < 300$



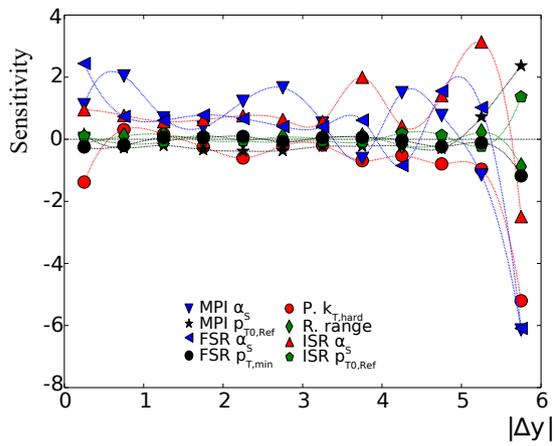
(b) τ_{32} for C/A jets, $200 < p_T/\text{GeV} < 300$



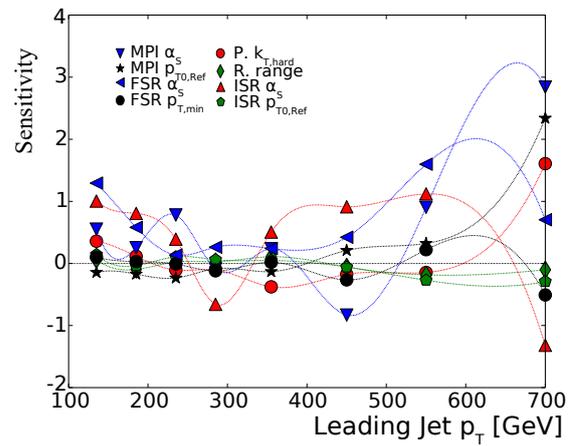
(c) Density $\rho_{\text{ch}}(r)$ for $4 < p_T/\text{GeV} < 6$, $|y| < 1.9$



(d) Dijet azim. decorr. for $160 < p_T^{\text{max}}/\text{GeV} < 210$



(e) $\overline{N_{\text{jets}}}$ vs. $|\Delta y|$ for $150 < p_T/\text{GeV} < 180$, Fwd/Bwd



(f) 3-to-2 jet ratio for jets with $p_T > 80$ GeV

Figure C.4.: Sensitivity plots for jet substructure, charged particle density, dijet azimuthal decorrelation, gap fraction and 3-to-2 jet ratio observables.

D. Other local recoil comparison plots

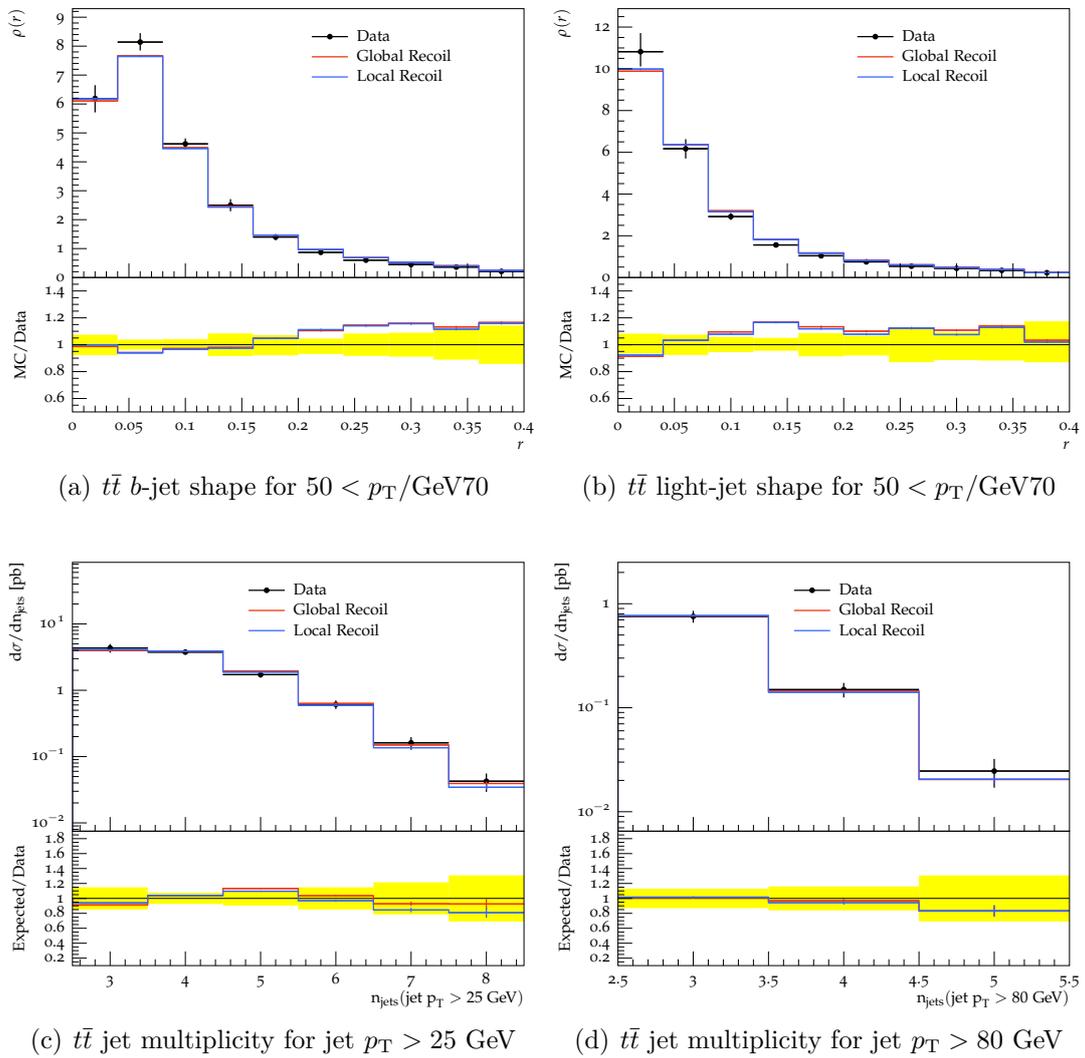


Figure D.1.: Comparison plots of A15_AG using global recoil or local recoil compared with ATLAS $t\bar{t}$ jet shapes and jet multiplicity distributions. The yellow shaded areas represent data uncertainty.

