Tuning of parameters sensitive to colour reconnection in the Pythia8 event generator using top-quark-antiquark pair production events recorded with the ATLAS detector

Master's Thesis

Maximilian Felix Caspar

BERGISCHE
UNIVERSITÄT
WUPPERTAL

BERGISCHE UNIVERSITÄT WUPPERTAL

Fakultät Mathematik und Naturwissenschaften
Fachgruppe Physik

Arbeitsgruppe Experimentelle Elementarteilchenphysik

Tuning of parameters sensitive to colour reconnection in the Pythia8 event generator using top-quark-antiquark pair production events recorded with the ATLAS detector

Thesis zur Erlangung des akademischen Grades
Master of Science (M.Sc.)
im Studiengang Physik
der Bergischen Universität Wuppertal

vorgelegt von

Maximilian Felix Caspar
1521165
Studiengang Physik

Wuppertal, den 18. Juni 2021

Erstprüfer:	Dr. Dominic Hirschbühl
Zweitprüfer:	Prof. Dr. Wolfgang Wagner

Kurzfassung

In dieser Masterarbeit wird die Anpassung von Monte-Carlo Modellen an experimentelle Daten durchgeführt, welche die Neuorientierung der QCD-Farbinteraktionen (”Colour Reconnection”) in einem Event beschreiben. Hierzu werden neue Techniken zur Evaluation der Sensitivität und der Umschließung der Referenzdaten erarbeitet und angewendet. Außerdem wird eine Anpassung der Gewichtung einzelner Observablen zur Reduktion des Einflusses redundanter Observablen untersucht. Die drei resultierenden ”Tunes” sind in Tabelle 1 dargestellt. Diese werden in der Arbeit anhand ihres Beitrags zur Anpassungsgüte χ^2 verglichen. Zusätzlich wird ein Vergleich zum verbreiteten A14-Tune vorgenommen. Die unter Verwendung der produzierten Tunes rekonstruierten Top- und W-Massen werden mit dem A14-Tune verglichen.

Abstract

In this master thesis, Monte Carlo models describing the reorientation of QCD colour interactions (”colour reconnection”) in an event are fitted to experimental data taken at ATLAS. For this purpose, new techniques for evaluating the sensitivity and envelopment of the reference data are developed and applied. Furthermore, an adjustment of the weighting of individual observables to reduce the influence of redundant observables is investigated. The three resulting ”tunes” are presented in Table 1. These are compared in the thesis on the basis of their contribution to the goodness-of-fit χ^2. In addition, a comparison is made to the widely used A14 tune. The masses of top quarks and W bosons reconstructed using the produced tunes are compared to A14.

Table 1: The three parameter tunes for CR2 achieved in this thesis.

Tune	Naive	Selected	Reweighted
χ^2/N.D.F.	25.5	16.3	18.3
MPI:pT0Ref	2.30	2.23	2.35
MPI:expPow	1.75	2.07	1.66
CR:m2Lambda	2.81	0.32	0.32
CR:fracGluon	1.00	0.78	0.82
CR:dLambdaCut	1.86	1.05	0.88

Danksagung

Ich danke den Mitgliedern der ATLAS-Gruppe der Bergischen Universität Wuppertal, im Besonderen Dr. Dominic Hirschbühl und Shayma Wahdan, für ihre Unterstützung in der Forschungsphase dieses Projektes. Herrn Dr. Holger Schulz von der Durham University danke ich für die Beantwortung zahlreicher Fragen zu der von ihm geschriebenen Apprentice-Software. Des Weiteren möchte ich Allen danken, die meine Arbeit korrekturgelesen haben, insbesondere meinen Kommilitonen Marvin Geyik und Johanna Kraus, und Dr. Mark McDonald von Whale Acoustics Inc.

Besondere Erwähnung soll an dieser Stelle Johanna Kraus dafür finden, dass sie das Kunstwerk auf der Titelseite der Arbeit gestaltet hat. Mein besonderer Dank gilt meinen Eltern, die mir dieses Studium ermöglicht und mich immer unterstützt haben und meiner Frau Lisa, die mich stets bekräftigt und moralisch unterstützt hat.

Contents

1

Introduction

During all of documented history, man has wondered about the composition and underlying laws of our universe. In the last century, our understanding of the world we find ourselves in has skyrocketed via the discovery of concepts like quantum mechanics, thermodynamics and quantum field theory. Today, even though many questions remain unanswered, we possess a powerfully predictive tool: The Standard Model (SM) of particle physics. It describes all constituents of matter known presently and their interactions. However powerful, we know the SM to be incomplete at best. Effects like dark matter or gravity are beyond its scope. Additionally, it includes a set of free parameters that cannot be inferred from first principle. As a result, the search for a "theory of everything" continues, representing the holy grail of theoretical physics.

In the meantime, many unanswered questions within the SM have to be addressed as well, one of which will be the topic of this thesis. In the SM there exists a set of particles that interact via three different charges capable of mixing or neutralising. These particles, called "quarks" and "gluons", represent the constituents of the atomic nuclei of all elements of the periodic table. Their interaction is described by a theory called Quantum Chromodynamics (QCD) and the quarks exhibit behaviour that seems kind of odd when compared to other physical systems. The strength of their interaction increases when they are pulled apart, to the point where new particles come into existence to ease the tension of this separation. QCD is very well understood at high energies via a perturbative treatment, but low energy or "soft" QCD remains inaccessible for now. Exactly this low energy QCD is required to describe many of the phenomena we observe in the contemporary high energy experiments at the Large Hadron Collider (LHC). In the current state of research, we need to compromise with phenomenological models that approximate the behaviour of such systems. However, the problem with models that are not derived from first principles is that they are always based on assumptions. Only comparisons to data can reveal which models describe a given area of physics well and multiple free parameters have to be determined. In this thesis, we will focus on models describing how different parts of a collision event at LHC rearrange themselves under the influence of QCD. This process, called "Colour Reconnection" (CR), is crucial for accurately modelling the entirety of the collision and therefore allowing for precise measurements of other quantities. We will describe ways of fixing the free parameters of such models in a process called "tuning" and introduce improvements to the procedures used in the past. In the end, we will arrive at parameter recommendations for a model of colour reconnection and compare its performance to data gathered at the LHC.

In this chapter, all fundamental principles are introduced. We start with a brief introduction to the LHC and the ATLAS detector. Next, all SM interactions with their associated particles are described and the processes relevant to this work are presented. Ultimately, the concept of Monte Carlo (MC) event generators and their tuning is discussed. Chapter 2 focusses on the concept of colour reconnection, the model under study and alternative models available in the studied MC generator. In Chapter 3, all physics analyses used in the MC tuning efforts are briefly presented. Chapter 4 introduces a new method of determining the sensitivity of MC observables to changes in input parameters. Comparisons between different observables of the "CR2" model are included. The following Chapter 5 includes tunes for the CR2 model using different approaches and compares their quality. Chapter 6 summarises the results of this work and attempts to provide an outlook on what still needs to be done.

1.1. LHC and the ATLAS Detector

The study of subatomic particles and their characteristics can seem quite counterintuitive: To study smaller and smaller systems, we need to build larger and larger accelerators. One can explain this by looking at these systems not in units of size, but in units of energy. Only in recent years have we been able to access center-of-mass energies of $\mathcal{O}$(TeV). This allows us to examine the high-energy frontier by using the largest machine ever built in human history: the Large Hadron Collider.

1.1.1. The Large Hadron Collider

The Large Hadron Collider is a synchrotron in collider configuration with an unprecedented circumference of $\sim$ 27 km [1]. It is located at the European Organisation for Nuclear Research (CERN) in Geneva and designed to reach center-of-mass energies of up to 14 TeV for two colliding beams of protons. It features four collision points at which the main experiments ATLAS, CMS, LHCb and ALICE are located. While ALICE is designed to examine the collision of heavy lead ions and LHCb specialises in the physics of the bottom quark, ATLAS and CMS are multipurpose detectors for pp collisions. Protons enter the LHC in bunches of approximately 10^{11} particles each and collide at the interaction points at a rate of once per 25 ns. However, before entering the main ring, particles are accelerated in several pre-accelerators since the main ring is optimised for injection energies of around 450 GeV. A schematic of the entire accelerator-configuration is provided in Figure 1.1. A particle collider is characterised by its center-of-mass energy and its luminosity $\mathcal{L}$. It measures the number of interactions occurring per unit of time and area. Together with the cross section σ, a specific measure for the probability of a particular process occurring, the integrated luminosity allows for a prediction of the aggregate number of events that have been measured in a given period of time

$$N = \sigma \int \mathcal{L} dt.$$

As of now the LHC has delivered 189 fb^{-1} since its first physics run[1] and is currently undergoing a longer shutdown [3].

1.1.2. The ATLAS detector

ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) [4] represents the largest experiment at the LHC. It was constructed to allow for precise measurements of SM physics as well as searches for physics

[1] A "barn" is a unit of measurement for the cross sectional area of subatomic reactions. 1 fb $\equiv 1 \times 10^{-15}$ b $\equiv 1 \times 10^{-39}$ cm^2

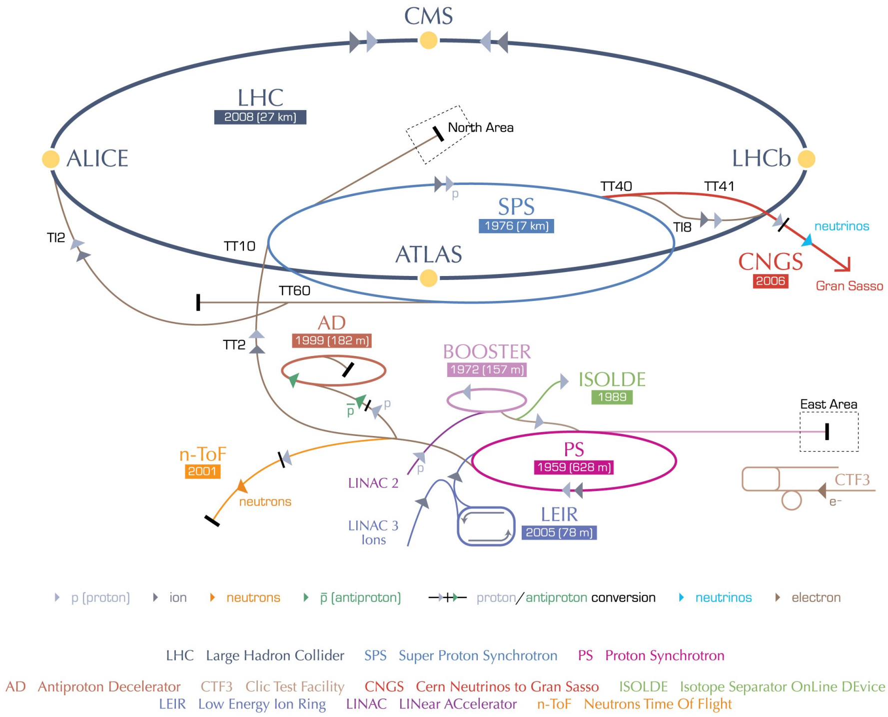


Figure 1.1: Schematic overview of the entire CERN accelerator complex [2].

beyond the Standard Model. The detector is laid out in a cylindrical shape with various subdetectors arranged in layers around the central point of interaction. It weighs over 7000 t, occupies a length of 46 m and is 25 m high. The innermost layer is responsible for tracking charged particles immediately after their collision. It is immersed in a solenoid magnetic field, which allows for the charge and also the type of particle to be determined. Around it, a calorimeter is placed allowing the energy of the collision products to be measured. A muon spectrometer is mounted on the very outside inside of a toroidal magnetic field. All of these components will be further discussed shortly below. An overview over ATLAS and all of its subdetectors is given in Figure 1.2.

ATLAS Coordinate System

The basic coordinate system used within ATLAS is a right handed Cartesian system originating in the interaction point (IP), in the middle of the detector. The x-axis points from the IP into the middle of the LHC while the y-axis points upwards into the sky. Therefore, the z-axis is directed along the direction of the beamline. One very important measure is the transverse momentum $\vec{p}_T$ of a particle, which represents the projection of its momentum vector onto the x-y-plane. It is invariant under a Lorentz boost in the z-direction. An alternative cylindrical coordinate system (z, r, ϕ) can be defined, where ϕ is the angle between the x-axis and transverse momentum vector of a particle. The polar angle θ of a particle is defined as the angle between momentum vector and beam-axis. The pseudorapidity makes use of this angle via

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

This expression is zero for particles traversing the detector volume outwards with no longitudinal momentum p_z and gets arbitrarily large for particles flying in the direction of the beam. An angular

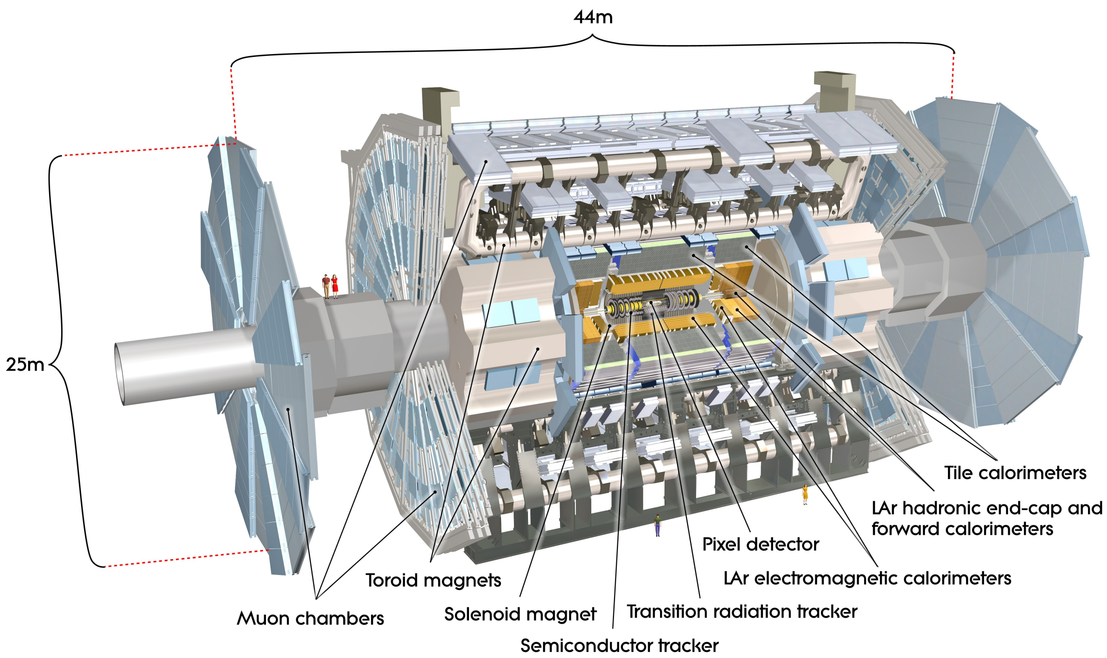


Figure 1.2: A complete overview of the ATLAS detector and all of its subdetectors [5].

distance between two particles is thus defined as

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

The Inner Detektor

Embedded in a solenoidal magnetic field of 2 T, the Inner Detector (ID) is the part of the ATLAS detector closest to the beam pipe. It consists of the innermost pixel-detector, a silicon microstrip detector and a transition radiation detector. These work together to track the movement of charged particles and determine their momentum and charge via the curvature of their paths. In addition, electron/pion separation is performed by the transition radiation detector. Since the pixel-detector is extremely close to the interaction point, it also allows for the identification of jets originating from a b-quark in a process called “b-tagging” (more on that in Section 1.3.1).

Calorimeters

The calorimeter section aims at measuring the energy of outgoing particles and jets. It subdivides into an electromagnetic section on the inside, and an outer hadronic section. The electromagnetic (EM) calorimeter is built as a sampling calorimeter with liquid argon as the active material. Lead and stainless steel is placed between the sampling layers as an absorber, optimised for the measurement of electrons and photons. The hadronic section is optimised to absorb particles that interact via the strong force and passed through the EM calorimeter. While the forward-regions are also built using liquid argon sampling layers, but with copper and tungsten in between, the main “barrel” part of this calorimeter comprises scintillating sampling layers with steel absorbers.

Muon Chambers
Due to their larger mass compared to electrons, muons are able to pass the calorimeter section with minimal bremsstrahlung loss. As not much energy is deposited in the inner layers, the outermost subdetectors of the ATLAS detector are dedicated to identifying and characterising muons. Enveloped in a toroidal magnetic field, 1200 muon chambers measure the tracks of traversing particles. In addition, a set of timing chambers are installed to allow for precisely timed selection of events containing muons.

1.2. The Standard Model of Particle Physics

Currently, there are four known fundamental forces which mediate all interactions between particles. Gravity represents the fundamental force giving rise to the large-scale structure of our universe. It affects all particles according to their mass and is always attractive, therefore leading to the formation of solar systems, galaxies and clusters. While the interaction of large masses is sufficiently described by the general theory of relativity, a theory of quantum gravity is still an active area of research.
Matter we encounter in our day-to-day lives is made up of atoms, which consist of a positively charged nucleus and negatively charged electrons that are bound to their nucleus by the electromagnetic interaction. The fundamental theory describing the electromagnetic interaction is called Quantum Electrodynamics (QED). The atomic nucleus itself is a compound system of positively charged protons and neutral neutrons. They are bound together by the strong nuclear force that only manifests itself on very short scales of a few fm. This allows for stable elements to exist despite the electromagnetic repulsion of the protons. Fundamentally, the strong nuclear interaction is described by the theory of Quantum Chromodynamics. While the proton as a particle has so far been observed to be completely stable, the neutron on the other hand is not. It decays by emission of an electron and a neutrino into a proton. This is conducted via a weak force which can change the type of a particle, also called "flavour". As it turns out, QED and weak interactions can be viewed as results from the same force described by the Electroweak Theory (EWT). This "unification" of forces remains an ultimate goal of theoretical particle physics. A potential combination of EWT with QCD is referred to as a Grand Unified Theory (GUT), while the unification of all four fundamental forces is called a "theory of everything" (TOE). A TOE is considered the "holy grail" of modern physics.

As of today, the most comprehensive description of fundamental interactions is the Standard Model of particle physics. It combines electroweak and strong interactions and their associated particles with the Higgs mechanism and a Yukawa interaction sector[2] to construct an incredibly effective theory that is capable of explaining most observations made by today's high-energy experiments. The model, however, does not include effects like neutrino oscillations. Also, it has 19 free parameters and is therefore considered more of an ad hoc model than a unified field theory. A comprehensive and in-depth introduction into SM physics is provided in [6], which this section is loosely based upon.

1.2.1. Fundamental particles

As of now, a total of 17 types of fundamental particles have been discovered. They can be divided into two categories: fermions and bosons, as presented below. A graphical overview with masses, spin and electric charges of the particles is provided in Figure 1.3.

[2] The Yukawa interaction of the SM describes the coupling between the Higgs-field and the massless quark and lepton fields.

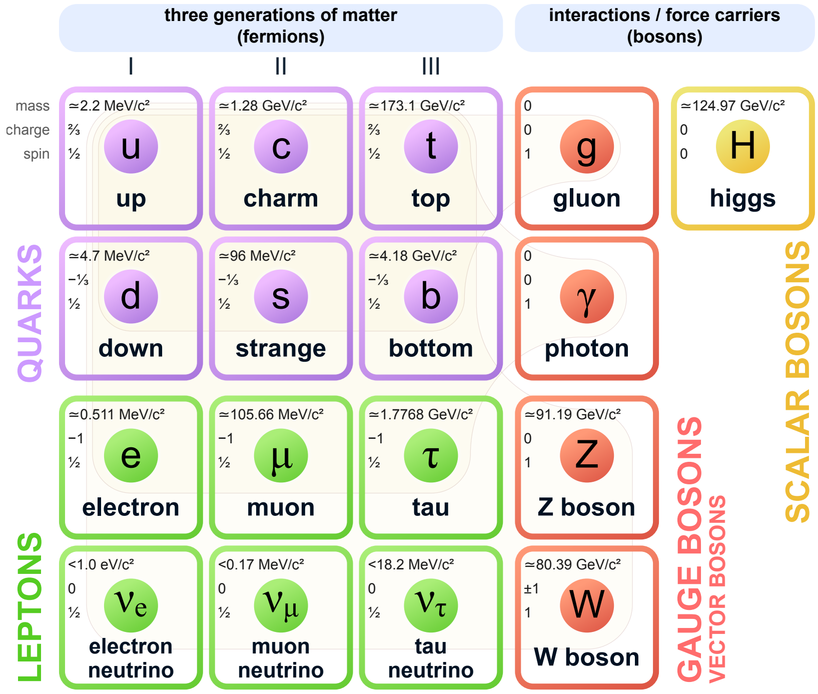


Figure 1.3: All elementary particles included in the SM with their masses, electric charges and spin values [7].

Fermions

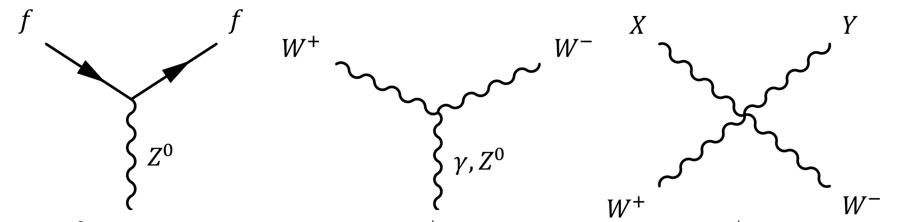
Fermions are massive particles with a half integer spin (½ , 3/2 , ...). As such they obey the Pauli exclusion principle, which says that in a quantum system, a quantum state can solely be occupied by a single fermion. This gives rise to effects like the atomic spectrum or the band structure of semiconductors. For each elementary SM fermion, there exists a corresponding antiparticle with opposing charges and otherwise identical properties. They subdivide into quarks and leptons. There are three families of leptons, each consisting of a charged lepton and a corresponding chargeless neutrino. The charged leptons are subject to both the electromagnetic and the weak interaction while the neutrinos can only interact weakly. In SM interactions, the number of leptons per family are always conserved (particles count as 1, antiparticles as -1). Quarks are electrically charged particles that interact with each other via the strong, the weak and the electromagnetic interaction. As with leptons, three families of quarks are known as of now. Each family contains a particle with a positive ⅔ charge and a counterpart with an electric charge of -⅓.

Bosons

Bosons represent particles with integer spin. When two elementary particles interact with each other via one of the fundamental interactions, they do so via the exchange of a gauge boson. All five types of SM gauge bosons have a spin of one. While gauge bosons are fundamentally massless, the Z and W bosons of the weak interaction are massive particles due to the Higgs mechanism. While a comprehensive description of the mechanism is beyond the scope of this thesis and can be found in [6], it is at least useful to know the Higgs mechanism gives rise to the Higgs particle as an additional spin zero boson. The Higgs boson has been successfully discovered by both ATLAS and CMS [8, 9] and therefore further justifies the inclusion of the Higgs mechanism in the SM .

Photons The photon (γ) is the neutral gauge boson of the electromagnetic interaction. It couples to any particle with an electric charge including other charged bosons, is massless and stable. QED has only one fundamental vertex, the corresponding Feynman diagram can be found in Figure A.1 along with example QED processes.

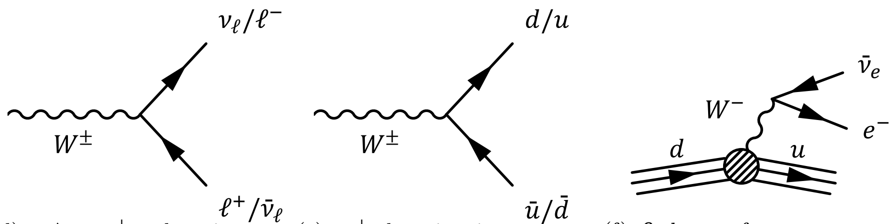
W and Z bosons The weak interaction is mediated by W and Z bosons. While the W bosons possess either a positive or negative charge of 1, the Z boson represents a neutral particle. W bosons in particular couple to fermions and their corresponding neutrinos and are also able to change a quarks flavour. For example, a W^- being emmited by a d-quark is responsible for the β^--decay of the neutron. For the relevant fundamental vertices and an example diagram for a β-decay, refer to Figure 1.4.



(a) A Z^0 interacting with any fermion f.

(b) Two $W^{\pm}$ interacting with a γ or a Z^0.

(c) Two $W^{\pm}$ interacting with any two electroweak bosons X, Y such that electric charge is conserved.



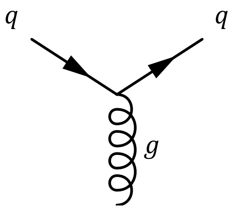
(d) A $W^{\pm}$ decaying into any (anti-)lepton and the appropriate (anti-)neutrino.

(e) $W^{\pm}$ decaying into an up-type (anti-)quark u and a down-type (anti-)quark d.

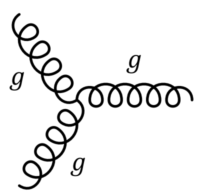
(f) β decay of a neutron (ddu) into a proton (udu) via a W^- particle.

Figure 1.4: Fundamental weak vertices and an example process. Note that all interactions conserve electric charge and lepton family numbers.

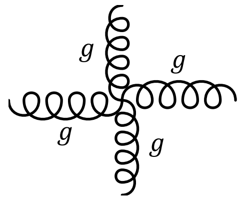
Gluons Unlike the other gauge bosons, gluons transport more complex “colour” - charges (three colours and anti-colours). A fundamental property of the underlying QCD theory (discussed further in [6, Chapter 10.4]) is that coloured particles cannot be observed in nature on their own, all QCD systems are therefore of neutral colour. The “colour singlet” state $(r\bar{r} + b\bar{b} + g\bar{g})/\sqrt{3}$ cannot exist in a single gluon, since the strong interaction is observed to be short ranged. This leaves a set of eight possible gluon states, which are referred to as the “colour octet”. Gluons can interact not only with quarks, but also with other gluons. All fundamental QCD Feynman diagram vertices are shown in Figure 1.5.



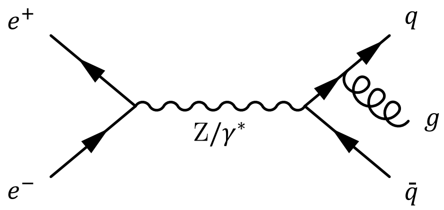
(a) A Gluon g interacting with any (anti-)quark q.



(b) Interaction of three gluons.



(c) Interaction of four gluons.



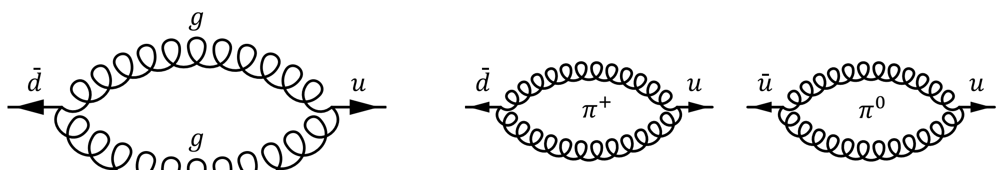
(d) Example of gluon bremsstrahlung in a quark pair production event, as observed by TASSO [10].

Figure 1.5: Fundamental QCD vertices and an example process.

Higgs boson Even though beyond the scope of this work, it is worth mentioning the Higgs boson for completeness. Since the gauge bosons of the electroweak interaction are not massless, the underlying gauge symmetry is broken. The introduction of the Higgs mechanism allows for preservation of local gauge invariance, making the SM a consistent theory. For a proper explanation of this mechanism, refer to [6]. The SM Higgs boson is uncharged and can interact with massive fermions and the intermediate massive vector bosons. Its discovery is remarkably compelling evidence for the Higgs mechanism and one of the most noteworthy successes of SM physics to date. The fundamental vertices and two dominant production processes of the Higgs boson are shown in Figure A.2

1.2.2. Hadrons and jet formation

Quarks have so far never been observed as singular stable particles, but only in bound hadronic states of two or three called “mesons” or “baryons”. While protons are (as observed until now) stable on their own, all other known hadrons are not. The leading explanation for this structure is the hypothesis of “colour confinement”. It states that only colour singlets can exist as free particles, believed to originate from gluons being able to interact with one another via the three-gluon and four-gluon vertices. At larger distances, the energy required to separate two quarks increases linearly with about $1\,\mathrm{GeV\,fm^{-1}}$. Two quarks rearranging into a colourless state therefore becomes more energetically favourable with increasing distance between them. When two quarks are drifting apart with sufficient kinetic energy, the potential energy of the string of gluons holding them together will at some distance be greater than the mass of a quark-antiquark pair. At this point, a pair will be produced from the vacuum and the string splits. Now two colourless mesons exist instead of one as illustrated in Figure 1.6. This process continues until the remaining kinetic energy of the “jet” of hadrons is no longer sufficient to produce more quark-antiquark combinations. Jet formation, also known as “hadronisation”, is still poorly understood from a QCD perspective. To recreate experimental data, we instead rely on highly phenomenological models which are presented later in this chapter.



(a) A "string" of gluons binding two quarks together into a π^+ meson.

(b) After the string has "snapped" due to separation, we end up with an additional π^0 meson.

Figure 1.6: Visualisation of how separation of quarks leads to the formation of more hadrons. Note that the "string" connecting the quarks can be much more complex due to three- and four-gluon vertices producing additional "junctions" in the string.

1.3. Processes under study

In this thesis, we will rely on several analyses from the ATLAS collaboration to include a variety of physical processes into our tunes of soft QCD models. The focus of each of these analyses is explained in Chapter 3. This section provides a brief introduction into the relevant hard processes analysed and explain their relevance to our tuning effort.

