GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Masterarbeit

Messung der Fehlidentifizierungsrate für hadronische Zerfälle von Tau Leptonen mit dem ATLAS Experiment

Measurement of the Fake Rate for Hadronic Tau Lepton Decays using the ATLAS Experiment

angefertigt von

Timo Dreyer

aus Aurich

am II. Physikalischen Institut

Arbeitsnummer: II.Physik-UniGö-MSc-2016/05

Bearbeitungszeit: 12. April 2016 bis 12. Oktober 2016

Erstgutachter/in: Prof. Dr. Stan Lai

Zweitgutachter/in: Priv.Doz. Dr. Jörn Große-Knetter

Contents

1. Introduction

The tau lepton is the heaviest lepton in the standard model of particle physics (SM) and an important probe of physics at high energy scales. For example, the joint observation of the $H \to \tau\tau$ signal in 2015 by the CMS and ATLAS experiments was the first direct observation of the Higgs boson coupling to fermions [1].

For signatures involving hadronically decaying tau leptons, it is important to have a good understanding of the performance of the tau reconstruction and identification algorithms. In particular, jets originating from quark- and gluon-emissions can be falsely identified as hadronically decaying tau leptons. The fraction of jets for which such a misidentification occurs (the so-called *fake rate*) is important to know in order to estimate the background from a variety of sources. This fake rate depends on the kinematic properties of the jet, as well as the quark-gluon composition of the jets in the chosen selection.

In this thesis, the performance of the tau identification algorithm is analyzed by measuring the fake rate on simulated Monte Carlo events (MC) and on data. The truth record available for MC events allows for the use of a *truth matching* algorithm to distinguish between real tau decays and misidentified jets. This algorithm has been improved and is used to verify the selection used for fake rate measurements.

The *tag and probe* method is used to measure the fake rate in $3.2\,\text{fb}^{-1}$ of data obtained in 2015 with the ATLAS detector at $\sqrt{s} = 13\,\text{TeV}$. This method utilizes the clean event signature of $Z \to ee$ decays to indirectly identify an object as a jet, which is subsequently used as a probe to study the performance of the tau identification algorithm. Using this method, fake rates are calculated on MC samples as well as on 2015 ATLAS data. These two sets of fake rates are compared by calculating the scale factor between them.

In addition, the data is separated into quark and gluon jet enriched regions, in which the fake rate is measured separately. A template fit method is used to estimate the relative amount of quark and gluon initiated jets in the two regions by comparing the distribution of the jet width in each region with templates obtained from MC events using truth matching information. With this information, the measured fake rates are unfolded into pure quark and gluon jet fake rates, which are compared to fake rates obtained from MC

using truth matching.

2. The Standard Model

The Standard Model of particle physics (SM) is a very successful theory that describes the interactions of all known fundamental particles. This chapter will give a short overview of the SM and its particle content with special focus on the Higgs mechanism. The success of the SM as a description of particle physics, but also known problems of the SM, are briefly discussed.

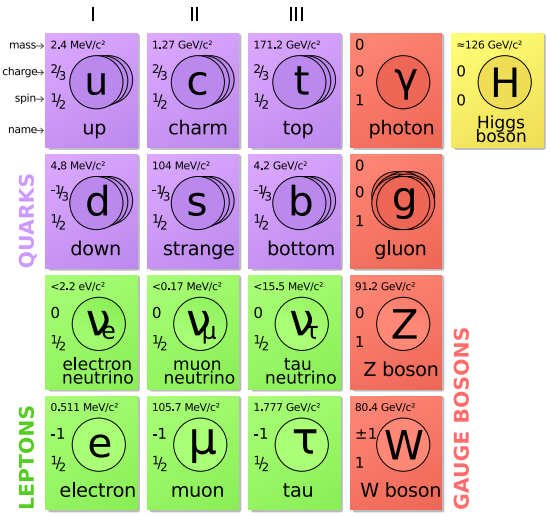
2.1. Overview

The SM ([2–4]) utilizes a locally gauge-invariant Lagrange density to describe the interactions of the fundamental particles via the strong, weak and electromagnetic forces. To ensure the local gauge invariance, additional fields, the so-called *gauge fields*, need to be introduced in the Lagrange density. This addition creates terms in the Lagrange density that relate to massless vector bosons (*gauge bosons*), which act as mediators for the fundamental forces.

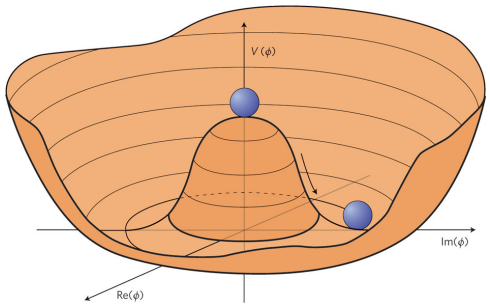
The strong, weak and electromagnetic forces are described in the SM by requiring local gauge-invariance under the symmetry groups $SU(3)_C \times SU(2)_L \times U(1)_Y$. Here the subscripts of the symmetry groups indicate the coupling of the respective force to colour charge (C), the weak isospin (L) or the hypercharge Y. The emerging gauge bosons are eight different *gluons* for the strong interaction, which couple to colour charge, and four bosons for the $SU(2)_L \times U(1)_Y$ symmetry groups, that manifest themselves as the charged W^+ and W^- bosons and the neutral Z^0 and γ bosons, the latter of which is also known as the *photon*.

All of the gauge bosons are *vector bosons*, i.e. spin-1 particles. Apart from these, the SM contains three "families" of spin-½ particles (fermions), each of which contains an *up-type* and a *down-type* quark, a neutrino and a charged lepton. Quarks are the only fermions that carry a colour charge and can therefore participate in the strong interaction. Neutrinos on the other hand carry neither electromagnetic nor colour charge and can only participate in the weak interaction.

The only scalar (spin-0) particle in the SM is the Higgs boson, which is introduced by



(a) Particle content of the SM.



(b) Potential of the Higgs field ϕ.

Figure 2.1.: The Standard Model and the Higgs mechanism.

the Higgs mechanism described below. An overview of the complete particle content of the SM and some properties of these particles is shown in Figure 2.1(a).

2.2. The Higgs Mechanism

The Z^0 and $W^\pm$ bosons, which mediate the weak interaction in the SM, have measured masses of $m_Z = 91.1876 \pm 0.0021\,\text{GeV}$ and $m_W = 80.385 \pm 0.015\,\text{GeV}$ [5]. However, the direct introduction of mass terms for vector bosons into the SM Lagrange density would spoil its local gauge invariance.

To solve this contradiction, a complex doublet of scalar fields is introduced into the Lagrange density:

$$\phi(\mathbf{x}) = \begin{pmatrix} \phi_1(\mathbf{x}) + i\phi_2(\mathbf{x}) \\ \phi_3(\mathbf{x}) + i\phi_4(\mathbf{x}) \end{pmatrix}, \tag{2.1}$$

where $\phi_i(\mathbf{x})$ are real-valued functions of the location $\mathbf{x}$. The potential of this field is given by:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \qquad \mu^2 < 0, \quad \lambda > 0. \tag{2.2}$$

The symmetric potential of this new field has a global minimum that is at distance

v from the origin, where v is known as the *vacuum expectation value* of the field. This leads to *spontaneous symmetry breaking* of the potential (Figure 2.1(b)). When the field is expressed as an expansion around this new ground state, terms for the masses of the Z^0 and $W^\pm$ bosons can be obtained without breaking the local gauge invariance of the Lagrange density. This process is known as the Brout-Englert-Higgs mechanism ([6–8]) and predicts an additional massive scalar particle, the *Higgs boson*, for which a candidate has been discovered in 2012 [9, 10].

This Higgs boson interacts with fermions and vector bosons via Yukawa coupling terms (Equation 2.3 and Figure 2.2). For the interaction with two fermions the coupling strength g_{Hff} is proportional to to the mass m_f of the involved fermion. The coupling strengths for vertices with one Higgs boson and two vector bosons g_{HVV} or with two Higgs bosons and two vector bosons g_{HHVV} are quadratically dependent on the vector boson mass m_V:

$$g_{Hff} = \frac{\sqrt{2}m_f}{v}, \qquad g_{HVV} = \frac{2m_V^2}{v}, \qquad g_{HHVV} = \frac{m_V^2}{v}. \tag{2.3}$$

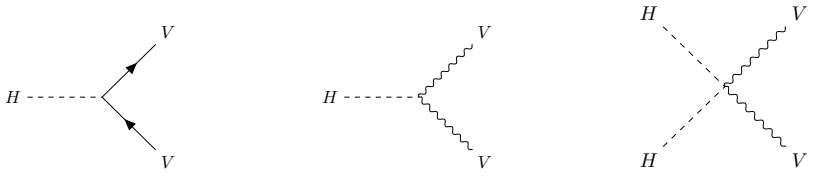


Figure 2.2.: Standard Model couplings of the Higgs boson to fermions and vector bosons.

2.3. SM Nature of the Higgs Boson

To confirm the SM nature of the Higgs boson discovered in 2012, its quantum numbers and especially the couplings to other particles need to be precisely measured. Deviations from the SM predictions in these quantities could give important hints at physics beyond the Standard Model. Since the Higgs mechanism predicts a scaling of the Higgs bosons couplings to other particles with the masses of these particles, it is important to verify this dependence.

Due to the mass dependence of the couplings, production mechanisms and decay channels of the Higgs boson that involve top quark loops and Z^0 and $W^\pm$ bosons are strongly favoured. These are also the channels in which the Higgs boson was first discovered. However, since the mass of the Higgs boson is lower than the mass of the top quark, decays

into a $t\bar{t}$ pair are kinematically forbidden. Therefore, the coupling strength to fermions had only been measured indirectly via top-loops in Feynman diagrams. The first direct measurement of the coupling to fermions was the measurement of the $H \to \tau^-\tau^+$ decay in 2015 [1].

2.3.1. The Decay $H \to \tau\tau$

Due the high mass of the tau lepton in comparison with other leptons, the branching ratio of a Higgs boson decaying into a $\tau^-\tau^+$ pair is the second highest branching ratio of all decays into fermions with BR$= 6.30 \pm 0.36\,\%$, while the dominant fermion decay produces a bottom quark pair with BR $57.5 \pm 1.9\,\%$ (Table 2.1) [11]. However, the decay mode into bottom quarks has background processes that are harder to suppress than the backgrounds for $H \to \tau\tau$ events. Therefore, this decay is a good probe for the Yukawa couplings of the Higgs boson to fermions and allowed for the first direct measurement of such a coupling at a significance of 5.5σ [1].

Decay Channel	Branching Ratio [%]
$H \to bb$	57.5 ± 1.9
$H \to WW$	21.6 ± 0.9
$H \to gg$	8.56 ± 0.86
$H \to \tau\tau$	6.30 ± 0.36
$H \to cc$	2.90 ± 0.35
$H \to ZZ$	2.67 ± 0.11
$H \to \gamma\gamma$	0.228 ± 0.011
$H \to Z\gamma$	0.155 ± 0.014
$H \to \mu\mu$	0.022 ± 0.001

Table 2.1.: SM predictions for the branching ratios of a Higgs boson with a mass of 125.09 GeV [11].

2.4. Success and Problems of the SM

The SM has been very successful at making predictions like the discoveries of the top quark [12] and the Higgs boson [9, 10]. Beyond these big discoveries, almost all particle physics measurements performed are compatible with the SM. Figure 2.3, for example, displays several measurements of total and fiducial cross sections performed by ATLAS experiment and compares them with the respective SM predictions. Even though the individual cross sections are distributed over almost 10 orders of magnitude, the measurements are compatible with the predictions within the statistical uncertainties.

Despite this enormous success, various problems suggest that the SM cannot be a complete theory. A selection of some of these problems are listed below [13–15].

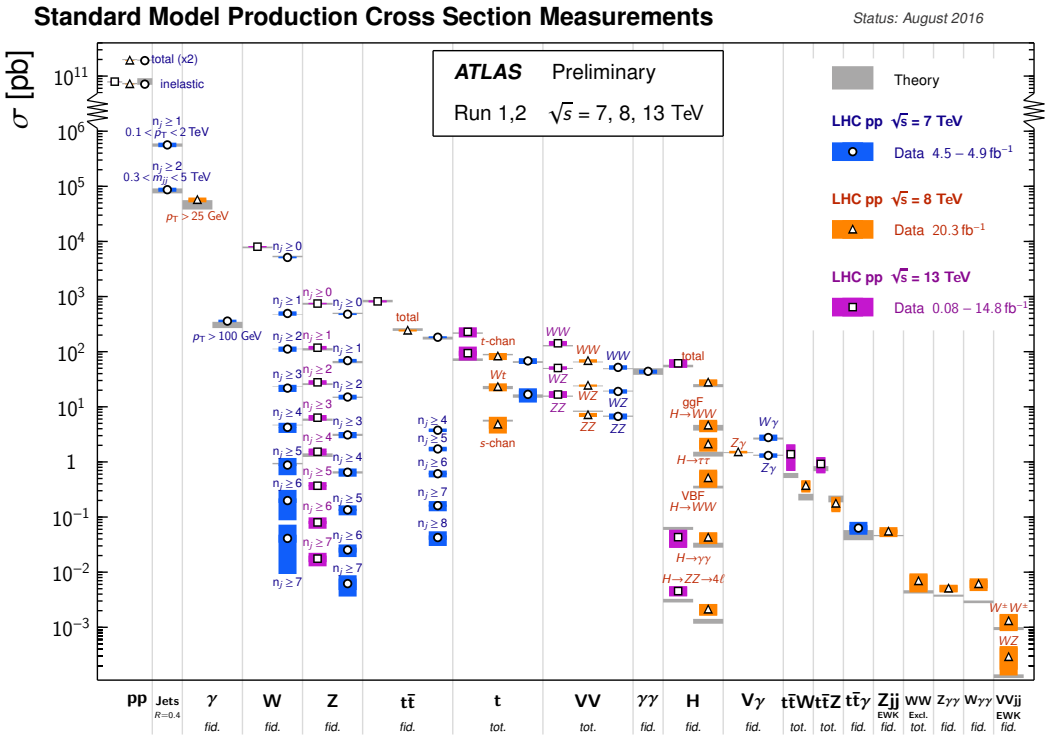


Figure 2.3.: Summary of several Standard Model total and fiducial production cross section measurements performed by the ATLAS experiment and comparison to theoretical predictions.

2.4.1. Gravity

Gravity is the only one of the four fundamental forces that is not included in the SM. It is the dominant force on large scales and very successfully described by general relativity for cosmological purposes. On particle physics scales its effects are negligible due to the small masses and high energies involved.

It is possible to introduce a fourth force in the SM that yields general relativity as the classical limit and introduces a spin-2 boson (the *graviton*) as its mediator. However, such a theory would not be renormalizable and therefore is not able to make any physically meaningful predictions.

2.4.2. Dark Matter

Cosmological observations, e.g. of rotational velocity profiles in galaxies or the cosmological microwave background (CMB), suggest that the universe contains significantly more mass than what can be directly observed from cosmic radiation. From these observations the existence of an unknown *dark matter* (DM) has been postulated.

Due to the lack of observable radiation from it, DM is thought to consist of massive, stable particles that interact only via the weak interaction, the so-called *weakly interacting massive particles* (WIMPs). The only particles in the SM that fit the description of WIMPs are neutrinos. However, the neutrino mass limits of $< 2\,\text{eV}$ [5] are to low to allow for them to explain all the DM in the universe.

Searches for DM at particle colliders like the LHC mainly focus on the creation of DM particles in the collision. Due to the weak interaction of DM particles, they would leave the detector without being measured, which would be visible as missing transverse energy in the event.

2.4.3. The Higgs Mass Hierarchy Problem

In the SM, the observed mass of a particle is determined by the behaviour of the particle's propagator. Since the SM is a quantum field theory, the propagator is a superposition of many possible diagrams, including the splitting of the particle into two other particles, that almost immediately recombine again to the initial particle (Figure 2.4).

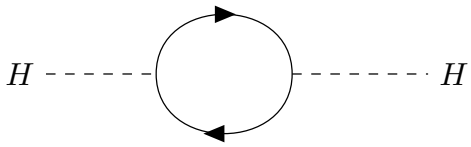


Figure 2.4.: First order loop correction to the Higgs propagator.

Due to these *loop corrections* to its propagator, the observed mass of the particle is different from its "bare mass". For the Higgs boson, as the only fundamental scalar particle in the SM, these corrections lead to a strong divergence of its observed mass in the form of:

$$M_h^2 = (M_h^{bare})^2 + \mathcal{O}(\Lambda^2), \tag{2.4}$$

where Λ is the limit up to which the momenta in the loop are integrated. Without this limit the integration would diverge. Such a limit is the energy scale at which new physics

phenomena become relevant. An upper boundary on this scale is given by the *Planck scale* in the order of 10^{18} GeV, where the gravitational force can no longer be neglected compared to the forces included in the SM.