D. Other local recoil comparison plots

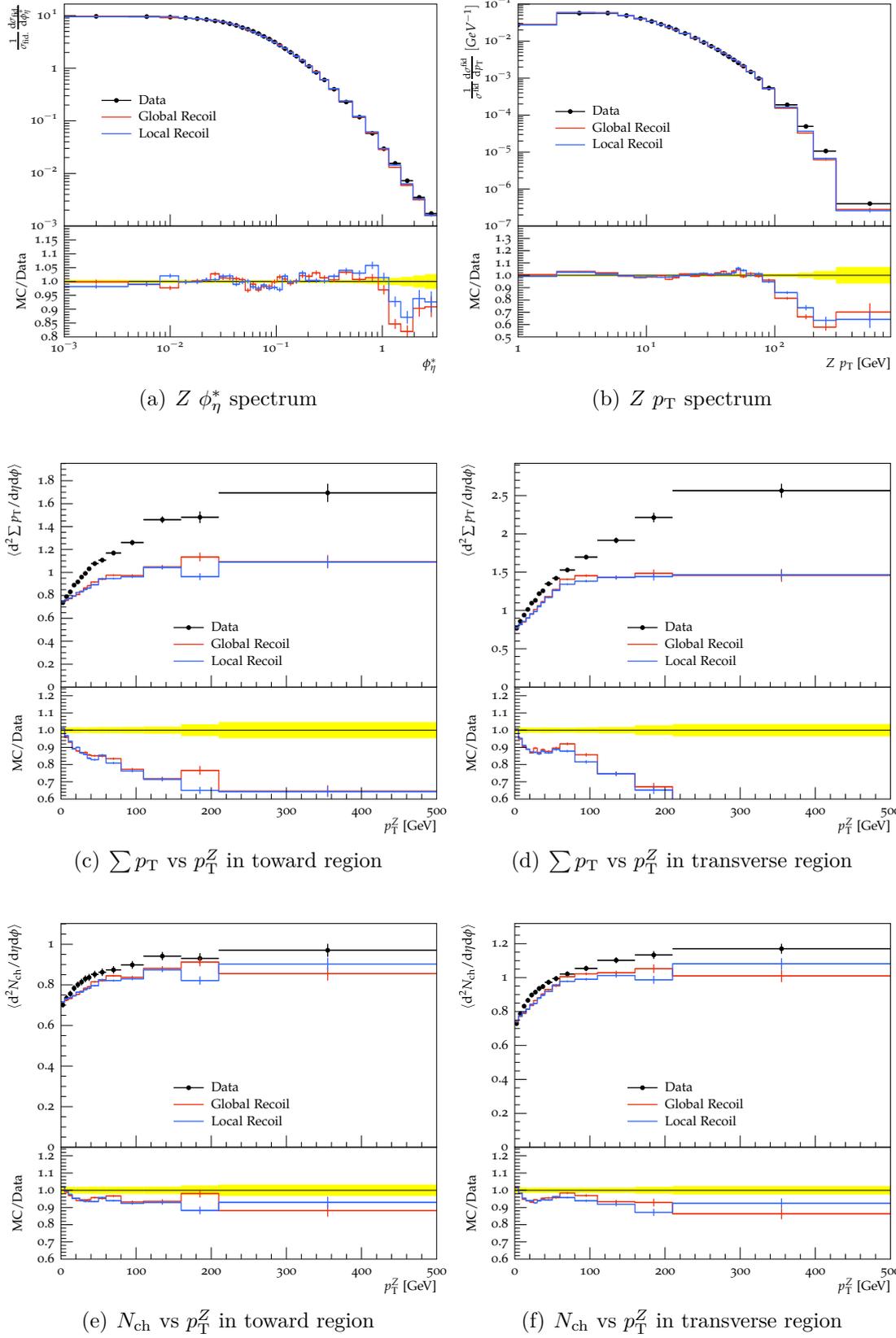
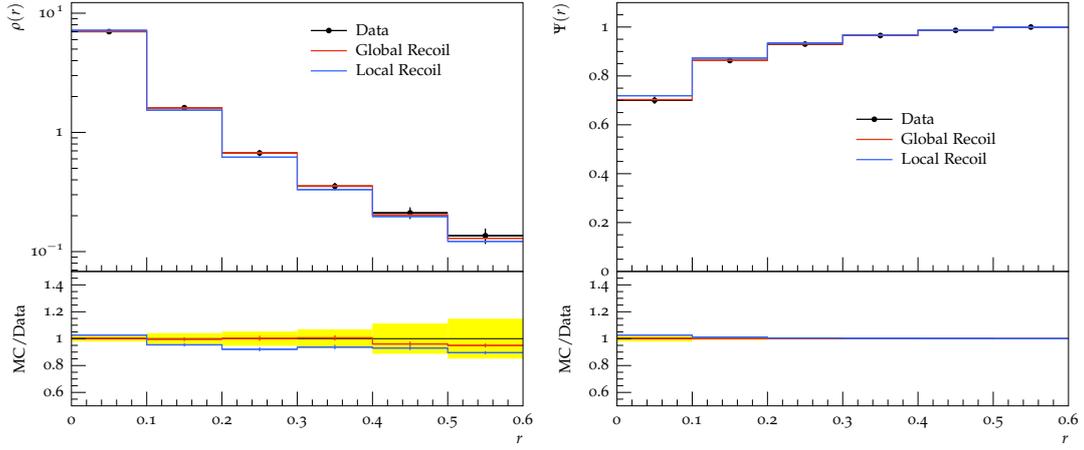
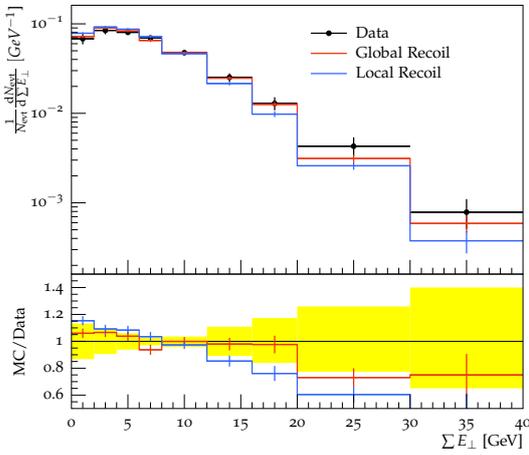


Figure D.2.: Comparison plots of A15_AG using global recoil or local recoil compared with ATLAS $Z \phi_\eta^*$, $Z p_T$ and Z UE distributions. The yellow shaded areas represent data uncertainty.