1.3.1. $t\bar{t}$

The top quark represents the most massive SM particle. Its remarkably short lifespan causes that there are no (known) bound states with other quarks. It decays after $\mathcal{O}(5 \times 10^{-25}\,\mathrm{s})$ or a maximum travel distance of $\mathcal{O}(10^{-16}\,\mathrm{m})$, even before a hadronisation process could have happened. This allows for the opportunity to probe a quark in a “bare” state, with no other effects from the binding forces of a hadron. It is predominantly produced in pairs of top and anti-top quarks, also referred to as $t\bar{t}$ or “TTBar”, via the conversion of highly energetic gluons that were created in the collision of two protons. Since the top quark must change its flavour to decay, the only known decay modes are associated with the weak interaction via $W^{\pm}$ bosons, leaving behind bottom-type quarks. While decays to s and d-quarks are also possible, they are suppressed by the CKM matrix elements, leaving the decay to bottom quarks as the dominant mode. The W bosons decay either leptonically into $\ell + \nu_\ell$ or hadronically into a quark pair, initiating the formation of a jet. Due to their short lifetimes, $\tau^{\pm}$ leptons decay almost immediately into other leptons or a quark-antiquark-pair. In this thesis, only decays to $e^{\pm}$ or $\mu^{\pm}$ are considered ”leptonic”. The bottom quarks left over from the decay of the top quarks allow for a direct way of identifying these events. Since the mass of the b-quark is lower than the top quark mass and other decays are suppressed via the CKM matrix, the b-quark is relatively stable ($\mathcal{O}(10^{-12}\,\mathrm{s})$). B-hadrons will therefore travel slightly further than other initiators of jets. Although not always unambiguous, this allows for identification of b-jets in a procedure called “b-tagging”. The number of b-tagged jets is therefore an important tool for identifying $t\bar{t}$ collisions. The decay modes of the W bosons allow for the production of zero, one or two leptons in their decay. Since there is inevitably an additional neutrino created for each lepton, the kinematics of these events are different due to the missing transverse momentum observed (because neutrinos go undetected). Therefore, the number of leptons is an essential selection criterion for most $t\bar{t}$ analyses. The decay modes of a $t\bar{t}$-pair are shown in Figure 1.7.

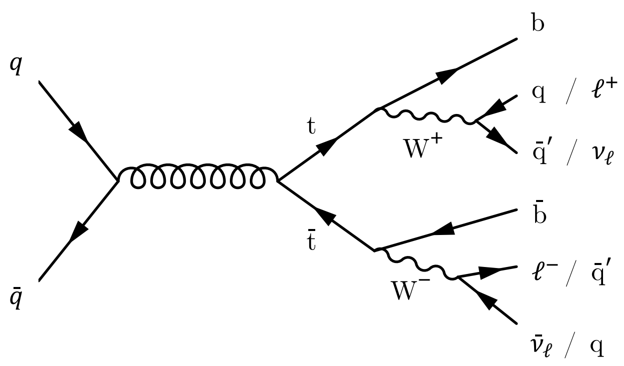


Figure 1.7: Possible decay modes for a $t\bar{t}$ pair. Due to the decay of the $W^{\pm}$ bosons, there are four distinct chanels: hadronic, ℓ^+, ℓ^- and $\ell^+\ell^-$.

1.3.2. $Z^0 \to \ell^+\ell^-$

In proton-proton scattering, it is possible for a quark-antiquark pair to annihilate into a Z^0-boson or photon, which in turn decays into a pair of charged leptons. This interaction represents a neutral current “Drell-Yan” process, named after Sidney Drell and Tung-Mow Yan [11] (shown in Figure 1.8). It can be mediated by either a photon or a Z boson, but in this thesis we will focus on events containing a Z boson only. Since electroweak bosons are unable to transport colour charge, it is possible to study hadronisation processes more precisely by selecting for Drell-Yan events and excluding the Z boson and its decay products. Since the rest of the event is unaffected by colour charges from the hard process, the UE is particularly easy to isolate from the rest of the event.

1.3.3. The underlying event

The underlying event (UE) is not so much a collision process as it is a concept helpful for understanding the structure of an event[3]. Most analysis efforts focus on a hard scattering process like for example $t\bar{t}$. However, the detector usually picks up a lot more particles than just the ones emitted by that process alone. Hadrons are complex objects, composed of sea- and valence quarks and gluons. As such, they will regularly produce more than one collision at a time. The resulting events will overlay to produce a single, larger event in the detector, as illustrated in Figure 1.9. All particles that are not emitted from the hard process, including the leftover beam remnants from the hadrons, make up the UE. While the physics of the UE is primarily determined by non-perturbative QCD and therefore poorly understood, phenomenological models exist that are effective at reproducing the distributions of the UE. Virtually pileup free samples of data have been captured by ATLAS using low luminosity runs, referred to as ”minimum bias” datasets (further described in Section 3.3).

[3] An "event" is a single pp-collision with all produced particles.

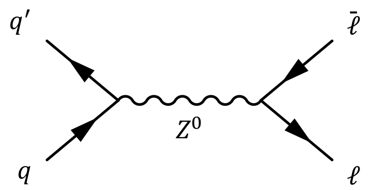
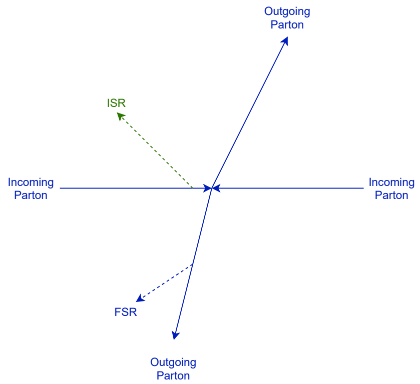
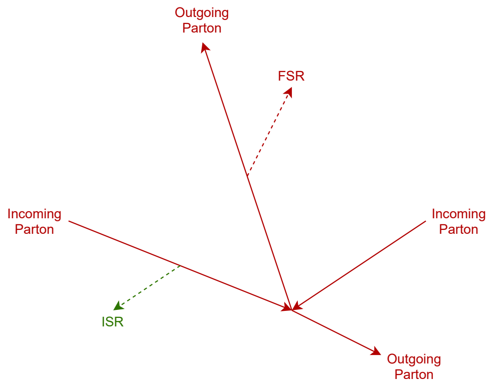


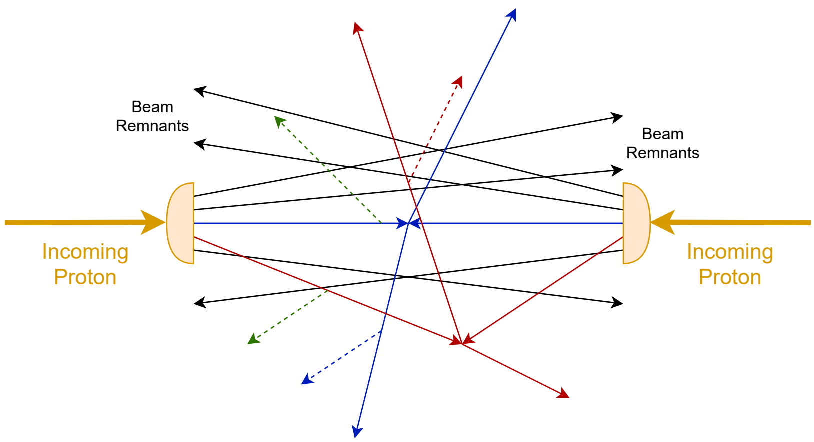
Figure 1.8: Drell-Yan event with an intermediate Z^0 boson decaying into a pair of charged leptons.



(a) The underlying hard process.



(b) Additional scattering of partons in the same event.



(c) Combination of hard process, interactions of other partons and beam remnants yield the event.

Figure 1.9: Schematic overview of how an event is comprised. Everything that is not associated with the hard process is part of the underlying event.

1.4. Monte Carlo Event Generators

The SM provides a description for all known elementary particles and their fundamental interactions. In large energy scale collisions like the one happening at LHC however, hundreds of particles are involved, making it impossible to directly calculate the resulting final states using SM physics. The first issue is with the phase space: each produced particle adds another three degrees of freedom to the phase space. Spin and flavour quantum numbers contribute as well. This makes the use of Monte Carlo methods almost inevitable, since the statistical accuracy of the technique improves with $1/\sqrt{N}$ regardless of the degrees of freedom[4]. Furthermore, soft QCD interactions are poorly understood via first principle approaches. At low energies of a few GeV, bound QCD states have a fundamentally non-perturbative nature which is mostly beyond our present computational abilities.

This is where Monte Carlo event generators are used to approximate the physics in these collisions. The purpose of an event generator is to generate particle physics processes as they occur in high-energy experiments, but without the effects of the imperfect detector. In order to achieve this, an event is split into several smaller processes which can be either directly simulated or approximated using phenomenological models.

The generator used in this thesis is called Pythia 8. Work on Pythia's predecessor JETSET [12] began in 1978 by the theory group at Lund University, the original Software was written in Fortran. Since Version 8.1, all code was migrated to C++. When generating an event, the underlying hard process is computed as a first step. Thereafter, other hard or soft processes that can happen in parallel to the initial hard process are added. In Pythia 8, the secondary hard processes can include all quarks up to bottom and gluons. These are called multi parton interactions (MPI) and can lead to major contributions in particle multiplicity in the final state. This is illustrated in [13] by comparing the charged-particle multiplicity predicted by Pythia 8 with and without MPI modelling to data from the ATLAS detector (as measured in [14]).

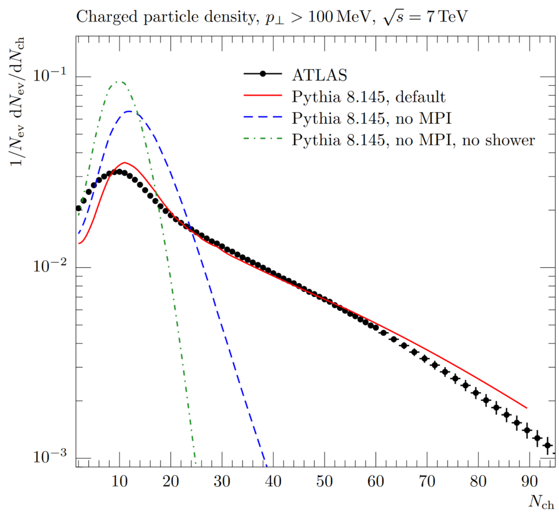


Figure 1.10: Charged-particle multiplicity at $\sqrt{s}$ = 7 TeV and p_T = 100 MeV for standard Pythia 8 settings and Pythia 8 without MPI compared to ATLAS data [14].

Figure 1.10 illustrates how the resulting particle multiplicity distribution falls off too quickly without a MPI model and can't be explained by shower modelling alone.

[4] This is a consequence of the central limit theorem.

The main idea of Pythia's MPI model was first explained in [15]. Essentially, each hadron is a complex object composed of several partons. When interacting with another hadron, there is a possibility of more than one parton pair interacting. These interactions take place in independence of each other. The total number of interactions in each event should therefore follow a Poisson distribution. Since hadrons are not point-like particles, one expects more central and more peripheral collisions to occur. Depending on the distribution of partons in the hadrons, this will widen the distribution of the number of interactions. The rate of parton-parton interactions as a function of transverse-momentum scale p_T is given by pertubative QCD, but in the region of $p_T \to 0$, this roughly diverges like $\alpha_s^2\left(p_T^2\right) dp_T^2/p_T^4$ [15, D], $\alpha_S\left(p_T^2\right)$ being the (running) strong coupling constant. Instead of regularising that divergence with a cutoff value, Pythia 8 uses a pre-factor of

$$p_T^4/\left(p_{T0}^2+p_T^2\right)^2. \tag{1.1}$$

Then, the argument in $\alpha_s(p_T^2)$ is replaced by $p_{T0}^2+p_T^2$. This avoids the divergence of the cross section, but it introduces an additional parameter that can not be inferred from first principle. Instead, the parameter has to be determined such that the event generator output most closely matches experimental data in a process called "tuning". This is further described in Section 1.5. In actuality, Pythia 8 models the p_{T0} parameter in Equation (1.1) depending on the center-of-mass energy of the simulation. This introduces a total number of three phenomenological parameters (see Table 1.1), p_{T0} is then obtained as follows:

$$p_{T0}\left(E_{CM}\right) = p_{T0,Ref}\cdot\left(\frac{E_{CM}}{E_{CM,Ref}}\right)^{E_{CM}^{Power}} \tag{1.2}$$

Table 1.1: MPI parameters that define the regularisation around $p_T \to 0$ [16].

Parameter	Pythia 8 Parameter Name	Default	Min	Max
$p_{T0,Ref}$	MultipartonInteractions:pT0Ref	2.15	0.5	10.0
$E_{CM,Ref}$	MultipartonInteractions:ecmRef	1800.0	1.0	
E_{CM}^{Power}	MultipartonInteractions:ecmPow	0.24	0.0	0.5

Even though all parameters in Equation (1.2) define the energy dependence of p_{T0}, many tuning efforts focus on only a single center-of-mass energy. The authors of Pythia 8 however recommend always tuning at least $p_{T0,Ref}$, since its value depends on a number of other parameter choices [16]. As mentioned before, hadrons are extended objects. This means that the number of collisions are dependent on the distribution of matter inside of the particle and on the impact parameter b. The impact parameter is defined as the smallest distance between the initial (straight) path of a particle and the centre of its collision partner. Different values of b result in different overlaps $\tilde{O}$ between the two initial particles. The overlap region for two particles with matter distributions ρ is proportional to the integral [17]

$$\tilde{O}(b) \propto \int dt \int d^3\vec{x}\rho(x,y,z,b)\rho(x+b,y,z+t,b). \tag{1.3}$$

The mean number of collisions is then proportional to that integral: $\langle n(b)\rangle = k\cdot\tilde{O}(b)$. Pythia 8 lets the user choose the shape of the particle density profile. There are five options available by setting the parameter *MultipartonInteractions:bProfile* (see Table 1.2).

Table 1.2: MultipartonInteractions:bProfile options affecting the proton mass distribution in Pythia 8 [16].

Option	Description	Free Parameters	Default	Min	Max
0	No impact parameter dependence				
1	Simple Gaussian distribution				
2	Double Gaussian distribution	MultipartonInteractions:coreRadius	0.4	0.1	1.0
		MultipartonInteractions:coreFraction	0.5	0.0	1.0
3	Exponential distribution	MultipartonInteractions:expPow	1.0	0.4	10.0
4	Gaussian distribution, variable width	MultipartonInteractions:a1	0.15	0.0	2.0

Since we later use the ATLAS A14 tune as a starting point [18], we use

$$\textit{MultipartonInteractions:bProfile} = 3$$

over the course of this thesis and include *MultipartonInteractions:expPow* in our own tunes. The colour configuration of the MPI systems is ambiguous. Therefore additional models describing the colour reconnection and hadronisation have to be applied afterwards. This is explained further in Chapter 2.

The system commonly used to validate and compare Monte Carlo event generators is called Rivet (Robust Independent Validation of Experiment and Theory) [19]. It offers an extensive set of analyses from most major HEP experiments including reference data. Rivet is purposefully built generator-agnostic by using a standardised input data format called HepMC event record. A HepMC record includes all details of an event but does not contain information about the generator. Rivet then provides a convenient way of applying a preexisting analysis to the data and exports the resulting histograms as YODA[5] files. Our generator of choice, Pythia 8, already comes with an interface to the HepMC standard, allowing for direct processing with Rivet.

1.5. Monte Carlo Tuning

Generating an appropriately large number of Monte Carlo samples takes a lot of time and computational power as well as disk space. Generating a statistically significant number of around 1 million events takes about $\mathcal{O}$(1 day) using a single CPU core. Usually, in fitting a model to data, one would employ some direct method like the Nelder-Mead-Algorithm to optimise a goodness of fit measure as the likelihood function or the sum of squares. In our case, however, each individual function evaluation using a slightly different set of input parameters would be incredibly expensive. Direct search methods tend to converge in $\mathcal{O}(1000)$ steps which are executed consecutively, which would result in several years of time required for the computation. Klaus Hamacher and Michael Weierstall were among the first to employ a different technique that allows all of the required samples to be calculated in parallel [20].

First, all points contained in the tune are randomly sampled from the parameter space. Next, the entirety of the MC samples is generated according to the parameterisation obtained in the first step. This can be performed in parallel, since each calculation is completely independent from the others. The usage of supercomputers like the Pleiades computational cluster at the University of Wuppertal we used over the course of this thesis makes it possible to obtain the entire sample in just a few days. Now we acquired a set of values for each bin of each histogram for a number of different input parameter choices. The values of a bin for different input parameters are now fitted to a function (we use polynomials in this thesis), as illustrated in Figure 1.11. This enables a fast evaluation of the response to an individual parametrisation, allowing for a

[5] YODA provides a set of classes for histogramming and is used for MC generator validation. More information is available at `https://yoda.hepforge.org/`.

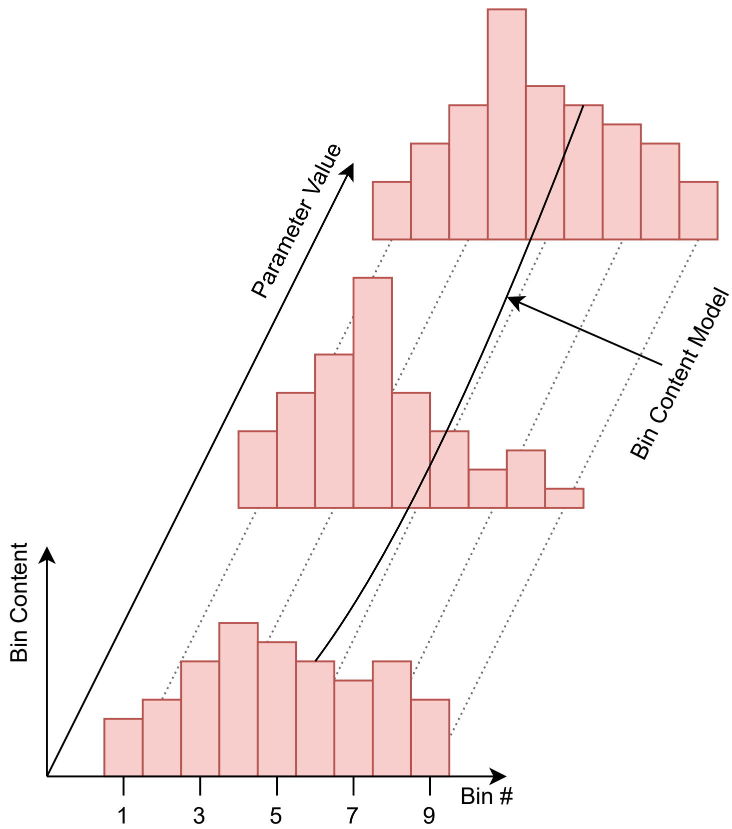


Figure 1.11: The evolution of each bin over the parameter space is interpolated by a model function to get an approximate generator behaviour that can be evaluated fast for fitting purposes (inspired by [21]).

direct search algorithm to seek the point of best fit. All of these tasks can be efficiently performed using a relatively new software package called Apprentice. It supersedes a tried and tested program called Professor [22] as a spin off to a new approach for calculating rational approximation functions to Monte Carlo data [23]. Since the amount of literature describing the use of the new Apprentice framework is lacking at the current time, the usage of the program is explained here alongside the underlying mathematics. An overview of this tuning process is provided in Figure 1.12. Additionally, an example tune is performed to visualise the tuning process in a one-dimensional parameter space.

1.5.1. Sampling of the parameter space

Most tunes deal with multiple parameters from a bounded n-dimensional parameter space

$$\mathcal{P} = [p_1^{\text{low}}, p_1^{\text{high}}] \times \cdots \times [p_n^{\text{low}}, p_n^{\text{high}}]. \tag{1.4}$$

This means that grid sampling of the entire parameter space becomes computationally expensive. The ATLAS A14 tunes, for example, include 10 parameters [18]. Therefore, creating just 10 subdivisions per parameter would have required 10^{10} runs, adding up to a CPU time of about 10^7 years. Instead, a different type of sampling is employed. The Professor framework provides four types of parameter sampling: grid, uniform, sobol and latin hypercube sampling. While the

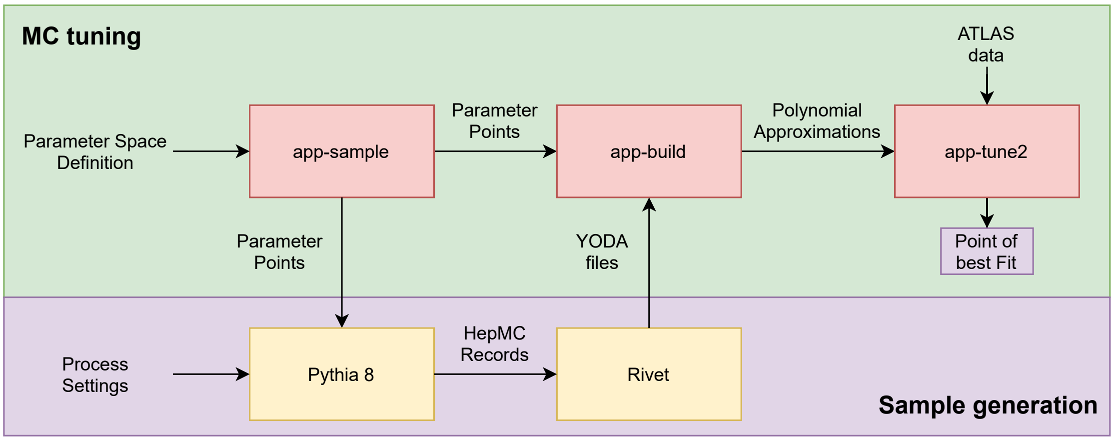


Figure 1.12: Schematic overview of the tuning process with apprentice.

sobol and latin hypercube distributions aim at covering the space more evenly for low sample sizes, Apprentice right now only supports uniformly distributed samples. A visual comparison between these techniques is provided in Figure 1.13. We use uniform sampling for our application with $n = 300$ parameter points.

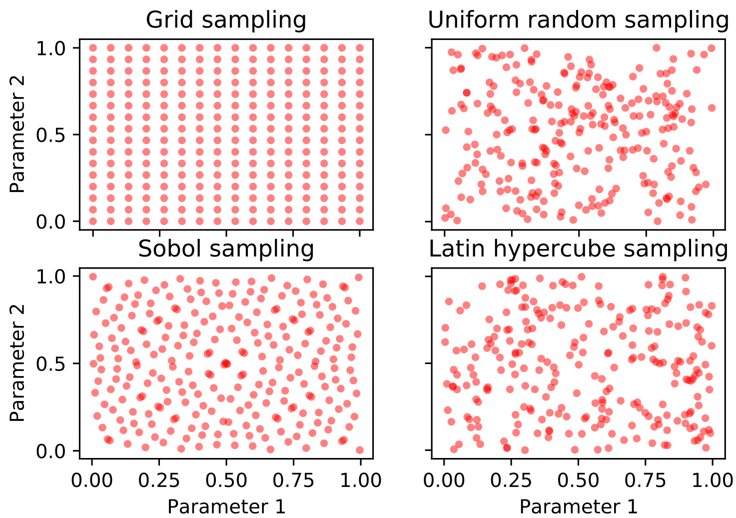


Figure 1.13: 256 samples of two parameters $P1, P2$ from the parameter space $\mathcal{P} = [0,1] \times [0,1]$ using grid, uniform, sobol and latin hypercube sampling.

The script *app-sample* is called the following way

app-sample parameterFile N template

where *parameterFile* is a file containing the parameter space definiton, *N* is the number of samples to produce and *template* is a template file for filling simulation scripts with the parameter values. The parameter file either has to be a *.json*-record containing the parameter box information or can be structured like in the older Professor framework, where each line contains the name, lower and upper bound of one parameter. The output of the program is a folder structure containing parameter point files. These subfolders are then filled with the YODA histogram files from the analysis performed in the sample generation step.

1.5.2. Conversion of the data

The Professor framework is loading the YODA files directly from the folder structure created during sample generation. This takes a lot of time, especially with a large number of YODA files. Since many tuning efforts require multiple executions of the tune, a conversion into a more efficient format is implemented into Apprentice. The usage of the Hierarchical Data Format (HDF5) [24] allows for quick load times and less disk space to be used. A conversion tool already comes with the Apprentice release, applying the command

app-yoda2h5 INPUTDIR -o data.h5

to the YODA folder structure *INPUTDIR* produces the desired *data.h5* file. Professor imports reference data from a folder containing YODA files. Apprentice has moved on to *.json* files, they can be viewed and modified by a much wider array of programs and offer at least basic human readability. For the task of converting a folder of reference data files to a *.json* record, the command

app-datadirtojson DATADIR -o data.json

is provided by Apprentice.

1.5.3. Fitting of polynomial models

Polynomials represent functions with extremely convenient properties that lend themselves well for applications like approximating data fast. Suppose that in a tuning problem with M free simulation parameters $p_1, \dots, p_M$, histograms are to be approximated to the order N. Appropriate polynomials are of the form

$$P(p_1, \dots, p_M) = \sum_{i_1,\dots,i_M=0}^{N} k_{i_1 \dots i_M} p_i^{i_1} \cdot \dots \cdot p_M^{i_M}. \tag{1.5}$$

These polynomials still need to be fitted to the Monte Carlo data by determining the coefficients k_i. For the polynomial in Equation (1.5), this means that there are $N_{Coeff} = \binom{M+N}{N}$ coefficients. Therefore, there also have to be at least this many Monte Carlo samples available in order to have an (over -) determined problem. Apprentice uses Numpy [25] to find the least square solution to a system of linear equations. For a square interpolation of a Monte Carlo sample $\overrightarrow{MC}$ of size S with M input parameters, the linear equations are of the form

$$\tilde{P} \cdot \vec{c} = \overrightarrow{MC} \Leftrightarrow \begin{pmatrix} 1 & p_1^1 & \dots & p_M^1 & p_1^1 p_1^1 & p_1^1 p_2^1 & \dots & p_M^1 p_M^1 \\ 1 & p_1^2 & \dots & p_M^2 & p_1^2 p_1^2 & p_1^2 p_2^2 & \dots & p_M^2 p_M^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & p_1^S & \dots & p_M^S & p_1^S p_1^S & p_1^S p_2^S & \dots & p_M^S p_M^S \end{pmatrix} \cdot \begin{pmatrix} k_{0\dots0} \\ \vdots \\ k_{2\dots2} \end{pmatrix} = \begin{pmatrix} MC_1 \\ \vdots \\ MC_S \end{pmatrix}, \tag{1.6}$$

where MC_i are the bin values corresponding to the sampling points $\vec{p}_i = (p_1^i, \dots, p_M^i)$. Approximation models are stored as *.json*-records by Apprentice and have to be generated for bin values and errors separately. Building approximations is done via the *app-build* command.

On top of polynomial regression, it also allows for the use of rational approximation functions. They take the form of

$$F(\vec{p})_{ab} = \frac{g_a(\vec{p})}{h_b(\vec{p})},$$

where g_a and h_b are polynomials of orders a and b. Two ways of efficiently building these rational approximations are presented in [23], but for the remainder of this thesis, we will limit ourselves to using only the polynomial modelling functionality. To calculate the coefficients for a file *data.h5*, we apply

app-build data.h5 –order 3,0 -o values.json

app-build data.h5 –order 2,0 -o errors.json –errs

to obtain value approximations to the third order stored in *values.json* and error model coefficients to the second order in *errors.json*. Over the course of this thesis, the content of the bins is approximated with 3rd order polynomials and the error of the bins is approximated with 2nd order polynomials, unless otherwise specified.

1.5.4. Goodness of fit function

While allowing custom goodness of fit (GoF) functions by means of an external *.py* file, the most common use case in Monte Carlo tuning is a χ^2 goodness of fit function. Since most tuning efforts are not focused on only one singular variable, a set of per variable weights $w_{\mathcal{O}}$ is introduced into the goodness of fit function to assign various degrees of importance to them. Getting the overall GoF measure is achieved by summing over all variables $\mathcal{O}$ and all bins $b \in \mathcal{O}$:

$$\chi^2(\vec{p}) = \sum_{\mathcal{O}} w_{\mathcal{O}}^2 \sum_{b\in\mathcal{O}} \frac{(P_b(\vec{p}) - \mathcal{R}_b)^2}{\Delta_b^2}, \tag{1.7}$$

where $\mathcal{R}_b$ is the reference value for bin b and Δ_b^2 is the associated variance consisting of the error of the interpolated bin and the error from the reference data bin. For fast evaluation of the GoF function, Apprentice offers a huge improvement over Professor by again using Numpy. A comparison between the evaluation frequency of Professor's and Apprentice's GoF function is given in Figure 1.14.