If no new physics exists below the Planck scale, it would seem natural for the Higgs mass to be in the order of magnitude of this upper boundary. Since the Higgs boson has an finite observed mass of about 125 GeV, the SM parameters that determine the loop corrections would need to be very *fine tuned* in such a way that different corrections cancel each other out. Such a fine tuning on the order of 16 magnitudes is deemed *unnatural* and seen as an indication for additional processes beyond the SM that can reduce the divergent behaviour.

2.5. Physics beyond the SM

To solve these problems of the SM, a great variety of theories for *physics beyond the Standard Model* (BSM) have been proposed. Due to the success of the SM in making verifiable predictions, these theories are required to include the SM as a valid approximation in the energy regime where it has been proven successful. Therefore most BSM theories introduce new effects that occur at high energies.

Searches for BSM physics can take two basic approaches: The search for new particles or phenomena at high energies or precision measurements in known processes. The latter approach searches for deviations in differential cross section measurements that can be caused by BSM couplings or particles in higher order loop corrections to the Feynman diagrams of the process. Since BSM physics is expected to appear at high energies, measurements in processes with heavy particles are most promising. The Higgs boson is especially interesting, since many BSM theories alter the Higgs sector of the SM, e.g. by introducing a second complex Higgs doublet into the SM Lagrange density.

3. The ATLAS Experiment

The ATLAS Collaboration maintains one of the four main Experiments hosted at the LHC in Geneva. An overview of the LHC, the ATLAS experiment and the data taking process is given in this chapter.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) at the CERN facilities in Geneva is a circular proton proton collider with a circumference of 26.7 km, which is located in a tunnel 175 m below the surface to shield it from the atmospheric radiation [16]. It uses multiple smaller accelerator rings as pre-accelerators for the protons, including its predecessors the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) (Figure 3.1(a)).

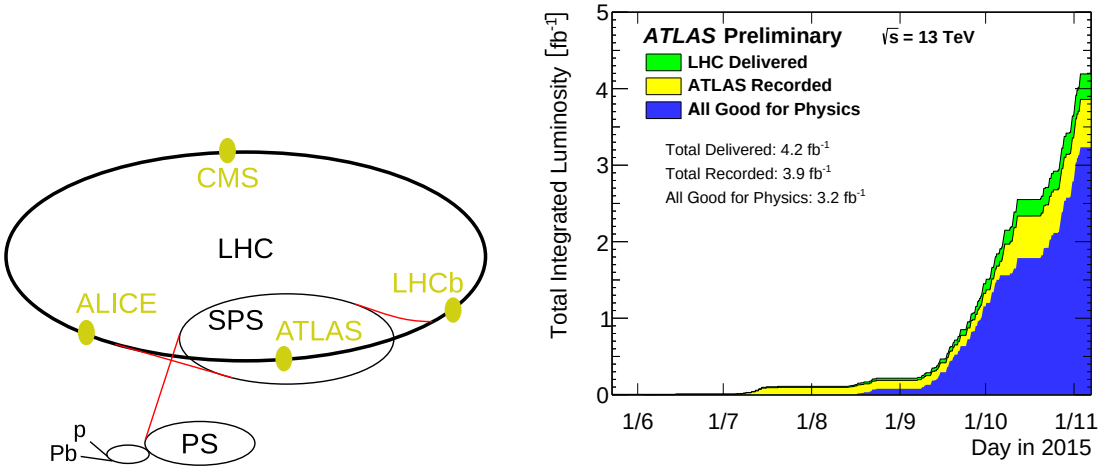
During Run I from 2010 to 2013, the LHC produced 5.46 fb^{-1} of data at a centre-of-mass energy of $\sqrt{s}$=7 TeV and 22.8 fb^{-1} at $\sqrt{s}$=8 TeV. After an upgrade, the LHC has started Run II in June 2015 with higher collision energies. In 2015, collisions have been taking place at a centre-of-mass energy of $\sqrt{s}$=13 TeV and the LHC delivered 4.2 fb^{-1} of which the ATLAS experiment uses 3.2 fb^{-1} for physics analyses (Figure 3.1(b)). It is planed to increase the centre-of-mass energy further to $\sqrt{s}$=14 TeV during Run II.

The LHC provides collisions for the six experiments ATLAS, CMS, LHCb, ALICE, LHCf and TOTEM. The ATLAS and CMS experiments each consist of a general-purpose detector that is designed to cover a wide range of physics.

3.1.1. Phenomenology of Hadron Collider Physics

During Run II of the LHC, bunches of up to 10^{11} protons collide every 25 ns at the four collision points. At each of these collisions, multiple interactions can occur between the protons of the colliding beams. Since usually only one of the occurring interaction events is interesting for a given analysis, the other, so-called *pileup* events represent an unwanted background for the analysis.

Since hadrons are not fundamental particles, their inner structure has to be taken into account when describing collisions at a hadron collider like the LHC. Protons consist



(a) Accelerator complex of the LHC.

(b) Total Integrated Luminosity and Data Quality in 2015 as recorded by the ATLAS experiment [17].

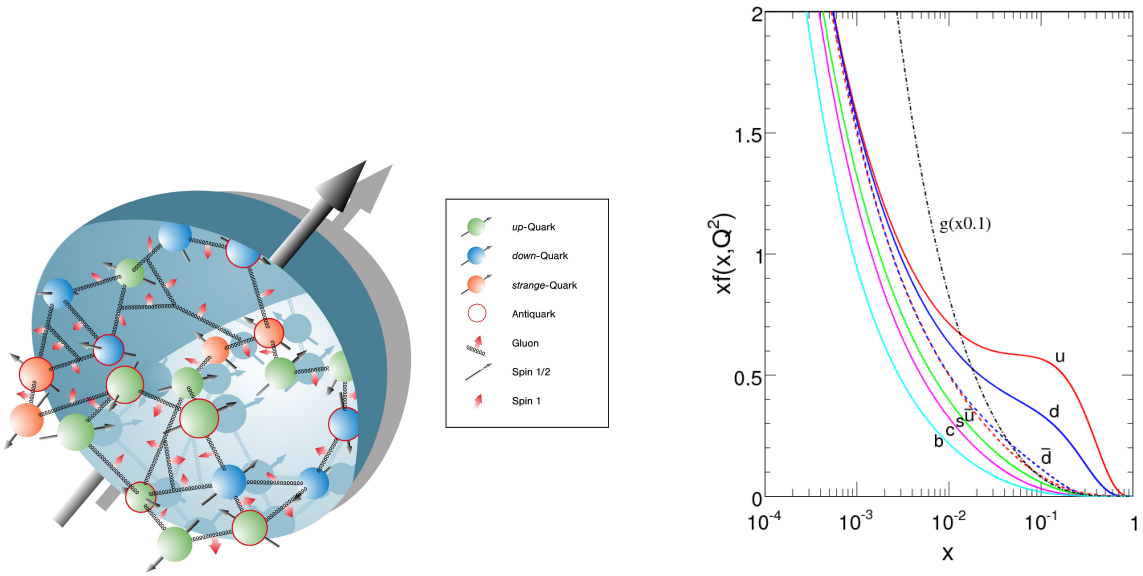
Figure 3.1.: On the Large Hadron Collider.

of three *valence quarks* (two up and one down quark) that are bound together due to the colour confinement of QCD. The strong interaction between these three quarks is mediated by gluons, which can temporarily split up into a quark-antiquark pair. This large number of QCD interactions inside of the proton leads to the presence many *sea quarks and gluons* (Figure 3.2(a)).

These virtual particles inside the proton can take part in the fundamental interaction of a collision event and will contribute a certain fraction x of the total momentum of the proton (the so-called *Bjorken x*). Depending on the centre-of-mass energy at which the collision takes place and the Bjorken x of the partons interacting in the fundamental process, the probability for certain partons to contribute changes. The corresponding probability distribution is given by the *parton distribution function* (Figure 3.2(b)).

3.2. The ATLAS Detector

The ATLAS (A Toroidal LHC ApparatuS, [19]) Experiment consists of a general-purpose particle detector build around one of the four collision points of the LHC (Figure 3.3). It can be divided into the inner detector, the calorimeters, the muon system and the magnet systems.



(a) Scheme of the quarks and gluons inside a proton.

(b) Parton distribution functions at an energy scale of Q^2=100 GeV^2. The gluon distribution is scaled by a factor of 0.1 [18].

Figure 3.2.: On the particle content of protons.

3.2.1. The ATLAS Coordinate System

The coordinate system of the ATLAS detector is based on the beampipe, which is defined as the z-axis of the system, where the positive z-direction points anti-clockwise along the LHC ring. The x-axis of the coordinate system is defined to point towards the centre of the LHC ring.

Alternative coordinates used in the ATLAS experiment are the azimuthal angle ϕ around the z-axis, which is defined with respect to the x-axis, and the pseudorapidity $\eta = -\ln\tan(\theta/2)$, where θ is the polar angle with respect to the z-axis. Distances between two objects reconstructed in the ϕ-η plane are typically given as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$.

3.2.2. Detector Components

The Inner Tracking Detector The inner tracking detector consists of 4 (3 in Run I) layers of silicon pixel detectors close to the beampipe surrounded by silicon strip detectors and an outer shell of a transition radiation tracker. The later allows for the differentiation of electrons and pion tracks by exploiting the dependence of the transition radiation on the relativistic γ factor.

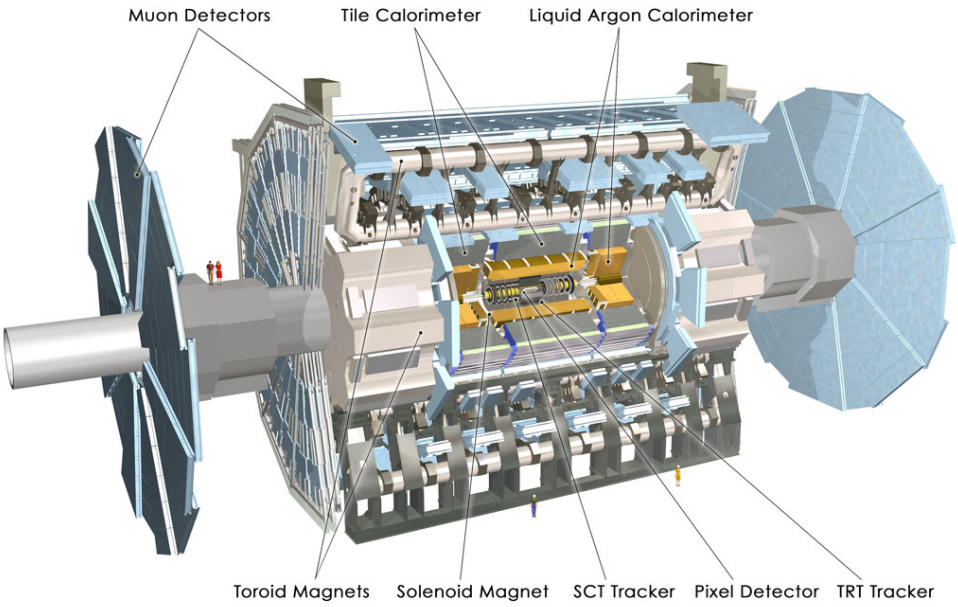


Figure 3.3.: The ATLAS detector [19].

This inner detector is embedded inside a 2 T solenoidal magnet system used for measurements of the transverse momentum p_T. These are performed by measuring the radius of curvature r of the tracks produced by charged particles.

The p_T resolution is dominated by the measurement of the radius of curvature r for a track, which has a higher uncertainty for high p_T particles, since these have straighter tracks. The resolution of the ATLAS experiment is:

$$\frac{\sigma_{p_T}}{p_T} = 5 \times 10^{-4}\, p_T[\text{GeV}] \oplus 0.01$$

The Calorimeter System Installed around the inner tracking detector is a calorimeter system consisting of an electromagnetic calorimeter (ECAL) surrounded by a hadronic calorimeter (HCAL). The electromagnetic calorimeter is build as a sampling calorimeter with lead as the active material and a liquid argon scintillator, while the hadronic calorimeter uses a plastic scintillator and an iron absorber.

The resolutions of the ATLAS calorimeters are:

$$\frac{\sigma_{E_T}}{E_T} = \frac{0.1}{\sqrt{E_T[\text{GeV}]}} \oplus 0.007 \qquad \text{(ECAL)}$$

$$\frac{\sigma_{E_T}}{E_T} = \frac{0.5}{\sqrt{E_T[\text{GeV}]}} \oplus 0.03 \qquad \text{(HCAL)}$$

A segmented strip layer in the electromagnetic calorimeter allows to separate the two photon produced in a typical pion decay from single electrons and photons.

The Muon System Since muons are *minimal ionising particles* at their expected energy scales, they are the only (measurable) particles that are expected to traverse the entire detector without being stopped. Therefore, the outermost part of the detector is the muon system. It surrounds the calorimeters and contains a toroidal magnetic system.

3.3. The ATLAS Analysis Data Flow

A simplified scheme of the data flow in a typical ATLAS analysis involving tau leptons is shown in Figure 3.4. The output of the ATLAS detector (or its simulated output for MC events) is processed by triggers that determine which events are being saved for analysis. These events are completely stored in the xAOD data format, which can produce file sizes in the order of petabytes.

To reduce the amount of data, so-called *derivations* are produced that only contain events and event information that is necessary for a certain analysis. This process results in a smaller *DxAOD* file, which use the same file format as the full xAOD but only have a size in the order of terabytes.

In an additional step, the information in the DxAOD is transformed into the variables that are used for the analysis, which are then stored in an *ntuple*, i.e. a ROOT [20] file containing a flat `TTree` object which holds the desired variables. On this file, which has a typical size in the order of gigabytes, the actual analysis is performed and plots are created.

3.3.1. Trigger

Due to the high collision rate of 40 MHz at the LHC, it is not possible to store each event the ATLAS detector records. To reduce the rate of data that is stored to disk, a *trigger system* is used, that consists of the hardware-based *Level-1* trigger and the software-

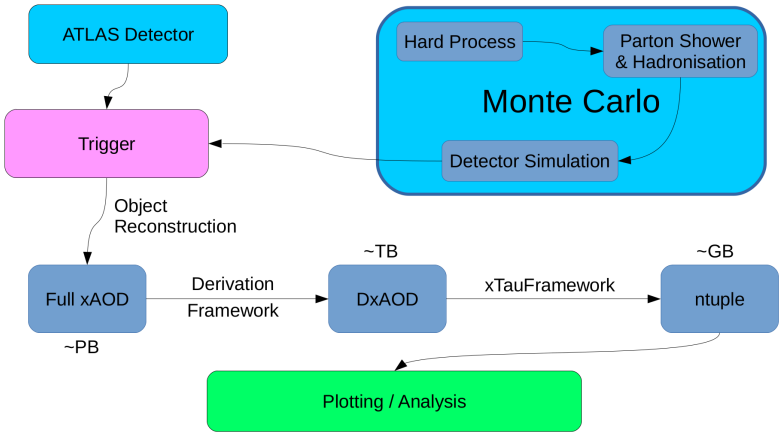


Figure 3.4.: Scheme of the data flow in an ATLAS analysis involving tau leptons.

based *high level trigger* (HLT) [21]. The purpose of this system is to identify events with topologies that are promising for the different analyses.

The Level-1 trigger uses calorimeter and muon detector information to define *Regions-of-Interest* (RoIs) in the detector. It takes 2.5 µs to decide whether an event is accepted by the Level-1 trigger. This decision is necessary for the further processing of the data measured by the individual front ends of the detector and already reduces the event rate down to 100 kHz.

Based on the Level-1 trigger decision and the RoIs defined by it, the HLT is activated. This level consists of sophisticated selection algorithms using more detailed event information than the previous level. In a decision time of 200 ms, it reduces the event rate further to roughly 1 kHz. With an event size of approximately 1.3 Mb, this results in a data rate of about 1.3 Gb/s.

4. Tau Leptons

With a mass of $1776.86 \pm 0.12\,\mathrm{MeV}$, the tau lepton is the third heaviest fermion and the heaviest lepton in the SM. Like all charged leptons, it carries an electric charge of -1 and its third isospin component is $-\frac{1}{2}$. It has a mean lifetime of $\tau_\tau = 290.3 \pm 0.5\,\mathrm{fs}$ which relates to a proper time of approximately $87.03\,\mathrm{\mu m}/c$ [5].

Tau leptons can either decay into a lighter lepton and the corresponding neutrinos or hadronically with mesons in the final state. The corresponding branching ratios are given in Table 4.1. In the following sections only hadronically decaying tau leptons are discussed in greater detail, since the leptonic decays are nearly indistinguishable from the direct production of the corresponding lepton.

τ^- Decay Mode	Branching Ratio [%]
$e^-\bar{\nu}_e\nu_\tau$	17.83 ± 0.04
$\mu^-\bar{\nu}_\mu\nu_\tau$	17.41 ± 0.04
$h^-\nu_\tau \geq 0$ neutrals	48.63 ± 0.12
$h^-h^-h^+\nu_\tau \geq 0$ neutrals	15.20 ± 0.08
Other decays	0.93 ± 0.16

Table 4.1.: Branching ratios for the different tau lepton decay channels [5]. The mesons $h^\pm$ can be either $\pi^\pm$ or $K^\pm$, "neutrals" are neutral mesons such as π^0. The branching ratio for "other decays" is derived from the values of the explicitly named decays in the table.