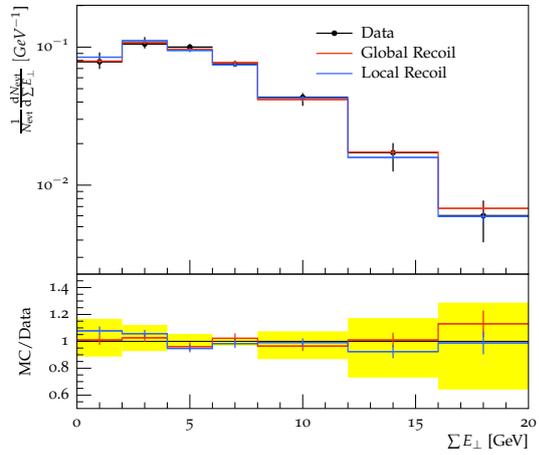


(a) Jet shape $\rho(r)$, $210 < p_T / \text{GeV} < 260$

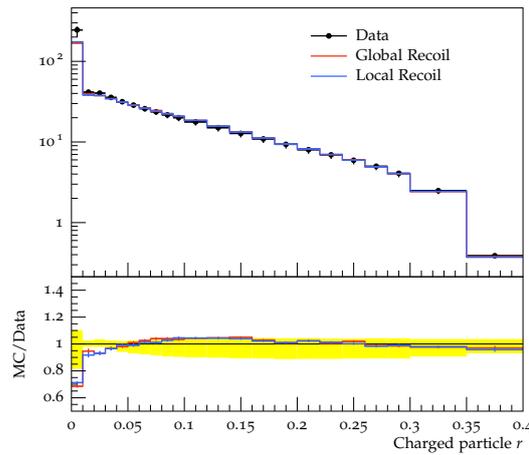
(b) Jet shape $\Psi(r)$, $210 < p_T / \text{GeV} < 260$



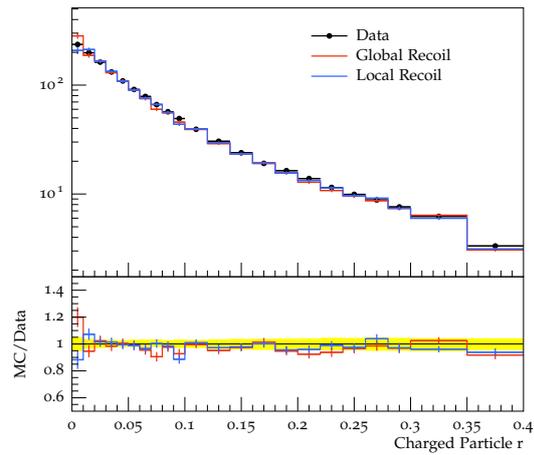
(c) $\sum E_T$ in $0.8 < |\eta| < 1.6$



(d) $\sum E_T$ in $3.2 < |\eta| < 4.0$



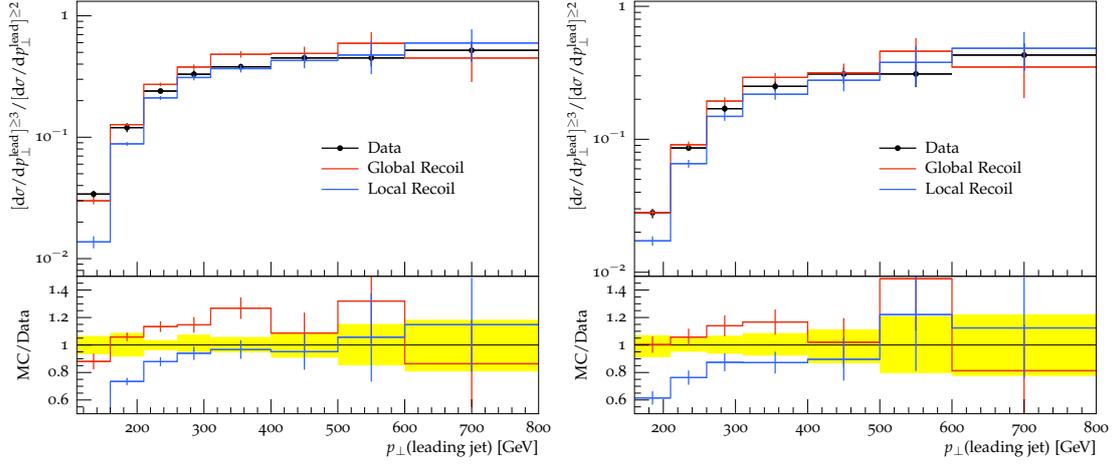
(e) Density $\rho_{\text{ch}}(r)$ for $4 < p_T / \text{GeV} < 6$, $|y| < 1.9$