1.5.5. Choice of weights

The definition of the GoF function is dependent on the choice of variable weights. Since there is no a priori method for finding an ideal set of weights for a tune, the tuning result will always depend on a more or less biased choice of prioritisation between all variables. We will try to identify variables insensitive to changes of the input parametrisation and exclude them from the tuning effort. We will also attempt to identify sensitive variables that describe the physics measured at ATLAS suboptimally regardless of the input and exclude them as well.
The tool app-ls provides a way of outputting a weight file (with $w_{\mathcal{O}} = 1$ for all variables $\mathcal{O}$) for a given parametrisation file *values.json*:

app-ls values.json -w > weights.txt

A critical aspect of determining the weights is excluding variables that are not well described by the Monte Carlo modelling. Including them harbours the danger of pushing the overall point of best fit in a direction that might not be ideal for other observables that are actually described by the model. We can therefore define a confidence interval round the mean value by assuming a Gaussian distribution. For normally distributed variables, this confidence interval is given by

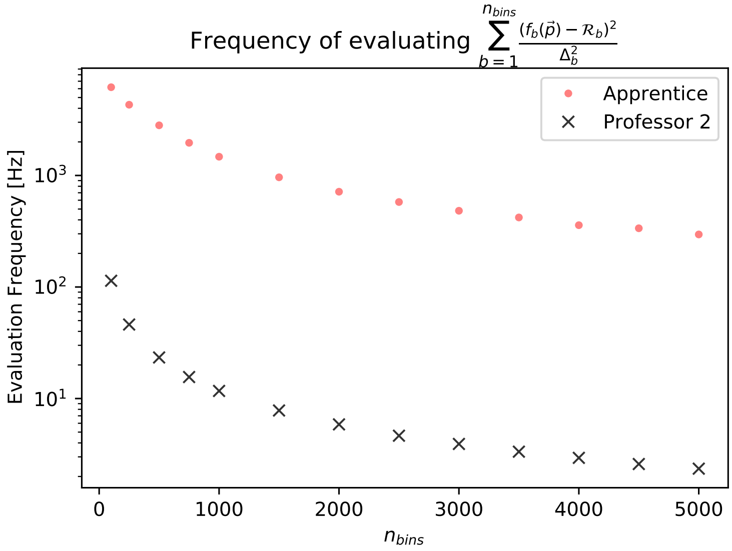


Figure 1.14: Comparison of χ^2 GoF evaluation frequency between the Professor and the Apprentice framework. Due to the extensive use of Numpy, Apprentice is almost one hundred times faster than it's predecessor[6].

$$[\mu - z\sigma, \mu + z\sigma] \ , z = \Phi^{-1}\left(\frac{p+1}{2}\right),$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution, μ and σ are the reference mean and error of that bin and p is a probability associated with the confidence interval. If the range of values of the polynomial model function in a bin does not (at least partially) intersect with that confidence interval, the model function and therefore the Monte Carlo modelling is ineffective at describing that bin of the observable (a schematic overview of this is provided in Figure 1.15). When testing for data envelopment, each observable is examined on a bin-by-bin basis for this agreement, and bins are classified based on whether they passed this test or not. When a certain ratio of all bins fail, the entire variable is considered to be non-descriptive of the data. The probability of the confidence interval and the threshold ratio are the defining parameters of this data quality test that have to be specified by the user during the tuning process.

1.5.6. Performing the tune

After all these steps, the χ^2 has to be minimised to find the point in parameter space where Monte Carlo simulation and reference data are in best agreement with each other. This represents a multidimensional optimisation problem that can be solved with the most common minima finding algorithms. Apprentice supports the use of multiple algorithms. We will however solely employ the default one called “TNC”, which is a truncated Newtonian algorithm that minimises a function inside a bounded parameter space based on [26]. TNC uses gradient information obtained from the approximation function, which is calculated inside Apprentice by invoking Numpy and just in time compilation to achieve better performance.

[6] Data from `https://iamholger.gitbook.io/apprentice/`

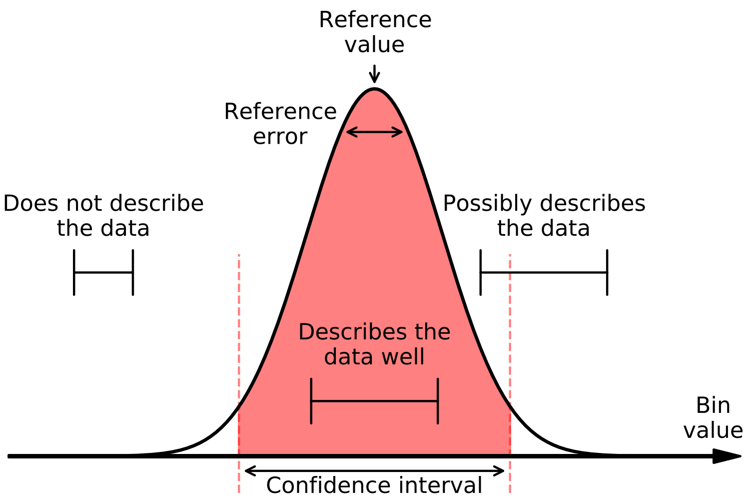
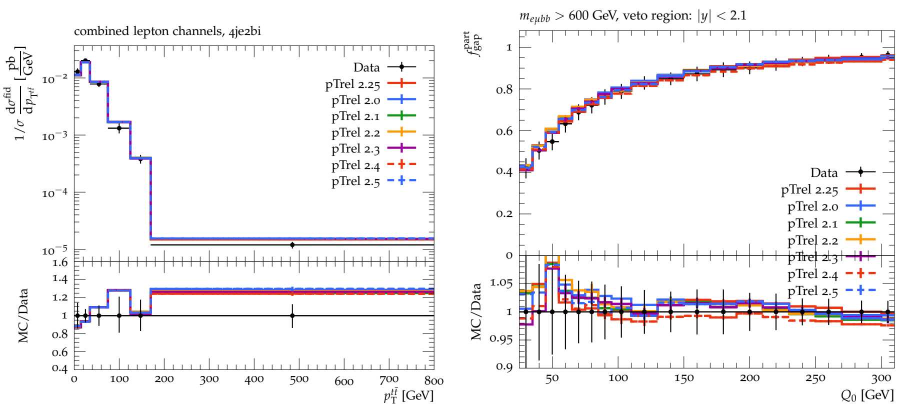


Figure 1.15: Visualisation of the bin-wise envelopment test. The polynomial model can partially or fully describe the data, this means that the interval set by the lowest and highest values of the model function must at least partially intersect with the confidence interval of the reference data. Otherwise, the bin is considered non-descriptive of the experimental data.

A tune is performed by using the command *app-tune2*: It requires the approximation files *values.json* and *errors.json*, the reference data stored in *data.json* and *weights.txt* containing the variable weights. Using the error estimates is optional.

app-tune2 weights.txt data.json values.json -e errors.json -o TUNE

outputs a weights file, an estimated YODA histogram file for the point of best fit and another text file with the detailed tuning results like parameter values and biased/unbiased χ^2.



(a) Poor envelopment of the reference data in two bins (differential cross sections as a function of $p_T^{t,had}$ from [27]).

(b) Good envelopment of the data (fraction of events with no jet activity for $m > 600$ GeV from [28])

Figure 1.16: Examples for good and bad envelopment of the data. Variables with visually bad envelopment were removed from the 1D tune.

1.5.7. The ATLAS A14 tunes

In 2014, the ATLAS collaboration released a series of four tunes for Pythia 8 to 7 TeV data, appropriately named as the A14 set of tunes [18]. Each of these tunes is based on a different parton density function (PDF). We will only be using NNPDF23LO in this thesis. All tunes include observables sensitive to the underlying event, jet structure and additional jet emissions above the lowest-order. The parameters shown in Table 1.3 were all tuned to the data on a random sample of 500 parameter points with one million events of dijet, $t\bar{t}$ and Z-boson hard processes each.

Table 1.3: Definition, parameter range and tuned value for the A14 NNPDF23LO tune [18].

Parameter	Definition	Range	A14 (NNPDF23LO)
SigmaProcess:alphaSvalue	α_S at scale $Q^2 = M_Z^2$	0.12 - 0.15	0.140
SpaceShower:pT0Ref	ISR p_T cutoff	0.75 - 2.5	1.56
SpaceShower:pTmaxFudge	Mult. factor on max ISR evolution scale	0.5 - 1.5	0.91
SpaceShower:pTdampFudge	Factorisation/renorm scale damping	1.0 - 1.5	1.05
SpaceShower:alphaSvalue	ISR α_S	0.10 - 0.15	0.127
TimeShower:alphaSvalue	FSR α_S	0.10 - 0.15	0.127
BeamRemnants:primordialKThard	Hard process primordial k_T	1.5 - 2.0	1.88
MultipartonInteractions:pT0Ref	MPI p_T dampening	1.5 - 3.0	2.09
MultipartonInteractions:alphaSvalue	MPI α_S	0.10 - 0.15	0.126
ColourReconnection:Range	CR strength	1.0 - 10.0	1.71

All other parameters were taken from the Monash tune included in Pythia 8 [29]. Details on the observables used and further discussions of the tune are provided in [18].

Table 1.4: Colour reconnection settings for our 1D example tune. These are the recommended settings for the gluon move model [30].

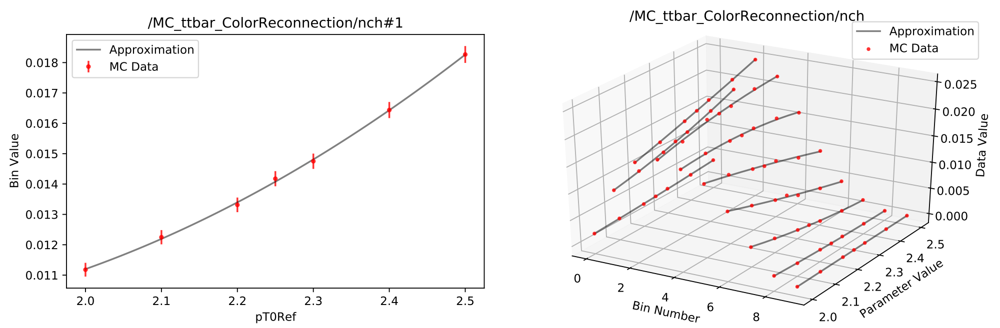
Parameter	Value
MultipartonInteractions:expPow	1.85
PartonLevel:earlyResDec	off
ColourReconnection:mode	2
ColourReconnection:reconnect	on
ColourReconnection:m2Lambda	1
ColourReconnection:fracGluon	1
ColourReconnection:dLambdaCut	0
ColourReconnection:flipMode	0

1.5.8. 1D example tune

In order to better visualise the entire tuning process, we have decided to first perform an example tune where only a single parameter is varied. This allows for easy visualisation of the data and makes the entire procedure more accessible. We decided to tune the gluon-move scheme (described in Section 2.3) based on the ATLAS A14 tune [18] with the settings detailed in Table 1.4. We choose *MultipartonInteractions:pT0Ref* as the tuning parameter with the values in Table 1.5.

Table 1.5: Parameter values for MultipartonInteractions:pT0Ref chosen for our 1D tune.

MC Run	1	2	3	4	5	6	7
$p_{\perp 0,Ref}$	2.0	2.1	2.2	2.25	2.3	2.4	2.5



(a) Bin content and modelling for the first bin.

(b) Bin content and modelling for the entire variable.

Figure 1.17: MC data and approximation function for the variable N_{Ch} from Section 3.4.

After generating the MC samples, we interpolated the data samples with second order polynomials as in Figure 1.17. For all variables for which the Monte Carlo samples properly enveloped the data and had reference data available, we chose $w_{\mathcal{O}} = 1$, excluding all other. Figure 1.16 gives a rough idea how variables were selected. A complete list of all variables used in this tune can be found in Appendix B.
This left us with 1474 bins included in the tune, distributed over 126 individual variables. We obtain a GoF value of $\chi^2 = 2912.9$ or $\chi^2/N_{df} = 1.98$ for

$$\textit{MultipartonInteractions:pT0Ref} = 2.08.$$

As visualised in Figure 1.18, there is a very clear optimisation valley at the point of best fit, no further minima are present in the parameter range.

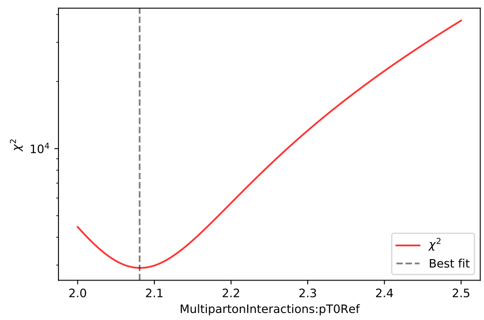


Figure 1.18: Visualisation of the optimisation result for the one-dimensional tune of MultipartonInteractions:pT0Ref.

Even though we only used $t\bar{t}$ hard-process samples and a different CR model to achieve this result, this agrees with the A14 tune value of $p_{T0,Ref} = 2.09$.

2

Colour Reconnection Models

Because finding solutions of non-perturbative QCD remains a problem that has yet to be solved, Pythia relies on phenomenological models instead to approximate a colour system's behaviour. The underlying description of hadronisation is based on the Lund string model, originally developed in 1983 [31]. A "string" represents a colour-connection between two or more partons that make up a colour system, with gluons being the "kinks" in the string (a visual example for this is provided in Figure 2.1). This picture is motivated by the fact that the QCD potential between two quarks increases linearly when they drift apart. The entirety of the colour field is approximated to be contained within a line of a constant energy density per unit length ($\kappa \approx 1\,\mathrm{GeV\,fm^{-1}}$). Self-attraction of the gluon field causes the field lines to collapse into a thin tube, warranting this approach. When two partons move in opposite directions, the string becomes elongated, thus more energetic. When creating a quark-antiquark pair from the vacuum becomes more energetically favourable, the string breaks apart into substrings. This process continues as long as there is energy available to fuel the creation of more quark-antiquark pairs. This point of view comes with a caveat: Since a string represents a confining attraction between partons, the way the partons are connected by the strings (referred to as "colour flow") affects the structure of the event. When e.g. two strings from different colour systems start to overlap due to the movement of the subsystem's constituents, these strings are able to reconnect into a more energetically favourable configuration. This may also attach strings together which where previously parts of different MPI systems.
The initial goal of this thesis was to compare three models for the reconnection of colour flow by tuning them to ATLAS data and evaluating their performance. While the respective approaches to the problem differ, these all try matching the resulting physics of the underlying event as closely as possible. The user can switch between the models by setting the parameter *ColourReconnection:mode*. Basic descriptions of these different CR models are available in the Pythia 8 CR documentation [30]. Due to time and computing limitations, we later decided to study just one of these models, described in Section 2.3.

2.1. MPI-based Colour Reconnection

This is the first model that was implemented in Pythia 8. It only has one single free parameter called *ColourReconnection:Range*, ranging between 0 and 10 with a default of 1.8. All partons are classified and sorted according to the MPI system they are associated with. The idea is to minimise the total length of colour strings by adding partons from the lower p_T systems to higher

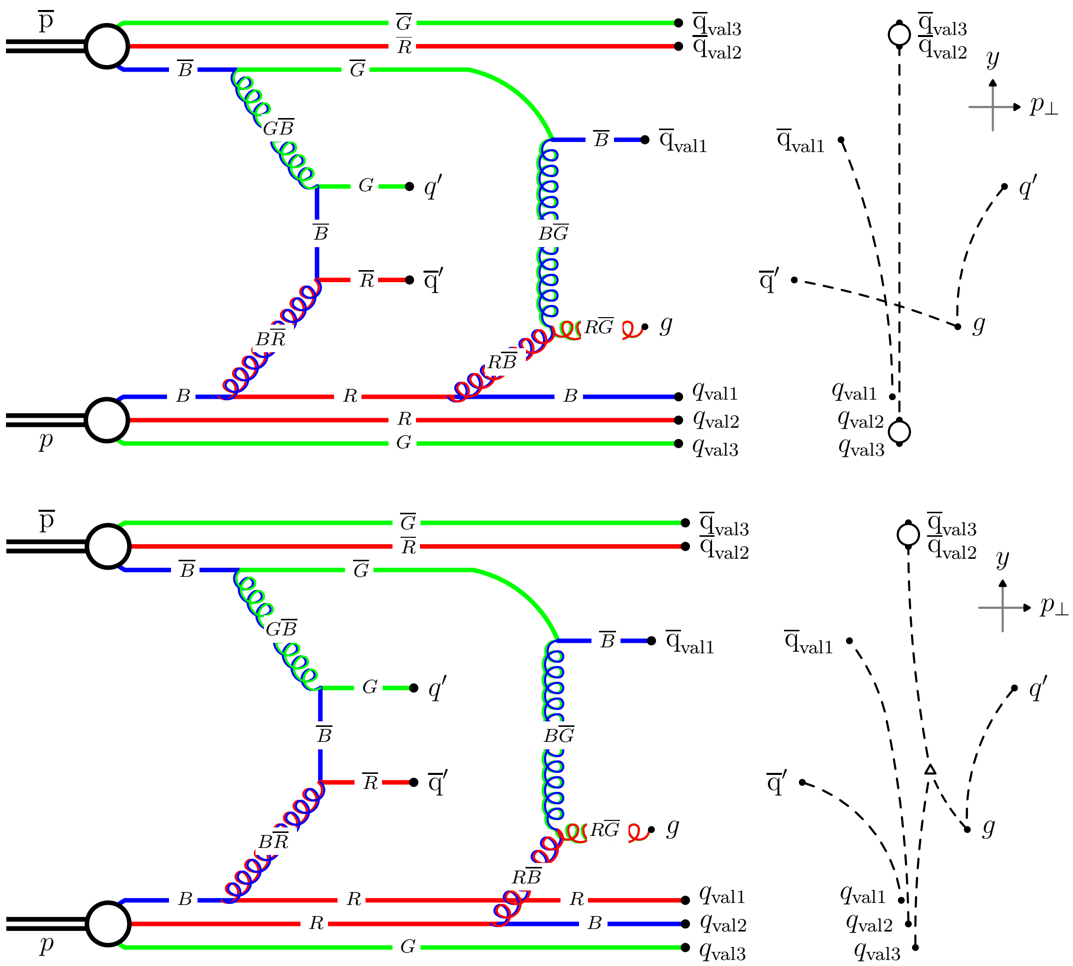
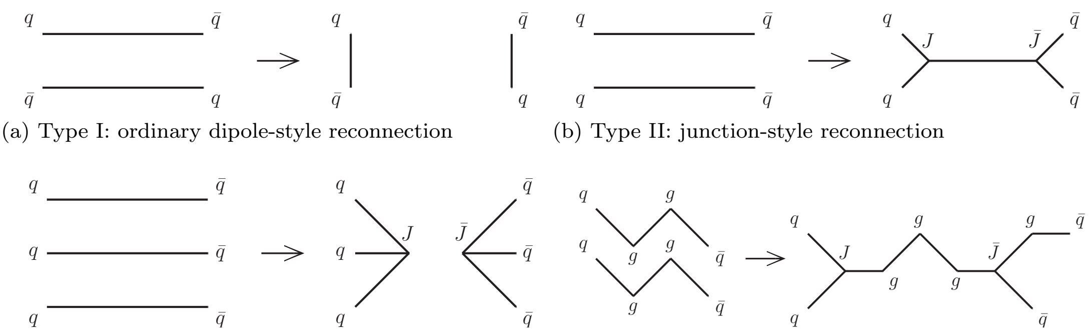


Figure 2.1: Two examples of possible colour assignments in the same two parton interaction final state of a $p\bar{p}$ collision and their corresponding string topologies in p_T, y space (taken from [13]).

p_T strings. Since most low p_T partons are gluons, this is what the model handles particularly well. After sorting for p_T, each system is assigned a reconnection probability

$$P = \frac{\left(p_{T,0} \cdot \mathit{range}\right)^2}{\left(p_{T,0} \cdot \mathit{range}\right)^2 + p_T^2} \tag{2.1}$$

with $p_{T,0}$ being the same parameter used in the MPI modelling (see Equation (1.2)). The output of the colour reconnection model is thus dependent on the modelling of the MPIs, therefore they should always be tuned in conjunction.
Starting with the lowest p_T system, a uniform random number from the interval [0,1] is generated for each higher p_T entry of the list. If this number is greater than the reconnection probability, the loop breaks and the systems are reserved to be reconnected. This is a classical example of a geometric distribution, where Bernoulli experiments are performed consecutively until a success is observed. The probability of the first system to reconnect with the m$^{\text{th}}$ is, therefore, $(1-P)^{m-1}P$, the probability for a system to reconnect to any of n others is given by $1-(1-P)^{n-1}$.
The merging step is performed after a reconnection decision has been made for all MPI systems, this time starting from the highest p_T system. Foremost, all colour dipoles are identified in the system. The gluons from the softer system are then inserted on the dipole that yields the lowest overall string length λ. "Inserting" a gluon in this context refers to the replacement of a colour dipole with two new ones. The $q\bar{q}$ pairs that arise from the splitting of a string are treated in just the same way, while quarks attached to the beam remnants are ignored in the reconnection process. By setting *PartonLevel:earlyResDec*, the user can choose to include the products of resonance decays in the colour reconnection scheme as well. The MPI-based colour reconnection scheme is activated by setting *ColourReconnection:mode* to 0 and is thus also referred to as CR0.



(a) Type I: ordinary dipole-style reconnection
(b) Type II: junction-style reconnection
(c) Type III: baryon-style junction reconnection
(d) Type IV: zipper-style junction reconnection

Figure 2.2: The four different allowed reconnection types in the CR1 model (taken from [32]).

2.2. QCD-based Colour Reconnection

Jesper Christiansen and Peter Skands developed a colour reconnection model based on minimising the string length while also including the colour rules from QCD [32] (also referred to as CR1). Additionally, this model introduces a mechanism for checking whether two colour systems are causally unconnected, preventing their colour strings from reconnecting. While the older model only connected systems via insertion into existing colour strings, junction formation is also permitted in this model.

Due to QCD colour rules, the four fundamental reconnections (see Figure 2.2 for their definition and nomenclature) are not always allowed. For before uncorrelated colour systems, they are only allowed in a fraction of cases (see [32] for detailed derivations of this). The probabilities for type I or II reconnections of a $q\bar{q}$ dipole to occur are marginally different than for gg dipoles. In this regard, authors made the design compromise to treat all colour dipoles the same. By assigning each dipole a random index between one and nine, these reconnection probabilities evaluate to $P_I = 1/9$, $P_{II} = 2/9$ and $P_{III} = 2/81$. The type IV reconnections are handled by demanding both dipoles having different but same-modulo-three indices. The overall probability then depends on the number of gluons involved. After evaluating all allowed reconfigurations according to their indices, the model carries out the colour reconnections that minimise the string length λ. Since the λ measure is not quite adequately understood for junction structures, several different measures are provided, along with a tunable parameter that yields a multiplicative correction for junctions. Unlike the CR0 model, the QCD-based approach requires more phenomenological parameters and various settings, some of which are subject to tuning. All tunable parameters are given in Table 2.1. Other settings that affect the models behaviour are listed in Table 2.2.

Table 2.1: All tunable parameters of the CR1 model.

Parameter	Default	Minimum	Maximum	Description
m0	0.3	0.1	5.0	Used in the λ measure for the string length
junctionCorrection	1.20	0.01	10.0	Multiplicative correction to m_0 for junctions
timeDilationPar	0.18	0	100	Influences the amount of CR and time dilation.

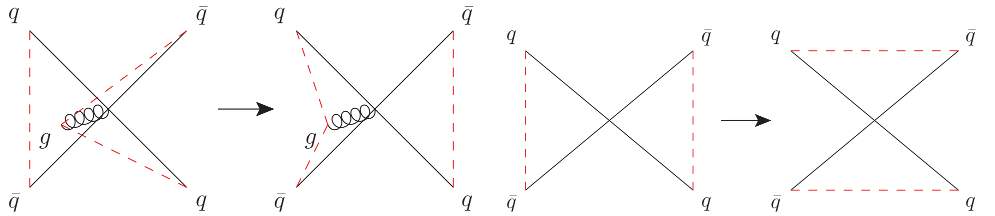
Table 2.2: All settings for the CR1 model. In the time dilation settings, λ is the boost factor between two strings, m_{Dip} is the mass of the dipole, m'_{Dip} is defined at formation of the dipole and $r = 1\,\text{fm}$.

Setting	Default	Description
nColour	9	Number of reconnection colours. Affects junction probabilities
sameNeighbourColours	off	Allows for same colour dipoles to reconnect (QCD does not allow that)
allowJunctions	on	Switches on the formation of junctions
lambdaForm	0	Switches between different forms of λ measure (contribution per dipole) 0: $\lambda = \ln(1 + \sqrt{2}E/m_0)$ (E is defined in the dipole/junction rest frame) 1: $\lambda = \ln(1 + 2E/m_0)$ 2: $\lambda = \ln(2E/m_0)$
allowDoubleJunRem	on	Always check if two junctions can be replaced by two strings
timeDilationMode	2	Criterion used to check whether two strings are causally connected 0: All strings are allowed to reconnect 1: $\gamma <$ timeDilationPar, all strings in causal contact 2: $\gamma <$ timeDilationPar $\cdot\, m_{Dip} \cdot r$, all strings in causal contact 3: $\gamma <$ timeDilationPar $\cdot\, m_{Dip} \cdot r$, at least one pair of dipoles in causal contact 4: $\gamma <$ timeDilationPar $\cdot\, m'_{Dip} \cdot r$, all strings in causal contact 5: $\gamma <$ timeDilationPar $\cdot\, m'_{Dip} \cdot r$, at least one pair of dipoles in causal contact

2.3. The Gluon-Move Scheme

Compared to the QCD-based scheme, the gluon-move method represents a relatively simplistic model for colour reconnection [33]. It solely focuses on minimising the total string length λ; neither colour algebra nor the event's geometry in space-time are considered at all. It encompasses two separate mechanisms to achieve this minimisation, a gluon-move and a gluon-flip step.

A gluon-move is carried out after all final gluons have been identified. If a gluon produces a shorter string length when connected to any other string piece of any other connected pair, the



(a) Example of a gluon-move by reconnecting a gluon to a different colour string.

(b) Example of a flip-type reconnection by reattaching two colour strings.

Figure 2.3: Example of gluon-move (a) and gluon-flip (b) reconnections in the gluon-move model (taken from [33]). The dashed lines represent the colour strings, the solid lines represent any QCD interaction that is compatible with the initial and final states.

move is carried out (see Figure 2.3 (a)). This process is continued until no further moves that minimise the total length are possible. A move step cannot change the colour configuration between quark endpoints, since it only removes and reinserts gluons into strings. An additional flip mechanism, that flips colour lines between two strings, is introduced to further minimise λ (see Figure 2.3(b)). Only one flip is allowed per string to prevent the formation of loops.

In this model, hereinafter referred to as CR2, the contribution to the string length of two partons i and j with invariant mass m_{ij} is defined as $\lambda = \ln(1 + m_{ij}^2/m_\lambda^2)$, m_λ^2 being a free parameter of the model. Mechanisms to manage the amount of reconnections are embedded into the model as well: The fraction of gluons participating in the move step can be reduced via a tunable parameter. In addition, a minimum reduction of λ can be required for any move or flip step, reducing the total number of steps taken per event. The tunable parameters for CR2 are given in Table 2.3, the settings for the handling of gluon-flips are shown in Table 2.4.

Table 2.3: All tunable parameters of the CR2 model.

Parameter	Default	Minimum	Maximum	Description
m2Lambda	1.	0.25	16.	m_λ^2 used in the definition of the λ measure
fracGluon	1.	0.	1.	Probability that a gluon will be considered for a move
dLambdaCut	0.	0.	10.	Moves and flips have to reduce λ by at least this amount

Table 2.4: The CR2 model only has one setting. It affects how flip and move steps are combined.

Setting	Default	Description
flipMode	0	If and where to perform the flip step 0: No flip steps 1: Allow flips only for strings without junctions 2: Allow flips for all strings 3: As 1, but no move steps 4: As 2, but no move steps

3

Analyses under Study

After an event record is generated by Pythia 8, one or multiple analyses are applied to it. Over the course of this thesis, we will be using Monte Carlo samples of $t\bar{t}$ processes at 13 TeV, minimum-bias samples at 7, 8 and 13 TeV as well as samples of Z^0 production processes at 7 TeV. All variables in the analyses provided by Rivet follow a naming convention:

/ANAYLSIS_NAME/VARIABLE_NAME

Individual bins are named

/ANALYSIS_NAME/VARIABLE_NAME#BIN_NUMBER

where *ANALYSIS_NAME* is the Rivet-specific identifier for an analysis. It usually consists of the name of the experiment, the year of the associated publication and a unique number. Users can either choose from over 900 analyses provided by Rivet [34], including reference data, or put together a Rivet version of their own analysis. This chapter serves as an overview of the analyses we used and documents the variables relevant to our tuning efforts. More detailed descriptions and original publications about these variables can be found in the respective references. Since the naming convention for *VARIABLE_NAME* can be rather cumbersome within Rivet, we included a reference for converting the Rivet-specific name to the physical meaning of the observable (see Appendix C). Each Rivet analysis comes with a set of unfolded data for all bins of all observables. The unfolded distributions are corrected for the effects of the detector and thus serve as experimental reference data to the output of MC generators.

3.1. Jet Shape Analyses

In proton-proton collisions with a high particle multiplicity, effects from multi-parton interactions and colour reconnection become more important. These processes therefore have a direct effect on the hadronisation that takes place inside of hadronic jets and can affect their shape [35]. The following analyses introduce measurements of jet shapes and are therefore considered in our colour reconnection tune.

3.1.1. Jet shapes in $t\bar{t}$ events at $\sqrt{s} = 7$ TeV (ATLAS_2013_I1243871)

In $t\bar{t}$ collisions, there are two types of jets that can be picked up by the detector: b-jets, which spawn from the b-quarks generated in the collision, and light jets, which originate from any other

quark lighter than the b-quark. Introducing the variable r as the minimum angular distance to the central axis of a jet, we can also define $p_T(r_1, r_2)$ as the scalar sum over all transverse momenta of the jet constituents with radii between r_1 and r_2.
The differential jet shape in an annulus[1] $\left(r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right)$ is defined as

$$\rho(r) = \frac{1}{\Delta r} \frac{p_\mathrm{T}\left(r - \frac{\Delta r}{2}, r + \frac{\Delta r}{2}\right)}{p_T(0, R)}$$

with $\Delta r = 0.04$. R is the outer radius of the jet. Since the energy density is highest around the centre of the jet, $\rho(r)$ is expected to decrease with larger values of r. The integrated jet shape variable is thus defined as

$$\Psi(r) = \frac{p_T(0, r)}{p_T(0, R)}$$

with $\Psi(r = R) = \frac{p_T(0,R)}{p_T(0,R)} = 1$. Intuitively, it represents the fraction of transverse momentum inside a cone-shaped region of radius r.

Table 3.1: All cuts applied in the $\sqrt{s} = 7\,\mathrm{GeV}$ jet shape analysis [36].