4.1. Hadronic Tau Decays

Hadronic decays of tau leptons produce mainly pions and kaons and are further classified by the number of charged tracks, or "prongs", that are produced in the decay. Due to charge conservation, the tau lepton can only decay into an odd number of charged particles. Since the phase space is larger for decays into less particles, the 1 prong decays occur about three times as often as 3 prong events. Higher prong numbers can also occur but are neglected, since their branching ratio is very low. In addition to the charged

mesons it is possible that neutral mesons are produced as well. The prong-classification is usually inclusive in the number of additional neutral particles.

Figure 4.1 shows a schematic for a possible 3 prong tau decay. A jet cone containing the decay products of the tau lepton is usually much narrower than a typical cone for a QCD jet and the leading particles inside the jet tend to carry higher fractions of the initial momentum of the tau lepton. Additionally, the mean lifetime of the tau lepton makes it possible to reconstruct the secondary vertex from which the jet originates. Along with the typical number of one or three charged tracks, these are the main quantities that allow for the identification of a hadronic tau decay.

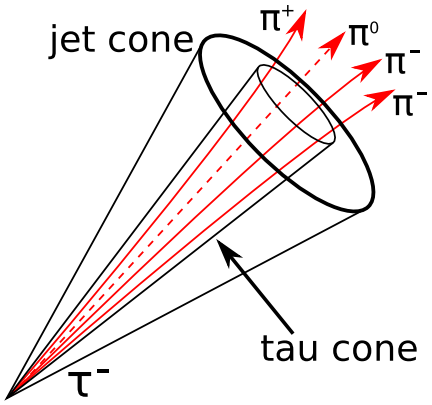


Figure 4.1.: Scheme of a 3 prong hadronic decay of the tau lepton.

4.2. Tau Lepton Reconstruction

The reconstruction of hadronic tau decays in the ATLAS experiment starts by forming candidates for the visible part of hadronically decaying tau leptons ($\tau_{\text{had-vis}}$ candidates) out of jets, that are formed from calorimeter cell clusters (Topocluster [22]) using the anti-k_t algorithm [23] with a distance parameter $R = 0.4$. To be further considered as $\tau_{\text{had-vis}}$ candidates, the formed jets are required to satisfy $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$.

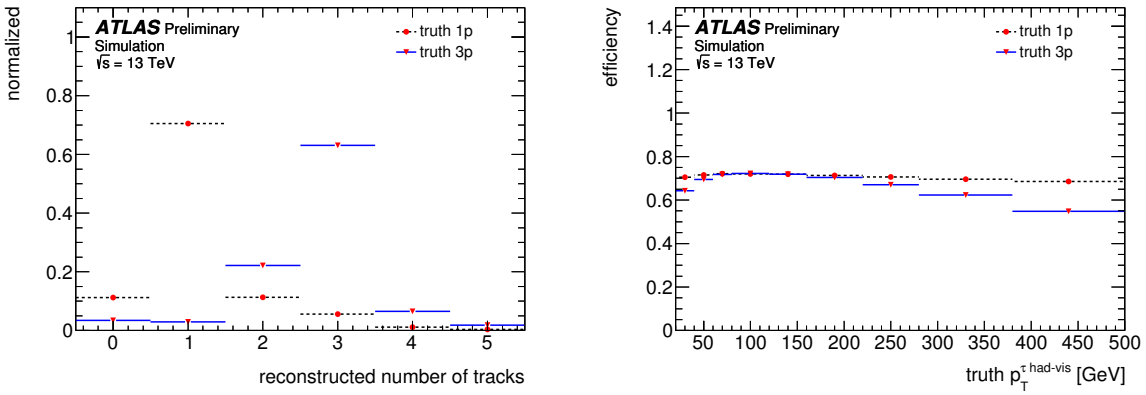
These jets are associated with one of the reconstructed vertices in the event, the so-called tau vertex (TV). For each primary vertex (PV), which is matched to tracks within the $\Delta R < 0.2$ region around the reconstructed tau candidate, the p_T sum of these tracks is calculated. The PV with the highest p_T sum is then chosen as the TV.

All tracks within the $\Delta R < 0.2$ region of the candidate that fulfil the following quality criteria are associated with it:

- $p_T > 1\,\text{GeV}$
- ≥ 2 hits in the pixel detector
- ≥ 7 hits in the pixel detector and the SCT detectors combined
- $|d_0| < 1.0\,\text{mm}$ (distance from the TV in the transverse plane)
- $|\Delta z_0 \sin\theta| < 1.5\,\text{mm}$ (longitudinal distance from the TV)

These track association criteria for the $\tau_{\text{had-vis}}$ candidate are optimised to maximize the number of candidates with correctly reconstructed charged particle multiplicity [24]. Figure 4.2(a) shows the distribution of the reconstructed number of tracks for true 1 and 3 prong events. Usually, only $\tau_{\text{had-vis}}$ candidates with exactly one or three associated tracks are considered for further analysis.

The total efficiency of the reconstruction algorithm, which is defined as the fraction of 1 or 3 prong hadronic tau decays that are reconstructed with the correct number of associated tracks, is shown in Figure 4.2(b) as a function of the true visible transverse momentum of the hadronically decaying tau lepton.



(a) Number of reconstructed charged tracks for true 1 and 3 prong tau decays.

(b) Reconstruction efficiency as a function of the true p_T of the tau.

Figure 4.2.: Efficiencies for the reconstruction of a $\tau_{\text{had-vis}}$ candidate for true 1 prong and 3 prong hadronic tau decays [24].

4.3. Tau Lepton Identification

The selection criteria applied in the tau lepton reconstruction described in Section 4.2 are also satisfied by a large fraction of QCD jets, which therefore provide a large background to the identification of hadronic tau decays. To distinguish between these jets and real

tau decays, the ATLAS experiment uses a *boosted decision tree* (BDT) for an additional tau lepton identification step [24].

4.3.1. Boosted Decision Trees

When trying to separate two classes of events ("signal" and "background"), a boosted decision tree is a possible method to combine multiple observables into a single variable that has a higher separation power than any of the individual observables.

A single decision tree is binary tree with a simple if-statement at each node. For each event the statement at the "trunk" of the tree is validated and the corresponding "branch" is followed to the next node. The statement at the new node is again validated and the procedure continues until an end node ("leaf") is reached. Each end node is labelled as either signal or background depending on which category the majority of candidates in this node belong to. A single decision tree can be interpreted as a set of simple rectangular cuts on the parameter space of the observables.

For a BDT, an entire set of trees (a "forest") is trained and the output of each tree is interpreted as a number (typically 1 if the output is "signal" and 0 if it is "background"). For each event the outputs of all trees are calculated and then combined into a single number (usually by averaging the outputs) that is used as the discriminant with higher values for signal-like and lower values for background-like events (Figure 4.3).

The training of the forest typically consists of creating and optimising one tree, checking its performance on a MC sample, giving a higher weight to falsely classified events and training a new tree on the reweighed MC events. All tree iterations produced in this process are then combined to build the forest.

A BDT gives the advantage that instead of one single very complicated tree, many sim-

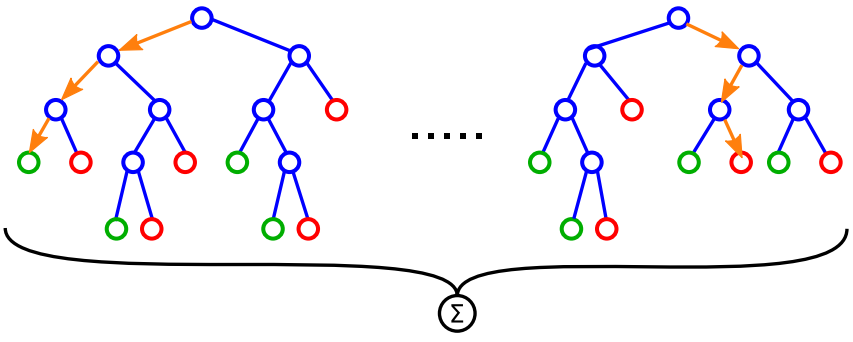


Figure 4.3.: Scheme of a boosted decision tree. In each node a binary decision decides the further path (arrows) until an end node corresponding to one of the possible categories (red or green) is reached. The (weighted) sum over the decisions of the entire forest yields the final BDT-score.

ple decision trees can be used, which allows for a highly parallelized computing approach and therefore *boosts* the calculation. Compared to simple rectangular cuts on the parameter space, it is possible to describe much more complicated acceptance regions with a BDT.

4.3.2. BDT Input Variables

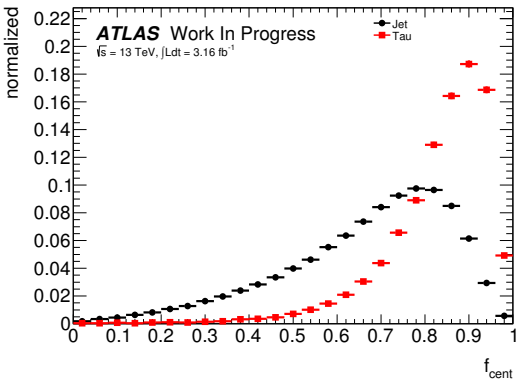
The tau identification BDT uses nine input variables for $\tau_{\text{had-vis}}$ candidates with 1 track and ten input variables for candidates with 3 tracks [24]. Table 4.2 lists the variables used for both numbers of tracks; detailed descriptions for each variable can be found in Section A.2.

Variable	1 track	3 track
f_{cent}	•	•
$f^{-1}_{\text{leadtrack}}$	•	•
$R^{0.2}_{\text{track}}$	•	•
$\lvert S_{\text{leadtrack}}\rvert$	•	
$f^{\text{track}}_{\text{iso}}$	•	
ΔR_{Max}		•
$S^{\text{flight}}_{\text{T}}$		•
m_{track}		•
$f^{\text{track-HAD}}_{\text{EM}}$	•	•
$f^{\text{EM}}_{\text{track}}$	•	•
$m_{\text{EM+track}}$	•	•
$p^{\text{EM+track}}_{\text{T}}/p_{\text{T}}$	•	•

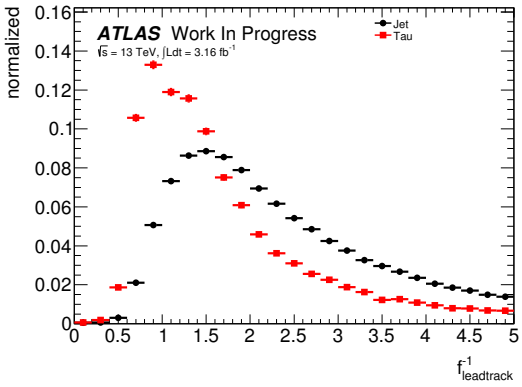
Table 4.2.: Input variables for the tau identification BDT. Detailed descriptions of the definitions of these variables are given in [24].

These twelve variables describe mainly the features of hadronic tau decays mentioned in Section 4.1, namely the narrowness of the produced jet, the secondary vertex and low particle multiplicity. The Figures 4.4 and 4.5 display the distribution of the BDT input variables for true tau decays and jets. For the distribution of real taus in these plots, the leading $\tau_{\text{had-vis}}$ candidate in a $Z \to \tau\tau$ MC sample is used. The jet distributions are produced from the leading $\tau_{\text{had-vis}}$ candidates in a $Z \to ee$ (+jets) MC sample.

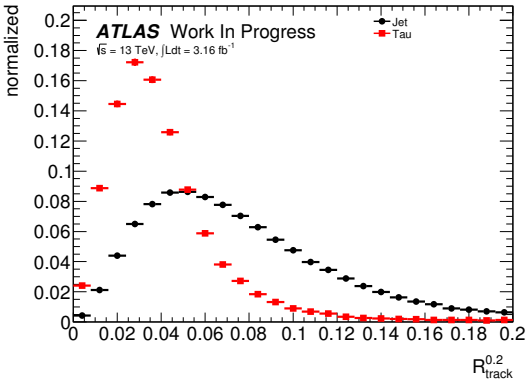
Both MC samples were generated with POWHEG [25] at a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ and interfaced with PYTHIA 8 [26] for parton showering using the AZNLO tune [27] and CTEQ6L1 [28] as the parameterization for parton distribution functions (see Section 5.1). To verify the identity of the used $\tau_{\text{had-vis}}$ candidates, the truth matching algorithm described in Section 5.2 has been applied.



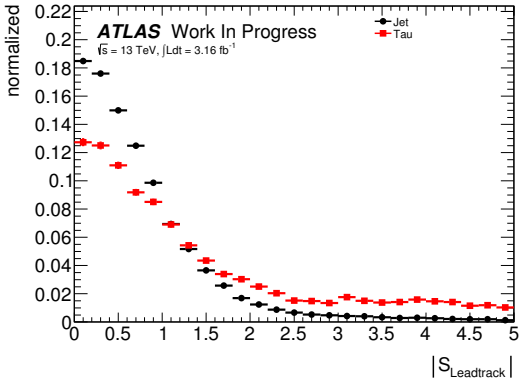
(a) Central energy fraction for tau candidates with 1 or 3 tracks.



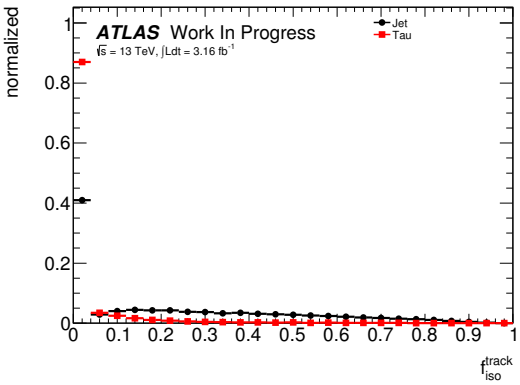
(b) Leading track momentum fraction for tau candidates with 1 or 3 tracks.



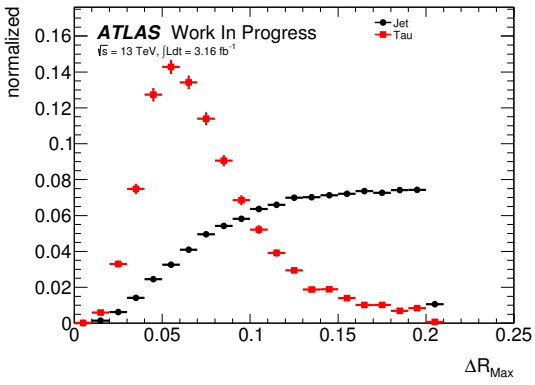
(c) Track radius for tau candidates with 1 or 3 tracks.



(d) Leading track IP significance for tau candidates with 1 track.

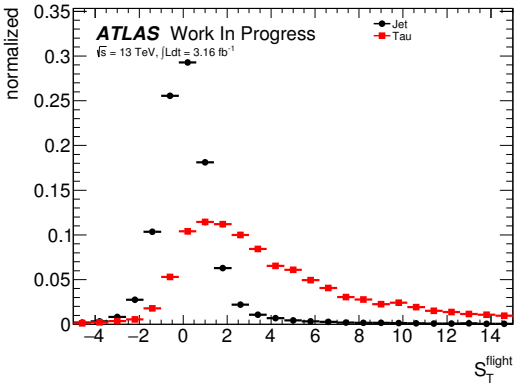


(e) Fraction of tracks p_T in the isolation region for tau candidates with 1 track.

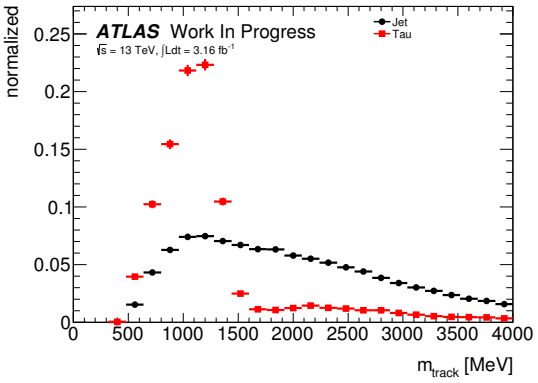


(f) Maximum ΔR for tau candidates with 3 tracks.

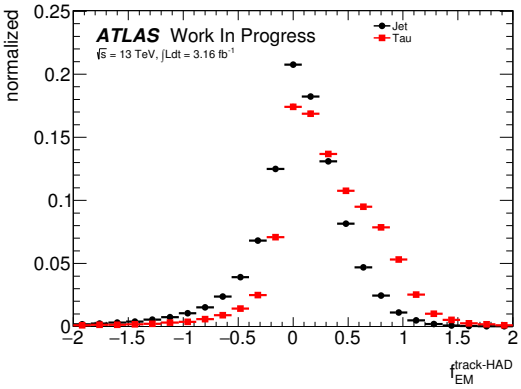
Figure 4.4.: Input variables for the tau identification BDT. The jet and tau distributions are obtained from the truth matched leading $\tau_{\text{had-vis}}$ candidates in a $Z \to ee$+jets and a $Z \to \tau\tau$ MC sample respectively.