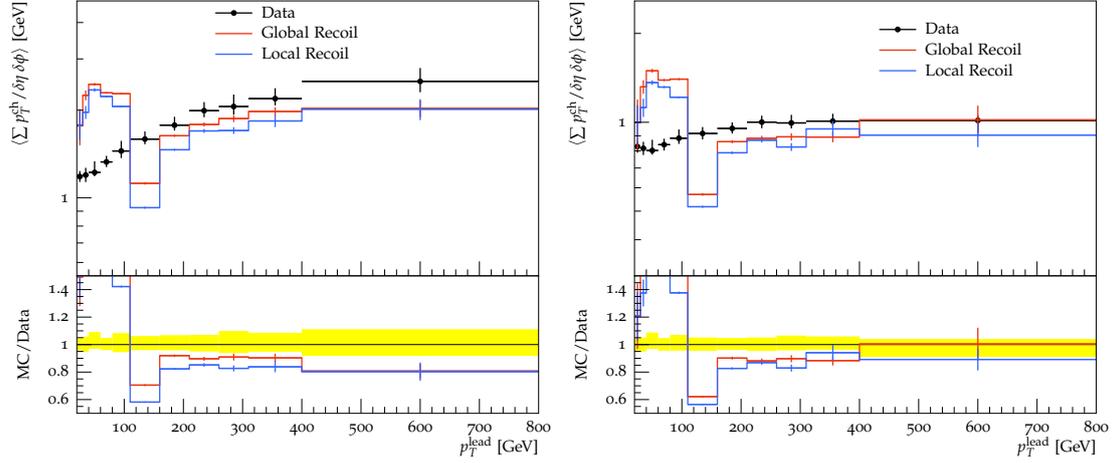
(f) Density $\rho_{\text{ch}}(r)$ for $24 < p_T / \text{GeV} < 40$, $|y| < 1.9$

Figure D.3.: Comparison plots of A15_AG using global recoil or local recoil compared with ATLAS dijet jet shapes, transverse energy flow and charged particle density distributions. The yellow shaded areas represent data uncertainty.

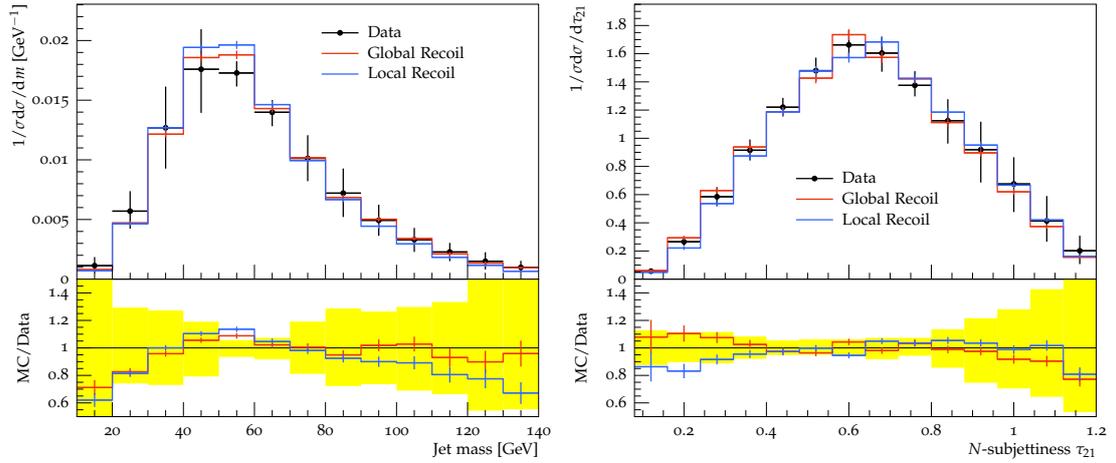
D. Other local recoil comparison plots



(a) 3-to-2 jet ratio for jets with $p_T > 80$ GeV (b) 3-to-2 jet ratio for jets with $p_T > 110$ GeV



(c) Transverse $\sum p_T^{\text{ch}}$ vs p_T^{lead} , $|y| < 2.5$, incl. (d) Trans-min $\sum p_T^{\text{ch}}$ vs p_T^{lead} , $|y| < 2.5$, incl.

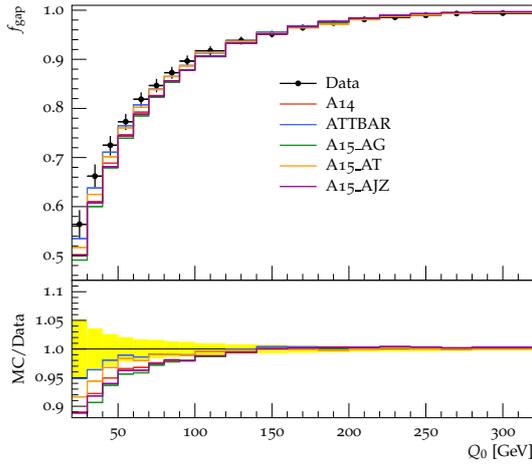


(e) Jet mass for anti- k_i jets, $200 < p_T/\text{GeV} < 300$ (f) τ_{21} for C/A jets, $200 < p_T/\text{GeV} < 300$

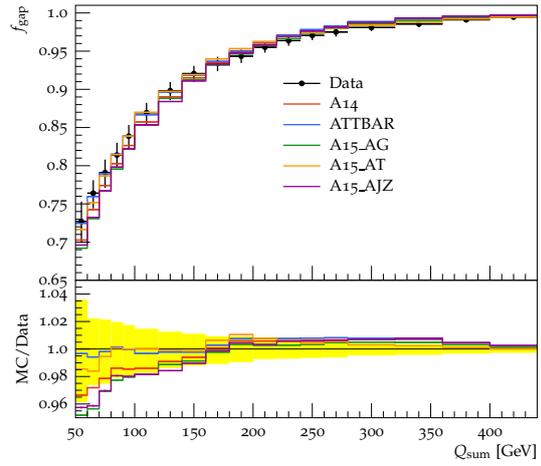
Figure D.4.: Comparison plots of A15_AG using global recoil or local recoil compared with ATLAS dijet 3-to-2 jet ratio, leading jet UE and jet substructure distributions. The yellow shaded areas represent data uncertainty.