	Parameter Value				
	Single-Lepton		Dilepton		
Parameter	e	μ	ee	$\mu\mu$	$e\mu$
Isolated Leptons	1		2, opposite charges		
Lepton pT	$> 25\,\mathrm{eV}$	$> 20\,\mathrm{eV}$	$> 25\,\mathrm{eV}$ for e, $> 20\,\mathrm{eV}$ for μ		
B-Tags	> 1		> 1		
Jet requirement	≥ 2 with $p_T >$, $\lvert\eta\rvert < 2.5$		≥ 4 with $p_T > 25\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.5$		
E_T^{Miss}	$> 35\,\mathrm{GeV}$	$> 20\,\mathrm{eV}$	$> 60\,\mathrm{eV}$	$> 60\,\mathrm{eV}$	
$\sum p_T$					$> 130\,\mathrm{GeV}$
Drell-Yan rejection			$m_{\ell\ell} > 15\,\mathrm{GeV}$, $\lvert m_{\ell\ell} - m_z\rvert \geq 10\,\mathrm{GeV}$		
Transverse Mass	$m_T > 25\,\mathrm{GeV}$	$E_T^{\mathrm{Miss}} + m_T > 60\,\mathrm{GeV}$			

This particular analysis considers both single and dilepton samples with the cuts in Table 3.1. The integrated and differential jet shapes are calculated for both b- and light jets at five different cuts of the jet's transverse momentum, producing a total of 20 variables in this analysis (see Table C.1).

3.1.2. Jet substructure at $\sqrt{s} = 13$ TeV (ATLAS_2019_I1724098)

This analysis at $\sqrt{s} = 13\,\mathrm{TeV}$ also studies the substructure of jets, but takes a different approach than the one presented in Section 3.1.1. Events are selected as either dijet, top or W events according to the criteria listed in Table 3.2. For each of these selections, multiple variables that are indicative of the jet substructures within the event are measured.

Foremost, the analysis examines the number of subjets N_{subjets} with $p_T > 10\,\mathrm{GeV}$ that are reconstructed from a larger-radius jet using the k_t algorithm described in [37] with $R = 0.2$. Also measured are the energy correlation functions $\mathrm{ECF}_2^{\mathrm{Norm}}$, $\mathrm{ECF}_3^{\mathrm{Norm}}$ and the related ratios C_2 and D_2. They are defined as

$$\mathrm{ECF}_1 = \sum_i p_{\perp,i}, \quad \mathrm{ECF}_2 = \sum_{i<j} p_{\perp,i} p_{\perp,j} (\Delta R_{ij})^{\beta^{\mathrm{ECF}}}, \quad \mathrm{ECF}_3 = \sum_{i<j<k} p_{\perp,i} p_{\perp,j} p_{\perp,k} (\Delta R_{ij} \Delta R_{jk} \Delta R_{ik})^{\beta^{\mathrm{ECF}}}$$

[1] In this context, an annulus refers to an interval of radii. Remember that r is an angle in the context of this analysis, not a measure of length.

$$\mathrm{ECF}_2^{\mathrm{Norm}} = \frac{\mathrm{ECF}_2}{(\mathrm{ECF}_1)^2}, \quad \mathrm{ECF}_3^{\mathrm{Norm}} = \frac{\mathrm{ECF}_3}{(\mathrm{ECF}_1)^3}$$
$$C_2 = \frac{\mathrm{ECF}_3^{\mathrm{Norm}}}{\left(\mathrm{ECF}_2^{\mathrm{Norm}}\right)^2}, \quad D_2 = \frac{\mathrm{ECF}_3^{\mathrm{Norm}}}{\left(\mathrm{ECF}_2^{\mathrm{Norm}}\right)^3}$$

Table 3.2: Event selection criteria (inclusive) for the $\sqrt{s} = 13\,\mathrm{TeV}$ [38].

Selection Criteria	Detector Level	Particle Level
Dijet selection:		
Two trimmed anti-k_t $R = 1.0$ jets	$p_T > 200\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.5$	
Leading-p_T anti-k_t $R = 1.0$ jet	$p_T > 450\,\mathrm{GeV}$	
Top and W selections:		
Exactly one muon	$p_T > 30\,\mathrm{GeV}$, $\lvert\eta\rvert < 2.5$ $\lvert z_0 \sin(\theta)\rvert < 0.5\,\mathrm{mm}$ and $\lvert d_0/\sigma(d_0)\rvert < 3$	
Anti-k_t $R = 0.4$ jets	$p_T > 25\,\mathrm{GeV}$, $\lvert\eta\rvert < 4.4$ JVT output > 0.5 (if $p_T < 60\,\mathrm{GeV}$)	
μ isolation criteria	$\Delta R(\mu, \mathrm{jet}) \geq 0.04 + 10\,\mathrm{GeV}/p_{\perp,\mu}$	
E_T^{Miss} and m_T^W	$E_T^{\mathrm{Miss}} > 20\,\mathrm{GeV}$, $E_T^{\mathrm{Miss}} + m_T^W > 60\,\mathrm{GeV}$	
Leptonic top	At least on small-radius jet with $0.4 < \Delta R(\mu, \mathrm{jet}) < 1.5$	
Top selection:		
Leading p_T trimmed anti-k_t $R = 1.0$ jet	$\lvert\eta\rvert < 1.5$, $p_T > 350\,\mathrm{GeV}$, mass $> 140\,\mathrm{GeV}$ ΔR(large radius jet, b-tagged jet) < 1 $\Delta\phi(\mu$, large-radius jet)> 2.3	
W selection:		
Leading p_T trimmed anti-k_t $R = 1.0$ jet	$\lvert\eta\rvert < 1.5$, $p_T > 350\,\mathrm{GeV}$, $60\,\mathrm{GeV} <$ mass $< 100\,\mathrm{GeV}$ $1 < \Delta R$(large radius jet, b-tagged jet) < 1.8 $\Delta\phi(\mu$, large-radius jet) > 2.3	

with the weights $\beta^{\mathrm{ECF}} = \beta^{\mathrm{NS}} = 1$ for the entire analysis and ΔR_{ij} being the angular distance between constituents i and j. Another class of variables, "subjettinesses", are defined as

$$\tau_0 = \sum_i p_{\perp,i} \Delta R_i^{\beta_{\mathrm{NS}}}, \quad \tau_j = \frac{1}{\tau_0} \sum_i p_{\perp,i} \min\left(\Delta R_{a_1,i}^{\beta_{\mathrm{NS}}}, \ldots, \Delta R_{a_j,i}^{\beta_{\mathrm{NS}}}\right),$$

where ΔR_i is the angular distance between the i^{th} constituent and the jet axis, while $\Delta R_{a_n,i}$ is the angular distance between the i^{th} constituent and the axis of the n^{th} subjet. From those, the ratios $\tau_{21} = \tau_2/\tau_1$ and $\tau_{32} = \tau_3/\tau_2$ are derived. This study uses a "winner-takes-all" (WTA) axis defined based on the direction of the hardest constituent, the ratios τ_{21}^{WTA} and τ_{32}^{WTA} calculated via that definition are the relevant variables in this analysis. As a measure for the broadness of a jet, the Les-Houches angularity (LHA) is used. It is defined as $\mathrm{LHA} = \sum_i z_i \theta_i^{0.5}$, where θ_i is the angle of the i^{th} constituent relative to the jet axis normalised by the jet radius and $z_i = p_{\perp,i}/\sum_j p_{\perp,j}$. As an extension of the split-filtering algorithm presented in [39] (these observables are referred to as "trimmed", $R_{\mathrm{sub}} = 0.2$ subjets with $p_T^i/p_T^{\mathrm{jet}} < 0.05$ being removed), another filtering condition called "soft-drop" is studied:

$$\frac{\min\left(p_{\perp,1}, p_{\perp,2}\right)}{p_{\perp,1} + p_{\perp,2}} > z_{\mathrm{Cut}} \left(\frac{\Delta R_{12}}{R}\right)^{\beta^{\mathrm{SD}}} \stackrel{!}{=} 0.1 \text{ in this analysis}$$

The names of all 44 observables in this analysis are given in Table C.2.

3.2. Event Shape Analyses

Just as one can study the shape of individual jets, in order to understand the properties of the underlying event, it's also possible to look at the shape of the overall event. Several measures can be introduced to characterise the distribution of the transverse momentum in the event.

3.2.1. Event shapes in leptonic Z^0-events (ATLAS_2016_I1424838)

In this analysis, event shape observables were measured in 7 TeV pp collisions in which a Z^0-boson candidate decayed either into an e^+e^- or an $\mu^+\mu^-$ pair. The rationale behind using the Z^0 boson is that it does not possess a colour charge, and thus all of its decay products are also colourless as a whole. Therefore, all variables in this analysis are calculated simply without the decay products of the Z boson to only be sensitive to the physics of the underlying event [40]. Events are selected via the criteria listed in Table 3.3. The origin of the track with the highest associated $\sum(p_T^{\text{trk}})^2$ that satisfies the criteria in Table 3.4 is considered as the primary vertex. The distributions $f_O = 1/N dN/dO$ are measured for the selected events for all observables O, but only particles with $p_T > 0.5$ GeV and $|\eta| < 2.5$ are included in the calculations.

Table 3.3: The cuts applied to the data in the $\sqrt{s} = 7$ TeV event shape analysis [40].

Requirement	e^+e^-	$\mu^+\mu^-$
Primary Vertex	Event must contain a primary vertex (as in Table 3.4)	
Leptons	Two oppositely charged leptons	
Invariant Mass	$m_{\ell^+\ell^-} \in [66, 116]$ GeV	
p_T, $\|\eta\|$	$p_T > 20$ GeV, $\|\eta\| < 2.4$	
Avoiding passive detector material	$\|\eta\| \notin (1.37, 1.52)$	
Transverse impact parameter	$\|d_0\| < 5$ mm	$\|d_0\| < 3 \times \sigma_{d_0}$
Longitudinal impact parameter		$\|z_0\| < 10$ mm

Table 3.4: Conditions for a track to be considered for the search of a "primary vertex" [38].

At least one hit in the ATLAS pixel detector
If the trajectory traverses a module of the innermost pixel layer, the module should have at least one hit
Overall, at least six hits in the semiconductor tracker
The total transverse momentum of the track should be $p_T^{\text{trk}} > 0.5$ GeV
For the pseudorapidity, the track should have $\|\eta^{\text{trk}}\| < 2.5$
It should have a transverse momentum greater than 400 MeV
Impact parameters $\|d_0\| < 1.5$ mm and $\|z_0\| \sin\theta < 1.5$ mm
A goodness-of-fit probability > 0.01 for tracks with $p_T^{\text{trk}} > 10$ GeV

The first relevant variable is the multiplicity of charged particles, N_{ch}. Also, the total transverse momentum for all of these selected particles $\sum p_T$ is measured. Four variables are directly related to the shape of the event. The beam thrust $\mathcal{B}$ is defined analogously to $\sum p_T$, but particles of higher pseudorapidity (travelling more into the direction of the beam) are exponentially suppressed via $\mathcal{B} = \sum_i p_{\perp,i} \cdot e^{-|\eta_i|}$. Defining a thrust axis $\vec{n}_T$, the transverse thrust

$$\mathcal{T} = \max_{\vec{n}_T} \frac{\sum_i |\vec{p}_{\perp,i} \cdot \vec{n}_T|}{\sum_i p_{\perp,i}}$$

over all charged particles is defined. The thrust axis $\vec{n}_T$ by definition maximises $\mathcal{T}$. It is found by implementing an algorithm [17] in which a start direction $\vec{n}_T^{(0)}$ is chosen and iteratively corrected via

$$\vec{n}_T^{(j+1)} = \frac{\sum_i \varepsilon\left(\vec{n}_T^{(j)} \cdot \vec{p}_{\perp,i}\right) \vec{p}_{\perp,i}}{||\sum_i \varepsilon\left(\vec{n}_T^{(j)} \cdot \vec{p}_{\perp,i}\right) \vec{p}_{\perp,i}||}$$

(where $\varepsilon(x) = 2\theta(x) - 1$, $\theta(x)$ being the Heaviside step function). The spheriocity

$$\mathcal{S} = \frac{\pi^2}{4} \min_{\vec{n}=(n_x,n_y,0)^T} \left(\frac{\sum_i |p_{\perp,i} \times \vec{n}|}{\sum_i p_{\perp,i}} \right)^2$$

is defined using a minimising vector $\vec{n}$, which conveniently coincides with one of the transverse momenta $p_{\perp,i}$. The last parameter

$$\mathcal{F} = \frac{\lambda_1}{\lambda_2}$$

is defined by the ratio of the small and large eigenvalues λ_1, λ_2 of the transverse momentum tensor

$$M^{\text{lin}} = \sum_i \frac{1}{p_{\perp,i}} \begin{pmatrix} p_{x,i}^2 & p_{x,i}p_{y,i} \\ p_{x,i}p_{y,i} & p_{y,i}^2 \end{pmatrix}$$

of the charged particles in the event. See Table C.3 for a complete list of variables with their Rivet identifiers.

3.3. UE Studies involving Minimum Bias Samples

To really isolate the physics of the underlying event, samples where the average number of collisions is very low were collected with the ATLAS detector. This allows for the creation of very clean, nearly pile-up free event records. Trigger settings that try to keep as many events as possible are called “minimum bias triggers”. The intention is to attain a representative sample of all collisions that occur in the detector. This is extremely valuable for studying the physics of the underlying event, since the UE is independent of the hard process.

3.3.1. Studies of the UE at 7 TeV with track-jets (ATLAS_2012_I1125575)

This analysis focuses on events that have passed the L1 minimum bias trigger[2] at a center-of-mass energy of 7 TeV. Primary vertices are reconstructed using at least two tracks with $p_T > 0.4$ GeV and $|\eta| < 2.5$, events with several primary vertices are rejected to reduce the contribution from pile-up. The cuts in Table 3.5 are applied, and the jets are reconstructed using the anti-kt algorithm with $R = 0.2, 0.4, 0.6, 0.8, 1.0$. The leading jet is defined as the jet with the largest p_T and $p_T^{\text{jet}} > 5$ GeV, $|\eta| < 1.5$.

Table 3.5: Cuts applied in the $\sqrt{s} = 7$ TeV UE analysis [42].

Requirement	Cut
Tracks in jet reconstruction	$p_T > 0.5$ GeV, $\|\eta\| > 2.5$
Transverse impact parameter	$\|d_0\| < 1.5$ mm
Longitudinal impact parameter	$\|z_0 \sin\theta\| < 1.5$ mm

The leading jet determines three ”regions” of the event. Particles moving in the opposite direction of the leading jet are in the “away” region, the areas perpendicular to the leading track are in the “transverse” region (see Figure 3.1). The ”toward” region is facing in the direction of the leading jet.
The first observable considered in this analysis is the number of charged tracks N_{Ch}. The scalar

[2]For more details on the L1 (Level 1) minimum bias trigger, refer to [41].

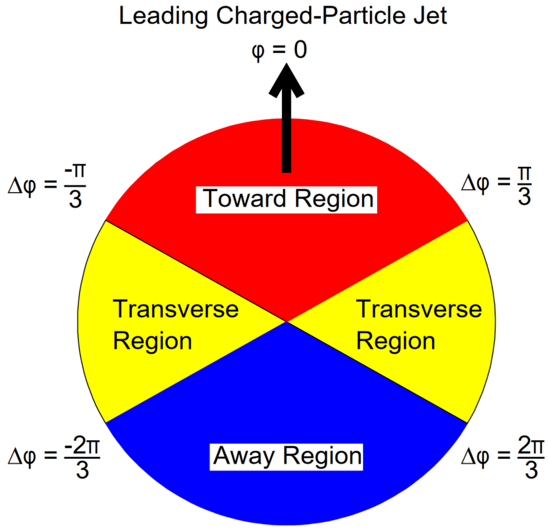


Figure 3.1: Definition of the regions relative to the leading jet in the $\sqrt{s}$ = 7 TeV analysis of the underlying event [42].

sum of the transverse momenta $\sum p_T$ and the average p_T of the tracks ($\sum p_T/N_{\text{Ch}}$) are also studied. They are examined in both the transverse and away regions for various cuts on the leading jet's p_T and as a function of the leading jets p_T. Due to this particular study having 330 observables, we do not provide a table detailing every single variable. Instead, we give instructions for decoding the variable names to their physical meaning (see Tables C.4 and C.5).

3.3.2. Track-based minimum bias at 8 TeV (ATLAS_2016_I1426695)

The $\sqrt{s} = 8$ TeV minimum bias dataset was analysed for the properties of the underlying event. Events were required to have exactly one primary vertex with two or more associated vertices originating from the luminous z-region with a track p_T of at least 100 MeV and a transverse approach to the beam spot closer than 4 mm. A track reconstruction algorithm and various quality criteria described in [43] were applied. The distributions of the charged particle multiplicity, the charged particle p_T, charged particle η and the average particle p_T for tracks with $|\eta| > 2.5$ were measured for various cuts in the phase space by placing requirements on the charged track multiplicity and the p_T of an individual track (see Table C.6).

3.3.3. Track-based UE at 13 TeV in ATLAS (ATLAS_2017_I1509919)

This analysis of $\sqrt{s} = 13$ TeV minimum-bias data is constructed similarly to the analysis of $\sqrt{s} =$ 7 TeV data already presented (see Section 3.3.1). Events are required to contain a primary vertex reconstructed from at least two tracks with $p_T > 100$ MeV and impact parameters below 1.5 mm. The vertex with the highest associated $\sum p_T$ was chosen as the primary. All tracks have to be reconstructed with $|\eta| < 2.5$ and events need to contain at least one track with $p_T > 1$ GeV [44]. The studied variables are listed in Table 3.6.

Table 3.6: Variables studied in the $\sqrt{s}$ = 13 TeV analysis of the UE [44].

Symbol	Description
Binned Variables:	
p_T^{lead}	Leading charged particle transverse momentum
N_{Ch}(Transverse)	Charged particle multiplicity in the transverse region
$\lvert\Delta\phi\rvert$	Absolute azimuthal angle to the leading charged particle
Averaged variables:	
$\langle N_{\text{Ch}}/\delta\eta\delta\phi\rangle$	Mean charged particle multiplicity per unit $\eta - \phi$
$\langle \sum p_T/\delta\eta\delta\phi\rangle$	Mean $\sum p_T$ of charged particles per unit $\eta - \phi$
$\langle$mean $p_T\rangle$	Per event average of the mean p_T of charged particles

Variables are analysed for various cuts on p_T^{lead} and for different regions of the detector, similar to the definition in Figure 3.1. In this analysis, the transverse region with the larger $\sum p_T$ is called ”Trans-max” region, the opposite region is called ”Trans-min”. All variable names are listed in Table C.7.

3.3.4. Minimum bias with low-p_T tracks at 13 TeV (ATLAS_2016_I1467230)

This additional analysis of $\sqrt{s} = 13$ TeV minimum bias data selects events according to the criteria in Table 3.7. Four different observables are measured:

- **Charged particle η**
  Distribution $1/N_{ev} \cdot dN_{ch}/d\eta$, N_{ev} being the number of events with $n_{ch} \geq 2$ and N_{ch} being the total number of charged particles within the kinematic acceptance.
- **Charged particle p_T**
  Distribution $1/N_{ev} \cdot 1/(2\pi p_T)d^2N_{ch}/d\eta dp_T$.
- **Charged multiplicity ≥ 2**
  Distribution $1/N_{ev} \cdot dN_{ev}/dn_{ch}$.
- **Average p_T vs. n_{ch}**
  2D combination of $\langle p_T\rangle$ and n_{ch}.

Table 3.7: Event selection criteria in the low p_T track analysis of $\sqrt{s}$ = 13 TeV minimum bias data [45].

Requirement	Cut
At least two tracks with:	
Transverse momentum	$p_T > 100$ MeV
Pseudorapidity	$\lvert\eta\rvert < 2.5$
Transverse impact parameter	$d_0 < 1.5$ mm
Longitudinal impact parameter	$z_0 \sin\theta < 1.5$ mm
Pixel detector	At least one hit
Innermost Layer	At least one hit if track passes module
If $p_T < 300$ MeV	At least two SCT hits
If $p_T < 400$ MeV	At least four SCT hits
If $p_T > 400$ MeV	At least six SCT hits

They are analysed for a minimum charged particle lifetime of $\tau > 30$ ps and $\tau > 300$ ps for a total of 8 observables in this analysis. Their Rivet identifiers and the associated kinematic acceptance limits can be found in Table C.8.

3.4. Observables Sensitive to CR in $t\bar{t}$ Dilepton Events

Disclaimer: This analysis is not yet published and therefore not included in Rivet. We worked with an internal version and a preliminary set of unfolded data. At the time of this thesis, it does not represent an official result from ATLAS and has yet to pass the review process.

Because the top quark has a very short average transverse decay length below 0.2 fm, it is possible that it interacts with other coloured partons before hadronisation. This particular interaction is described by colour reconnection. In this analysis, the goal is to measure observables particularly sensitive to colour reconnection effects, to revise various models for CR and improve their performance [46]. $t\bar{t}$ events are selected from $\sqrt{s} = 13$ TeV ATLAS data according to these selection criteria:

Table 3.8: Event selection criteria in the $\sqrt{s} = 13$ TeV CR analysis of $t\bar{t}$ dilepton events [46]

Requirement	Cut
Exactly one electron and one muon, opposite charges	
Lepton p_T	> 25 GeV
Number of hadronic jets	$\in \{2, 3\}$
Jet p_T	> 25 GeV
Number of b-tags	$= 2$
Dilepton invariant mass $m_{\ell\ell}$	> 15 GeV
Total number of tracks	≤ 100

Additionally, detector level requirements are applied (described in detail in [46]).
The sensitive observables involved are:

- **Charged particle multiplicity**
  After all tracks associated with the charged leptons and identified jets are removed, the number of charged tracks n_{ch} is counted. The removal of jets is achieved by excluding all particles within $\Delta R < 0.4$ of the jet axis. This observable is calculated per-event and is referred to as *MC_ttbar_ColorReconnection/nch* in the context of the Rivet analysis.

- **Scalar sum of transverse momenta**
  The sum of transverse momenta ($\sum p_T$) of the tracks used in the definition of N_{ch}. In Rivet, it is called *MC_ttbar_ColorReconnection/sumPt*.

- **$\sum p_T$ vs. n_{ch}**
  This observable is the 2D combination of $\sum p_T$ and n_{ch} and is reffered to as *MC_ttbar_ColorReconnection/sumPt_vs_nch*. At the time of this thesis, unfolded distributions for this variable are not yet available. We thus do not include it in our tunes.

Note that as of now, unfolded distributions to "particle-level" are only available for n_{ch} and $\sum p_T$. Therefore, all other observables in this analysis are not included in the tunes, even though they might be covered in the sensitivity analyses performed.

3.5. K_S^0 and Λ^0 Production at 7 TeV (ATLAS_2019_I1746286)

In this analysis, a data set of dileptonic $t\bar{t}$ events at a centre-of-mass energy of 7 TeV was studied. The focus is on observables of the production of K_S^0 and Λ^0 particles. The criteria for event selection can be found in [47]. The observables of the K_S^0 particles break down according to the regions in which the particles were observed. Observables were measured for particles in b-jets, non-b-tagged jets, outside of jets and in the entire event. The transversal momentum p_T, the energy fraction $x = E_{\text{particle}}/E_{\text{jet}}$, the particle energy E_{particle}, the pseudorapidity $|\eta|$ and the multiplicity were analysed. The observables of the Λ^0 particles were only recorded in the entire event. An overview over the observables provided by this analysis is given in Table 3.9 along with their Rivet identifiers.

Table 3.9: Observables of the $\sqrt{s} = 7$ TeV analysis of K_S^0 and Λ^0 particles [48].

	K_S^0				
Region	p_T	x_K	Energy	$\|\eta\|$	Multiplicity
Inside b-jets	d01-x01-y01	d02-x01-y01	d03-x01-y01	d04-x01-y01	d05-x01-y01
Inside non-b-jets	d06-x01-y01	d07-x01-y01	d08-x01-y01	d09-x01-y01	d10-x01-y01
Outside jets	d11-x01-y01	-	d12-x01-y01	d13-x01-y01	d14-x01-y01
Total Sample	d15-x01-y01	-	d16-x01-y01	d17-x01-y01	d18-x01-y01
Region	Λ^0				
Total Sample	d19-x01-y01	-	d20-x01-y01	d21-x01-y01	-

3.6. Aliases

Since the rivet identifiers of the analyses are based on the Inspire IDs of the associated papers, the assignment is very unintuitive. In order to improve the readability of this tesis, we have introduced our own aliases from which the physics dealt with in the analyses should become clear. These can be found in Table 3.10.

Table 3.10: Aliases for the Rivet analyses used in this thesis.

Process	Analysis	Energy	Alias
Minimum bias	ATLAS_2012_I1125575	7 TeV	Minimum bias @ 7 TeV
	ATLAS_2016_I1426695	8 TeV	Minimum bias @ 8 TeV
	ATLAS_2017_I1509919	13 TeV	Minimum bias @ 13 TeV
	ATLAS_2016_I1467230	13 TeV	Low p_T minimum bias
Z	ATLAS_2016_I1424838	7 TeV	Z + jets
$t\bar{t}$	ATLAS_2013_I1243871	7 TeV	Jet shapes
	ATLAS_2019_I1746286	7 TeV	$K_S^0 + \Lambda^0$
	ATLAS_2019_I1724098	13 TeV	Jet substructure
	MC_ttbar_ColorReconnection	13 TeV	$t\bar{t}$ CR

4

Model Sensitivity Studies

When tuning a mathematical model to a dataset, it is very helpful to evaluate the model's sensitivity to its input parameters. In MC tuning, however, each histogram bin is modelled individually. Since most variable histograms consist of several bins, each of which may have a very different sensitivity to the parameterisation, commonly used sensitivity evaluation techniques like the Morris method [49] are not a good fit for our application. For ranking the entire number of variables by their sensitivity, the effect of all bins should be taken into account. We also have to consider that individual variables can have different mean values and numbers of bins.

4.1. Proposed Method

Because the models under investigation are evaluated on a bound parameter space (Equation (1.4)), the range of values of the j$^{\text{th}}$ modelling polynomial of an observable is also a bound interval

$$D_j = \left[\min\left(\{P_j(\vec{p}) : \forall \vec{p} \in \mathcal{P}\}\right), \max\left(\{P_j(\vec{p}) : \forall \vec{p} \in \mathcal{P}\}\right)\right].$$

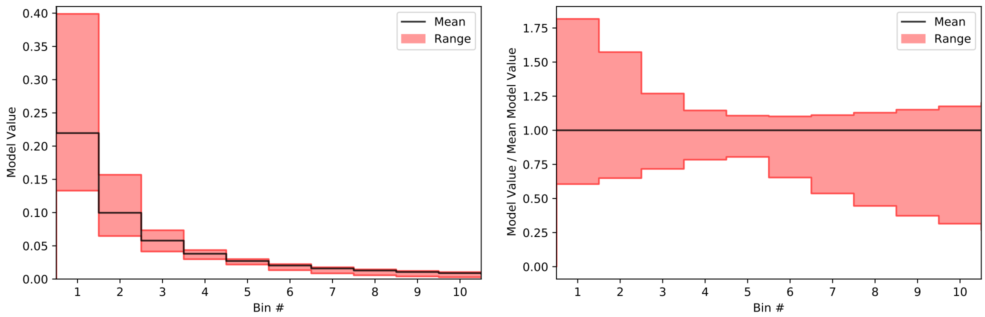
The magnitude of the values of the bin relevant for the tune are indicated by the value of the polynomial at the point of best fit $\vec{p}^{\text{BF}}$. We therefore define for an observable with m bins the total sensitivity

$$c = \frac{1}{m}\sum_{j=1}^{m} \frac{|D_j|}{P_j\left(\vec{p}^{\text{BF}}\right)}. \tag{4.1}$$

As an academic example, assume a histogram with 10 bins with a content that is modelled via the lognormal distribution function

$$f(n,\sigma) = \frac{1}{\sqrt{2\pi}\sigma n}\exp\left(-\frac{\ln(n)^2}{2\sigma^2}\right), \tag{4.2}$$

n being the bin number and $\sigma \in [1, 3]$ a free parameter of the model. In this case, the magnitude of each bin is determined by the mean model value. We observe that the sensitivity rises for low-variation bins that also have a low model output, as can be seen in Figure 4.1 (b). Because many MC observables tend to be statistical distributions, a low bin value is often associated with low statistics, which is in turn associated with a higher relative statistical uncertainty. Therefore, the range of values of the modelling function might not be very meaningful in low valued bins.



(a) Mean, minimum and maximum model values under variation of σ.

(b) The resulting binwise contribution to c_σ is the range of values relative to the mean value of each bin.

Figure 4.1: Model and ratio plot from Equation (4.2). The resulting sensitivity measure (Equation (4.1)) evaluates to $c_\sigma = 0.667$.