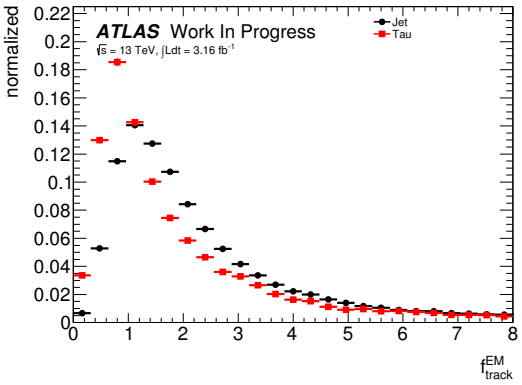
(a) Transverse flight path siginificance for tau candidates with 3 tracks.



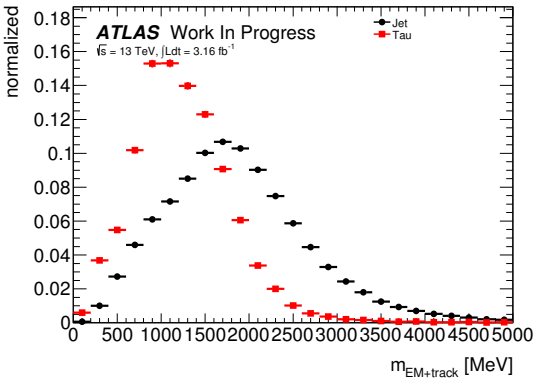
(b) Track mass for tau candidates with 3 tracks.



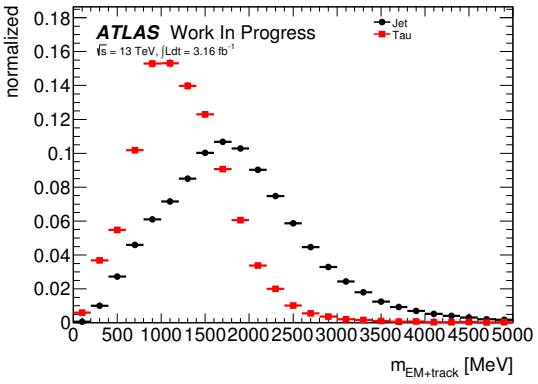
(c) Fraction of EM energy from charged pions for tau candidates with 1 or 3 tracks.



(d) Ratio of EM energy to track momentum for tau candidates with 1 or 3 tracks.



(e) Track-plus-EM-system mass for tau candidates with 1 or 3 tracks.

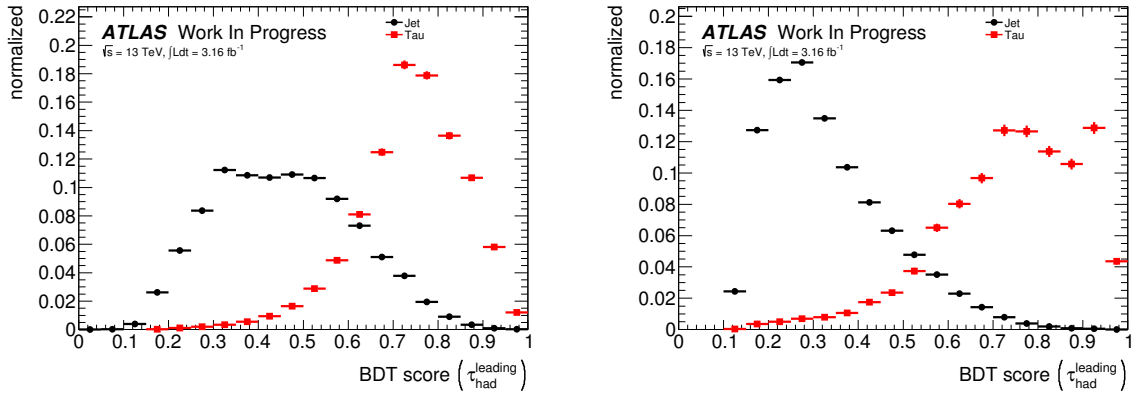


(f) Ratio of track-plus-EM-system to p_T for tau candidates with 1 or 3 tracks.

Figure 4.5.: Input variables for the tau identification BDT. The jet and tau distributions are obtained from the truth matched leading $\tau_{\text{had-vis}}$ candidates in a $Z \to ee$+jets and a $Z \to \tau\tau$ MC sample respectively.

4.3.3. BDT Working Points and Efficiency

The tau identification BDT provides a score between 0 and 1 for each reconstructed $\tau_{\text{had-vis}}$ candidate (Figure 4.6). When the score of a candidate is higher than a certain minimum score, it passes the identification criterion. Three different working points (*loose*, *medium* and *tight*) of the tau identification algorithm are defined by requiring different minimum BDT scores. These BDT thresholds are tuned as a function of the $\tau_{\text{had-vis}}$ transverse momentum to gain a relatively constant efficiency value for the combined tau reconstruction and identification. The efficiencies are defined as the fraction of 1 or 3 prong hadronic tau decays that are reconstructed with the correct number of associated tracks and are passing the corresponding BDT threshold. The desired signal efficiencies for the different tunes are listed in Table 4.3, while the actually achieved efficiencies as functions of $\tau_{\text{had-vis}}$ p_T are shown in Figure 4.7.



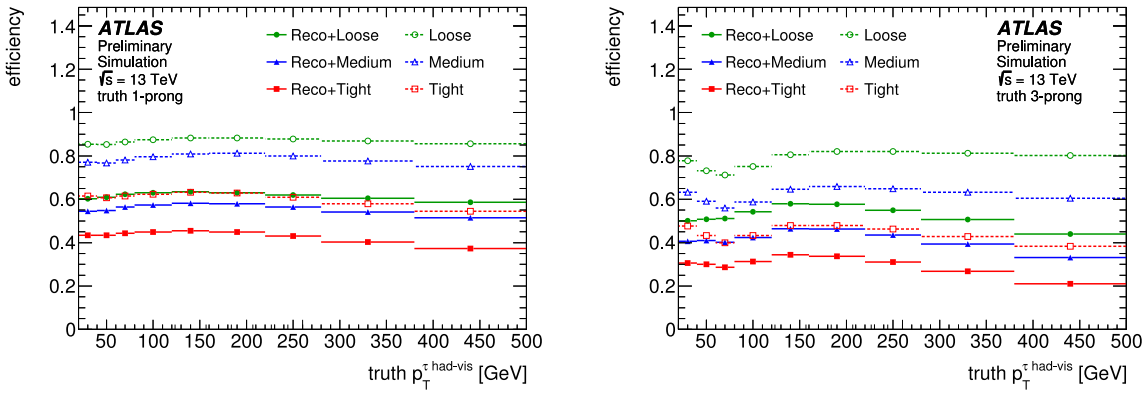
(a) 1 reconstructed track.

(b) 3 reconstructed track.

Figure 4.6.: BDT scores for true hadronic tau decays and QCD jets with 1 or 3 reconstructed tracks. The jet and tau distributions are obtained from the truth matched leading $\tau_{\text{had-vis}}$ candidates in a $Z \to ee$+jets and a $Z \to \tau\tau$ MC sample respectively.

Working Point	1 prong	3 prong
Loose	60 %	50 %
Medium	55 %	40 %
Tight	45 %	30 %

Table 4.3.: Combined signal efficiencies for tau reconstruction and identification the BDT is tuned to achieve.



(a) 1 prong decays.

(b) 3 prong decays.

Figure 4.7.: Efficiency for the tau identification algorithm at different working points for true hadronic 1 and 3 prong decays of tau leptons [24].

5. Monte Carlo Studies

This chapter introduces the Monte Carlo (MC) samples used in this thesis and the selection criteria which are applied for the further analysis. The truth matching algorithm for the $\tau_{\text{had-vis}}$ candidates is discussed. The MC predictions for kinematic distributions under the applied selection criteria are compared to the 2015 ATLAS data and a reweighting of the MC samples is applied to improve the modelling of the data.

5.1. Datasets and Selection

The analysis in this thesis is performed on data collected in 2015 with the ATLAS detector, which corresponds to an integrated luminosity of $3.2\,\text{fb}^{-1}$. The MC samples used for this thesis were generated with POWHEG [25] and interfaced with PYTHIA 8 [26] for parton showering using the AZNLO tune [27] and CTEQ6L1 [28] as the parametrisation for parton distribution functions. The samples are produced for the inclusive $Z \to ee$ process in proton proton collisions with additional jets in the final state at a centre-of-mass energy of 13 TeV. Additional MC samples of other processes with similar final states are used to estimate the background contribution in the final selection. Table 5.1 lists all MC samples used in this theses.

Each MC event also contains simulation of the so-called *pileup*, which is produced at the LHC due to the fact that for each bunch crossing, multiple collisions occur inside the ATLAS detector. The particles produced in the different collisions of one bunch crossing and the neighbouring crossings will be detected essentially simultaneously with the desired event and therefore influence the reconstruction and identification efficiencies.

5.1.1. Object Selection and Overlap Removal

The following analysis requires the definition of electron and tau objects ($\tau_{\text{had-vis}}$ candidates) for the tag and probe method. In addition, muons object are defined, which are necessary to remove other objects in the same detector region that are likely to be caused by the muons (*overlap removal*) and to apply a veto on the identification of electrons. The object definitions used in this thesis are the following:

Dataset ID	Description	Events	Cross Section [pb]	$\int \mathcal{L}dt$ [fb^{-1}]
361106	$Z \to ee$	20.0 M	1901.2	10.52
361107	$Z \to \mu\mu$	20.0 M	1901.2	10.51
361108	$Z \to \tau\tau$	39.5 M	1901.2	20.77
361100	$W^+ \to e^+\nu_e$	30.0 M	11306.0	2.65
361101	$W^+ \to \mu^+\nu_\mu$	30.0 M	11306.0	2.65
361102	$W^+ \to \tau^+\nu_\tau$	30.0 M	11306.0	2.65
361103	$W^- \to e^-\bar{\nu}_e$	40.0 M	8282.6	4.83
361104	$W^- \to \mu^-\bar{\nu}_\mu$	20.0 M	8282.6	2.41
361105	$W^- \to \tau^-\bar{\nu}_\tau$	20.0 M	8282.6	2.41
410000	$t\bar{t}$	50.0 M	696.11	132.21

Table 5.1.: Overview of the MC samples used for the analysis. All listed samples were generated with POWHEG [25] and interfaced with PYTHIA 8 [26] for parton showering using the AZNLO tune [27] and CTEQ6L1 [28] as the parametrisation for parton distribution functions. The last column states the integrated luminosity corresponding to the number of events in the sample, which can be obtained by dividing the number of events by the cross section.

- Electrons are required to pass the medium identification criteria for electrons that are commonly used in the ATLAS experiment [29]. Additionally they also need to fulfil $p_T > 20\,\text{GeV}$, $|\eta| < 2.5$ and they should not overlap with an object that passes the muon definition.
- For muons, quality criteria are applied as well as the cuts $p_T > 10\,\text{GeV}$ and $|\eta| < 2.0$. They also need to fulfill a loose isolation criterion [30].
- Candidates for hadronic decays of tau leptons ($\tau_{\text{had-vis}}$ candidates) must carry an absolute charge equal to 1 and fulfil $p_T > 20\,\text{GeV}$ and $|\eta| < 2.5$. They are required to have one or three associated tracks. Additionally they should not overlap with muons or electrons that pass the definitions above. Tau candidates within the crack region of the detector at $1.37 < |\eta| < 1.52$ are excluded from the analysis.

5.1.2. Event Selection

The $Z \to ee$ events for the analysis are preselected by using a single electron trigger with the additional requirement that the particle activating the trigger needs to be matched to the leading reconstructed lepton ℓ_{lead} in the event. Since the low-threshold L1 trigger includes an absolute isolation criterion, this results in a relative isolation criterion that decreases with the transverse momentum of the electron candidate. Therefore the triggers with a p_T threshold of 24 GeV and 60 GeV are only used for p_T values of the leading lepton below 65 GeV and 135 Gev respectively (Table 5.2). The efficiencies of the different triggers

applied to MC and data for a $p_T(\ell_{\text{lead}})$ are corrected by a scale factor (see Section 5.3.1).

The events are further required to contain at least two electrons and at least one $\tau_{\text{had-vis}}$ candidate. Any event containing muons is rejected.

In accordance with the tag and probe method described in Section 6.1.2, the two leading electrons are further required to pass a medium electron identification criterion, be well isolated and carry opposite charge to each other. The p_T threshold of the leading electron is raised to 26 GeV, while the threshold for the sub-leading electron remains at 20 GeV. The reconstructed invariant mass of the two electrons needs to be compatible with the known Z^0 boson mass of about 91.19 GeV within ± 5 GeV, which corresponds to roughly two times the full decay width $\Gamma_Z = 2.4952 \pm 0.0023$ GeV of the Z^0 boson [5].

After this event selection, the leading $\tau_{\text{had-vis}}$ candidate is used for the fake rate measurement.

Type	$p_T(\ell_{\text{lead}})$ Cut	Trigger
MC	< 65 GeV	`HLT_e24_lhmedium_L1EM18VH`
Data	< 65 GeV	`HLT_e24_lhmedium_L1EM20VH`
MC and Data	< 135 GeV	`HLT_e60_lhmedium`
MC and Data	unlimited	`HLT_e120_lhloose`

Table 5.2.: Single electron triggers applied for the analysis. In addition, a trigger match to the leading lepton is required. Each event needs to pass at least one of these criteria.

5.2. Truth Matching

One benefit of MC samples compared to data is the access to the so-called *truth-level information*. While on data it is only possible to access the measurements taken with the detector, which need to be combined in order to reconstruct and identify from which particle/process it was originally created, MC events allow to directly access the "true" particles with all their properties *before* they interact with the detector material (or its simulation).

The process of associating a reconstructed object in the detector with the corresponding truth particle is known as *truth matching*. For this thesis, a truth matching algorithm of the `xTauFramework` has been used to match candidates for hadronically decaying tau leptons. This algorithm was extensively edited as part of the investigations presented in this thesis.

5.2.1. The Initial Matching Algorithm

The initial truth matching algorithm of the `xTauFramework` works as follows: For each reconstructed $\tau_{\text{had-vis}}$ candidate, a ΔR search cone is defined around the direction of its momentum ($\Delta R < 0.2$ for leptons, $\Delta R < 0.4$ for partons). Only truth particles with a momentum vector within this cone are considered for the truth matching (Figure 5.1). Out of these particles, the one with the highest transverse momentum is chosen as the match. For tau truth particles, no such selection based on the transverse momentum is performed. If tau truth particles are present within the ΔR search cone, this leads to a "random" match to the tau truth particle that holds the last position on the MC truth particle list.

This algorithm matches the leading $\tau_{\text{had-vis}}$ candidate to a tau truth particle in approximately 0.05 % of the $Z \to ee$ MC events (over 9.4 million). In these events, however, no prompt hadronic tau decays should be present. The truth particles to which the candidates are matched registered in the truth record of the MC events as daughter particles of different mesons, most of which contain a b quark. Thus these truth particles are part of a QCD jet to which the $\tau_{\text{had-vis}}$ candidate should have been matched instead.

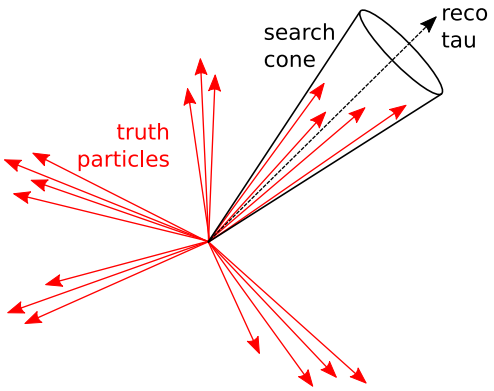


Figure 5.1.: Scheme for the truth matching algorithm. Only particles within a search cone of $\Delta R < 0.2$ for leptons and $\Delta R < 0.4$ for partons are considered in the matching algorithm.

5.2.2. Improvements to the Tau Truth Matching

To improve the truth matching, the following changes have been implemented:

Parent Particle Check To avoid the algorithm to match to tau truth particles that are part of a QCD jet, the parent particle of a tau truth particle is required to be a $W^{\pm}$, Z^0 or Higgs boson. Since the truth record of the parton showering does not store mediating gauge bosons in cases where the tau lepton originates from the decay of hadrons, this

additional check ensures the tau truth particles to be *prompt* taus, i.e. they originate from the simulated hard process.

With this modification no more matches to tau lepton are found in the $Z \to ee$ MC sample.

Strict Matching Priority Matches to tau truth particles are preferred, if possible, to avoid a random selection of possible matches depending on the order of execution. If no matches to tau leptons are possible, the priority to find a truth match moves to other leptons and finally to quarks and gluons.

Least-ΔR Matching Particles passing the matching criteria, the one with the lowest ΔR towards the $\tau_{\text{had-vis}}$ candidate is chosen as the match. This principle can be overruled by the matching priority described above.