E. Other dedicated tuning plots

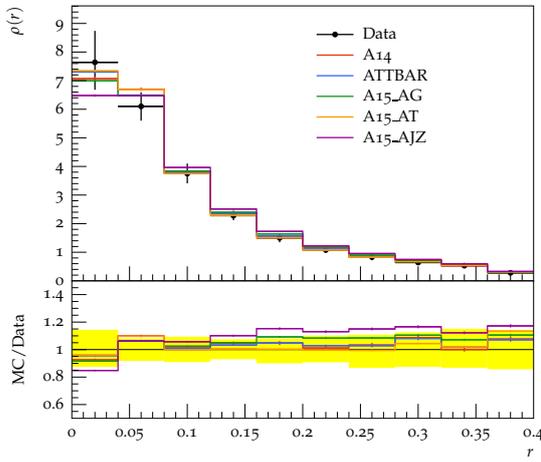
E. Other dedicated tuning plots



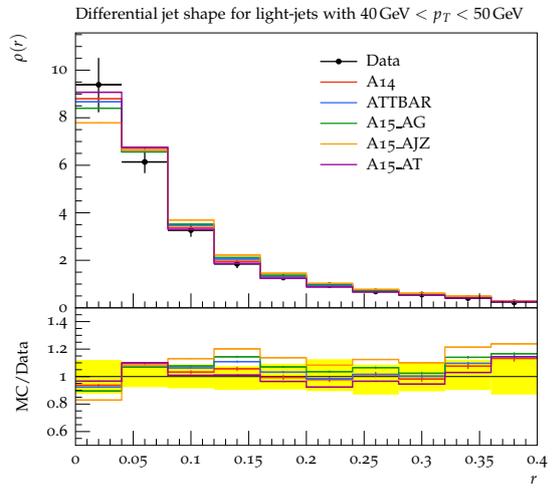
(a) $t\bar{t}$ gap fraction vs. Q_0 in $|y| < 2.1$



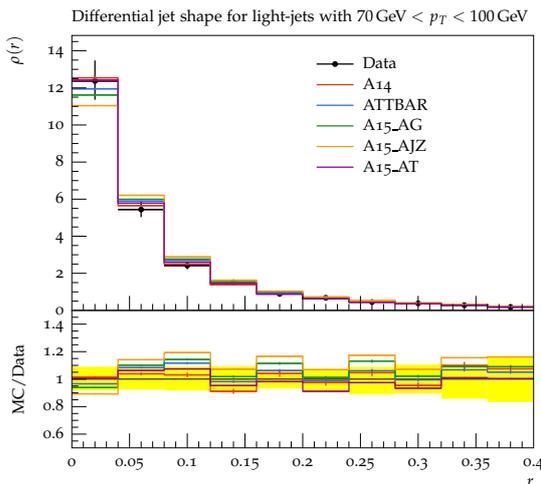
(b) $t\bar{t}$ gap fraction vs. Q_{sum} in $|y| < 2.1$



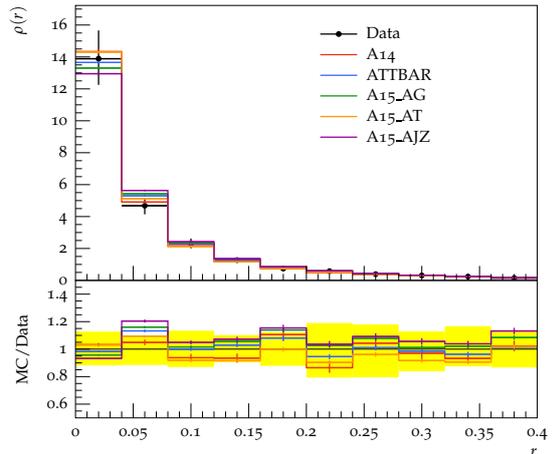
(c) $t\bar{t}$ light-jet shape for $30 < p_T/\text{GeV} < 40$