We have therefore decided to introduce a multiplicative penalty term that limits the impact of such bins on the calculated sensitivity. With

$$\overline{P\left(\vec{p}^{\text{BF}}\right)} = \frac{1}{m}\sum_{j=1}^{m} P_j\left(\vec{p}^{\text{BF}}\right)$$

being the mean magnitude of the observable, we define the linear penalty term

$$S_j = \frac{P_j\left(\vec{p}^{\text{BF}}\right)}{\overline{P\left(\vec{p}^{\text{BF}}\right)}}.$$

From this follows the definition of a second measure of sensitivity

$$c^* = \frac{1}{m}\sum_{j=1}^{m} S_j \cdot \frac{|D_j|}{P_j\left(\vec{p}^{\text{BF}}\right)} = \frac{1}{m}\sum_{j=1}^{m} \frac{P_j\left(\vec{p}^{\text{BF}}\right)}{\overline{P\left(\vec{p}^{\text{BF}}\right)}} \cdot \frac{|D_j|}{P_j\left(\vec{p}^{\text{BF}}\right)} = \frac{1}{m}\sum_{j=1}^{m} \frac{|D_j|}{\overline{P\left(\vec{p}^{\text{BF}}\right)}}. \tag{4.3}$$

Evaluating the toy model (Equation (4.2)) a second time, we observe that the effect of low value bins is now mitigated under the c^* measure (see Figure 4.2). In order to also assess the sensitivity to changes in individual input parameters, we extend the definitions 4.1 and 4.3. Previously, we considered the range of values over the entire phase space of the input parameters. Now we change our approach by using a subspace

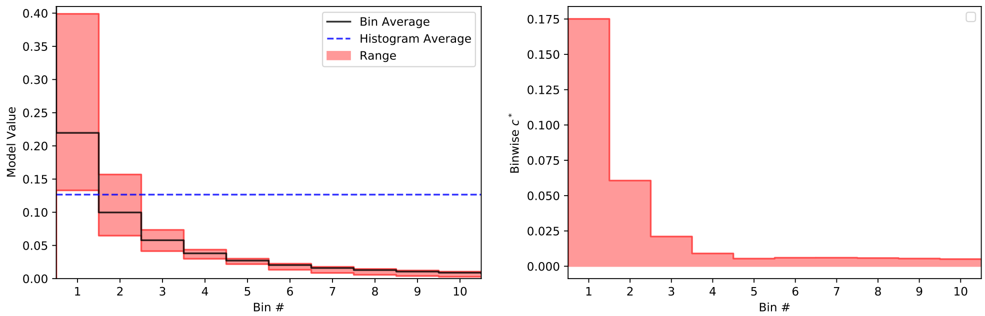
$$\mathcal{L}_i = \left\{\left(p_1^{\text{BF}}, \dots, p_{i-1}^{\text{BF}},\ x,\ p_{i+1}^{\text{BF}}, \dots, p_n^{\text{BF}}\right) : \forall x \in \left[p_i^{\text{low}}, p_i^{\text{high}}\right]\right\} \subseteq \mathcal{P}$$

of the parameter space in which all parameters are fixed except for the one under investigation, given by the index i. This subspace $\mathcal{L}_i$ represents a one-dimensional line in the n-dimensional parameter space with length

$$|\mathcal{L}_i| = p_i^{\text{high}} - p_i^{\text{low}}.$$

The sensitivity of the model to changes in only one input parameter is now dependent on the range of values

$$D_j^{\mathcal{L}_i} = \left[\min\left(\left\{P_j(\vec{p}) : \forall \vec{p} \in \mathcal{L}_i\right\}\right), \max\left(\left\{P_j(\vec{p}) : \forall \vec{p} \in \mathcal{L}_i\right\}\right)\right]$$



(a) Mean, minimum and maximum model values under variation of σ.

(b) The resulting binwise contribution to c^*.

Figure 4.2: Model and binwise c^* plot from Equation (4.2). The resulting sensitivity measure (Equation (4.3)) is calculated by summing over the binwise plot and evaluates to $c^* = 0.30$.

of the model on this line. The magnitude of a modelling polynomial on the line is given by

$$\overline{P_j}_{\mathcal{L}_i} = \frac{1}{|\mathcal{L}_i|}\int_{\mathcal{L}_i} P_j(\vec{x})d\vec{x} = \frac{1}{p_i^{\text{high}} - p_i^{\text{low}}}\int_{p_i^{\text{low}}}^{p_i^{\text{high}}} P_j(p_1^{\text{BF}}, \ldots, x, \ldots, p_n^{\text{BF}})dx,$$

we define the magnitude of the entire histogram as

$$\overline{P}_{\mathcal{L}_i} = \frac{1}{m}\sum_{j=1}^{m}\overline{P_j}_{\mathcal{L}_i}.$$

Thus we define the parameter-wise sensitivities as

$$c_i = \frac{1}{m}\sum_{j=1}^{m}\frac{|D_j^{\mathcal{L}_i}|}{\overline{P_j}_{\mathcal{L}_i}} \tag{4.4}$$

and

$$c_i^* = \frac{1}{m}\sum_{j=1}^{m}\frac{|D_j^{\mathcal{L}_i}|}{\overline{P}_{\mathcal{L}_i}}. \tag{4.5}$$

4.2. 1D Example Sensitivity Study

Using the one-dimensional tuning problem from Section 1.5.8, we can also use the data set to check whether our sensitivity measures work as intended. Foremost, we found that some bins have statistics which are so low that the model prediction is consistently zero. We therefore define these bins to be entirely insensitive to changes in the input parameters. The top ten variables ordered by c_i are listed in Table 4.1. The variable with the highest c_i is */ATLAS_2019_I1724098 /d29-x01-y01*, which is the normalized cross section of the $t\bar{t}$ process depending on the number of subjets. The variable peaks at $N_{\text{subjets}} = 3, 4$ and is very low at other values of N_{subjets} (see Figure 4.3).

Table 4.1: Top 10 most sensitive variables from the 1D tuning exercise in Section 1.5.8 when sorted by c_i.

Variable	c_i	c_i^*
/ATLAS_2019_I1724098/d29-x01-y01	1.2858	0.0536
/MC_ttbar_ColorReconnection/nch	0.8501	0.4066
/ATLAS_2019_I1724098/d20-x01-y01	0.5561	0.1117
/MC_TTBAR:TTMODE=ANYLEP/anylep_jetb_1_l2_deta	0.5255	0.0290
/ATLAS_2019_I1724098/d42-x01-y01	0.4861	0.0718
/ATLAS_2019_I1724098/d07-x01-y01	0.4565	0.0462
/ATLAS_2019_I1724098/d17-x01-y01	0.3929	0.1249
/MC_ttbar_ColorReconnection/sumPt	0.3923	0.3070
/ATLAS_2019_I1724098/d21-x01-y01	0.3919	0.1324
/ATLAS_2019_I1724098/d12-x01-y01	0.3831	0.1183

Since the low bin values are mitigated by the penalty term, the importance of the variable decreases significantly when sorting for c_i^*. The new dominant observable is

/MC_ttbar_ColorReconnection/nch,

which measures the number of charged particles outside of jets. The ten most sensitive variables are shown in Table 4.2, a visual comparison between both top places is provided by Figure 4.3.

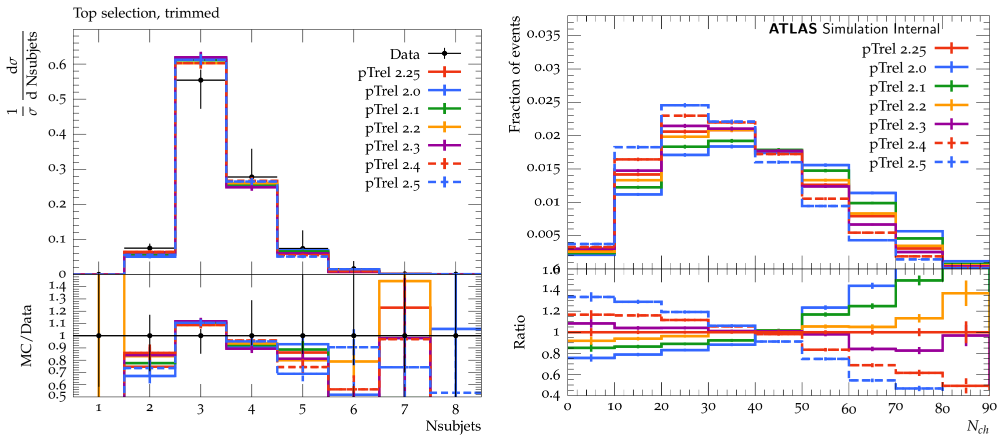
Table 4.2: Top 10 most sensitive variables from the 1D tuning exercise in Section 1.5.8 when sorted by c_i^*.

Variable	c_i	c_i^*
/MC_ttbar_ColorReconnection/nch	0.8501	0.4066
/MC_ttbar_ColorReconnection/sumPt	0.3923	0.3070
/ATLAS_2019_I1724098/d14-x01-y01	0.1910	0.1563
/ATLAS_2019_I1724098/d09-x01-y01	0.3678	0.1480
/MC_ttbar_ColorReconnection/avgPt	0.2658	0.1457
/ATLAS_2019_I1724098/d36-x01-y01	0.1873	0.1331
/ATLAS_2019_I1724098/d21-x01-y01	0.3919	0.1324
/ATLAS_2019_I1724098/d33-x01-y01	0.2489	0.1314
/ATLAS_2019_I1724098/d30-x01-y01	0.2565	0.1256
/ATLAS_2019_I1724098/d17-x01-y01	0.3929	0.1249

From these results we can already see that our approach suppresses low content bins as intended. We therefore will continue to use c_i^* as our measure of sensitivity and for excluding observables from the tunes.

4.3. Sensitivities to the CR2 Model Parametrisation

Monte Carlo samples of Z^0 decays at $\sqrt{s} = 7$ TeV, soft QCD events with $\sqrt{s} = 7$, 8 and 13 TeV, and $t\bar{t}$ events at $\sqrt{s} = 7$ and 13 TeV were generated as described in Section 5.1. To obtain the sensitivity of a variable, we need to know the point of best fit on which the sensitivity measure is based (see Section 4.1). Since the analysis of the sensitivities is a prerequisite for the selection of the observables for the final tune, we approximate the point of best fit by means of a ”naive” tune. This tune is presented in Section 5.2. In the following subsections, we break down the sensitivities of the variables of the individual analyses and compare them with each other.



(a) Most sensitive variable using the c_i measure (/ATLAS_2019_I1724098/d29-x01-y01).

(b) Most sensitive variable under the c_i^* measure (/MC_ttbar_ColorReconnection/nch).

Figure 4.3: Most sensitive variables under both the unpenalised as well as the penalised sensitivity measure.

4.3.1. Sensitivities of event shapes in leptonic Z^0-events

The analysis of event-shapes in leptonic Z^0-events described in Section 3.2.1 has 60 different observables. We therefore only present the statistics of the sensitivity analysis performed, which can be found in Table 4.3.

Table 4.3: Sensitivities of the Z-event-shape analysis (described in Section 3.2.1) to the input parameters under the c^* measure.

Parameter	Mean sensitivity	Standard deviation	Minimum	Maximum
Combined	1.48	0.21	1.17	1.97
pt0Ref	1.36	0.32	0.94	2.02
expPow	0.42	0.08	0.31	0.60
m2Lambda	0.084	0.04	0.05	0.19
fracGluon	0.11	0.05	0.05	0.22
dLambdacut	0.12	0.04	0.06	0.20

4.3.2. Sensitivities of observables involving minimum bias samples

In total, Monte Carlo samples were produced for four different minimum bias analyses. For $\sqrt{s} = 7$ TeV, the sensitivities of the analysis presented in Section 3.3.1 are as in Table 4.4. The results at $\sqrt{s} = 8$ TeV are presented in Table 4.5. The combined statistics for both analyses at $\sqrt{s} = 13$ TeV are given in Table 4.6.

Table 4.4: Statistics on the sensitivities of the $\sqrt{s} = 7\,\text{TeV}$ minimum bias analysis presented in Section 3.3.1. This analysis has 330 observables.

Parameter	Mean sensitivity c^*	Standard deviation	Minimum	Maximum
Combined	2.67	2.14	0.29	17.33
pt0Ref	1.38	0.47	0.10	2.28
expPow	0.66	0.54	0.04	3.91
m2Lambda	0.31	0.39	0.03	3.35
fracGluon	0.31	0.37	0.05	3.46
dLambdacut	0.41	0.49	0.08	4.26

Table 4.5: Statistics on the sensitivities of the $\sqrt{s} = 8\,\text{TeV}$ minimum bias analysis presented in Section 3.3.2. This analysis has 14 observables.

Parameter	Mean sensitivity c^*	Standard deviation	Minimum	Maximum
Combined	1.59	0.88	0.28	3.03
pt0Ref	0.62	0.23	0.21	0.88
expPow	0.27	0.18	0.01	0.50
m2Lambda	0.07	0.04	0.02	0.13
fracGluon	0.08	0.05	0.01	0.18
dLambdacut	0.14	0.09	0.02	0.30

Table 4.6: Statistics on the sensitivities of the $\sqrt{s} = 13\,\text{TeV}$ minimum bias analyses.

(a) Sensitivity statistics for the track-based analysis of the UE described in Section 3.3.3. This analysis has 32 observables.

Parameter	Mean sensitivity c^*	Standard deviation	Minimum	Maximum
Combined	1.48	1.17	0.21	3.89
pt0Ref	0.63	0.43	0.04	1.34
expPow	0.47	0.46	0.02	1.36
m2Lambda	0.12	0.09	0.02	0.36
fracGluon	0.07	0.03	0.02	0.15
dLambdacut	0.12	0.04	0.04	0.24

(b) Sensitivity statistics for the track-based analysis with low-p_T tracks described in Section 3.3.4. This analysis has 8 observables.

Parameter	Mean sensitivity c^*	Standard deviation	Minimum	Maximum
Combined	1.77	1.23	0.29	3.14
pt0Ref	0.62	0.29	0.22	0.90
expPow	0.20	0.16	0.01	0.46
m2Lambda	0.04	0.02	0.02	0.06
fracGluon	0.12	0.04	0.08	0.18
dLambdacut	0.19	0.08	0.10	0.31

4.3.3. Sensitivities of observables involving $t\bar{t}$ samples

The analysis of jet shapes in $\sqrt{s} = 7$ TeV $t\bar{t}$ events described in Section 3.1.1 has 20 observables. The statistics on the sensitivities obtained with our method is found in Table 4.7. Statistics on the 21 observables of the $\sqrt{s} = 7$ TeV production of K_S^0 and Λ^0 (see Section 3.5) are shown in Table 4.8.

Two analyses relevant to our tuning efforts were applied to the $t\bar{t}$ samples at $\sqrt{s} = 13$ TeV. The sensitivities of the observables of the jet shape analysis described in Section 3.1.2 are given in Table 4.9. Finally, the analysis of observables sensitive to CR in $t\bar{t}$ dilepton events described in Section 3.4 only has two observables, $\sum p_T$ and n_{ch}. Their respective sensitivities are given in Table 4.10.

Table 4.7: Statistics for the $\sqrt{s} = 7$ TeV analysis of jet shapes in $t\bar{t}$ events described in Section 3.1.1.

Parameter	Mean sensitivity c^*	Standard deviation	Minimum	Maximum
Combined	0.058	0.04	0.01	0.13
pt0Ref	0.041	0.03	0.006	0.10
expPow	0.013	0.01	0.002	0.03
m2Lambda	0.004	0.003	0.001	0.01
fracGluon	0.003	0.002	0.0003	0.01
dLambdacut	0.003	0.003	0.0005	0.01

Table 4.8: Statistics for the $\sqrt{s} = 7$ TeV analysis of K_S^0 and Λ^0 production in $t\bar{t}$ events described in Section 3.5.

Parameter	Mean sensitivity c^*	Standard deviation	Minimum	Maximum
Combined	0.91	0.85	0.00	2.37
pt0Ref	0.52	0.44	0.00	1.22
expPow	0.27	0.24	0.00	0.64
m2Lambda	0.03	0.03	0.00	0.10
fracGluon	0.04	0.04	0.00	0.12
dLambdacut	0.08	0.07	0.00	0.21

Table 4.9: Statistics for the $\sqrt{s} = 13$ TeV analysis of jet shapes in $t\bar{t}$ events described in Section 3.1.2.

Parameter	Mean sensitivity c^*	Standard deviation	Minimum	Maximum
Combined	0.73	0.39	0.14	1.41
pt0Ref	0.21	0.14	0.02	0.59
expPow	0.22	0.12	0.02	0.41
m2Lambda	0.18	0.10	0.03	0.33
fracGluon	0.13	0.07	0.02	0.23
dLambdacut	0.16	0.09	0.03	0.29

Table 4.10: Sensitivities for the $\sqrt{s}$ = 13 TeV analysis of observables sensitive to CR in $t\bar{t}$ events described in Section 3.4.

Observable	Combined	pt0Ref	expPow	m2Lambda	fracGluon	dLambdacut
n_{ch}	1.09	0.89	0.43	0.05	0.14	0.12
$\sum p_T$	1.02	0.71	0.35	0.08	0.10	0.14

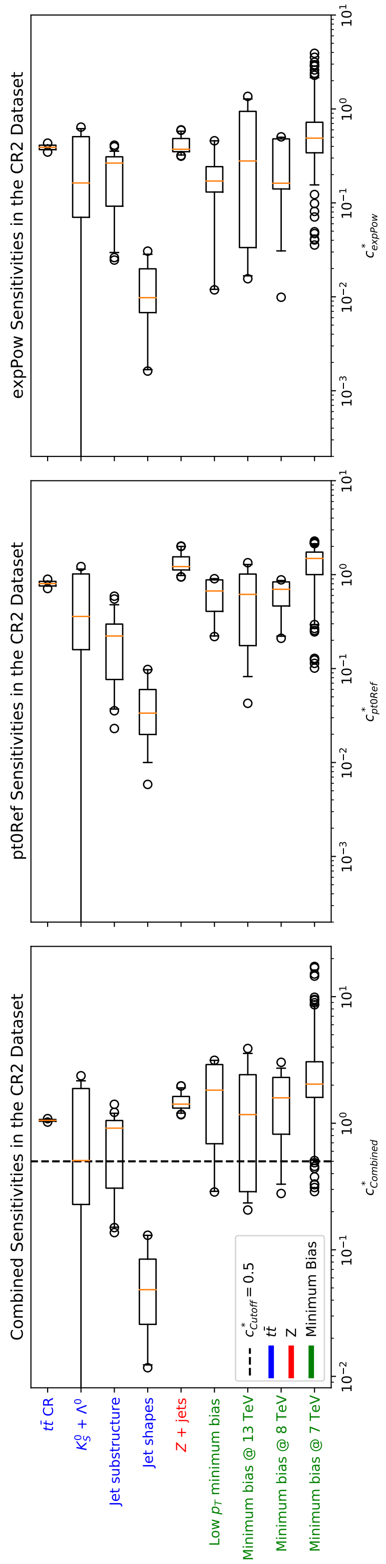
4.3.4. Comparison of the analyses used

A visual comparison between the different analyses is provided in Figure 4.4. Minimum bias events at $\sqrt{s}$ = 7 TeV are particularly sensitive to all input parameters under the CR2 model. The associated analysis includes 330 different observables, but many of them represent the same physics (see Section 3.3.1). It is therefore advisable to limit the number of observables used to a few selected ones or limit their influence by choosing a lower weight. Otherwise, there is a possibility that this one process is overweighted in the final tune and other processes have an insufficient influence on the point of best fit. Apart from this, the observables of the minimum bias and Z^0 analyses seem to be similarly sensitive. The observables of the $t\bar{t}$ processes at $\sqrt{s}$ = 7 TeV react in the median by more than one order of magnitude less sensitively to changes in the model parameterisation than observed in the other analyses. Overall, the analyses of the minimum bias processes are more sensitive in the median ($\tilde{c^*}_{\text{Minimum Bias}} = 2.01$) than the analyses of the Z-decays ($\tilde{c^*}_{Z^0} = 1.42$) or $t\bar{t}$-decays ($\tilde{c^*}_{t\bar{t}} = 0.37$).

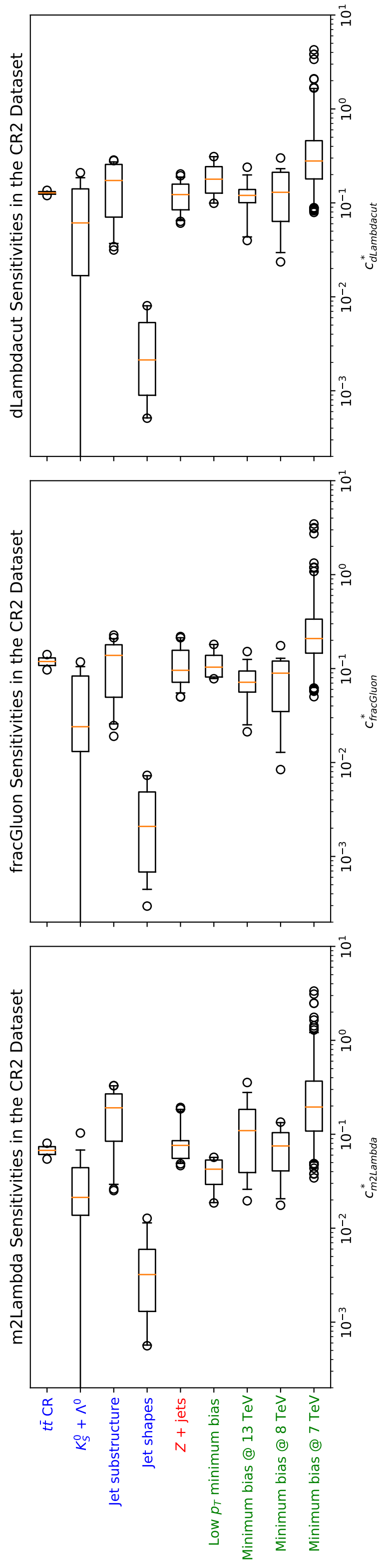
Table 4.11: Median sensitivities sorted by process and parameter.

Analyses	Combined	pt0Ref	expPow	m2Lambda	fracGluon	dLambdacut
All	1.71	1.16	0.43	0.13	0.16	0.20
Minimum Bias	2.01	1.35	0.48	0.18	0.18	0.23
Z^0	1.42	1.22	0.37	0.08	0.10	0.12
$t\bar{t}$	0.37	0.19	0.12	0.05	0.05	0.08

As shown in Table 4.11, the sensitivity of all processes to *MultipartonInteractions:pT0Ref* is greater than to other individual parameters. All analyses show a lower sensitivity to the parameters of the colour reconnection model (*ColourReconnection:m2Lambda*, *fracGluon* and *dLambdaCut*) than to the parameterisation of the multi parton interactions (*MultipartonInteractions:pt0Ref* and *expPow*).



(a) Sensitivities for changes in all input parameters, pT0Ref and expPow in the CR2 dataset.



(b) Sensitivities for changes m2Lambda, fracGluon and dLambdaCut in the CR2 dataset.

Figure 4.4: Distribution of sensitivities for all analyses of the CR2 Monte Carlo samples. Whiskers are drawn such that the lowest and highest 2.5% of data points are treated as outliers (black circles). The yellow lines represent the median sensitivities. The box represents the inner 50 % of data points.

5

Parameter Tunes

Due to time constraints, we have chosen to examine only the CR2 model in this thesis. In this chapter, we discuss the samples and the method we used. Finally, we present three parameter tunes and compare their performance.

5.1. Monte Carlo Samples

The Pythia 8 default settings of the A14 (NNPDF23LO) tune were adopted for all processes. In addition, the settings of the colour reconnection model were set to the values in Table 5.1. Samples of minimum bias, $t\bar{t}$ and Z^0 processes at different energies were generated for this study. We uniformly sampled 300 points from the parameter space shown in Table 5.2.

Table 5.1: Our settings for the CR2 model (described in Section 2.3), in addition to the A14 NNPDF23LO settings [18].

Parameter	Value	A14 (NNPDF23LO)
ColourReconnection:mode	2	0
ColourReconnection:flipMode	0	-
PartonLevel:earlyResDec	off	off

Table 5.2: Sampling ranges defining our CR2 parameter space and the associated Pythia 8 defaults.

Parameter	Minimum	Maximum	Default	A14 (NNPDF23LO)
ColourReconnection:fracGluon	0.00	1.00	1.00	-
ColourReconnection:dLambdaCut	0.00	10.00	0.00	-
ColourReconnection:m2Lambda	0.25	16.00	1.00	-
MultipartonInteractions:pT0Ref	1.20	3.00	2.15	2.09
MultipartonInteractions:expPow	0.40	3.00	1.00	-

5.1.1. Minimum bias

Pythia 8 was set to produce only soft inelastic QCD interactions by setting

SoftQCD:inelastic = on.

2 million events were generated for each parameter point at three collision energies and subjected to the Rivet analyses listed in Table 5.3.

Table 5.3: The analyses and collision energies used in the minimum bias CR2 dataset.

$\sqrt{s}$	Analyses	Documentation
7 TeV	ATLAS_2012_I1125575	Section 3.3.1
8 TeV	ATLAS_2016_I1426695	Section 3.3.2
13 TeV	ATLAS_2017_I1509919	Section 3.3.3
	ATLAS_2016_I1467230	Section 3.3.4

5.1.2. Z boson decays

Five hundred thousand decays of Z bosons at a pp collision energy of $\sqrt{s} = 7$ TeV were simulated per parameter point and processed by the event shape analysis discussed in Section 3.2.1. In addition to the A14 (NNPDF23LO) settings, the Pythia 8 settings from Table 5.4 were used. The PHOTOS algorithm [50] with *InfraRedCutOff* $= 10^{-7}$ was additionally integrated for the QED corrections.

Table 5.4: Settings for the generation of Z^0 events.

Parameter	Value
WeakSingleBoson:ffbar2gmZ	on
23:onMode	off
23:onIfAny	11 -11
23:onIfAny	13 -15
23:onIfAny	15 -15
PhaseSpace:mHatMin	60.0
TimeShower:QEDshowerByL	off
TimeSHower:QEDshowerByOther	off

5.1.3. Top-quark-antiquark events

In the case of the $t\bar{t}$ events, the quark-antiquark pair and the associated decay were simulated by using matrix elements calculated via the `POWHEG` algorithm [51] at next-to-leading-order. The entire hard process and the first gluon emmision were calculated by `POWHEG` and then passed to Pythia 8. `POWHEG` was called with the parameters listed in Table 5.6 (a). Additionally, the settings in Table 5.6 (b) were applied to Pythia 8. One million events at centre-of-mass energies of 7 TeV and 13 TeV were generated per parameter point. They were processed with the corresponding analyses listed in Table 5.5.

5.1.4. Timing issues

At first, we used values of *MultipartonInteractions:pT0Ref* $\in [0.5, 3.0]$. During the generation of the Monte Carlo samples, it was noticeable that some parameter space points had significantly

Table 5.5: The analyses and collision energies used in the $t\bar{t}$ CR2 dataset.

$\sqrt{s}$	Analyses	Documentation
7 TeV	Jet shapes	Section 3.1.1
7 TeV	$K_S^0 + \Lambda^0$	Section 3.5
13 TeV	Jet substructure	Section 3.1.2
	$t\bar{t}$ CR	Section 3.4

Table 5.6: Settings used for the Monte Carlo sample generation of $t\bar{t}$ events.

(a) POWHEG settings for the generation of $t\bar{t}$ events.

Parameter	Value
decay_mode	t t~ > all
hdamp	258.75
mu_F	1.0
mu_R	1.0
PDF	260000

(b) Pythia 8 settings for the generation of $t\bar{t}$ events.

Parameter	Value
SpaceShower:pTmaxMatch	2
TimeShower:pTmaxMatch	2
Powheg:pTHard	0
Powheg:NFinal	2
Powheg:pTdef	2
Powheg:veto	1
Powheg:vetoCount	3
Powheg:pTemt	0
Powheg:emitted	0
Powheg:MPIveto	0

longer computation times than the rest of the sample. In order to find out whether the choice of input parameters correlates with the calculation time, additional timing information was collected, as shown in Figure 5.1. The parameter-wise correlations are listed in Table 5.7.

Table 5.7: Correlation between parameter value and computation time for all of our samples in the CR2 tuning effort.

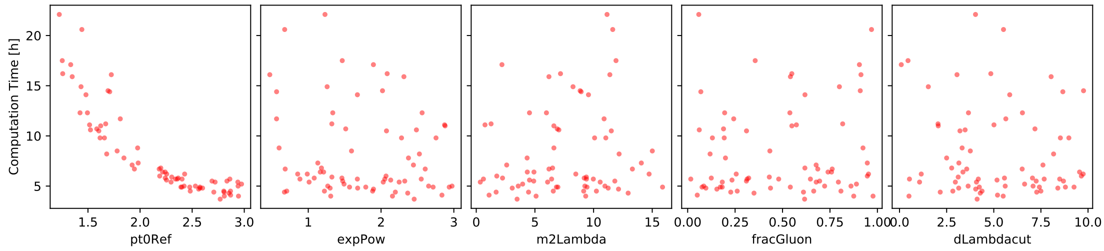
	MPI:pT0Ref	MPI:expPow	CR:m2Lambda	CR:fracGluon	CR:dLambdacut
Minimum bias at 7 TeV	-0.87	-0.16	0.14	0.07	-0.11
Minimum bias at 8 TeV	-0.72	-0.25	0.01	-0.09	-0.17
Minimum bias at 13 TeV	-0.72	-0.29	0.04	-0.07	-0.12
Z^0 decays at 7 TeV	-0.60	-0.47	0.01	-0.05	-0.15
$t\bar{t}$ at 7 TeV	-0.67	-0.35	-0.13	-0.11	-0.19
$t\bar{t}$ at 13 TeV	-0.84	-0.20	0.17	-0.61	0.12

Both visually and from the calculated correlation coefficients, it is clear that low values of *MultipartonInteractions:pT0Ref* in particular result in high computation times. In order to reduce the total calculation time, we have therefore limited ourselves to the parameter space in Table 5.2 with *MultipartonInteractions:pT0Ref* $\in [1.2, 3.0]$. Points that did not satisfy that requirement were resampled and computed again.

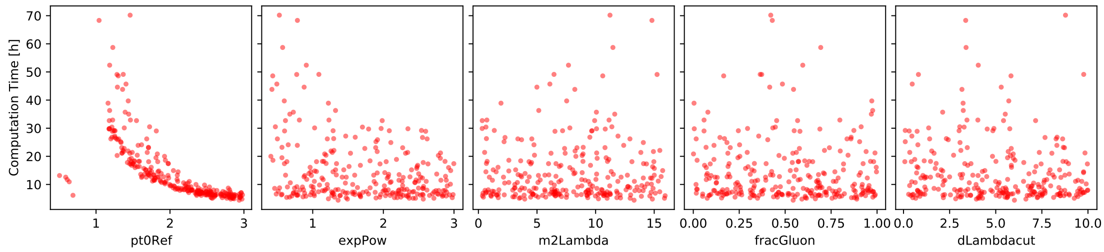
5.2. Naive Tune

In order to get a first overview of the data set and an approximate point of best fit for the analysis of the sensitivities (see Section 4.3), a tune was first performed including all observables from all generated analyses. The weights of all observables were chosen as $\omega_{\mathcal{O}} \equiv 1$. Specifically, all observables of the analyses in Table 3.10 were used.