Visible Momentum Previously the ΔR between the tau truth particles and the reconstructed $\tau_{\text{had-vis}}$ candidate was calculated with the full four momentum of the truth tau. However, in a hadronic decay of a tau lepton at least one neutrino is produced which carries away a certain fraction of the tau lepton's momentum. The modified truth matching corrects for this fact by subtracting the momenta of neutrino daughter particles from the truth tau momentum before the matching is performed. This is important, since only the visible decay products determine the reconstructed momentum direction of the $\tau_{\text{had-vis}}$ candidate.

This calculation can be validated by comparing the reconstructed transverse momentum of hadronic tau decays p_T^{reco} in $Z \to \tau\tau$ MC events to the calculated visible momentum $p_T^{vis-truth}$ of the tau truth particle to which the candidate is matched. The distribution of the relative difference $(p_T^{reco} - p_T^{vis-truth})/p_T^{vis-truth}$ between the two values (Figure 5.2) peaks at zero, as expected. The positive tail is slightly larger than the negative one, which is likely caused by the application of a p_T threshold on the reconstructed $\tau_{\text{had-vis}}$ candidate which is higher than the threshold on the truth particles to which they are matched.

5.3. Data/MC Comparison

Before the MC samples are used in the analysis, different weights are applied to the events, which scale the number of events in the MC sample to the expected number in data and are supposed to improve the modelling of the data by the MC.

This section describes the applied weighting to the MC events and compares the resulting distributions to data from the 2015 data taking period of the ATLAS detector.

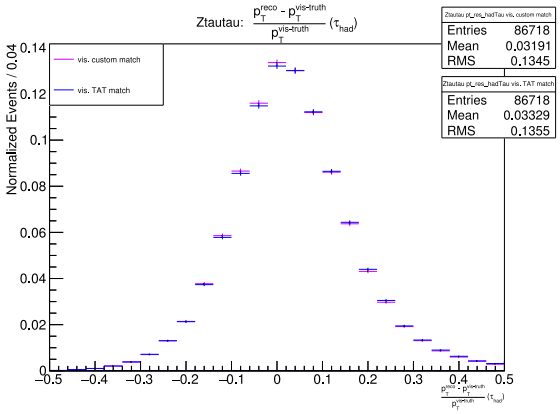


Figure 5.2.: Validiation plots for the tau truth matching. The distribution is calculated from the leading candidate for a hadronic tau lepton decay in the $Z \to \tau\tau$ MC sample.

5.3.1. Weighting of MC Events

The amount of pileup during a data taking period is influenced by many different experimental factors, which makes it nearly impossible to predict the exact distribution. Therefore, the events of MC samples that are simulated with a predicted pileup distribution before the data taking period are weighted to match the measured pileup distribution of the data that needs to be modelled.

The applied selection of events can also influence the modelling of the data. Additional scale factors are applied to the MC events to compensate for these effects. Since the primary criteria in the tag and probe selection are based on the electrons in the event, scale factors are used to correct the simulated trigger, reconstruction, identification and isolation efficiencies to the ones measured in data. These are binned in the p_T and η of the electron and are applied to each event individually.

After the distributions of an MC sample is corrected by applying a weight w_i to each event i, the total number of events must be scaled to the expected number of events N_{exp}^X of the given process X in data. This number is given by the product of the integrated luminosity $\int \mathcal{L}dt$ of the data and the cross section σ_X of the process.

Since the total number of modelled events is changed by the weights applied to the MC sample, the expected number N_{exp}^X should be compared to the sum of these weights to obtain the scale factor W_{lumi}^X:

$$W_{lumi}^X = \frac{N_{exp}^X}{N_{MC}^X} = \frac{\sigma_X \cdot \int \mathcal{L}dt}{\sum_i w_i} \tag{5.1}$$

5.3.2. Reweighting to the Z^0 Momentum

After the weights are applied to the MC events, the number of events passing the selection described in Section 5.1 are determined for both the different MC samples and the 2015 ATLAS data which corresponds to $3.2\,\text{fb}^{-1}$ (Table 5.3). The selection is obviously dominated by the signal events. However, the combined number of MC events is significantly higher than the number of data events.

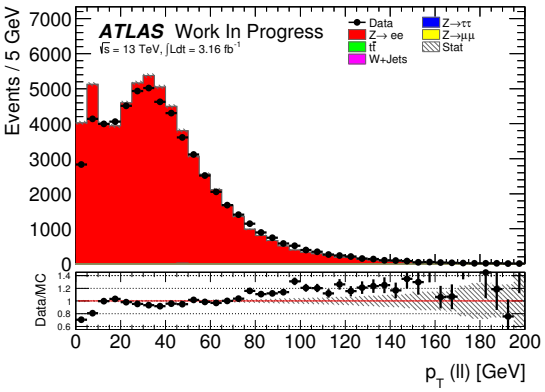
Process	Before Reweighting	After Reweighting
$Z \to ee$	$63\,220 \pm 220$	$60\,480 \pm 210$
$t\bar{t}$	137 ± 8	137 ± 8
$Z \to \tau\tau$	6.4 ± 2.3	5.8 ± 2.1
$W \to \ell\ell$	3.5 ± 2.5	3.5 ± 2.5
$Z \to \mu\mu$	0	0
Sum MC:	$63\,360 \pm 220$	$60\,630 \pm 210$
Data:	$58\,950 \pm 240$	$58\,950 \pm 240$

Table 5.3.: Event yields for the different MC samples and the 2015 ATLAS data under the selection criteria described in Section 5.1. The MC samples are weighted to the data luminosity of $3.2\,\text{fb}^{-1}$. A reweighting is applied on the MC samples of Z^0 decays to improve the modelling.

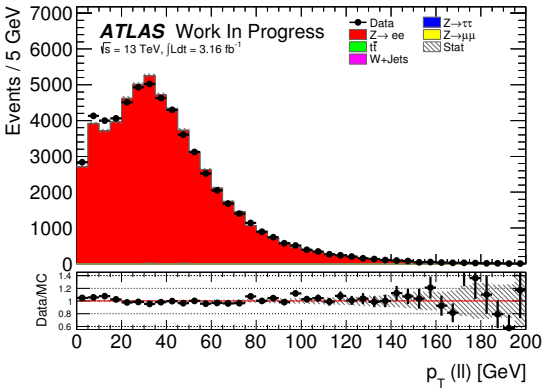
To verify the shape modelling of the weighted MC events, their kinematic distributions are compared to the distributions from data. The Figure 5.3(a) displays this comparison in the distribution of the transverse momentum of the recombined Z^0 boson. The left-hand plots in Figure 5.4 also shows the ΔR distribution of the two leading leptons and the transverse momentum of the leading lepton and the leading $\tau_{\text{had-vis}}$ candidate. All four distributions show (in addition to the normalisation problem) a significant slope in the ratio of data to MC events.

Since the kinematic distribution of the three objects are correlated, a mismodelling of the p_T of the Z^0 boson could also produce a mismodelling in the other distributions. Therefore, the data/MC ratio of the transverse momentum distribution of the Z^0 boson has been applied as an additional scale factor to reweight the MC events. Figure 5.3(c) shows the applied scale factors as a function of the reconstructed Z^0 boson p_T in each event and Figure 5.3(d) displays the final distribution of the weights for all MC events passing the selection criteria.

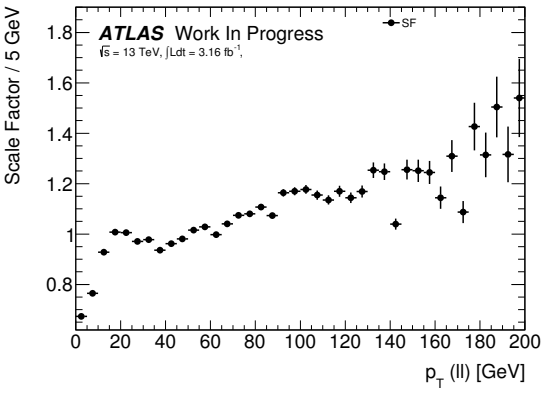
The Figure 5.3(b) and the right-hand plots in Figure 5.4 display the same comparisons of data to MC samples after this reweighting has taken place. The normalisation is now by design in good agreement with the data, but also the slope in the data/MC ratio of the distributions has vanished for the Z^0 boson and the leading lepton. The p_T distribution of



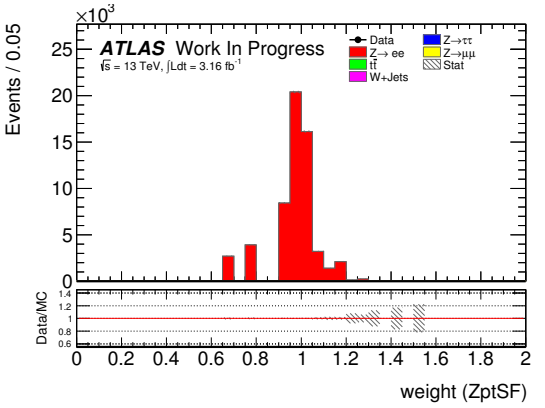
(a) Transverse momentum p_T of the Z^0 boson as reconstructed from the two leading leptons in the event (Before reweighting).



(b) Transverse momentum p_T of the Z^0 boson as reconstructed from the two leading leptons in the event (After reweighting).



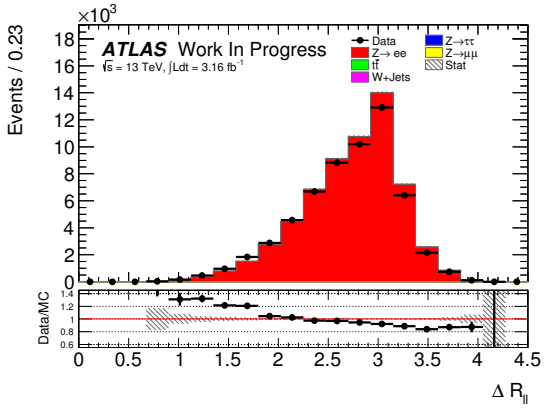
(c) Weights used for the reweighting of the MC samples as a function of the reconstructed p_T of the Z boson.



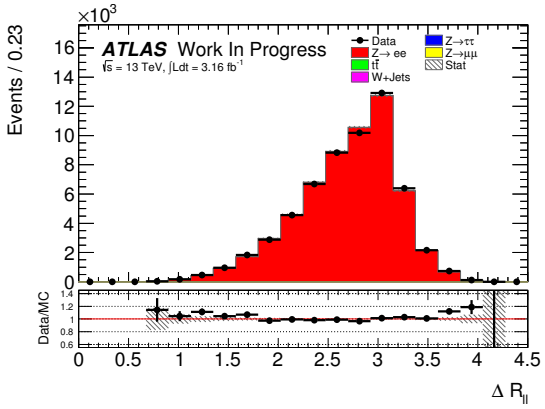
(d) Distribution of applied weights for the MC events that are passing the event selection.

Figure 5.3.: Weights used for the reweighting of the MC samples.

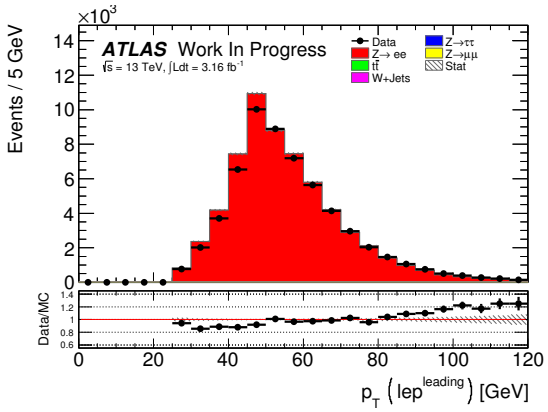
the leading $\tau_{\text{had-vis}}$ candidate still shows a small slope, but since the further analysis will consider the measurements that are binned in this observable, this slight mismodelling is deemed acceptable and the effect is neglected.



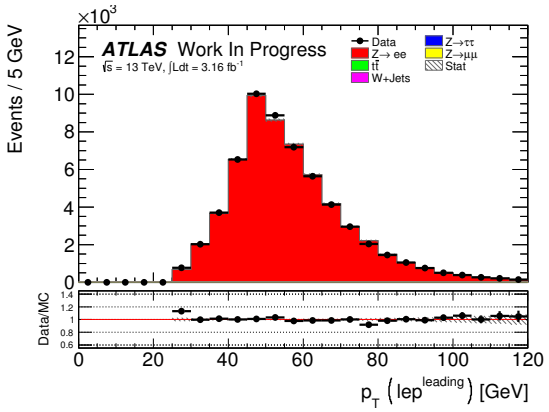
(a) ΔR between the two leading leptons in the event (Before reweighting).



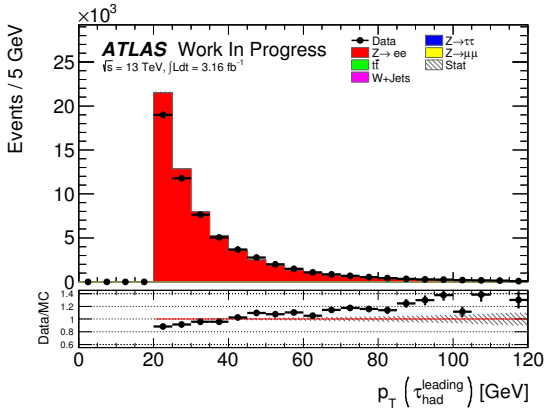
(b) ΔR between the two leading leptons in the event (After reweighting).



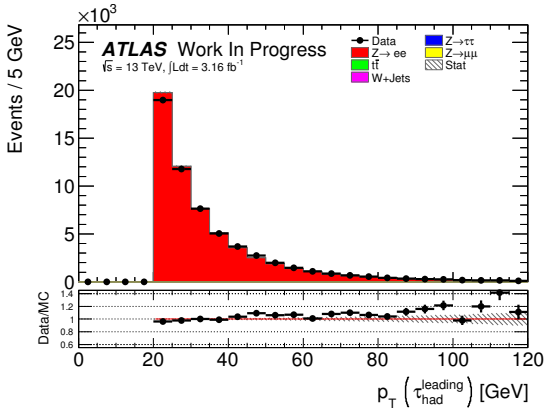
(c) Transverse momentum p_T of the leading lepton in the event (Before reweighting).



(d) Transverse momentum p_T of the leading lepton in the event (After reweighting).



(e) Transverse momentum p_T of the leading $\tau_{\text{had-vis}}$ candidate in the event (Before reweighting).



(f) Transverse momentum p_T of the leading $\tau_{\text{had-vis}}$ candidate in the event (After reweighting).

Figure 5.4.: Comparison of data/MC agreement in different variables before and after the MC events have been reweighted. The contributions of MC samples other than $Z \rightarrow ee$ are too small to be visible after the tag-and-probe selection has been applied.

6. Fake Rate Measurements

This chapter discusses the measurement of fake rates on data and MC samples with the tag-and-probe method. At first, systematic uncertainties are calculated for the MC samples. The fake rates are then determined on the $Z \to ee$ MC sample and on data. For the fake rate measurements on data, the MC samples are used to estimate the amount of background events in the selection. Finally, the fake rates on data and on the MC samples are compared and scale factors are calculated to correct the MC fake rates to the data fake rates.

6.1. Fake Rate

As motivated in Section 2.3.1, the performance of the tau lepton identification algorithm is very important in all analyses involving hadronic tau lepton decays, including measurements involving $H \to \tau\tau$ decays. One of the major backgrounds for this analysis, beyond the irreducible background from $Z \to \tau\tau$ decays, results from the misidentification of hadronic jets (QCD jets) as hadronically decaying tau leptons.

One way to quantify this background is the measurement of the misidentification probability or *fake rate* FR, which is defined as the fraction of QCD jets that pass the tau identification algorithm out of the total number of jets that are reconstructed as $\tau_{\text{had-vis}}$ candidates while passing the selection criteria of the analysis:

$$FR = \frac{\#\text{Jets } (\tau\text{-reco, selection, } \tau\text{-ID})}{\#\text{Jets } (\tau\text{-reco, selection})}. \tag{6.1}$$

6.1.1. Previous Results

A 2011 ATLAS study measured fake rates in three different physics processes (γ + jets, Z^0 + jets and dijet) at a centre-of-mass energy of $\sqrt{s} = 7\,\text{GeV}$ [31]. While the total fake rate was found to differ between the different processes, the study also differentiated between quark and gluon initiated jets. Examples of the Feynman diagrams for the production of quark and gluon jets in $Z \to ee$+jet events are shown in Figure 6.1.

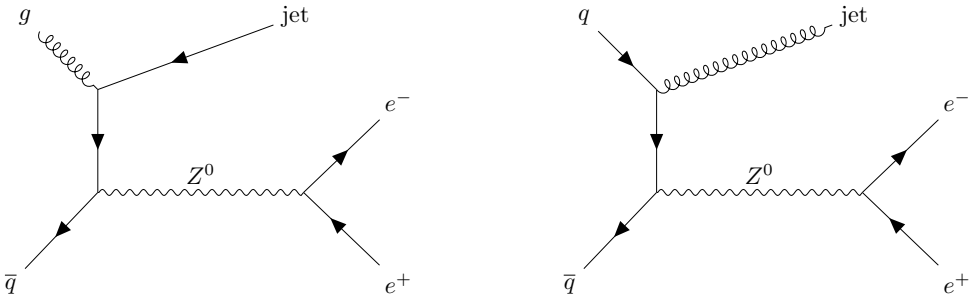


Figure 6.1.: Possible Feynman diagram for a $Z \to ee$ + jet events with a quark or gluon initiated jet.