(d) $t\bar{t}$ light-jet shape for $40 < p_T/\text{GeV} < 50$

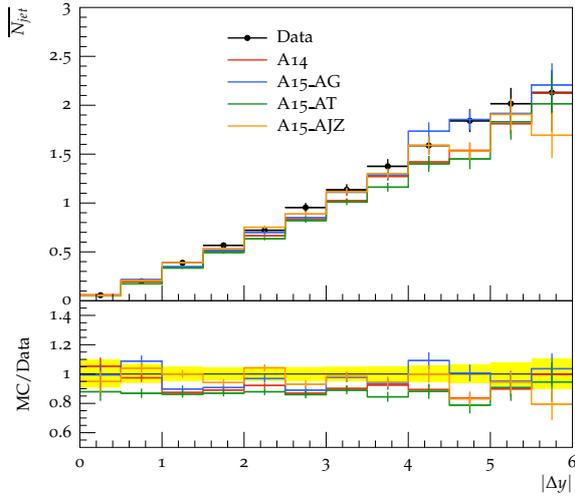


(e) $t\bar{t}$ light-jet shape for $70 < p_T/\text{GeV} < 100$

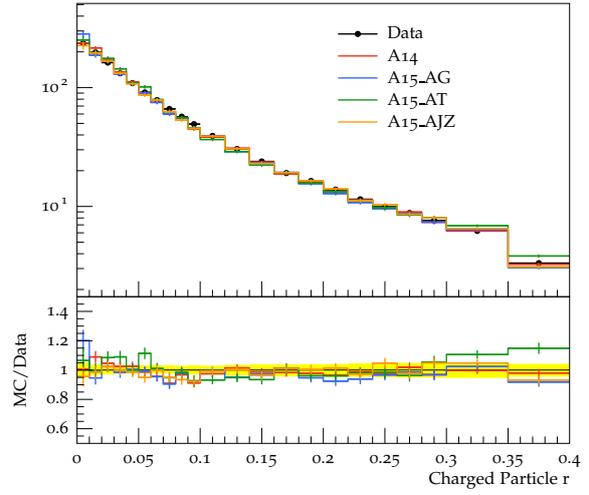


(f) $t\bar{t}$ light-jet shape for $100 < p_T/\text{GeV} < 150$

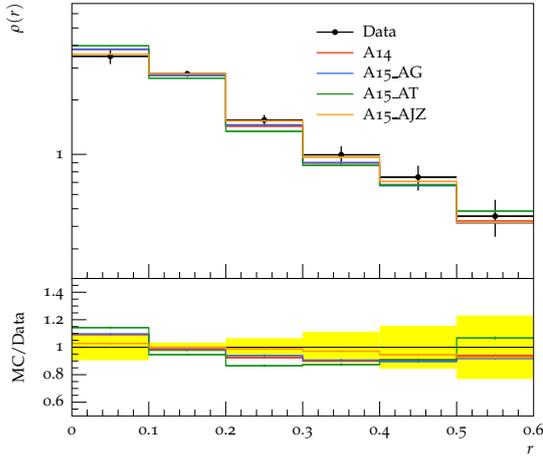
Figure E.1.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (A14, ATTBAR, A15_AG, A15_AT, A15_AJZ) compared with ATLAS $t\bar{t}$ gap fraction and jet shape distributions. The yellow shaded areas represent data uncertainty.



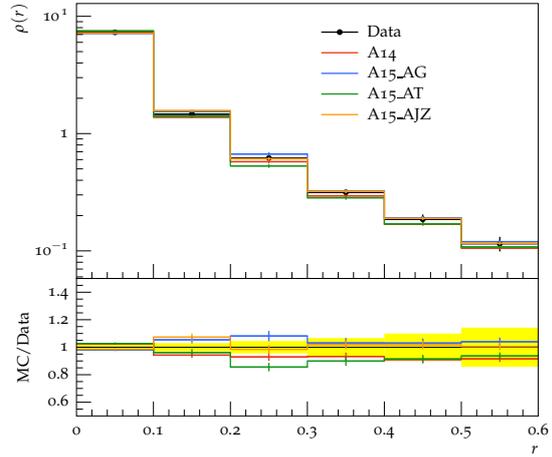
(a) $\overline{N}_{\text{jets}}$ vs $|\Delta y|$ for $120 < p_T / \text{GeV} < 150$, Fwd/Bwd



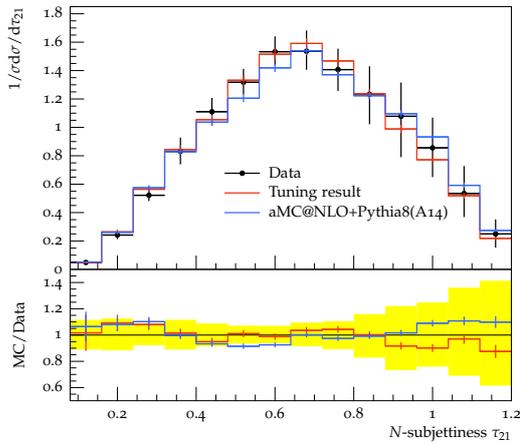
(b) Density $\rho_{\text{ch}}(r)$, $24 < p_T / \text{GeV} < 40$, $|y| < 1.9$



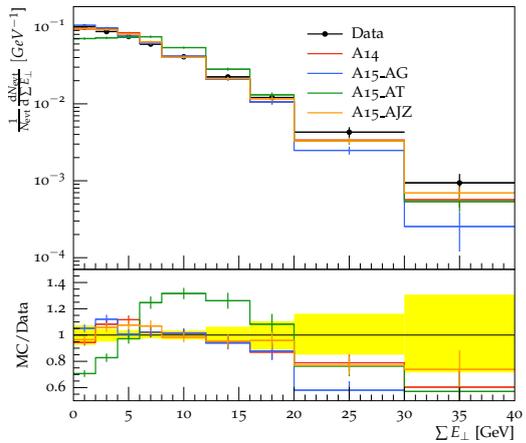
(c) Jet shape $\rho(r)$ for $30 < p_T / \text{GeV} < 40$



(d) Jet shape $\rho(r)$ for $260 < p_T / \text{GeV} < 310$



(e) τ_{21} for anti- k_t jets, $200 < p_T / \text{GeV} < 300$



(f) $\sum E_T$ in $|\eta| < 0.8$

Figure E.2.: The MC predictions of MADGRAPH5_aMC@NLO+PYTHIA8 (A14, A15_AG, A15_AT, A15_AJZ) compared with ATLAS dijet E_T flow, charged particle density, jet shape, jet substructure and gap fraction distributions. The yellow shaded areas represent data uncertainty. 91

F. Parameter studies with POWHEG+Pythia8

As was shown in section 3.1, POWHEG+PYTHIA8 using the A14 tune and $h_{\text{damp}} = \infty$ does not agree with ATLAS 7 TeV data, while PYTHIA8 stand-alone provides the best prediction of the data of the gap fraction distribution. To bring POWHEG+PYTHIA8 closer to PYTHIA8, the parameter *POWHEG:pTdef* can be set such that the initial and final state radiation use the p_T definitions from PYTHIA. Fig. F.1 and F.2 show the result of this parameter change compared to PYTHIA8 stand-alone. As can be seen, this has only a tiny effect and does not reduce the deviation from data significantly. The proper change of the h_{damp} parameter has a much larger effect and h_{damp} has to be set to m_{top} for the tuning to give better agreement with data.