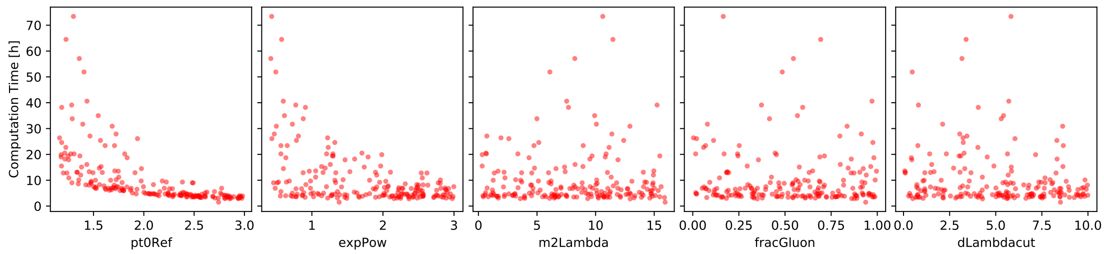
A total of 485 observables were included in the naive tune, resulting in 7585 degrees of freedom. Our goodness of fit function (see Equation (1.7)) takes a value of $\chi^2 = 193785.82$ at the point of best fit, neglecting correlations between the individual bins and the various observables. The point of best fit in Table 5.8 is therefore reached with $\chi^2/\text{N.D.F.} = 25.5$. Note that the value of *CR:fracGluon* is at the boundary of the parameter space. In the CR2 model this means that under the naive tune, all gluons participate in the colour reconnection process.



(a) Minimum bias at $\sqrt{s} = 7\,\text{TeV}$



(b) Minimum bias at $\sqrt{s} = 13\,\text{TeV}$



(c) Z^0 decays at $\sqrt{s} = 7\,\text{TeV}$

Figure 5.1: Computation times for multiple processes. Note that the dataset is incomplete, because the timing information was not collected from the very beginning of the sample generation.

Table 5.8: Point of best fit for the naive tune on the CR2 data set.

Parameter	MPI:pt0Ref	MPI:expPow	CR:m2Lambda	CR:fracGluon	CR:dLambdacut
Value	2.30	1.75	2.81	1.00	1.86

5.3. Selected Tune

Instead of including all available observables in the tune, we now set minimum requirements for the sensitivity and quality of modelling. For this purpose, we first apply the bin-wise envelopment test described in Section 1.5.5 and require that for at least 80% of the bins of an observable, the 95% confidence interval of the reference data is intersected by the value range of the model function. In this way, we avoid including observables in the tune for which the CR model has no predictive power. We therefore limit the tune to the physical observables that the Monte Carlo model can actually reproduce. In addition, we require a minimum sensitivity of the observables. If an observable is insensitive, it makes no variable contribution to the goodness of fit function, more or less independently of the parameterisation. For this reason, observables with $c^*_{\text{Combined}} < c^*_{\text{Cutoff}} = 0.5$ are also not considered (as illustrated in Figure 4.4 (a)). Based on these criteria, 98 observables were excluded from this ”selected” tune. The fraction of remaining observables per analysis and process is shown in Table 5.9. These were each weighted with $\omega_{\mathcal{O}} = 1$. An overview over all observables of the tune is provided in Appendix D. The resulting fit has 6996 degrees of freedom and minimises with $\chi^2 = 114033.87$. The point of best fit (see Table 5.10) is therefore reached with $\chi^2/\text{N.D.F.} = 16.3$.

Table 5.9: Statistics on observable usage in the CR2 "selected" tune.

Process	Comment	Energy	Observables used	Observables excluded	Percentage used
Minimum bias		7 TeV	300	30	90.9 %
		8 TeV	12	2	85.7 %
		13 TeV	20	12	62.5 %
	Low p_T	13 TeV	6	2	75.0 %
Z		7 TeV	58	2	96.7 %
$t\bar{t}$	Jet shapes	7 TeV	0	20	00.0 %
	$K^0_S + \Lambda^0$	7 TeV	7	14	33.3 %
	Jet substructure	13 TeV	28	16	63.6 %
	CR	13 TeV	2	0	100.0 %

Table 5.10: Point of best fit for the selected tune on the CR2 data set.

Parameter	MPI:pt0Ref	MPI:expPow	CR:m2Lambda	CR:fracGluon	CR:dLambdacut
Value	2.23	2.07	0.32	0.78	1.05

5.4. Reweighting

As shown in Table 5.9, the number of observables in each analysis that meet our criteria is highly variable. This results in individual analyses having a disproportionate influence on our tune. The minimum bias analysis at 7 TeV (ATLAS_2012_I1125575, see Section 3.3.1) in particular stands out here, as it has many observables and the modeled physics is quite redundant. Each variable is present several times here, but only the radius parameter and transverse momentum interval differ in each case. To increase the influence of analyses in which only a few variables are suitable for tuning, the weights can be adjusted accordingly. Let N_{Tune} be the number of observables used in the tune and N_{Analysis} be the number of observables in a particular analysis, then choose the weights for all observables in an analysis as

$$\omega_{\text{Analysis}} = \frac{N_{\text{Tune}}}{\sum\limits_{\mathcal{A}\in\text{Analyses}} \sqrt{N_{\mathcal{A}}}} \cdot \frac{1}{\sqrt{N_{\text{Analysis}}}}. \tag{5.1}$$

This choice satisfies the normalisation criterion $N_{\text{Tune}} = \sum_{\mathcal{A}\in\text{Analyses}} N_{\mathcal{A}}\omega_{\mathcal{A}}$ [1], and each observable now has a weight proportional to $1/N_{\text{Analysis}}$[2].
This approach results in the weights shown in Table 5.11 for the observables from the selected tune (Section 5.3). Performing the tune yields the point of best fit in Table 5.12 with $\chi^2 = 128293.25$ (with unit weights for all included observables), i.e. $\chi^2/\text{N.D.F.} = 18.3$.

Table 5.11: Individual analysis weights used in the CR2 "reweighted" tune.

Process	Observables used	ω_{Analysis}
Minimum bias @ 7 TeV	300	0.56
Minimum bias @ 8 TeV	12	2.80
Minimum bias @ 13 TeV	20	2.17
Low p_T minimum bias	6	3.96
Z + jets	58	1.27
Jet shapes	0	0.00
$K_S^0 + \Lambda^0$	7	3.66
Jet substructure	28	1.83
$t\bar{t}$ CR	2	6.85

Table 5.12: Point of best fit for the reweighted tune on the CR2 data set.

Parameter	MPI:pt0Ref	MPI:expPow	CR:m2Lambda	CR:fracGluon	CR:dLambdacut
Value	2.35	1.66	0.32	0.82	0.88

5.5. Comparison of the Tunes

By removing non-sensitive or poorly described observables, there is a significant improvement in performance from the naive to the selected tune. This becomes particularly clear by comparing the average (unweighted) χ^2 contribution per bin of the selected observables (see Table 5.13). It is obvious that the average goodness of fit for the observables of the $\sqrt{s} = 7$ TeV minimum bias analysis decreases significantly in each case due to selection and reweighting. From this it can be deduced that the influence of analyses with a smaller number of observables increases in each case as we planned. It can be observed that the performance of the selected observables clearly improves or remains approximately the same in the other analyses. Reweighting the observables leads to a better goodness of fit of the minimum bias and jet substructure $\sqrt{s} = 13$ TeV observables compared to selection alone, even for the ones that did not pass the selection step. At the same time, however, the quality of the tune suffers in all other analyses. These relationships are shown graphically in Figure 5.2. Selected observables from all analyses are presented graphically and discussed in Figures 5.4 to 5.8. Data envelopes are calculated by finding the minimum and maximum values of the polynomial model on the parameter space for each bin.

[1] Because $\sum_{\mathcal{B}\in\text{Analyses}} N_{\mathcal{B}}\omega_{\mathcal{B}} = \sum_{\mathcal{B}\in\text{Analyses}} \frac{N_{\text{Tune}}}{\sum_{\mathcal{A}\in\text{Analyses}}\sqrt{N_{\mathcal{A}}}} \cdot \frac{N_{\mathcal{B}}}{\sqrt{N_{\mathcal{B}}}} = \frac{N_{\text{Tune}}}{\sum_{\mathcal{A}\in\text{Analyses}}\sqrt{N_{\mathcal{A}}}} \cdot \sum_{\mathcal{B}\in\text{Analyses}} \sqrt{N_{\mathcal{B}}} = N_{\text{Tune}}$. □

[2] By squaring, which is done in the goodness of fit function (Equation (1.7)).

Table 5.13: Mean χ^2 contribution per observable for all performed tunes per analysis considering all observables in the generated samples (left table) and only observables that passed the criteria of the "selected" tune (right table).

	All Observables				Selected Observables			
Analysis	Naive	Selected	Reweighted	A14	Naive	Selected	Reweighted	A14
Minimum bias @ 7 TeV	69.74	101.64	124.12	10.28	71.70	107.57	132.73	9.77
Minimum bias @ 8 TeV	69.90	56.34	75.28	16.49	66.03	50.51	78.09	14.88
Minimum bias @ 13 TeV	173.59	154.84	117.75	39.42	43.86	42.99	33.84	36.68
Low p_T minimum bias	177.86	134.39	154.17	34.32	226.63	169.21	198.32	43.59
Z + jets	215.32	173.98	256.54	141.26	202.67	163.83	243.69	134.33
Jet shapes	0.43	0.42	0.42	0.43	-	-	-	-
$K_S^0 + \Lambda^0$	32.49	32.51	32.56	32.55	0.66	0.68	0.71	0.61
Jet substructure	122.05	218.19	169.14	94.43	7.38	5.20	5.16	4.74
$t\bar{t}$ CR	2029.90	1846.04	2057.39	465.11	2029.90	1846.04	2057.39	465.11

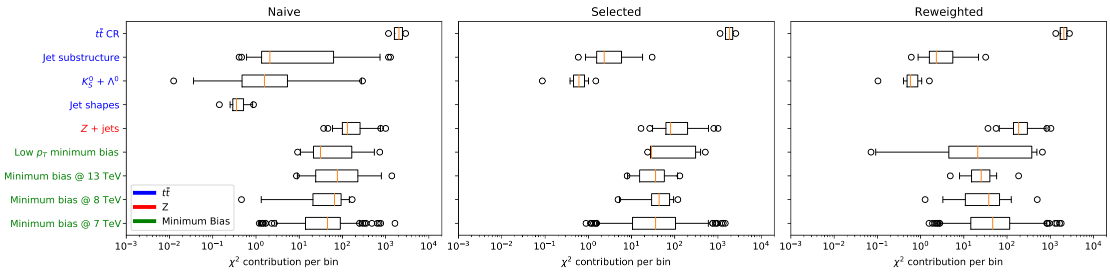


Figure 5.2: Visual comparison between the mean binwise contributions to the χ^2 measure per observable included in the respective tune. Whiskers are drawn such that the lowest and highest 2.5 % of data are treated as outliers, the yellow lines represent the median contribution in this analysis. The box represents the inner 50 % of data points.

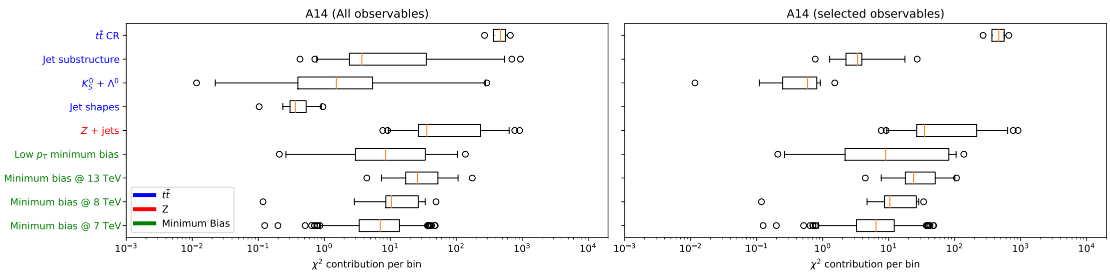
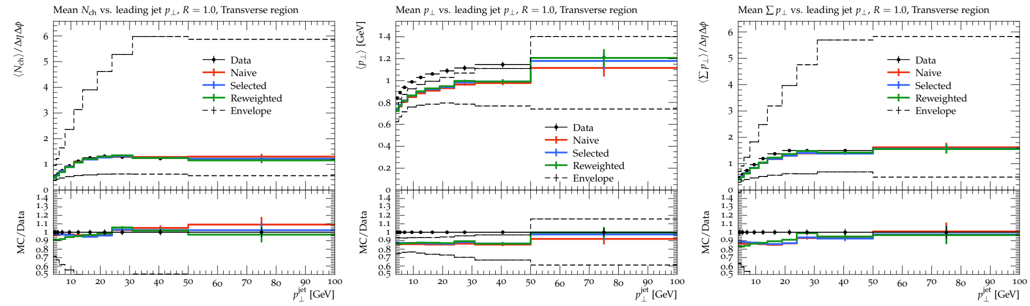
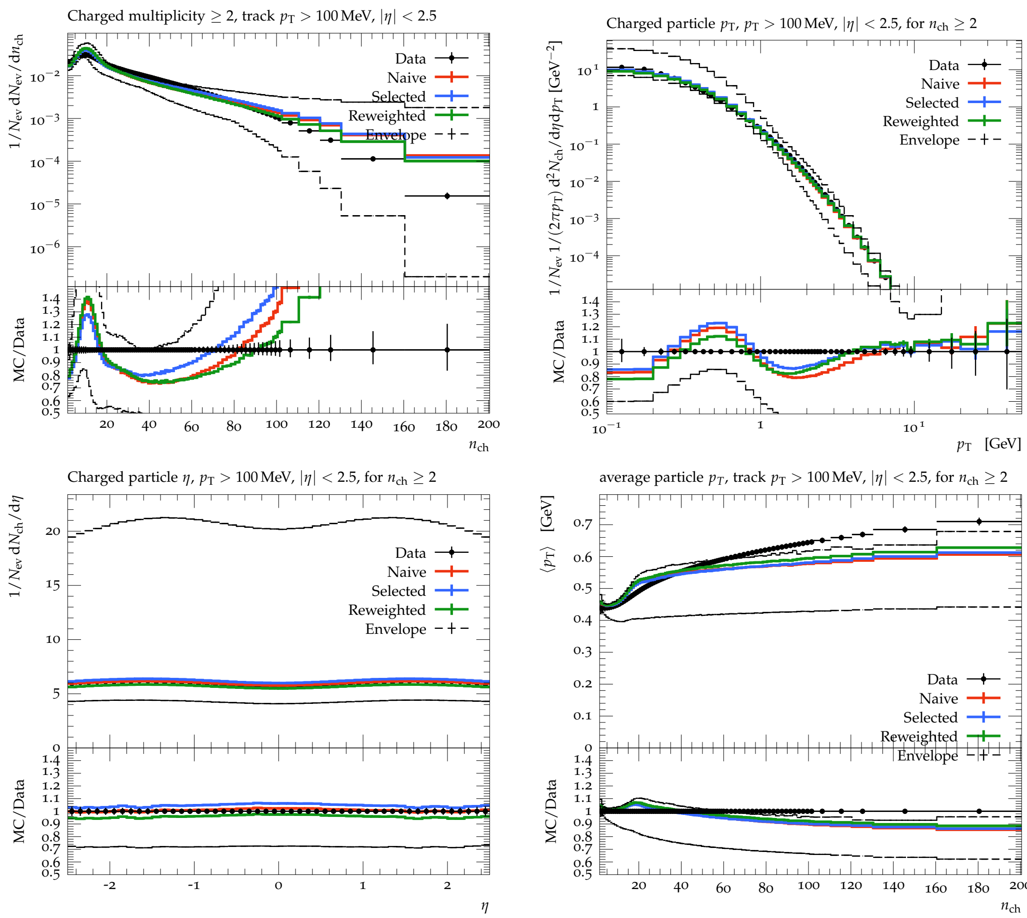


Figure 5.3: Visual comparison between the mean binwise contributions to the χ^2 measure per observable of the A14 tune, for all and CR2 selected variables. Whiskers are drawn such that the lowest and highest 2.5 % of data are treated as outliers, the yellow lines represent the median contribution in this analysis. The box represents the inner 50 % of data points.

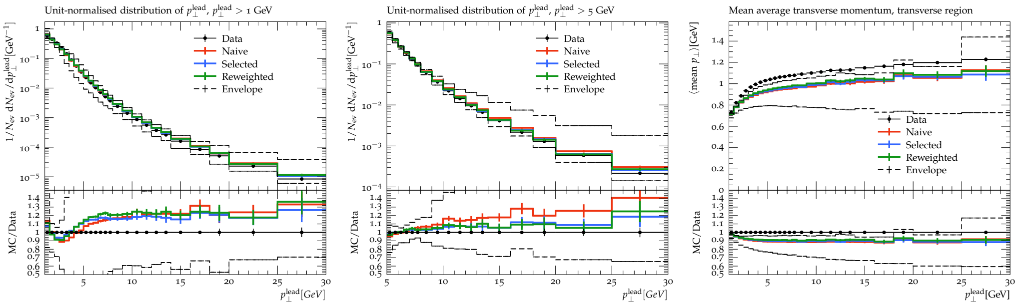


(a) Three example plots for minimum bias at 7 TeV. Shown are (from left to right) mean N_{ch}, mean p_T and mean $\sum p_\perp$ vs. leading jet p_T for $R = 1.0$ in the transverse detector region. We note that the mean p_T is not well described by the CR model in our parameter space at low values of P_T^{jet} . This can be generalised to all values of R.

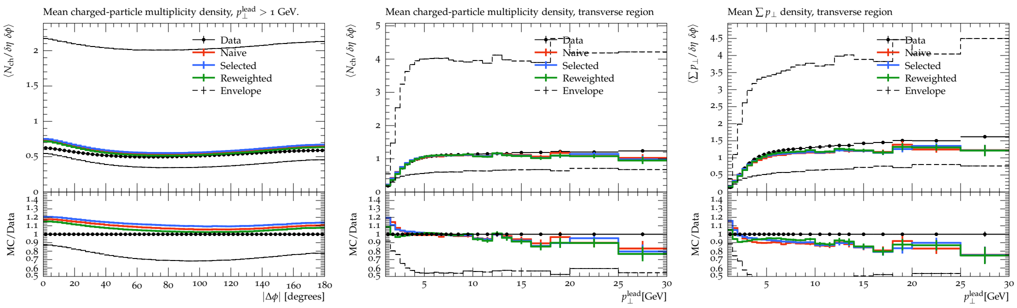


(b) Four representative plots for minimum bias at 8 TeV. Shown are the distributions of n_{ch}, p_T, η and $\langle p_T \rangle$ for $p_T > 100$ MeV, $n_{ch} \geq 2$ and $|\eta| < 2.5$. It is striking that only η is correctly described by the CR model. $\langle p_T \rangle$ is not correctly enveloped in all observables.

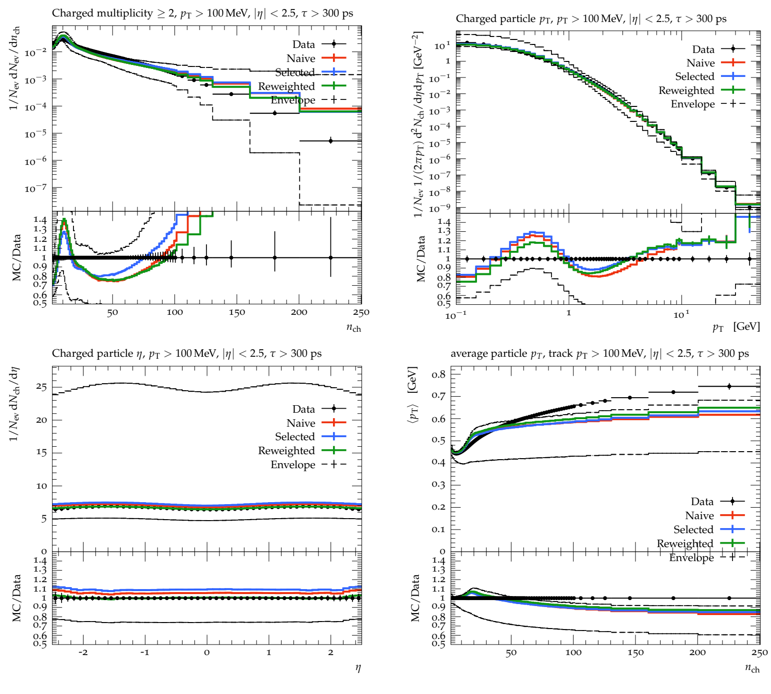
Figure 5.4: Represenstatively selected plots for minimum bias observables at $\sqrt{s} = 7$, 8 TeV.



(a) From left to right: Unit-normalised distribution of p_T^{lead} for $p_T^{\text{lead}} > 1$ GeV and 5 GeV. Furthermore, the mean average transverse momentum in the transverse region of the event. For the third observable, it is easy to see that the data on our parameter space are poorly described by CR2.



(b) From left to right: mean charged-particle multiplicity density for $p_T^{\text{lead}} > 1$ GeV, mean charged-particle multiplicity densitiy in the transverse region and mean $\sum p_T$ in the transverse region. The first distribution shows the influence of the reweighting on the tune.



(c) Four distributions of the analysis with low-p_T tracks. These are defined similarly to the distributions of the 8 TeV minimum bias analysis, these are shown in Figure 5.4 (b). The tune behaves the same in both cases, a high redundancy between these two analyses can be assumed.

Figure 5.5: Selected plots of standard minimum bias ((a) and b) and low-p_T minimum bias at $\sqrt{s} = 13$ TeV.

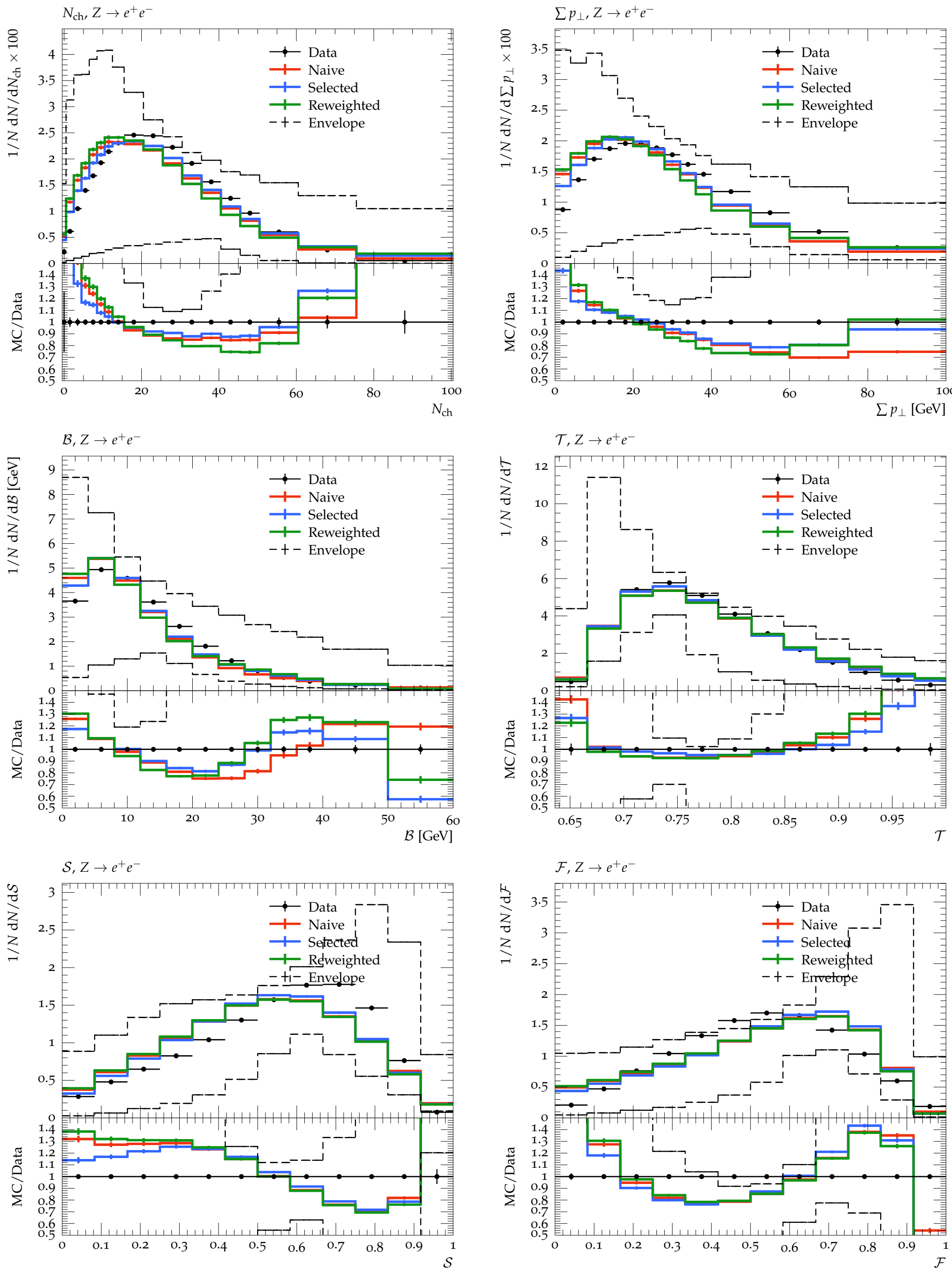
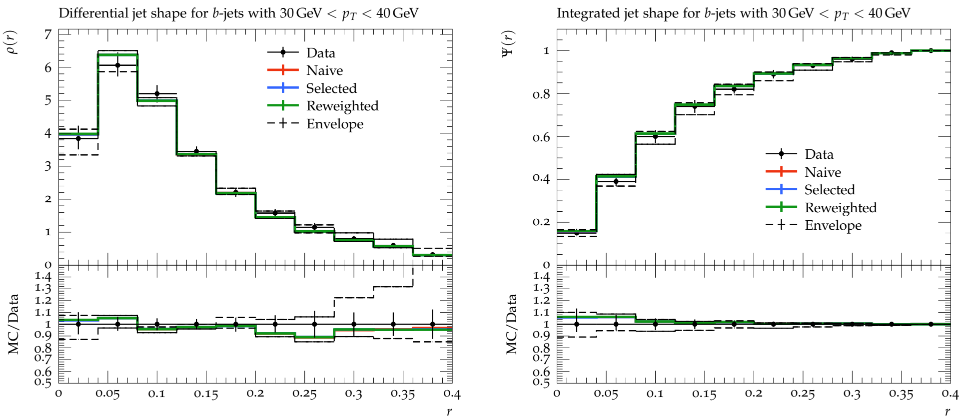
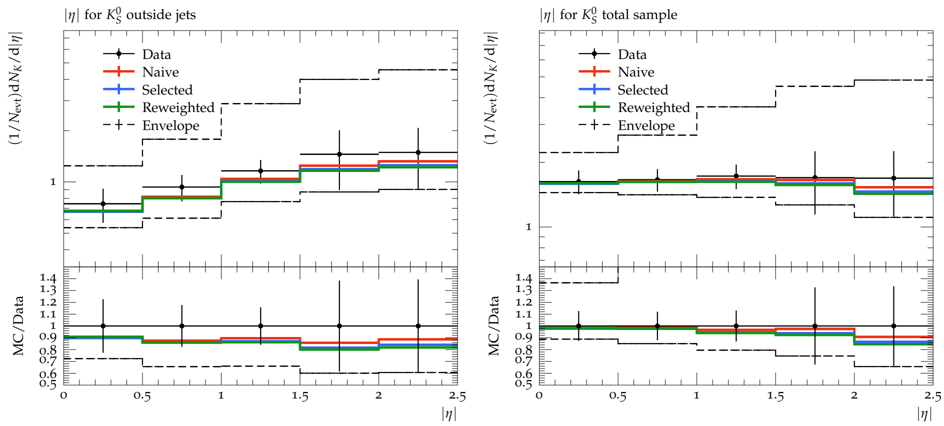


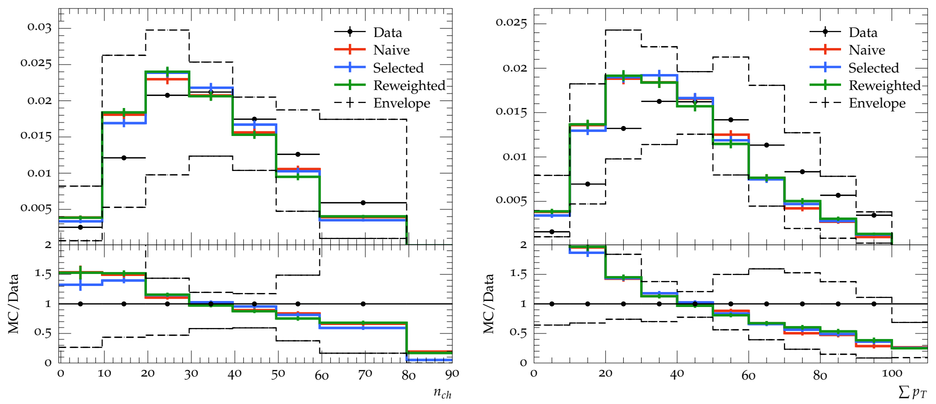
Figure 5.6: Distributions of N_{ch}, $\sum p_T$, $\mathcal{B}$, $\mathcal{T}$, $\mathcal{S}$ and $\mathcal{F}$ in the $Z \to e^+e^-$ channel without further restrictions for $p_T(Z)$. At $\sqrt{s} = 7$ TeV, the reference data are well enclosed by the CR2 model but, with a few exceptions, are not well described within the envelopes.