When the fake rates were calculated separately for $\tau_{\text{had-vis}}$ candidates originating from quark and gluon initiated jets (using MC samples), the fake rates in the different processes were compatible with each other within statistical uncertainties. These fake rates are shown in Figure 6.2 as functions of the $\tau_{\text{had-vis}}$ candidate's transverse momentum.

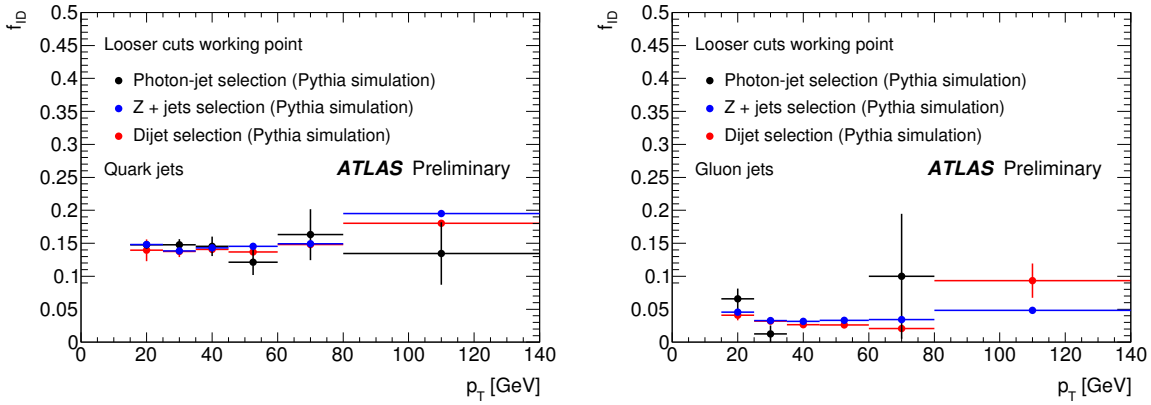


Figure 6.2.: Fake rates for 7 TeV MC simulations of different physics processes calculated separately for quark and gluon initiated jets. [31].

Due to this observation, it is suspected that the incompatible fake rates between different processes can be traced back to the different ratio of quark initiated to gluon initiated jets in the applied selections. Gluons tend to produce wider jets, which are easier to distinguish from hadronically decaying tau leptons than quark jets. Figure 6.3 displays the fraction of quark initiated jets contributing to the $\tau_{\text{had-vis}}$ candidates originating from jets in the different processes as a function of the transverse momentum of the jets. Since quark jets suffer from a higher fake rate than gluon jets, processes with a higher fraction of quark jets should have a higher total fake rate.

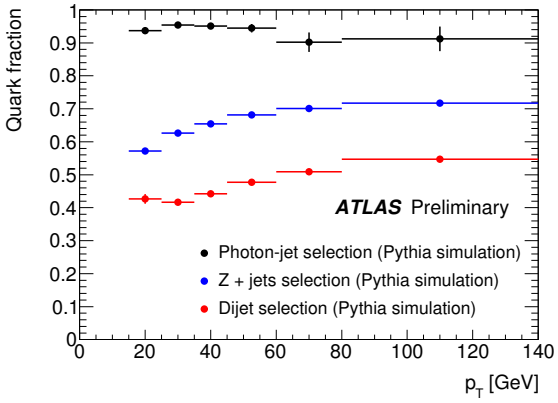


Figure 6.3.: Fraction of fake $\tau_{\text{had-vis}}$ candidates originating from quark initiated jets in 7 TeV MC simulations of different physics processes [31].

6.1.2. Measurement with the Tag and Probe Method

To measure the fake rate as defined in Equation 6.1, it is necessary to select a pure sample of $\tau_{\text{had-vis}}$ candidates produced from jets that can be probed. While a truth matching algorithm (see Section 5.2) could be used for this task, this is not possible on data. The tag and probe method uses a clean channel, which can easily be identified ("tagged") and contains some object that is of interest for the given analysis and can be "probed".

In the case of this thesis, the decay of a Z^0 boson into an electron positron pair is used as the event "tag". Any additional $\tau_{\text{had-vis}}$ candidates in the same event are then very likely to be caused by QCD jets instead of real tau decays and can be used for the measurement of the fake rate (Figure 6.4). The exact selection criteria are given in Section 5.1.

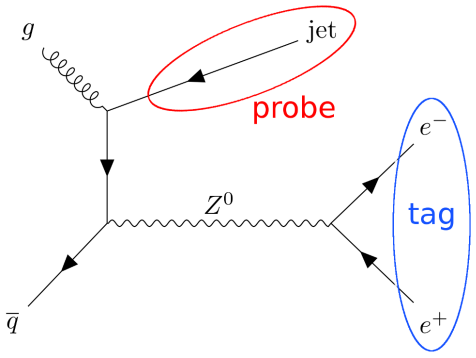


Figure 6.4.: A possible Feynmann diagram for a $Z \to ee$ event with an additional jet. These processes are used for the presented fake rate measurement with the tag and probe method.

6.2. Estimation of Systematic Uncertainties

For each of the following sources of systematic uncertainties, two variations are applied on the MC samples, which are called *up* and *down*. It is important to note, that the naming of a variation as "up" or "down" does not necessarily correspond to its effect on the MC distributions, but is derived from the way in which the variation is created. Therefore, it is possible for both variations to cause deviations of the same sign and the absolute values of the up and down deviations can have different magnitudes.

The sources of systematic uncertainties considered for the fake rate measurement are listed below.

Electron Scale Factor To improve their modelling of the data, multiple scale factors are applied to the MC events (compare Section 5.3.1). Since these scale factors are determined from a data to MC comparison, each of them has an assigned uncertainty, which determines the up and down variation. Systematic uncertainties have been determined for the scale factors associated with the choice of identification requirements on the leading lepton in the event (*MediumLLH*), the modeling of the *track reconstruction*, the *isolation* criterion applied to electrons and the choice of the *trigger* [32].

Electron Energy Scale The tag-and-probe method applied relies on the reconstruction of the invariant mass of the two electrons produced in a $Z \to ee$ decay to separate signal from background events. Therefore, the measurement of the energy and momentum of electrons plays a significant role for the amount of background in the selection. The MC modelling of the interpretation of the detector output for the electrons is reflected in two systematic uncertainties for the *resolution* and the *scale* of the calibration.

Z Mass Window For the tag-and-probe method a $\pm 5\,\text{GeV}$ window around the Z^0 boson mass is defined, in which the invariant mass of the two electrons is required to fall. Since the choice of this width is to some extent arbitrary and could effect the fake rate measurement, a variation of the window's width up to $\pm 8\,\text{GeV}$ and down to $\pm 4\,\text{GeV}$ is considered as an additional source of uncertainty.

Tau Energy Scale The reconstruction of the $\tau_{\text{had-vis}}$ candidate in the event uses the *tau energy scale* (TES), which is derived from simulated events [33]. The calibration of the TES introduces systematic uncertainties on the measured properties of the $\tau_{\text{had-vis}}$ candidate arising from the simulation of the *detector*, the choice of the used MC *model* and the comparison to *in-situ* measurements.

The direct effect of the different systematic variations on the denominator in the fake rate calculation are shown in Figures A.1 and A.2 in the appendix.

The effects of each systematic uncertainty on the fake rate measurement have been estimated by computing the binned fake rate separately on the nominal MC sample and on the up and down variation. For both variations the difference to the nominal value have been calculated. To obtain a symmetric estimation of the uncertainty, the arithmetic mean of the two variations is used as the systematic uncertainty on the fake rate.

Table 6.1 lists the obtained uncertainties on the fake rates. While the statistical uncertainties on the fake rates dominate, the choice of the Z mass window width and the electron energy measurements also result in non-negligible systematic uncertainties. These systematic variations are the only ones that directly impact the applied tag and probe selection by varying either the width of the Z mass window or the energy measurement of the electrons and therefore their invariant mass. This results in the selection of a slightly different set of events, which affects the fake rate in the selection.

The variations of the electron scale factors and the tau energy scale, on the other hand, mainly result in a bin-to-bin migration of some events in $p_T(\tau_{\text{had-vis}})$ binned distributions. These migrations cancel each other out almost completely when the fake rate is calculated.

	1 Prong			3 Prong		
Source of Uncertainty	Loose	Medium	Tight	Loose	Medium	Tight
Statistics	8.9 %	14 %	22 %	18 %	30 %	41 %
Z Mass Window	4.1 %	5.9 %	5.7 %	5.1 %	3.8 %	11 %
Electron Energy Resolution	1.7 %	6.8 %	9.4 %	3.5 %	4.5 %	9.8 %
Electron Energy Scale	2.4 %	6.3 %	9.4 %	3.4 %	4.7 %	8.8 %
Electron SF MediumLLH	0.05 %	0.03 %	0.03 %	0.04 %	0.07 %	0.09 %
Electron SF Trigger	0.04 %	0.02 %	0.03 %	0.08 %	0.10 %	0.08 %
Electron SF Reco Track	0.02 %	0.02 %	0.02 %	0.03 %	0.03 %	0.06 %
Electron SF Isolation	0.01 %	0.005 %	0.008 %	0.03 %	0.04 %	0.03 %
Tau Energy Scale In-situ	0.05 %	0.08 %	0.007 %	0.30 %	0.01 %	0.01 %
Tau Energy Scale Model	0.04 %	0.08 %	0.007 %	0.15 %	0.01 %	0.01 %
Tau Energy Scale Detector	0.0 %	0.0 %	0.0 %	0.01 %	0.01 %	0.01 %

Table 6.1.: Size of the effect of the considered uncertainties on the fake rates for the three different working points of the tau identification algorithm. Listed is always the uncertainty of the $p_T(\tau_{\text{had-vis}})$ bin with the largest absolute uncertainty.

6.3. Fake Rates in MC

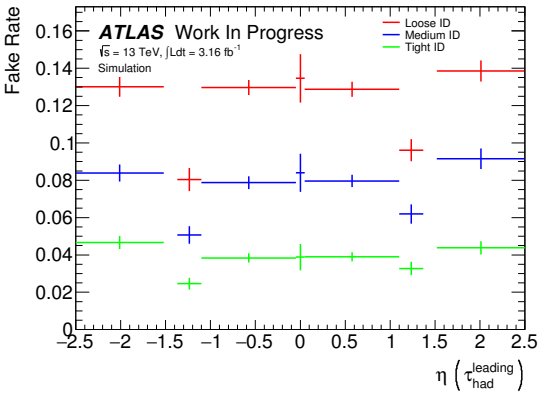
Figure 6.5 presents fake rates obtained by applying the tag and probe method described in Section 6.1.2 to the $Z \to ee$ MC sample, where the leading $\tau_{\text{had-vis}}$ candidate in each event is probed. Since the tau identification algorithm uses two independent BDTs for the 1 prong and 3 prong decays, the fake rates are measured for $\tau_{\text{had-vis}}$ candidates with one and three associated tracks separately. Candidates with other numbers of tracks are not taken into account, as they are usually not considered as $\tau_{\text{had-vis}}$ candidates for most analyses. For each number of tracks, fake rates have been computed for the three different working points of the identification algorithm (Figure 6.5).

The $\eta(\tau_{\text{had-vis}})$ dependence of the fake rate is shown in Figures 6.5(a) and 6.5(b). Both distributions are flat within statistical uncertainties, with the exception of the bins close to the crack region in Figure 6.5(a). This deviation is likely caused by a contribution of jets with more than one track, for which only one track does not lie within the crack region. Those jets would be reconstructed as having one associated track, but exhibit the lower fake rate of jets with multiple tracks (compare Figure 6.5(b)).

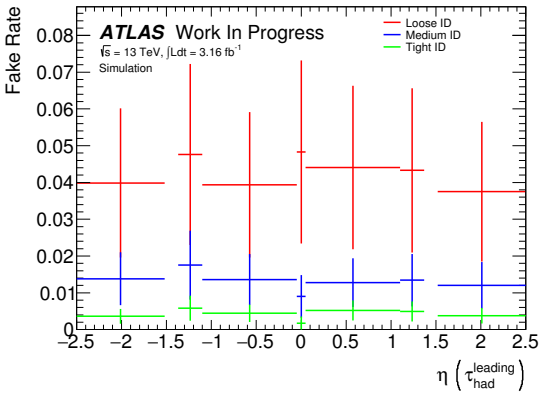
Due to the rather flat distribution, these plots are especially well suited to illustrate the dependence of the fake rates on the efficiencies, which the BDT thresholds are tuned to achieve (Table 4.3). Higher efficiency requirements result in looser thresholds, which in turn allows for more background to pass the identification, thus also raising the fake rate.

Figures 6.5(c) and 6.5(d) present the fake rate as a function of the $p_T(\tau_{\text{had-vis}})$. The p_T dependence of the fake rates is mainly governed by the tuning of the different tau identification working points, which use p_T dependent thresholds on the BDT score as identification criteria. These thresholds have been tuned in order to flatten the identification efficiency for real tau leptons as a function of p_T. A side effect of this tuning is the dependence of the fake rates as a function of p_T.

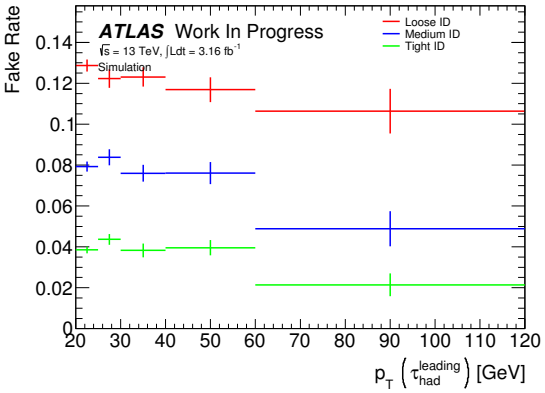
Figures 6.5(e) and 6.5(f) show the fake rate dependence on the average number of interactions per bunch crossing, related to the amount of pileup in the event. Both distributions appear to be flat within the statistical uncertainties, indicating that the additional overlaid tracks from pileup do not influence the reconstruction efficiency in a significant way.



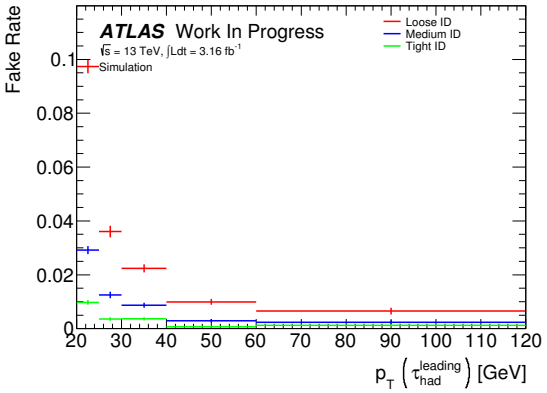
(a) η dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



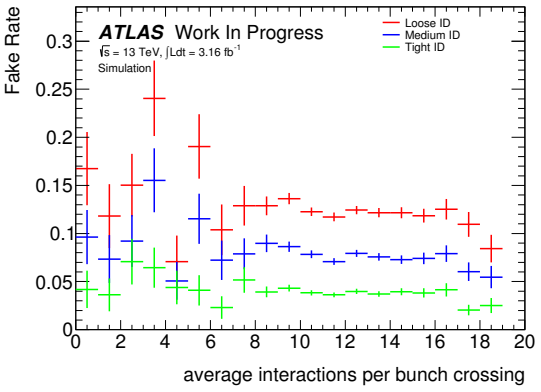
(b) η dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.



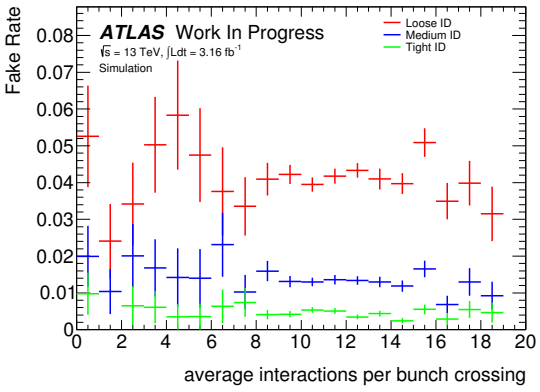
(c) p_T dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



(d) p_T dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.



(e) Pileup dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



(f) Pileup dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.

Figure 6.5.: Fake rates determined from a $Z \to ee$ MC sample using the tag and probe method.

6.4. Fake Rates in Data

For the measurement of fake rates on data, the ATLAS data set collected in 2015, which corresponds to $3.2\,\mathrm{fb}^{-1}$, is used. The selection discussed in Section 5.1 is applied to the data events before they are used for further analysis.