F. Parameter studies with POWHEG+Pythia8

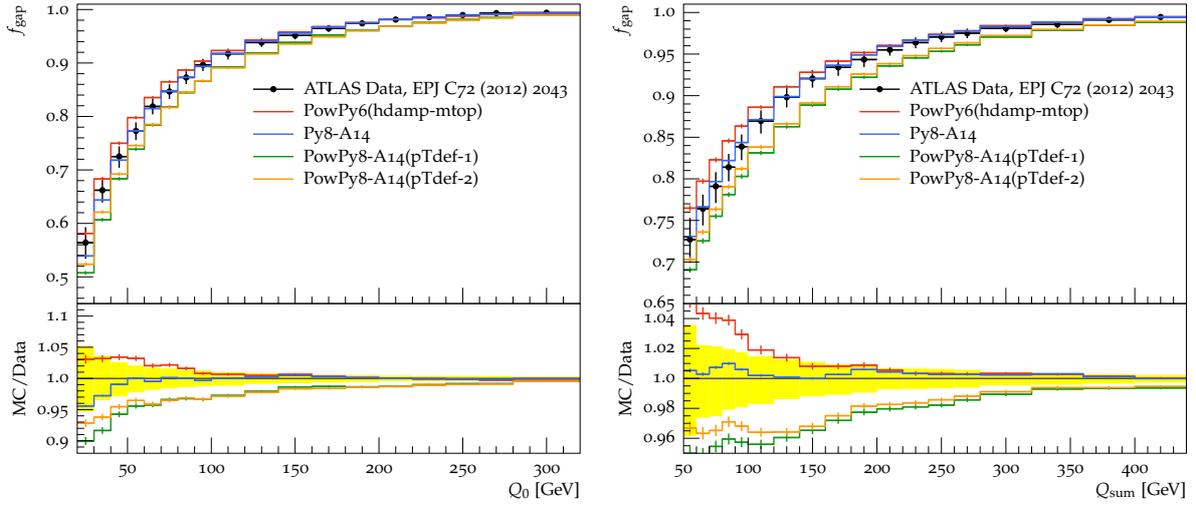


Figure F.1.: Gap fractions are plotted vs. Q_0 (left) and vs. Q_{sum} (right) for a rapidity region of $|y| < 2.1$.

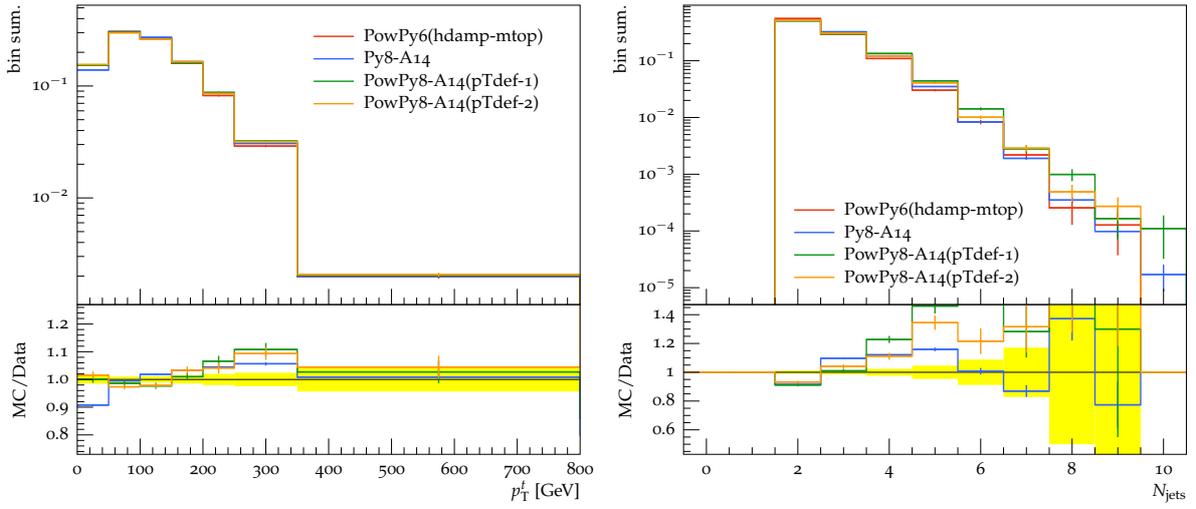


Figure F.2.: Top p_T (left) and jet multiplicity (right).

Bibliography

- [1] J. Alwall, et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **1407**, 079 (2014)
- [2] T. Sjostrand, S. Mrenna, P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05**, 026 (2006)
- [3] T. Sjostrand, S. Mrenna, P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput.Phys.Commun. **178**, 852 (2008)
- [4] The ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva (2014)
- [5] A. Buckley, et al., *Rivet user manual*, Comput.Phys.Commun. **184**, 2803 (2013)
- [6] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716**, 1 (2012)
- [7] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716**, 30 (2012)
- [8] M. Thomson, *Modern Particle Physics*, Cambridge University Press (2013)
- [9] K. Olive, et al. (Particle Data Group), *Review of Particle Physics*, Chin.Phys. **C38**, 090001 (2014)
- [10] LHC Higgs Cross Section Working Group (2013)
- [11] A. Bredenstein, et al., *NLO QCD Corrections to Top Anti-Top Bottom Anti-Bottom Production at the LHC: 2. full hadronic results*, JHEP **1003**, 021 (2010)
- [12] R. K. Ellis, W. J. Stirling, B. R. Webber, *QCD and Collider Physics*, Cambridge University Press (2003)

Bibliography

- [13] The D0 Collaboration, *Observation of the top quark*, Phys.Rev.Lett. **74**, 2632 (1995)
- [14] The CDF Collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, Phys.Rev.Lett. **74**, 2626 (1995)
- [15] The ATLAS, CDF, CMS and D0 Collaborations, *First combination of Tevatron and LHC measurements of the top-quark mass (ATLAS-CONF-2014-008, CDF-NOTE-11071, CMS-PAS-TOP-13-014, D0-NOTE-6416, FERMILAB-TM-2582-E)* (2014), arXiv:1403.4427 [hep-ex]
- [16] The CDF and D0 Collaborations, *Combination of measurements of the top-quark pair production cross section from the Tevatron Collider*, Phys.Rev. **D89(7)**, 072001 (2014)
- [17] The CMS Collaboration, *Combination of ATLAS and CMS top-quark pair cross section measurements using proton-proton collisions at $\sqrt{s} = 7$ TeV (CMS-PAS-TOP-12-003)* (2013)
- [18] The ATLAS Collaboration, *Combination of ATLAS and CMS top-quark pair cross-section measurements using proton-proton collisions at $\sqrt{s} = 8$ TeV (ATLAS-CONF-2014-054)* (2014)
- [19] L. Evans, P. Bryant, *LHC Machine*, JINST **3(08)**, S08001 (2008)
- [20] The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3(08)**, S08003 (2008)
- [21] The ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19, CERN, Geneva (2010)
- [22] T. Gleisberg, et al., *Event generation with SHERPA 1.1*, JHEP **0902**, 007 (2009)
- [23] M. Bahr, et al., *Herwig++ Physics and Manual*, Eur.Phys.J. **C58**, 639 (2008)
- [24] A. Buckley, et al., *Systematic event generator tuning for the LHC*, Eur.Phys.J. **C65**, 331 (2010)
- [25] The ATLAS Collaboration, *Underlying event characteristics and their dependence on jet size of charged-particle jet events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys.Rev. **D86**, 072004 (2012)