(a) Two example plots of the analysis of the 7 TeV jet shapes in $t\bar{t}$ events (differential and integrated jet shape for b-jets with 30 GeV $< p_T <$ 40 GeV). All observables of this analysis have in common that the reference data are well described by the MC, but the output of the MC is not sensitive to changes in the parameterisation.



(b) The two most sensitive observables of the K^0_S analysis are the $|\eta|$ outside of jets and for the total sample.



(c) Both observables of the $\sqrt{s}$ = 13 TeV CR analysis of $t\bar{t}$ events are by design very sensitive. They are well enveloped but poorly modelled by CR2 on the parameter space studied.

Figure 5.7: Two example plots each of the 7 TeV jet shapes & K^0_S analyses and both observables of the 13 TeV CR analysis.

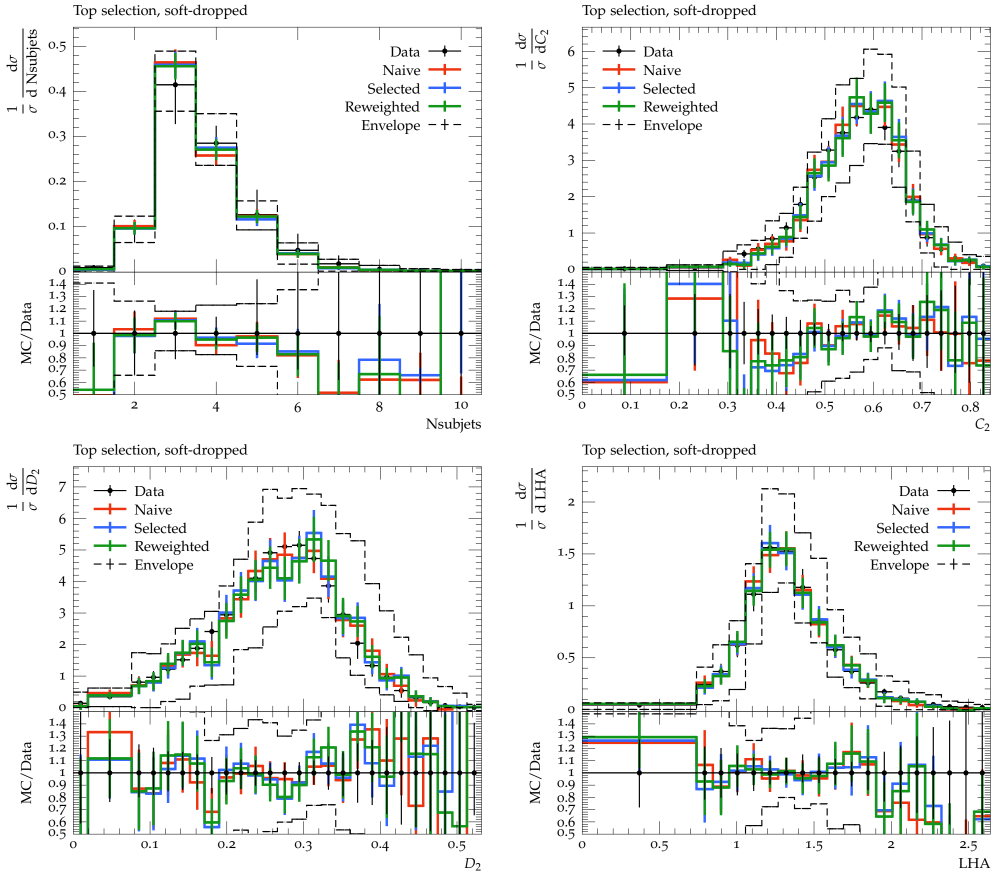


Figure 5.8: Four plots of the 13 TeV jet substructure analysis. Shown are the distributions of the number of subjets, C_2, D_2 and the Les-Houches angularity LHA for the soft-dropped top-selection. All observables of the analysis are well described by the tunes. However, the errors are so high that differences in the tunes are hardly comprehensible. The dijet-associated observables were not included in the tunes because no corresponding samples were generated.

5.5.1. Comparison to the A14 tune

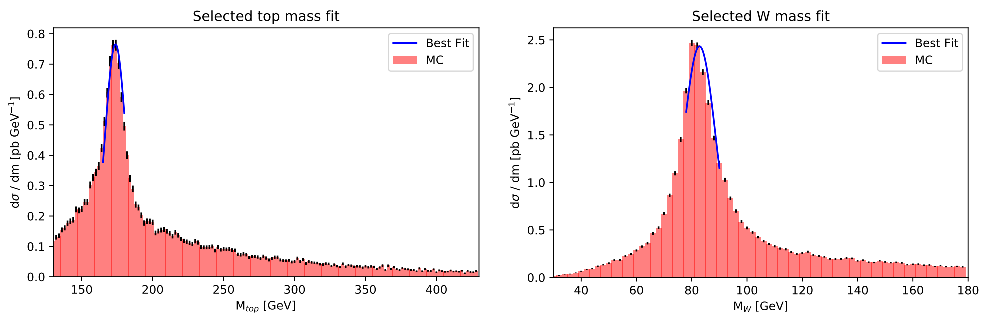
Since the selected tune has the best goodness of fit, we compare it with the standard A14 tune (see Section 1.5.7) used in many ATLAS activities. For this purpose, a sample was produced with all settings described in Section 5.1.1 to Section 5.1.3, the settings of the colour reconnection model were taken from A14. The mean contribution to the χ^2 measure per observable is lower or equal for the A14 tune compared to our own tunes for all analyses, as documented in Table 5.13. This is also true when only the observables of the CR2 selected tune are considered, with the exception of minimum bias at 13 TeV. The distribution of binwise χ^2 contributions for A14 is visualised in Figure 5.3 for comparison with Figure 5.2. All in all, the A14 tune (using the CR0 model) provides a better description of the experimental data than our tunes of the CR2 model. This is in line with previous research by the ATLAS collaboration [52].

5.5.2. Description of the top-quark and W-boson masses

In addition to the included analyses of large particle physics experiments, Rivet also contains some pure MC studies. One of these is MC_TTBAR [53]. This study provides, among other things, distributions of the W-boson and top-quark mass in $t\bar{t}$ events. These are well suited to investigate a possible influence of the CR settings on a correct reconstruction of the top quarks. First, we fitted a Gaussian curve to the distribution of W-boson masses in the interval $m_W \in [75\,\text{GeV}, 90\,\text{GeV}]$ to obtain an estimator for the mass of the W boson. The top-quark mass is then reconstructed from events with $m_W \in [75\,\text{GeV}, 85\,\text{GeV}]$. This distribution was also fitted with a Gaussian curve, this time in the interval $m_t \in [165\,\text{GeV}, 180\,\text{GeV}]$. The results for all fits can be found in Table 5.14, and the fits for the selected CR2 tune are graphically represented in Figure 5.9. No significant differences between the different generator settings were observed in the W-boson mass reconstruction. The reconstructed top-quark masses for CR2 default and reweighted settings were observed to be lower than in the A14 tune.

Table 5.14: Resulting masses from the gaussian fits to MC data for the different tunes. "CR2" refers to the default CR settings for CR2 provided by the authors of the model [30] (see Table 5.2).

Tune	m_t [GeV]	m_W [GeV]	$(m_t^{\text{A14}} - m_t)$ [GeV]	$(m_W^{\text{A14}} - m_W)$ [GeV]
A14	173.801(93)	82.80(18)	-	-
Naive	173.67(20)	82.76(26)	0.13(22)	0.05(32)
Selected	173.692(85)	82.81(23)	0.11(13)	-0.01(29)
Reweighted	173.46(15)	82.72(20)	0.34(17)	0.09(27)
CR2	173.55(17)	82.65(25)	0.26(19)	0.16(31)



(a) Fit to the top-quark mass distribution. (b) Fit to the W-boson mass distribution.

Figure 5.9: Plots of both fits for our selected CR2 tune.

6

Conclusion and Outlook

The goal of this thesis was to create one or more Monte Carlo tunes with a focus on colour reconnection models under inclusion of $t\bar{t}$ events. For this purpose, a new technique for analysing the sensitivities of individual observables was developed and presented. It was achieved to calculate the sensitivities of analyses of different processes and to compare them directly. In addition, a test for correct description of the data by the output of the MC generator was developed. Together with a minimum requirement for the sensitivity of the observables, a criterion was derived with which certain observables can be excluded from a tune. We were able to show that a tune with the observables selected in this way can achieve a significantly better fit to the reference data than one with naive inclusion of all available observables. In addition, general weights were given that can be used to reduce the influence on the tune of analyses that include many redundant observables. In our case, this reweighting technique led to a certain reduction in the goodness of fit, but at the same time it was able to strengthen the performance for an analysis (minimum bias at $\sqrt{s}$ = 13 TeV) that previously had less influence. We have produced a total of three tunes of the CR2 model, of which the ”selected” one exhibits the best goodness of fit to the experimental data (see Table 6.1).

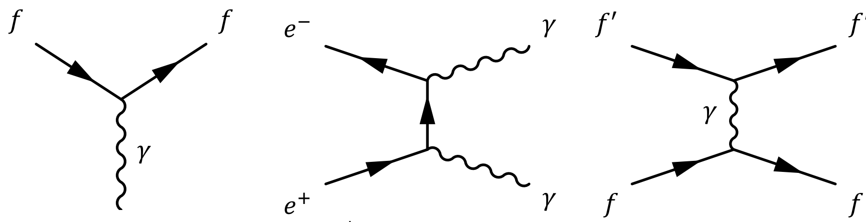
Table 6.1: The three tunes we produced for the CR2 model. "Naive" indicates the tune that contains all available observables, "Selected" contains those selected according to our criterion. The "Reweighted" tune was created by applying reweighting to the observables of the "Selected" tune.

Tune	Naive	Selected	Reweighted
χ^2 (unweighted)	193785.82	114033.87	128293.25
Degrees of Freedom	7585	6996	6996
χ^2/N.D.F.	25.5	16.3	18.3
MPI:pT0Ref	2.30	2.23	2.35
MPI:expPow	1.75	2.07	1.66
CR:m2Lambda	2.81	0.32	0.32
CR:fracGluon	1.00	0.78	0.82
CR:dLambdaCut	1.86	1.05	0.88

One of the findings of the sensitivity analysis is that distributions of the jet shapes in $t\bar{t}$ events we considered are not suitable for tuning the CR2 model. While these observables are consistently well described by the MC, they hardly react to changes in the parameterisation. Furthermore, the study of the jet substructure at $\sqrt{s} = 13$ TeV turned out to have only limited tuneability, as the individual tunes are indistinguishable within the model errors. The search for further analyses of $t\bar{t}$ events with observables sensitive to CR parameters should therefore be worthwhile in order to more strongly incorporate top-quark physics into the tunes. Due to time constraints, we were only able to investigate one CR model. A sensible next step would be to provide tunes for the other two models discussed. By comparing the sensitivity analyses carried out in this process, it might also be possible to identify observables that are particularly sensitive to colour reconnection among all of the available models. Additionally, the two comparison values for the selection criterion (proportion of well-described bins and minimum sensitivity) were chosen by eye. By examining different settings, the criterion could possibly prove to be even more effective. Finally, an extension of the concept of reweighting would be conceivable. Observables could be weighted not only according to their multiplicity in the analyses, but also according to the χ^2 contributions achieved. In this way, the result of the tune would depend on the actual performance of the variables, and not so much on the definition of the Rivet analyses.

A

Additional Feynman Diagrams

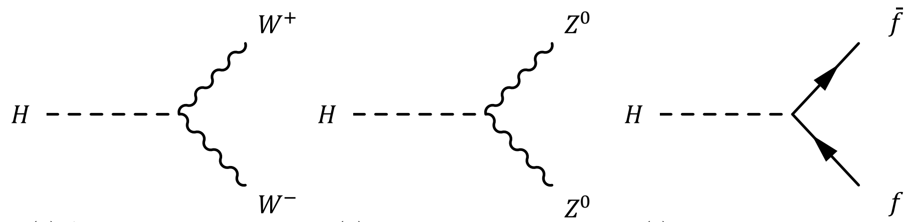


(a) The fundamental QED vertex with one γ and a charged fermion f.

(b) e^+e^- annihilating into two γ.

(c) Compton scattering of two charged fermions.

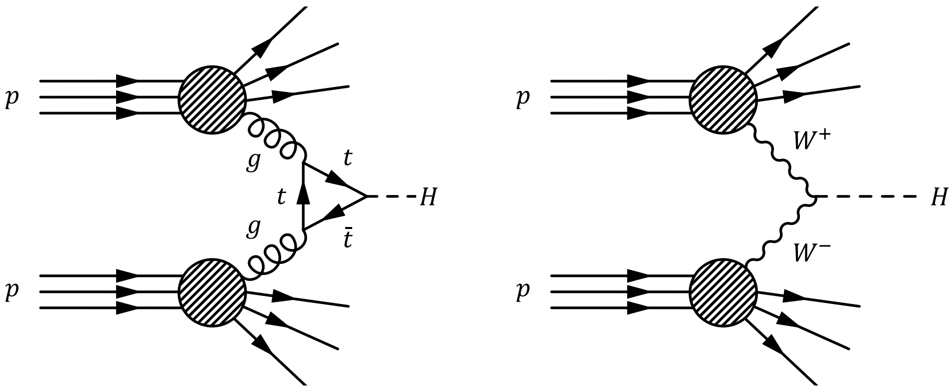
Figure A.1: Fundamental QED vertex and two example processes.



(a) A Higgs coupling to a $W^{\pm}$ pair.

(b) The Higgs boson coupling to a Z^0 pair.

(c) Higgs boson coupling to a fermion / antifermion pair.



(d) Gluon-gluon fusion process producing a Higgs boson via a top quark loop.

(e) Vector boson fusion process producing a Higgs boson.

Figure A.2: Fundamental Standard Model vertices involving the Higgs and the two dominant production channels at LHC.

B

One-dimensional Tune for PT0Ref

B.1. Variables included in the tune

- /ATLAS_2017_I1495243/d01-x01-y01
- /ATLAS_2017_I1495243/d02-x01-y01
- /ATLAS_2017_I1495243/d03-x01-y01
- /ATLAS_2017_I1495243/d04-x01-y01
- /ATLAS_2017_I1495243/d05-x01-y01
- /ATLAS_2017_I1495243/d06-x01-y01
- /ATLAS_2017_I1495243/d07-x01-y01
- /ATLAS_2017_I1495243/d08-x01-y01
- /ATLAS_2017_I1495243/d09-x01-y01
- /ATLAS_2017_I1495243/d10-x01-y01
- /ATLAS_2017_I1495243/d11-x01-y01
- /ATLAS_2017_I1495243/d12-x01-y01
- /ATLAS_2017_I1495243/d13-x01-y01
- /ATLAS_2017_I1495243/d14-x01-y01
- /ATLAS_2017_I1495243/d15-x01-y01
- /ATLAS_2017_I1495243/d16-x01-y01
- /ATLAS_2017_I1495243/d17-x01-y01
- /ATLAS_2017_I1495243/d18-x01-y01
- /ATLAS_2017_I1495243/d19-x01-y01
- /ATLAS_2017_I1495243/d20-x01-y01
- /ATLAS_2017_I1495243/d21-x01-y01
- /ATLAS_2017_I1495243/d22-x01-y01
- /ATLAS_2017_I1495243/d23-x01-y01
- /ATLAS_2018_I1656578/d100-x01-y01
- /ATLAS_2018_I1656578/d101-x01-y01
- /ATLAS_2018_I1656578/d102-x01-y01
- /ATLAS_2018_I1656578/d103-x01-y01
- /ATLAS_2018_I1656578/d104-x01-y01
- /ATLAS_2018_I1656578/d105-x01-y01
- /ATLAS_2018_I1656578/d106-x01-y01
- /ATLAS_2018_I1656578/d107-x01-y01
- /ATLAS_2018_I1656578/d108-x01-y01
- /ATLAS_2018_I1656578/d109-x01-y01
- /ATLAS_2018_I1656578/d110-x01-y01
- /ATLAS_2018_I1656578/d111-x01-y01
- /ATLAS_2018_I1656578/d112-x01-y01
- /ATLAS_2018_I1656578/d113-x01-y01
- /ATLAS_2018_I1656578/d114-x01-y01

- /ATLAS_2018_I1656578/d115-x01-y01
- /ATLAS_2018_I1656578/d96-x01-y01
- /ATLAS_2018_I1656578/d97-x01-y01
- /ATLAS_2018_I1656578/d98-x01-y01
- /ATLAS_2018_I1656578/d99-x01-y01
- /ATLAS_2018_I1705857/d01-x01-y01
- /ATLAS_2018_I1705857/d02-x01-y01
- /ATLAS_2018_I1705857/d03-x01-y01
- /ATLAS_2018_I1705857/d04-x01-y01
- /ATLAS_2018_I1705857/d05-x01-y01
- /ATLAS_2018_I1705857/d06-x01-y01
- /ATLAS_2018_I1705857/d07-x01-y01
- /ATLAS_2018_I1705857/d08-x01-y01
- /ATLAS_2018_I1705857/d09-x01-y01
- /ATLAS_2018_I1705857/d10-x01-y01
- /ATLAS_2018_I1705857/d11-x01-y01
- /ATLAS_2018_I1705857/d12-x01-y01
- /ATLAS_2018_I1705857/d13-x01-y01
- /ATLAS_2018_I1705857/d14-x01-y01
- /ATLAS_2018_I1705857/d15-x01-y01
- /ATLAS_2018_I1705857/d16-x01-y01
- /ATLAS_2018_I1705857/d17-x01-y01
- /ATLAS_2018_I1705857/d18-x01-y01
- /ATLAS_2018_I1705857/d19-x01-y01
- /ATLAS_2018_I1705857/d20-x01-y01
- /ATLAS_2018_I1705857/d21-x01-y01
- /ATLAS_2018_I1705857/d22-x01-y01
- /ATLAS_2018_I1705857/d23-x01-y01
- /ATLAS_2018_I1705857/d24-x01-y01
- /ATLAS_2018_I1705857/d25-x01-y01
- /ATLAS_2018_I1705857/d26-x01-y01
- /ATLAS_2018_I1705857/d27-x01-y01
- /ATLAS_2018_I1705857/d28-x01-y01
- /ATLAS_2018_I1705857/d29-x01-y01
- /ATLAS_2018_I1705857/d30-x01-y01
- /ATLAS_2018_I1705857/d31-x01-y01
- /ATLAS_2018_I1705857/d32-x01-y01
- /ATLAS_2018_I1705857/d33-x01-y01
- /ATLAS_2018_I1705857/d34-x01-y01
- /ATLAS_2018_I1705857/d35-x01-y01
- /ATLAS_2018_I1705857/d36-x01-y01
- /ATLAS_2018_I1705857/d37-x01-y01
- /ATLAS_2018_I1705857/d38-x01-y01
- /ATLAS_2018_I1705857/d39-x01-y01
- /ATLAS_2018_I1705857/d40-x01-y01
- /ATLAS_2018_I1705857/d41-x01-y01
- /ATLAS_2018_I1705857/d42-x01-y01
- /ATLAS_2018_I1705857/d43-x01-y01
- /ATLAS_2018_I1705857/d44-x01-y01
- /ATLAS_2018_I1705857/d45-x01-y01
- /ATLAS_2018_I1705857/d46-x01-y01
- /ATLAS_2018_I1705857/d47-x01-y01
- /ATLAS_2018_I1705857/d48-x01-y01
- /ATLAS_2018_I1705857/d50-x01-y01
- /ATLAS_2019_I1724098/d07-x01-y01
- /ATLAS_2019_I1724098/d08-x01-y01
- /ATLAS_2019_I1724098/d09-x01-y01
- /ATLAS_2019_I1724098/d10-x01-y01
- /ATLAS_2019_I1724098/d11-x01-y01
- /ATLAS_2019_I1724098/d12-x01-y01
- /ATLAS_2019_I1724098/d13-x01-y01
- /ATLAS_2019_I1724098/d14-x01-y01

- /ATLAS_2019_I1724098/d15-x01-y01
- /ATLAS_2019_I1724098/d16-x01-y01
- /ATLAS_2019_I1724098/d17-x01-y01
- /ATLAS_2019_I1724098/d18-x01-y01
- /ATLAS_2019_I1724098/d19-x01-y01
- /ATLAS_2019_I1724098/d20-x01-y01
- /ATLAS_2019_I1724098/d21-x01-y01
- /ATLAS_2019_I1724098/d22-x01-y01
- /ATLAS_2019_I1724098/d29-x01-y01
- /ATLAS_2019_I1724098/d30-x01-y01
- /ATLAS_2019_I1724098/d31-x01-y01
- /ATLAS_2019_I1724098/d32-x01-y01
- /ATLAS_2019_I1724098/d33-x01-y01
- /ATLAS_2019_I1724098/d34-x01-y01
- /ATLAS_2019_I1724098/d35-x01-y01
- /ATLAS_2019_I1724098/d36-x01-y01
- /ATLAS_2019_I1724098/d37-x01-y01
- /ATLAS_2019_I1724098/d38-x01-y01
- /ATLAS_2019_I1724098/d39-x01-y01
- /ATLAS_2019_I1724098/d40-x01-y01
- /ATLAS_2019_I1724098/d41-x01-y01
- /ATLAS_2019_I1724098/d42-x01-y01
- /ATLAS_2019_I1724098/d43-x01-y01
- /ATLAS_2019_I1724098/d44-x01-y01
- /MC_ttbar_ColorReconnection/nch
- /MC_ttbar_ColorReconnection/sumPt

B.2. Variables not included in the tune

- /ATLAS_2018_I1705857/d49-x01-y01
- /ATLAS_2019_I1724098/d01-x01-y01
- /ATLAS_2019_I1724098/d02-x01-y01
- /ATLAS_2019_I1724098/d03-x01-y01
- /ATLAS_2019_I1724098/d04-x01-y01
- /ATLAS_2019_I1724098/d05-x01-y01
- /ATLAS_2019_I1724098/d06-x01-y01
- /ATLAS_2019_I1724098/d23-x01-y01
- /ATLAS_2019_I1724098/d24-x01-y01
- /ATLAS_2019_I1724098/d25-x01-y01
- /ATLAS_2019_I1724098/d26-x01-y01
- /ATLAS_2019_I1724098/d27-x01-y01
- /ATLAS_2019_I1724098/d28-x01-y01
- Any variable from /MC_TTBAR
- /MC_ttbar_ColorReconnection/Pt_vs_nch
- /MC_ttbar_ColorReconnection/avgPt
- /MC_ttbar_ColorReconnection/avgPt_vs_nch
- /MC_ttbar_ColorReconnection/bjet1Eta
- /MC_ttbar_ColorReconnection/bjet1Pt
- /MC_ttbar_ColorReconnection/bjet2Eta
- /MC_ttbar_ColorReconnection/bjet2Pt
- /MC_ttbar_ColorReconnection/deltaRtrkJet
- /MC_ttbar_ColorReconnection/deltaRtrkLep
- /MC_ttbar_ColorReconnection/eleEta
- /MC_ttbar_ColorReconnection/elePt
- /MC_ttbar_ColorReconnection/mlb_1
- /MC_ttbar_ColorReconnection/mlb_2
- /MC_ttbar_ColorReconnection/muonEta
- /MC_ttbar_ColorReconnection/muonPt
- /MC_ttbar_ColorReconnection/nEle

- /MC_ttbar_ColorReconnection/nJets
- /MC_ttbar_ColorReconnection/nMuons
- /MC_ttbar_ColorReconnection/sumPt_vs_nch
- /MC_ttbar_ColorReconnection/trkEta
- /MC_ttbar_ColorReconnection/trkPt

C

Tables of Variable Names

This appendix provides a reference for Rivet's variable naming scheme for the analyses used throughout the thesis.

C.1. ATLAS_2013_I1243871

Table C.1: Variable names in the Rivet version of the $\sqrt{s}$ = 7 TeV jet shape analysis [54].

	Rivet variable name			
	b-jets		Light jets	
Jet $p_\perp$ cut	Differential jet shape	Integrated jet shape	Differential jet shape	Int. jet shape
30 GeV < $p_\perp$ < 40 GeV	d01-x01-y01	d01-x01-y02	d01-x02-y01	d01-x02-y02
40 GeV < $p_\perp$ < 50 GeV	d02-x01-y01	d02-x01-y02	d02-x02-y01	d02-x02-y02
50 GeV < $p_\perp$ < 70 GeV	d03-x01-y01	d03-x01-y02	d03-x02-y01	d03-x02-y02
70 GeV < $p_\perp$ < 100 GeV	d04-x01-y01	d04-x01-y02	d04-x02-y01	d04-x02-y02
100 GeV < $p_\perp$ < 150 GeV	d05-x01-y01	d05-x01-y02	d05-x02-y01	d05-x02-y02

C.2. ATLAS_2019_I1724098

Table C.2: Variable names in the Rivet version of the $\sqrt{s}$ = 13 TeV jet substructure analysis [55]

	Algorithm						
	Trimmed			Soft-dropped			
Variable	Dijet	W	Top	Dijet	W	Top	Selection
N_{subjets}	d01-x01-y01	d37-x01-y01	d29-x01-y01	d23-x01-y01	d15-x01-y01	d07-x01-y01	
C_2	d02-x01-y01	d38-x01-y01	d30-x01-y01	d24-x01-y01	d16-x01-y01	d08-x01-y01	
D_2	d03-x01-y01	d39-x01-y01	d31-x01-y01	d25-x01-y01	d17-x01-y01	d09-x01-y01	
LHA	d04-x01-y01	d40-x01-y01	d32-x01-y01	d26-x01-y01	d18-x01-y01	d10-x01-y01	
$\text{ECF}_2^{\text{Norm}}$	d05-x01-y01	d41-x01-y01	d33-x01-y01	d27-x01-y01	d19-x01-y01	d11-x01-y01	
$\text{ECF}_3^{\text{Norm}}$	d06-x01-y01	d42-x01-y01	d34-x01-y01	d28-x01-y01	d20-x01-y01	d12-x01-y01	
τ_{21}		d43-x01-y01	d35-x01-y01		d21-x01-y01	d13-x01-y01	
τ_{32}		d44-x01-y01	d36-x01-y01		d22-x01-y01	d14-x01-y01	

C.3. ATLAS_2016_I1424838

Table C.3: Variable names in the Rivet version of the $\sqrt{s} = 7\,\text{TeV}$ event shape analysis [56]

Channel	$p_\perp$ - Cut	N_{ch}	$\sum p_\perp$	$\mathcal{B}$	$\mathcal{T}$	$\mathcal{S}$	$\mathcal{F}$
e^+e^-	-	d01-x01-y01	d02-x01-y01	d03-x01-y01	d04-x01-y01	d05-x01-y01	d06-x01-y01
	$p_\perp \in [0,6)$ GeV	d07-x01-y01	d08-x01-y01	d09-x01-y01	d10-x01-y01	d11-x01-y01	d12-x01-y01
	$p_\perp \in [6,12)$ GeV	d13-x01-y01	d14-x01-y01	d15-x01-y01	d16-x01-y01	d17-x01-y01	d18-x01-y01
	$p_\perp \in [12,25)$ GeV	d19-x01-y01	d20-x01-y01	d21-x01-y01	d22-x01-y01	d23-x01-y01	d24-x01-y01
	$p_\perp > 25$ GeV	d25-x01-y01	d26-x01-y01	d27-x01-y01	d28-x01-y01	d29-x01-y01	d30-x01-y01
$\mu^+\mu^-$	-	d31-x01-y01	d32-x01-y01	d33-x01-y01	d34-x01-y01	d35-x01-y01	d36-x01-y01
	$p_\perp \in [0,6)$ GeV	d37-x01-y01	d38-x01-y01	d39-x01-y01	d40-x01-y01	d41-x01-y01	d42-x01-y01
	$p_\perp \in [6,12)$ GeV	d43-x01-y01	d44-x01-y01	d45-x01-y01	d46-x01-y01	d47-x01-y01	d48-x01-y01
	$p_\perp \in [12,25)$ GeV	d49-x01-y01	d50-x01-y01	d51-x01-y01	d52-x01-y01	d53-x01-y01	d54-x01-y01
	$p_\perp > 25$ GeV	d55-x01-y01	d56-x01-y01	d57-x01-y01	d58-x01-y01	d59-x01-y01	d60-x01-y01

C.4. ATLAS_2012_I1125575

Table C.4: Variables depending on the leading jet's $p_\perp$ [57].

	R=0.2	R=0.4	R=0.6	R=0.8	R=1.0	
Mean N_{ch} vs leading jet $p_\perp$	d01-x01-y01	d01-x02-y01	d01-x03-y01	d01-x04-y01	d01-x05-y01	Away Region
	d01-x01-y02	d01-x02-y02	d01-x03-y02	d01-x04-y02	d01-x05-y02	Transverse Region
Mean $p_\perp$ vs leading jet $p_\perp$	d02-x01-y01	d02-x02-y01	d02-x03-y01	d02-x04-y01	d02-x05-y01	Away Region
	d02-x01-y02	d02-x02-y02	d02-x03-y02	d02-x04-y02	d02-x05-y02	Transverse Region
Mean $\sum p_\perp$ vs leading jet $p_\perp$	d03-x01-y01	d03-x02-y01	d03-x03-y01	d03-x04-y01	d03-x05-y01	Away Region
	d03-x01-y02	d03-x02-y02	d03-x03-y02	d03-x04-y02	d03-x05-y02	Transverse Region

Table C.5: All other variable names are constructed like d04-x01-y02 according to the encoding in this table [57].

(a) The observable gives the first component of the variable name.

Variable	Name Component
N_{Ch}	d04
$p_\perp$	d05
$\sum p_\perp$	d06

(b) The second component of the name is given by the radius parameter.

Radius parameter	Name Component
$R = 0.2$	x01
$R = 0.4$	x02
$R = 0.6$	x03
$R = 0.8$	x04
$R = 1.0$	x05

(c) The third component is given by the $p_\perp$ interval selected.