The MC samples listed in Table 5.1 are used to estimate the amount of background events in the selection. As can be seen in Table 5.3 and Figures 5.3 and 5.4, the main background contribution originates from $t\bar{t}$ events. The number of MC background events passing the selection is subtracted from the amount of data to estimate the number of signal events in the data. This process is applied to both the nominator and denominator of the fake rate (Equation 6.1).

Figure 6.6 displays the fake rates obtained this way. Overall, they are distributed very similarly to the MC fake rates discussed in Section 6.3.

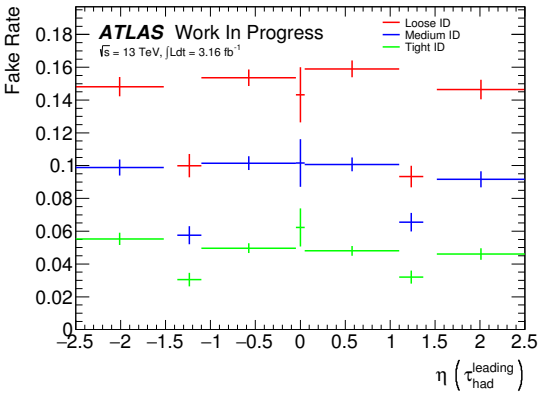
6.5. Scale Factors for Fake Rates

For an easier comparison of the fake rate measured on MC to the data fake rates, the *scale factor* s between them is calculated according to Equation 6.2. The corresponding estimation of the uncertainty is derived in Section A.3.2.

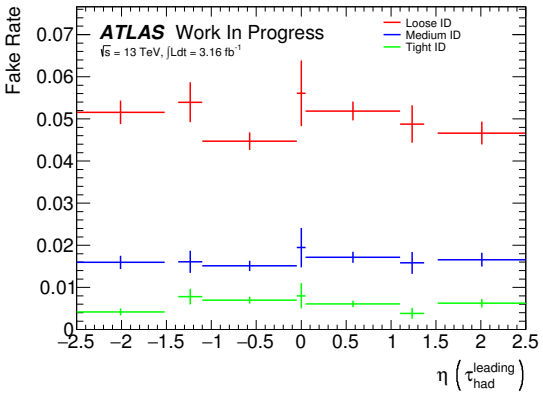
$$s = \frac{FR^{Data}}{FR^{MC}} \tag{6.2}$$

Figure 6.7 displays the scale factors calculated from the distributions in Figures 6.5 and 6.6. While the scale factors are compatible with 1 within one or two σ in almost all bins of the distributions, the $\eta(\tau_{\text{had-vis}})$ (Figures 6.7(a) and 6.7(b)) and the pileup distributions (Figures 6.7(e) and 6.7(f)) show that scale factors above 1 dominate in the case of $\tau_{\text{had-vis}}$ candidates with one associated track. In addition Figures 6.7(c) and 6.7(d) exhibit high scale factors at low $p_T(\tau_{\text{had-vis}})$. However, overall the data and MC fake rates are compatible within the given uncertainties.

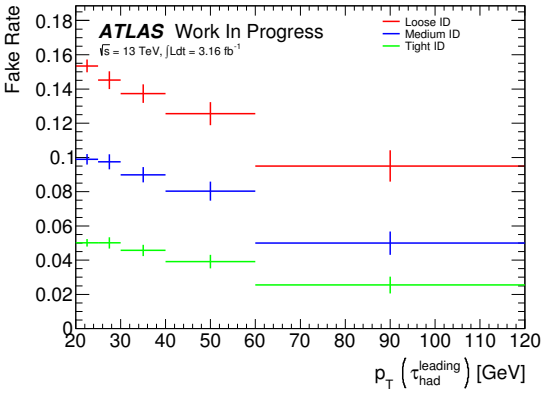
The slight tendency towards scale factors above 1 would mean that the fake rate in data tends towards higher values than the MC fake rate. Under the hypothesis that fake rates can be interpreted as mixed from pure quark and gluon fake rates, this would hint at a higher quark fraction in data than simulated in MC, since quark jets suffer from higher fake rates than gluon jets. However, this assumes a correct modelling of the pure quark and gluon fake rates in the simulation, which is investigated in the next chapter.



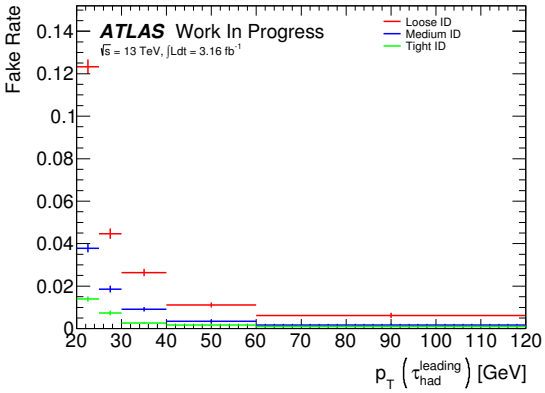
(a) η dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



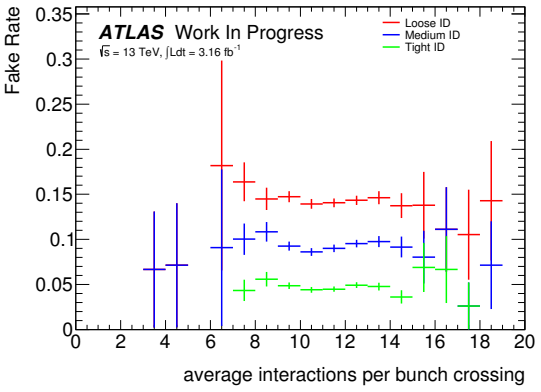
(b) η dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.



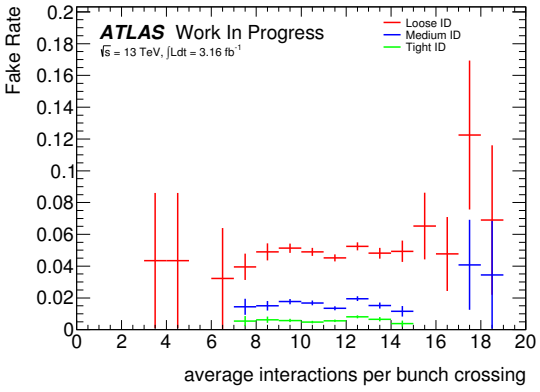
(c) p_T dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



(d) p_T dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.

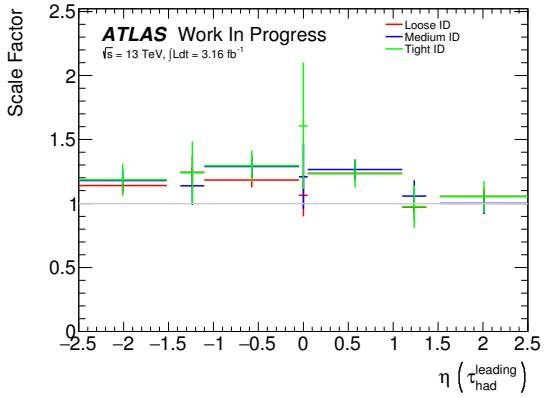


(e) Pileup dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.

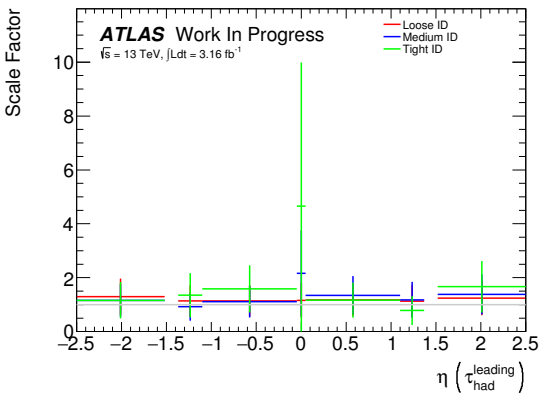


(f) Pileup dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.

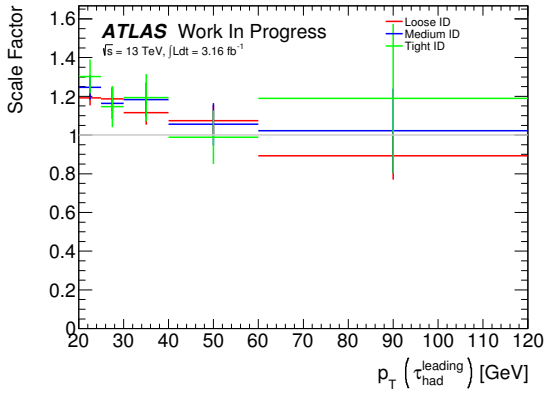
Figure 6.6.: Fake rates determined from 2015 ATLAS data using the tag and probe method.



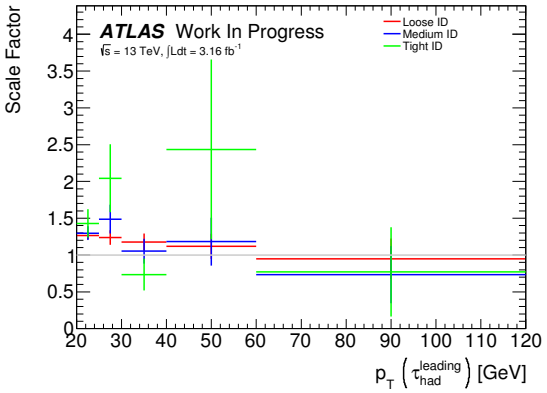
(a) η dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



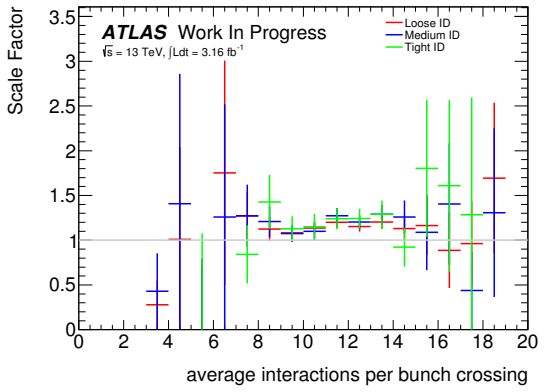
(b) η dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.



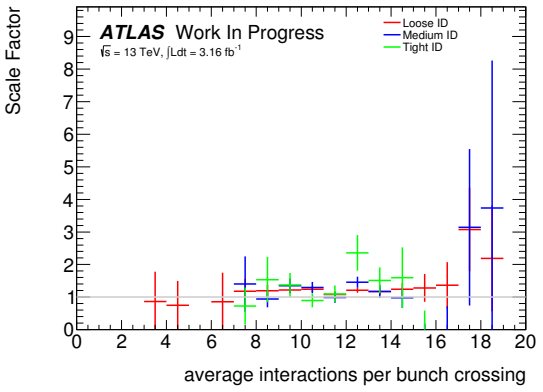
(c) p_T dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



(d) p_T dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.



(e) Pileup dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with one charged track.



(f) Pileup dependence of the fake rate for $\tau_{\text{had-vis}}$ candidates with three charged tracks.

Figure 6.7.: Scale factors determined from a $Z \to ee$ MC sample using the tag and probe method.

7. Extraction of Quark Jet and Gluon Jet Fake Rates

As discussed in Section 6.1.1, a previous ATLAS study suggests that the differences in the fake rate distributions in different processes is mainly caused by the different ratio of quark to gluon initiated jets [31]. Consequently, it should be possible to extract "pure" fake rate distributions FR_q and FR_g of quark and gluon initiated jets from two fake rate measurements FR_i, in regions i=1,2 with different, known quark and gluon fractions q_i and g_i, by solving the linear system $FR_i = q_i \cdot FR_q + g_i \cdot FR_g$.

In this chapter, the template fit method is used to measure the quark and gluon fraction in data. The $Z \to ee$ MC sample is used to define two regions with different quark/gluon ratios using truth matching information. The fake rates and quark fractions in these two regions are measured in data. From the measurements, an extraction of "pure" quark and gluon fake rates is attempted and compared to predictions from truth tagged MC events.

7.1. Template Fit

To measure the relative contributions of quark and gluon initiated jets to the $\tau_{\text{had-vis}}$ candidates, which are selected by the tag and probe method (Section 6.1.2), the template fit method is used.

This method utilises a variable that exhibits different distributions for the two classes of jets. *Templates* of the distributions of the two classes are obtained from the leading $\tau_{\text{had-vis}}$ candidate in the $Z \to ee$ MC sample by requiring a truth match to either quarks or gluons. The distribution of the data with the same selection criteria is then fitted using these templates.

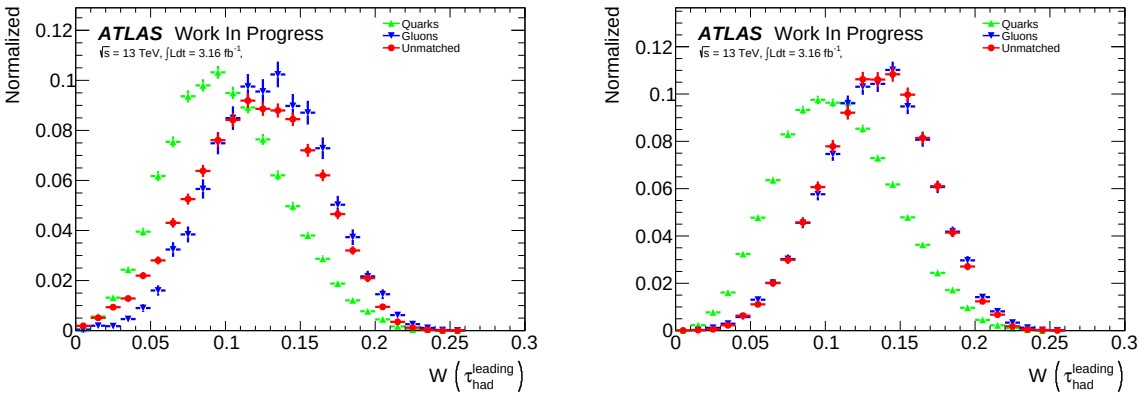
To perform the fit, the ROOT [20] function `TFractionFitter` is used, which implements the fitting method described in [34]. This fitting method takes into account the finite statistics in the MC sample used for the templates.

7.1.1. Template Fit Variable

For the template fit method, a variable is needed that possesses different distributions for the two classes of jets. Since gluons tend to produce wider jets than quarks, the width w of the $\tau_{\text{had-vis}}$ candidate has been chosen. This variable is defined as the weighted average ΔR of all objects within the jet, where the weights are given by the transverse momenta p_T of the objects:

$$w = \frac{\sum_i \Delta R^i p_T^i}{\sum_i p_T^i} \tag{7.1}$$

Figure 7.1 displays the w distribution of for $\tau_{\text{had-vis}}$ candidates in the $Z \to ee$ MC sample, which are truth matched to quarks and gluons. For some $\tau_{\text{had-vis}}$ candidates no truth match could be found. The "unmatched" candidates will be considered as a systematic uncertainty on the template fit (see Section 7.1.3) While the width and shape of the two distributions is similar, there exists a clear difference in their mean values. As expected, the distribution of the gluon jets is shifted to higher values of w when compared to the quark jets. This behaviour is visible in both the 1 prong and 3 prong cases (Figure 7.1).



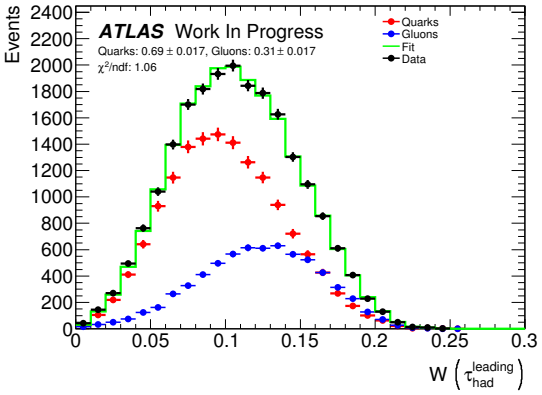
(a) $\tau_{\text{had-vis}}$ candidates with one associated reconstructed track.

(b) $\tau_{\text{had-vis}}$ candidates with three associated reconstructed tracks.

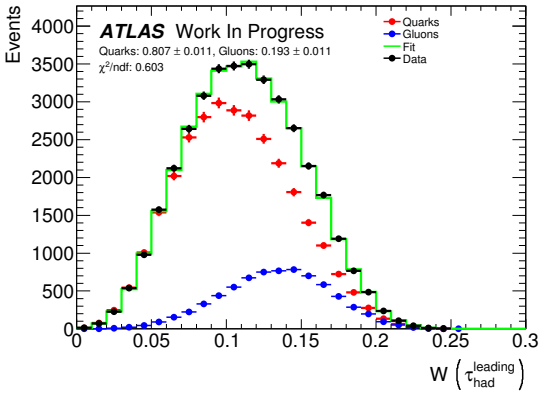
Figure 7.1.: Width w of $\tau_{\text{had-vis}}$ candidates in the $Z \to ee$ MC sample separated by the truth match. The "unmatched" candidates will be considered as a systematic uncertainty on the template fit (see Section 7.1.3).