- [26] The ATLAS Collaboration, *Measurement of the underlying event in jet events from 7 TeV proton-proton collisions with the ATLAS detector*, Eur.Phys.J. **C74(8)**, 2965 (2014)
- [27] The ATLAS Collaboration, *Properties of jets measured from tracks in proton-proton collisions at center-of-mass energy $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys.Rev. **D84**, 054001 (2011)
- [28] The ATLAS Collaboration, *Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment*, JHEP **1205**, 128 (2012)
- [29] The ATLAS Collaboration, *Study of Jet Shapes in Inclusive Jet Production in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector*, Phys.Rev. **D83**, 052003 (2011)
- [30] The ATLAS Collaboration, *Measurement of jet shapes in top-quark pair events at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Eur.Phys.J. **C73(12)**, 2676 (2013)
- [31] The ATLAS Collaboration, *Measurement of Dijet Azimuthal Decorrelations in pp Collisions at $\sqrt{s} = 7$ TeV*, Phys.Rev.Lett. **106**, 172002 (2011)
- [32] The ATLAS Collaboration, *Measurement of $t\bar{t}$ production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Eur.Phys.J. **C72**, 2043 (2012)
- [33] The ATLAS Collaboration, *Measurement of multi-jet cross sections in proton-proton collisions at a 7 TeV center-of-mass energy*, Eur.Phys.J. **C71**, 1763 (2011)
- [34] The ATLAS Collaboration, *Measurement of the transverse momentum distribution of Z/gamma* bosons in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys.Lett. **B705**, 415 (2011)
- [35] The ATLAS Collaboration, *Measurement of the Z/gamma* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **1409**, 145 (2014)
- [36] J. R. Andersen, et al., *Les Houches 2013: Physics at TeV Colliders: Standard Model Working Group Report* (2014), [arXiv:1405.1067](https://arxiv.org/abs/1405.1067) [hep-ph]
- [37] S. Carrazza, S. Forte, J. Rojo, *Parton Distributions and Event Generators*, in *Proceedings, 43rd International Symposium on Multiparticle Dynamics (ISMD 13)* (2013), [arXiv:1311.5887](https://arxiv.org/abs/1311.5887) [hep-ph]

Bibliography

- [38] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys.Rev. **D82**, 074018 (2010)
- [39] The ATLAS Collaboration, *Comparison of Monte Carlo generator predictions to ATLAS measurements of top pair production at 7 TeV*, Technical Report ATL-PHYS-PUB-2015-002, CERN, Geneva (2015)
- [40] C. Oleari, *The POWHEG-BOX*, Nucl.Phys.Proc.Suppl. **205-206**, 36 (2010)
- [41] The ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section as a function of jet multiplicity and jet transverse momentum in 7 TeV proton-proton collisions with the ATLAS detector*, JHEP **01**, 020 (2015)
- [42] M. Cacciari, G. P. Salam, G. Soyez, *The Anti- $k(t)$ jet clustering algorithm*, JHEP **0804**, 063 (2008)
- [43] R. Corke, T. Sjostrand, *Improved Parton Showers at Large Transverse Momenta*, Eur.Phys.J. **C69**, 1 (2010)
- [44] J. Pumplin, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07**, 012 (2002)
- [45] The ATLAS Collaboration, *Measurement of angular correlations in Drell-Yan lepton pairs to probe Z/γ^* boson transverse momentum at $\sqrt{s}=7$ TeV with the ATLAS detector*, Phys. Lett. **B720**, 32 (2013)
- [46] The ATLAS Collaboration, *Measurement of distributions sensitive to the underlying event in inclusive Z-boson production in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. **C74(12)**, 3195 (2014)
- [47] The ATLAS Collaboration, *Measurement of dijet production with a veto on additional central jet activity in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, JHEP **09**, 053 (2011)
- [48] The ATLAS Collaboration, *Measurements of the pseudorapidity dependence of the total transverse energy in proton-proton collisions at $\sqrt{s} = 7$ TeV with ATLAS*, JHEP **11**, 033 (2012)
- [49] Y. L. Dokshitzer, et al., *Better jet clustering algorithms*, JHEP **08**, 001 (1997)
- [50] R. Corke, T. Sjostrand, *Interleaved Parton Showers and Tuning Prospects*, JHEP **03**, 032 (2011)

- [51] P. Skands, S. Carrazza, J. Rojo, *Tuning PYTHIA 8.1: the Monash 2013 Tune*, Eur. Phys. J. **C74(8)**, 3024 (2014)

- [52] The ATLAS Collaboration, *A study of the sensitivity to the Pythia8 parton shower parameters of $t\bar{t}$ production measurements in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS experiment at the LHC*, Technical Report ATL-PHYS-PUB-2015-007, CERN, Geneva (2015)

Acknowledgement

I want to thank everyone who has supported me in this past year that I have worked on my master's thesis. I want to thank my family and friends for believing in me and taking away all my worries. I also want to thank all my colleagues from the II. institute of Physics at the university of Göttingen, you are the kindest and most helpful people I had the pleasure to meet.

Special thanks goes out to my supervisors Prof. Dr. Arnulf Quadt and Dr. Maria Moreno Llacer. If it was not for their unyielding care, support and commitment, I would neither possess the knowledge and skills that I do now, nor would I be as involved in the ATLAS experiment.

Erklärung

nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen:

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 28. April 2016

(Jannik Geisen)