$p_\perp$ Interval	Away region	Transverse region
$[4-5)$ GeV	y01	y02
$[5-6)$ GeV	y03	y04
$[6-8)$ GeV	y05	y06
$[8-11)$ GeV	y07	y08
$[11-14)$ GeV	y09	y10
$[14-19)$ GeV	y11	y12
$[19-24)$ GeV	y13	y14
$[24-31)$ GeV	y15	y16
$[31-50)$ GeV	y17	y18
$[50-100)$ GeV	y19	y20

C.5. ATLAS_2016_I1426695

Table C.6: Variable names in the $\sqrt{s} = 8\,\mathrm{TeV}$ analysis of the underlying event with minimum-bias data [58].

Variable	Track $p_\perp \geq$	Charged Track Multiplicity $\geq$	Name
N_{Ch}	100 MeV	2	d02-x01-y01
N_{Ch}	500 MeV	1	d03-x01-y01
Charged particle $p_\perp$	100 MeV	2	d04-x01-y01
Charged particle $p_\perp$	500 MeV	1	d05-x01-y01
Charged particle $p_\perp$	500 MeV	6	d06-x01-y01
Charged particle $p_\perp$	500 MeV	20	d07-x01-y01
Charged particle $p_\perp$	500 MeV	50	d08-x01-y01
Charged particle η	100 MeV	2	d09-x01-y01
Charged particle η	500 MeV	1	d10-x01-y01
Charged particle η	500 MeV	6	d11-x01-y01
Charged particle η	500 MeV	20	d12-x01-y01
Charged particle η	500 MeV	50	d13-x01-y01
Average particle $p_\perp$	100 MeV	2	d14-x01-y01
Average particle $p_\perp$	500 MeV	1	d15-x01-y01

C.6. ATLAS_2017_I1509919

Table C.7: All variable names in the $\sqrt{s} = 13$ TeV analysis of the minimum bias analysis [59].

(a) Averaged variables by region.

Variable	Region	Name
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-min	d02-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-max	d03-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-diff	d04-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-min	d05-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-max	d06-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-diff	d07-x01-y01
$\langle$mean $p_\perp\rangle$ vs N_{Ch}	Transverse	d08-x01-y01
$\langle$mean $p_\perp\rangle$ vs N_{Ch}	Trans-min	d09-x01-y01
$\langle$mean $p_\perp\rangle$ vs N_{Ch}	Trans-max	d10-x01-y01
$\langle$mean $p_\perp\rangle$ vs $p_\perp^{\mathrm{lead}}$	Transverse	d11-x01-y01
$\langle$mean $p_\perp\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-min	d12-x01-y01
$\langle$mean $p_\perp\rangle$ vs $p_\perp^{\mathrm{lead}}$	Trans-max	d13-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Transverse	d21-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Towards	d22-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Away	d23-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Transverse	d24-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Towards	d25-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $p_\perp^{\mathrm{lead}}$	Away	d26-x01-y01
$\langle$mean $p_\perp\rangle$ vs N_{Ch}	Towards	d27-x01-y01
$\langle$mean $p_\perp\rangle$ vs N_{Ch}	Away	d28-x01-y01
$\langle$mean $p_\perp\rangle$ vs $p_\perp^{\mathrm{lead}}$	Towards	d29-x01-y01
$\langle$mean $p_\perp\rangle$ vs $p_\perp^{\mathrm{lead}}$	Away	d30-x01-y01
$\langle$mean $p_\perp\rangle$ vs $p_\perp^{\mathrm{lead}}$	trans-min	d31-x01-y01
$\langle$mean $p_\perp\rangle$ vs $p_\perp^{\mathrm{lead}}$	trans-max	d32-x01-y01

(b) Variables by cut in $p_\perp^{\mathrm{lead}}$.

Variable	$p_\perp^{\mathrm{lead}} >$	Name
Normalized $p_\perp^{\mathrm{lead}}$ distribution	1 GeV	d01-x01-y01
Normalized $p_\perp^{\mathrm{lead}}$ distribution	5 GeV	d14-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $\Delta\phi$	1 GeV	d15-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $\Delta\phi$	5 GeV	d16-x01-y01
$\langle N_{\mathrm{Ch}}/\delta\eta\delta\phi\rangle$ vs $\Delta\phi$	10 GeV	d17-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $\Delta\phi$	1 GeV	d18-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $\Delta\phi$	5 GeV	d19-x01-y01
$\langle \sum p_\perp/\delta\eta\delta\phi\rangle$ vs $\Delta\phi$	10 GeV	d20-x01-y01

C.7. ATLAS_2016_I1467230

Table C.8: All variable names of the low-p_T minimum bias analysis at $\sqrt{s} = 13$ TeV. Track are required to have $p_T > 100$ MeV and $|\eta| < 2.5$ [45].

Lifetime	Charged particle η	Charged particle p_T	Charged multiplicity ≥ 2	Average p_T vs. n_{ch}
$\tau > 300$ ps	d03-x01-y01	d02-x01-y01	d01-x01-y01	d04-x01-y01
$\tau > 30$ ps	d03-x01-y02	d02-x01-y02	d01-x01-y02	d04-x01-y02

D

Variables selected for CR2

D.1. ATLAS_2012_I1125575

- d01-x01-y01
- d01-x01-y02
- d01-x02-y01
- d01-x02-y02
- d01-x03-y01
- d01-x03-y02
- d01-x04-y01
- d01-x04-y02
- d01-x05-y01
- d01-x05-y02
- d03-x01-y01
- d03-x01-y02
- d03-x02-y01
- d03-x02-y02
- d03-x03-y01
- d03-x03-y02
- d03-x04-y01
- d03-x04-y02
- d03-x05-y01
- d03-x05-y02
- d04-x01-y01
- d04-x01-y02
- d04-x01-y03
- d04-x01-y04
- d04-x01-y05
- d04-x01-y06
- d04-x01-y07
- d04-x01-y08
- d04-x01-y09
- d04-x01-y10
- d04-x01-y11
- d04-x01-y12
- d04-x01-y13
- d04-x01-y14
- d04-x01-y15
- d04-x01-y16
- d04-x01-y17
- d04-x01-y18
- d04-x01-y19
- d04-x01-y20
- d04-x02-y01
- d04-x02-y02
- d04-x02-y03
- d04-x02-y04
- d04-x02-y05
- d04-x02-y06
- d04-x02-y07
- d04-x02-y08
- d04-x02-y09
- d04-x02-y10
- d04-x02-y11
- d04-x02-y12
- d04-x02-y13
- d04-x02-y14
- d04-x02-y15
- d04-x02-y16
- d04-x02-y17
- d04-x02-y18
- d04-x02-y19
- d04-x02-y20
- d04-x03-y01
- d04-x03-y02
- d04-x03-y03
- d04-x03-y04
- d04-x03-y05
- d04-x03-y06
- d04-x03-y07
- d04-x03-y08
- d04-x03-y09
- d04-x03-y10
- d04-x03-y11
- d04-x03-y12
- d04-x03-y13
- d04-x03-y14
- d04-x03-y15
- d04-x03-y16
- d04-x03-y17

- d04-x03-y18
- d04-x03-y19
- d04-x03-y20
- d04-x04-y01
- d04-x04-y02
- d04-x04-y03
- d04-x04-y04
- d04-x04-y05
- d04-x04-y06
- d04-x04-y07
- d04-x04-y08
- d04-x04-y09
- d04-x04-y10
- d04-x04-y11
- d04-x04-y12
- d04-x04-y13
- d04-x04-y14
- d04-x04-y15
- d04-x04-y16
- d04-x04-y17
- d04-x04-y18
- d04-x04-y19
- d04-x04-y20
- d04-x05-y01
- d04-x05-y02
- d04-x05-y03
- d04-x05-y04
- d04-x05-y05
- d04-x05-y06
- d04-x05-y07
- d04-x05-y08
- d04-x05-y09
- d04-x05-y10
- d04-x05-y11
- d04-x05-y12
- d04-x05-y13
- d04-x05-y14
- d04-x05-y15
- d04-x05-y16
- d04-x05-y17
- d04-x05-y18
- d04-x05-y19
- d04-x05-y20
- d05-x01-y01
- d05-x01-y03
- d05-x01-y04
- d05-x01-y05
- d05-x01-y06
- d05-x01-y07
- d05-x01-y08
- d05-x01-y09
- d05-x01-y10
- d05-x01-y11
- d05-x01-y12
- d05-x01-y13
- d05-x01-y14
- d05-x01-y15
- d05-x01-y16
- d05-x01-y17
- d05-x01-y18
- d05-x01-y19
- d05-x01-y20
- d05-x02-y03
- d05-x02-y05
- d05-x02-y06
- d05-x02-y07
- d05-x02-y08
- d05-x02-y09
- d05-x02-y10
- d05-x02-y11
- d05-x02-y12
- d05-x02-y13
- d05-x02-y14
- d05-x02-y15
- d05-x02-y16
- d05-x02-y17
- d05-x02-y18
- d05-x02-y19
- d05-x02-y20
- d05-x03-y05
- d05-x03-y06
- d05-x03-y07
- d05-x03-y08
- d05-x03-y09
- d05-x03-y10
- d05-x03-y11
- d05-x03-y12
- d05-x03-y13
- d05-x03-y14
- d05-x03-y15
- d05-x03-y16
- d05-x03-y17
- d05-x03-y18
- d05-x03-y19
- d05-x03-y20
- d05-x04-y05
- d05-x04-y07
- d05-x04-y08
- d05-x04-y09
- d05-x04-y10
- d05-x04-y11
- d05-x04-y12
- d05-x04-y13
- d05-x04-y14
- d05-x04-y15
- d05-x04-y16
- d05-x04-y17
- d05-x04-y18
- d05-x04-y19
- d05-x04-y20
- d05-x05-y07
- d05-x05-y09
- d05-x05-y10
- d05-x05-y11
- d05-x05-y12
- d05-x05-y13
- d05-x05-y14
- d05-x05-y15
- d05-x05-y16
- d05-x05-y17
- d05-x05-y18
- d05-x05-y19
- d05-x05-y20

- d06-x01-y01
- d06-x01-y02
- d06-x01-y03
- d06-x01-y04
- d06-x01-y05
- d06-x01-y06
- d06-x01-y07
- d06-x01-y08
- d06-x01-y09
- d06-x01-y10
- d06-x01-y11
- d06-x01-y12
- d06-x01-y13
- d06-x01-y14
- d06-x01-y15
- d06-x01-y16
- d06-x01-y17
- d06-x01-y18
- d06-x01-y19
- d06-x01-y20
- d06-x02-y01
- d06-x02-y02
- d06-x02-y03
- d06-x02-y04
- d06-x02-y05
- d06-x02-y06
- d06-x02-y07
- d06-x02-y08
- d06-x02-y09
- d06-x02-y10
- d06-x02-y11
- d06-x02-y12
- d06-x02-y13
- d06-x02-y14
- d06-x02-y15
- d06-x02-y16
- d06-x02-y17
- d06-x02-y18
- d06-x02-y19
- d06-x02-y20
- d06-x03-y01
- d06-x03-y02
- d06-x03-y03
- d06-x03-y04
- d06-x03-y05
- d06-x03-y06
- d06-x03-y07
- d06-x03-y08
- d06-x03-y09
- d06-x03-y10
- d06-x03-y11
- d06-x03-y12
- d06-x03-y13
- d06-x03-y14
- d06-x03-y15
- d06-x03-y16
- d06-x03-y17
- d06-x03-y18
- d06-x03-y19
- d06-x03-y20
- d06-x04-y01
- d06-x04-y02
- d06-x04-y03
- d06-x04-y04
- d06-x04-y05
- d06-x04-y06
- d06-x04-y07
- d06-x04-y08
- d06-x04-y09
- d06-x04-y10
- d06-x04-y11
- d06-x04-y12
- d06-x04-y13
- d06-x04-y14
- d06-x04-y15
- d06-x04-y16
- d06-x04-y17
- d06-x04-y18
- d06-x04-y19
- d06-x04-y20
- d06-x05-y01
- d06-x05-y02
- d06-x05-y03
- d06-x05-y04
- d06-x05-y05
- d06-x05-y06
- d06-x05-y07
- d06-x05-y08
- d06-x05-y09
- d06-x05-y10
- d06-x05-y11
- d06-x05-y12
- d06-x05-y13
- d06-x05-y14
- d06-x05-y15
- d06-x05-y16
- d06-x05-y17
- d06-x05-y18
- d06-x05-y19
- d06-x05-y20

D.2. ATLAS_2016_I1424838

- d01-x01-y01
- d02-x01-y01
- d03-x01-y01
- d04-x01-y01
- d05-x01-y01
- d06-x01-y01
- d07-x01-y01
- d08-x01-y01
- d09-x01-y01
- d10-x01-y01
- d11-x01-y01
- d12-x01-y01
- d13-x01-y01
- d14-x01-y01
- d15-x01-y01
- d16-x01-y01
- d17-x01-y01
- d18-x01-y01
- d19-x01-y01
- d20-x01-y01
- d21-x01-y01
- d22-x01-y01
- d23-x01-y01
- d24-x01-y01
- d25-x01-y01
- d26-x01-y01
- d27-x01-y01
- d28-x01-y01
- d29-x01-y01
- d31-x01-y01
- d32-x01-y01
- d33-x01-y01
- d34-x01-y01
- d35-x01-y01
- d36-x01-y01
- d37-x01-y01
- d38-x01-y01
- d39-x01-y01
- d40-x01-y01
- d41-x01-y01
- d42-x01-y01
- d43-x01-y01
- d44-x01-y01
- d45-x01-y01
- d46-x01-y01
- d47-x01-y01
- d48-x01-y01
- d49-x01-y01
- d50-x01-y01
- d51-x01-y01
- d52-x01-y01
- d53-x01-y01
- d54-x01-y01
- d55-x01-y01
- d56-x01-y01
- d57-x01-y01
- d58-x01-y01
- d59-x01-y01

D.3. ATLAS_2016_I1426695

- d02-x01-y01
- d03-x01-y01
- d04-x01-y01
- d05-x01-y01
- d06-x01-y01
- d07-x01-y01
- d08-x01-y01
- d09-x01-y01
- d10-x01-y01
- d11-x01-y01
- d12-x01-y01
- d13-x01-y01

D.4. ATLAS_2016_I1467230

- d01-x01-y01
- d01-x01-y02
- d02-x01-y01
- d02-x01-y02
- d03-x01-y01
- d03-x01-y02

D.5. ATLAS_2017_I1509919

- d01-x01-y01
- d02-x01-y01
- d03-x01-y01
- d04-x01-y01
- d05-x01-y01
- d06-x01-y01
- d07-x01-y01
- d15-x01-y01
- d16-x01-y01
- d17-x01-y01
- d18-x01-y01
- d19-x01-y01
- d20-x01-y01
- d21-x01-y01
- d22-x01-y01
- d23-x01-y01
- d24-x01-y01
- d25-x01-y01
- d26-x01-y01
- d29-x01-y01

D.6. ATLAS_2019_I1724098

- d08-x01-y01
- d09-x01-y01
- d10-x01-y01
- d11-x01-y01
- d12-x01-y01
- d13-x01-y01
- d14-x01-y01
- d16-x01-y01
- d17-x01-y01
- d18-x01-y01
- d19-x01-y01
- d20-x01-y01
- d21-x01-y01
- d22-x01-y01
- d30-x01-y01
- d31-x01-y01
- d32-x01-y01
- d33-x01-y01
- d34-x01-y01
- d35-x01-y01
- d36-x01-y01
- d38-x01-y01
- d39-x01-y01
- d40-x01-y01
- d41-x01-y01
- d42-x01-y01
- d43-x01-y01
- d44-x01-y01

D.7. ATLAS_2019_I1746286

- d11-x01-y01
- d12-x01-y01
- d13-x01-y01
- d15-x01-y01
- d16-x01-y01
- d17-x01-y01
- d21-x01-y01

D.8. MC_ttbar_ColorReconnection

- nch
- sumPt

Bibliography

All links listed in this bibliography were last retrieved on the 2nd of June 2021. The documentation is scheduled to move to `https://pythia.org/` in the not so distant future.

[1] L. Evans and P. Bryant, "LHC machine," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 001–S08 001, aug 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/s08001

[2] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN," Dec 2008. [Online]. Available: https://cds.cern.ch/record/1260465

[3] R. Steerenberg, "Lhc report: Another run is over and ls2 has just begun…," 2018. [Online]. Available: https://home.cern/news/news/accelerators/lhc-report-another-run-over-and-ls2-has-just-begun

[4] The ATLAS Collaboration, "The ATLAS experiment at the CERN large hadron collider," *Journal of Instrumentation*, vol. 3, no. 08, pp. S08 003–S08 003, aug 2008. [Online]. Available: https://doi.org/10.1088/1748-0221/3/08/s08003

[5] J. Pequenao, "Computer generated image of the whole ATLAS detector," Mar 2008. [Online]. Available: https://cds.cern.ch/record/1095924

[6] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.

[7] Wikipedia, "Standard model of elementary particles," 2019. [Online]. Available: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg

[8] The ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037026931200857X

[9] The CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the cms experiment at the lhc," *Physics Letters B*, vol. 716, no. 1, p. 30–61, Sep 2012. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2012.08.021

[10] R. Brandelik *et al.*, "Evidence for planar events in e^+e^- annihilation at high energies," *Physics Letters B*, vol. 86, no. 2, pp. 243–249, 1979. [Online]. Available: https://doi.org/10.1016/0370-2693(79)90830-X

[11] S. D. Drell and T.-M. Yan, "Massive lepton-pair production in hadron-hadron collisions at high energies," *Phys. Rev. Lett.*, vol. 25, pp. 316–320, Aug 1970. [Online]. Available: https://doi.org/10.1103/PhysRevLett.25.316

[12] T. Sjostrand, "The Lund Monte Carlo for Jet Fragmentation," *Comput. Phys. Commun.*, vol. 27, p. 243, 1982. [Online]. Available: https://doi.org/10.1016/0010-4655(82)90175-8

[13] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Höche, H. Hoeth, F. Krauss, L. Lönnblad, E. Nurse, P. Richardson, and et al., “General-purpose event generators for lhc physics,” *Physics Reports*, vol. 504, no. 5, p. 145–233, Jul 2011. [Online]. Available: http://dx.doi.org/10.1016/j.physrep.2011.03.005

[14] G. Aad, B. Abbott, J. Abdallah, A. A. Abdelalim, A. Abdesselam, O. Abdinov, B. Abi, M. Abolins, H. Abramowicz, H. Abreu, and et al., “Charged-particle multiplicities in pp interactions measured with the atlas detector at the lhc,” *New Journal of Physics*, vol. 13, no. 5, p. 053033, May 2011. [Online]. Available: http://dx.doi.org/10.1088/1367-2630/13/5/053033

[15] T. Sjöstrand and M. van Zijl, “A multiple-interaction model for the event structure in hadron collisions,” *Phys. Rev. D*, vol. 36, pp. 2019–2041, Oct 1987. [Online]. Available: https://doi.org/10.1103/PhysRevD.36.2019

[16] T. Sjöstrand *et al.*, “Pythia 8 documentation: Multiparton interactions.” [Online]. Available: http://home.thep.lu.se/~torbjorn/pythia81html/MultipartonInteractions.html

[17] T. Sjöstrand, S. Mrenna, and P. Skands, “Pythia 6.4 physics and manual,” *Journal of High Energy Physics*, vol. 2006, no. 05, p. 026–026, May 2006. [Online]. Available: http://dx.doi.org/10.1088/1126-6708/2006/05/026

[18] T. A. Collaboration, “Atlas pythia 8 tunes to 7 TeV data,” Nov 2014. [Online]. Available: https://cds.cern.ch/record/1966419/files/ATL-PHYS-PUB-2014-021.pdf

[19] B. M. W. et al., “Hztool and rivet: Toolkit and framework for the comparison of simulated final states and data at colliders,” May 2006. [Online]. Available: https://arxiv.org/abs/hep-ph/0605034

[20] M. W. Klaus Hamacher, “The next round of hadronic generator tuning heavily based on identified particle data,” November 1995. [Online]. Available: https://arxiv.org/abs/hep-ex/9511011

[21] H. Schulz, “Systematic event generator tuning with professor,” May 2009. [Online]. Available: https://professor.hepforge.org/diplomathesis_h_schulz.pdf

[22] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, and J. E. von Seggern, “Systematic event generator tuning for the lhc,” *The European Physical Journal C*, vol. 65, no. 1, p. 331, Nov 2009. [Online]. Available: https://doi.org/10.1140/epjc/s10052-009-1196-7

[23] A. P. Austin, M. Krishnamoorthy, S. Leyffer, S. Mrenna, J. Müller, and H. Schulz, “Practical algorithms for multivariate rational approximation,” *Computer Physics Communications*, p. 107663, 2020. [Online]. Available: https://doi.org/10.1016/j.cpc.2020.107663

[24] “The hdf group website.” [Online]. Available: https://www.hdfgroup.org/

[25] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[26] S. G. Nash, “Newton-type minimization via the lanczos method,” *SIAM Journal on Numerical Analysis*, vol. 21, no. 4, pp. 770–788, 1984. [Online]. Available: https://doi.org/10.1137/0721052

[27] F. L. Ruffa and C. Gutschow, “Rivet analyses reference: Atlas_2018_i1656578,” 2018. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2018_I1656578.html

[28] C. Bertsche *et al.*, “Rivet analyses reference: Atlas_2017_i1495243,” 2017. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2017_I1495243.html

[29] P. Skands, S. Carrazza, and J. Rojo, “Tuning pythia 8.1: the monash 2013 tune,” Apr 2014. [Online]. Available: https://arxiv.org/abs/1404.5630

[30] T. Sjöstrand *et al.*, “Pythia 8 documentation: Colour reconnection.” [Online]. Available: http://home.thep.lu.se/~torbjorn/pythia82html/ColourReconnection.html

[31] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton fragmentation and string dynamics,” *Physics Reports*, vol. 97, no. 2, pp. 31–145, 1983. [Online]. Available: https://doi.org/10.1016/0370-1573(83)90080-7

[32] J. R. Christiansen and P. Z. Skands, “String formation beyond leading colour,” *Journal of High Energy Physics*, vol. 2015, no. 8, Aug 2015. [Online]. Available: http://dx.doi.org/10.1007/JHEP08(2015)003

[33] J. R. Christiansen and T. Sjöstrand, “Colour reconnection at future $\mathrm{e}^{+}\mathrm{e}^{-}$ colliders,” 2015. [Online]. Available: https://doi.org/10.1140/epjc/s10052-015-3674-4

[34] “Rivet analyses reference.” [Online]. Available: https://rivet.hepforge.org/analyses.html

[35] Z. Varga, R. Vértesi, and G. G. Barnaföldi, “Jet structure studies in small systems,” *Universe*, vol. 5, no. 5, p. 132, May 2019. [Online]. Available: http://dx.doi.org/10.3390/universe5050132

[36] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. A. Abdelalim, O. Abdinov, R. Aben, B. Abi, and et al., “Measurement of jet shapes in top-quark pair events at $\sqrt{s} = 7$ TeV using the atlas detector,” *The European Physical Journal C*, vol. 73, no. 12, Dec 2013. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-013-2676-3

[37] G. P. S. Matteo Cacciari, “Dispelling the n^3 myth for the k_t jet-finder,” 2006. [Online]. Available: https://arxiv.org/abs/hep-ph/0512210

[38] M. Aaboud, G. Aad, B. Abbott, D. C. Abbott, O. Abdinov, A. Abed Abud, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, and et al., “Measurement of jet-substructure observables in top quark, w boson and light jet production in proton-proton collisions at $\sqrt{s} = 13$ TeV with the atlas detector,” *Journal of High Energy Physics*, vol. 2019, no. 8, Aug 2019. [Online]. Available: http://dx.doi.org/10.1007/JHEP08(2019)033

[39] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new higgs-search channel at the large hadron collider,” *Phys. Rev. Lett.*, vol. 100, p. 242001, Jun 2008. [Online]. Available: https://doi.org/10.1103/PhysRevLett.100.242001

[40] G. Aad *et al.*, “Measurement of event-shape observables in $Z \to \ell^+\ell^-$ events in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC,” *Eur. Phys. J. C*, vol. 76, no. 7, p. 375, 2016. [Online]. Available: https://doi.org/10.1140/epjc/s10052-016-4176-8

[41] R. Kwee, “Studies for the development of a minimum bias trigger at low luminosities in the atlas-experiment at the largehadron collider,” July 2007. [Online]. Available: http://atlas-archiv.desy.de/theses/Kwee_dipl.pdf

[42] G. Aad, B. Abbott, J. Abdallah, S. Abdel Khalek, A. A. Abdelalim, O. Abdinov, B. Abi, M. Abolins, O. S. AbouZeid, H. Abramowicz, and et al., “Underlying event characteristics and their dependence on jet size of charged-particle jet events inppcollisions at $\sqrt{s} = 7$ TeV with the atlas detector,” *Physical Review D*, vol. 86, no. 7, Oct 2012. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.86.072004

[43] G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, M. Abolins, O. S. AbouZeid, N. L. Abraham, and et al., “Charged-particle distributions in pp interactions at $\sqrt{s} = 8$ TeV measured with the atlas detector,” *The European Physical Journal C*, Jul 2016. [Online]. Available: https://doi.org/10.1140/epjc/s10052-016-4203-9

[44] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, and et al., “Measurement of charged-particle distributions sensitive to the underlying event in $\sqrt{s} = 13$ TeV proton-proton collisions with the atlas detector at the lhc,” *Journal of High Energy Physics*, vol. 2017, no. 3, Mar 2017. [Online]. Available: http://dx.doi.org/10.1007/JHEP03(2017)157

[45] M. Aaboud, G. Aad, B. Abbott, J. Abdallah, O. Abdinov, B. Abeloos, R. Aben, O. S. AbouZeid, N. L. Abraham, and et al., “Charged-particle distributions at low transverse momentum in $\sqrt{s} = 13$ tev pp interactions measured with the atlas detector at the lhc,” *The European Physical Journal C*, vol. 76, no. 9, Sep 2016. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-016-4335-y

[46] D. Hirschbuehl, S. Wahdan, W. Wagner, A. Knue, T. Kuhl, C. Helsens, T. Moskalets, and F. Deliot, “Measurements of observables sensitive to colour reconnection in $t\bar{t}$ dilepton events with the ATLAS detector at $\sqrt{s} = 13$ TeV (Internal note),” CERN, Geneva, Tech. Rep. ATL-COM-PHYS-2019-888, Jul 2019. [Online]. Available: https://cds.cern.ch/record/2682602

[47] G. Aad, B. Abbott, D. C. Abbott, O. Abdinov, A. A. Abud, K. Abeling, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, and et al., “Measurement of k_s^0 and λ^0 production in $t\bar{t}$ dileptonic events in pp collisions at $\sqrt{s} = 7$ tev with the atlas detector,” *The European Physical Journal C*, vol. 79, no. 12, Dec 2019. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-019-7512-y

[48] J. L. Sergio Calvente, Fernando Barreiro, “Rivet analyses reference: Atlas_2019_i1746286,” 2019. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2019_I1746286

[49] M. D. Morris, “Factorial sampling plans for preliminary computational experiments,” *Technometrics*, vol. 33, no. 2, pp. 161–174, 1991. [Online]. Available: http://dx.doi.org/10.1080/00401706.1991.10484804

[50] P. Golonka and Z. Was, “Photos monte carlo: a precision tool for qed corrections in z and w decays,” *The European Physical Journal C*, vol. 45, no. 1, p. 97–107, Jan 2006. [Online]. Available: http://dx.doi.org/10.1140/epjc/s2005-02396-4

[51] J. M. Campbell, R. K. Ellis, P. Nason, and E. Re, “Top-pair production and decay at nlo matched with parton showers,” *Journal of High Energy Physics*, vol. 2015, no. 4, Apr 2015. [Online]. Available: http://dx.doi.org/10.1007/JHEP04(2015)114

[52] The ATLAS Collaboration, "A study of different colour reconnection settings for Pythia8 generator using underlying event observables," May 2017. [Online]. Available: https://cds.cern.ch/record/2262253/files/ATL-PHYS-PUB-2017-008.pdf

[53] H. Hoeth, A. Buckley, C. Gutschow, D. Mallows, and M. Kawalec, "Rivet analyses reference: Mc_ttbar." [Online]. Available: https://rivet.hepforge.org/analyses/MC_TTBAR

[54] J. Llorente, "Rivet analyses reference: Atlas_2013_i1243871," 2013. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2013_I1243871

[55] D. Kar, A. Vaidya, and C. Gutschow, "Rivet analyses reference: Atlas_2019_i1724098," 2019. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2019_I1724098

[56] H. Schulz, "Rivet analyses reference: Atlas_2016_i1424838," 2016. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2016_I1424838.html

[57] K. Joshi, "Rivet analyses reference: Atlas_2012_i1125575," 2012. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2012_I1125575.html

[58] R. Lysak, "Rivet analyses reference: Atlas_2016_i1426695," 2016. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2016_I1426695.html

[59] ——, "Rivet analyses reference: Atlas_2017_i1509919," 2017. [Online]. Available: https://rivet.hepforge.org/analyses/ATLAS_2017_I1509919.html

Name, Vorname: ....

Erklärung

gem. § 14 Abs. 7 Prüfungsordnung vom 15.04.2013

Hiermit erkläre ich, dass ich die von mir eingereichte Abschlussarbeit (Master-Thesis) selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Stellen der Abschlussarbeit, die anderen Werken dem Wortlaut oder Sinn nach entnommen wurden, in jedem Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

....
Datum Unterschrift

Erklärung

Hiermit erkläre ich mich damit einverstanden, dass meine Abschlussarbeit (Master-Thesis) wissenschaftlich interessierten Personen oder Institutionen und im Rahmen von externen Qualitätssicherungsmaßnahmen des Studienganges zur Einsichtnahme zur Verfügung gestellt werden kann.

Korrektur- oder Bewertungshinweise in meiner Arbeit dürfen nicht zitiert werden.

....
Datum Unterschrift