The result of a fit of the quark and gluon templates for the w distributions obtained from the truth matched $Z \to ee$ MC sample to the data is shown in Figure 7.2. The template distributions shown in the fit results are *post-fit*, i.e. they include statistical

Poisson fluctuations performed by the `TFractionFitter` and the distributions are scaled in such a way that the relative integrals of the histograms correspond to the results of the fit.



(a) $\tau_{\text{had-vis}}$ candidates with one associated reconstructed track.



(b) $\tau_{\text{had-vis}}$ candidates with three associated reconstructed track.

Figure 7.2.: Fit of the quark and gluon templates to data. The templates are obtained from the $Z \to ee$ MC sample using truth matching.

7.1.2. Corrections on Fit Uncertainties

To validate the error estimation on the quark fraction q that is given by the `TFractionFitter` function, the *pull* of q has been calculated. The pull is defined as follows:

$$\text{Pull}(q_i) = \frac{q_i - \bar{q}}{\sigma_{q_i}}, \tag{7.2}$$

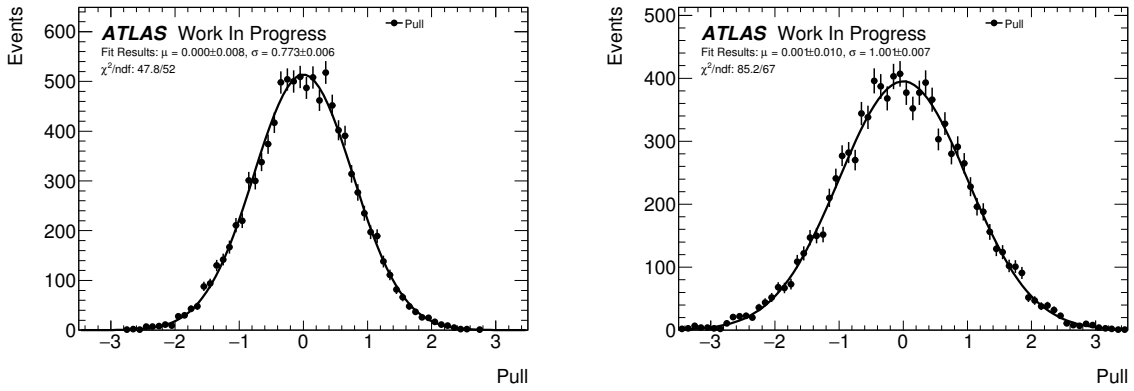
where $\bar{q}$ is the average quark fraction over an ensemble of template fits in 10 000 toy experiments. The symbols q_i and σ_{q_i} denote the quark fraction and its uncertainty in a specific experiment out of the 10 000. For the toy experiments, templates have been extracted from the $Z \to ee$ MC sample as usual, but the fit is performed to a data distribution that is randomly fluctuated in each bin according to the Poisson distribution.

For a correctly estimated error σ_{q_i}, the pull distribution over all toy experiments should yield a Gaussian distribution with a mean value μ of 0 and a standard deviation σ of 1. The obtained pull distribution for $\tau_{\text{had-vis}}$ candidates with one associated track is shown in Figure 7.3(a) together with a fitted Gaussian. While the shape and mean value matches the expectations, the standard deviation is significantly lower than 1*. This corresponds

* While the exact value of the fitted σ varies between different selections, it is consistently below 1.

to an *overestimation* of σ_{q_i} by the `TFractionFitter`.

To correct this behaviour, a pull distribution is calculated for every performed template fit and the σ obtained from a Gaussian fit to the pull distribution is multiplied onto the σ_q given by the `TFractionFitter`. Figure 7.3(b) displays a recalculation of the pull plot shown in Figure 7.3(a) after this correction has been applied. The standard deviation of the new pull distribution is in perfect agreement with the expected value of 1.



(a) Before correction.

(b) After correction.

Figure 7.3.: Pull of the quark fraction given by the `TFractionFitter` for $\tau_{\text{had-vis}}$ candidates with one associated track before and after the error estimation has been corrected.

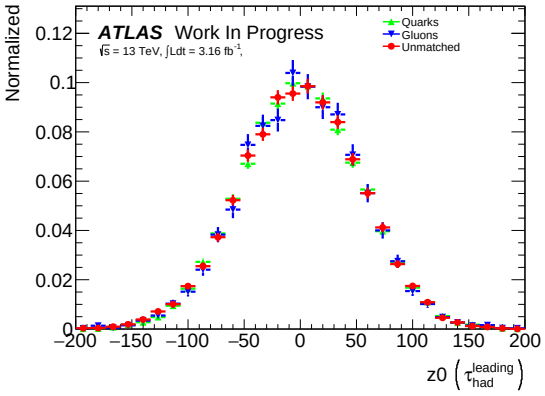
7.1.3. Systematic Uncertainty from Unmatched

In addition to the truth matches to quarks (56 %) and gluons (15 %), 30 % of the $\tau_{\text{had-vis}}$ candidates in the $Z \to ee$ MC sample could not be matched to a truth particle. Thus, the truth record for these events contains no truth particles within the $\Delta R < 0.4$ search cone of the truth matching algorithm. A widening of the search cone to $\Delta R < 0.6$ could not resolve this issue.

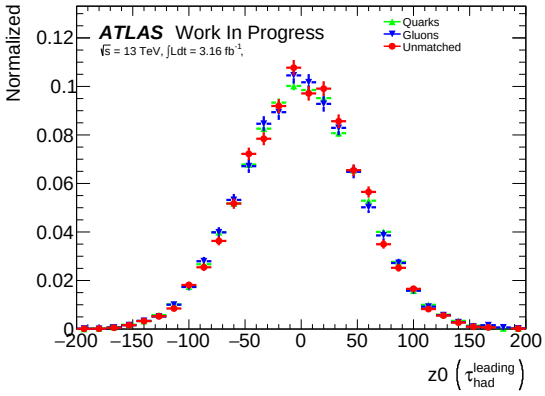
A plausible explanation for the presence of these unmatched $\tau_{\text{had-vis}}$ candidates in the MC sample is that they originate from pileup radiation, since pileup is not stored in the truth record. $\tau_{\text{had-vis}}$ candidates from pileup would be expected to have a wider z_0[†] distribution than candidates originating from the primary vertex of the event. However, as Figure 7.4 shows, the z_0 distribution of the unmatched $\tau_{\text{had-vis}}$ candidates does not differ significantly from the distribution of candidates matched to quarks or gluons. It is therefore unlikely, that the unmatched candidates are originating from pileup.

[†] z_0 is defined as the longitudinal distance of the reconstructed origin of the jet from the primary vertex of the event.

Further investigation of the origin of these unmatched candidates is outside the scope of this thesis.



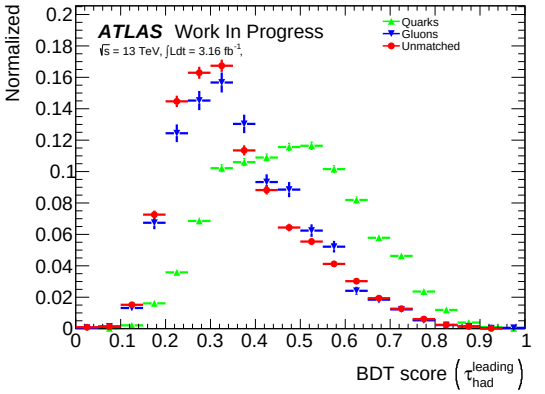
(a) $\tau_{\text{had-vis}}$ candidates with one associated reconstructed track.



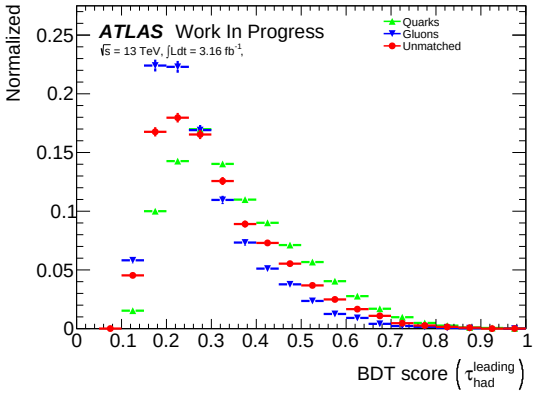
(b) $\tau_{\text{had-vis}}$ candidates with three associated reconstructed track.

Figure 7.4.: z_0 (in mm) distribution of the leading $\tau_{\text{had-vis}}$ candidates in the $Z \to ee$ MC sample separated by truth match.

In a comparison of the jet width distribution of the different truth matches (Figure 7.1), the unmatched $\tau_{\text{had-vis}}$ candidates display a very similar behaviour to gluon jets. The same holds for most of the input variables of the tau identification BDT. Figure 7.5 shows the distributions of the BDT score for the different matches.



(a) $\tau_{\text{had-vis}}$ candidates with one associated reconstructed track.



(b) $\tau_{\text{had-vis}}$ candidates with three associated reconstructed track.

Figure 7.5.: Distribution of the tau identification BDT score of the leading $\tau_{\text{had-vis}}$ candidates in the $Z \to ee$ MC sample separated by truth match.

Based on these observations, the unmatched $\tau_{\text{had-vis}}$ candidates are grouped together with the gluon matched candidates to form the gluon jet w template for the fit. An

additional fit is performed using only the gluon matched candidates as the gluon template and the difference between the two fit results is added as a systematic uncertainty on the measured quark fraction.

7.2. Definition of Enriched Regions

For the extraction of quark and gluon fake rates from the measured fake rates, two regions with different quark fractions need to be defined. Figure 7.6 shows the distribution of truth matches to the leading $\tau_{\text{had-vis}}$ candidate on the $Z \to ee$ MC sample as a function of $p_T(\ell\ell)$. When the unmatched candidates are counted as gluons, as discussed in Section 7.1.3, the distribution for quarks dominates at higher values than the gluon distribution. This allows for the definition of the two regions by splitting the $p_T(\ell\ell)$ spectrum in a low and a high p_T region. The high p_T region is expected to have a larger quark fraction than the low p_T region.

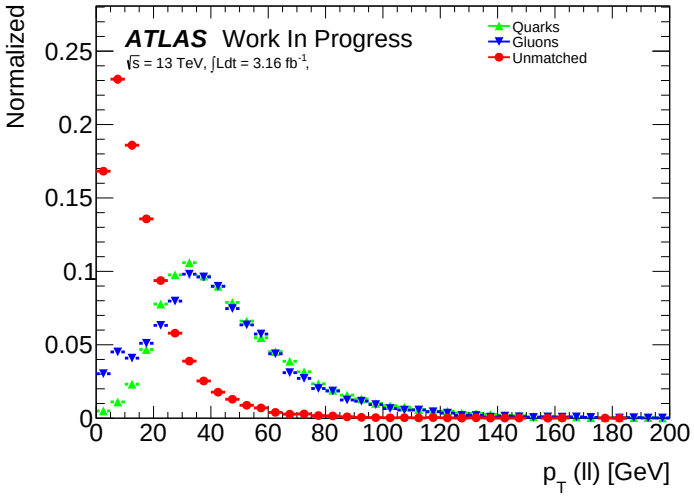


Figure 7.6.: $p_T(\ell\ell)$ dependent distribution of $\tau_{\text{had-vis}}$ candidates in the $Z \to ee$ MC sample that are matched to quarks or gluons or are unmatched.

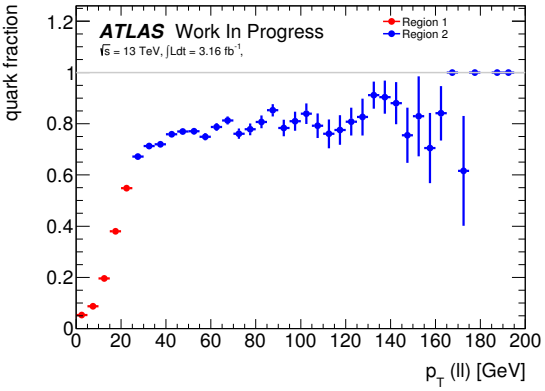
Table 7.1 lists the quark fractions in the two regions for different choices of $p_T(\ell\ell)$ cuts for the region definition. While lower cuts yield regions with a higher difference in the quark fraction, low cuts also strongly reduce the number of events in the low p_T region. As a compromise between a good separation and statistics, a cut of $p_T(\ell\ell) = 25\,\text{GeV}$ has been chosen to separate the two regions.

Figures 7.7(a) and 7.7(c) displays the quark fraction as a function of $p_T(\ell\ell)$. The chosen cut of $p_T(\ell\ell) = 25\,\text{GeV}$ separates the sample where the quark fraction suddenly drops off. At higher $p_T(\ell\ell)$ it is roughly constant. The $p_T(\tau_{\text{had-vis}})$ binned quark fractions for the

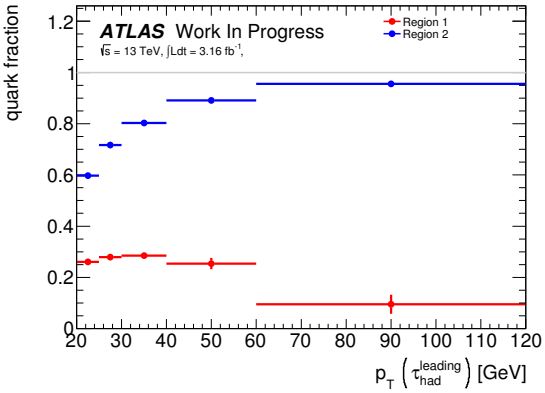
Cut	Region 1 (Low p_T)		Region 2 (High p_T)		
	q_1	g_1	q_2	g_2	Δq
20 GeV	0.21	0.79	0.74	0.26	0.53
25 GeV	0.30	0.70	0.75	0.25	0.45
30 GeV	0.38	0.62	0.76	0.24	0.38
35 GeV	0.44	0.56	0.77	0.23	0.33
40 GeV	0.48	0.52	0.78	0.22	0.29

Table 7.1.: Quark fraction as calculated from truth matched MC events for different region definitions.

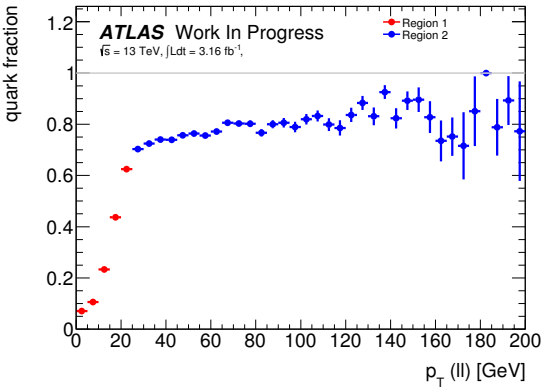
two regions can be seen in Figures 7.7(b) and 7.7(d). In all bins, the quark fractions of the two regions are well separated by a difference of at least 0.3.



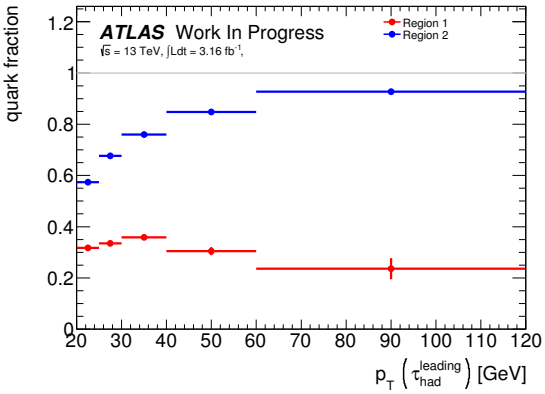
(a) $\tau_{\text{had-vis}}$ candidates with one associated track as a function of $p_T(\ell\ell)$.



(b) $\tau_{\text{had-vis}}$ candidates with one associated track as a function of $p_T(\tau_{\text{had-vis}})$.



(c) $\tau_{\text{had-vis}}$ candidates with three associated tracks as a function of $p_T(\ell\ell)$.



(d) $\tau_{\text{had-vis}}$ candidates with three associated tracks as a function of $p_T(\tau_{\text{had-vis}})$.

Figure 7.7.: Quark fraction in the regions below and above $p_T(\ell\ell) = 25\,\text{GeV}$ as calculated from truth matched MC events.

7.3. Template Fit Results

Figure 7.8 displays the fitted w distributions for $\tau_{\text{had-vis}}$ candidates with one or three associated tracks in the full $p_T(\ell\ell)$ spectrum as well as in the two defined regions. The quark fractions obtained from the fits are also listed in Tables 7.2, 7.3 and 7.4 in comparison with quark fractions calculated from the truth matched $Z \to ee$ MC sample.

While the systematic uncertainties clearly dominate for the quark fractions obtained via truth matching, the fractions from template fits are dominated by the statistical uncertainties of the fit, as soon as the sample is divided into the two regions. In the $p_T(\tau_{\text{had-vis}})$-binned fit result many quark fractions are given uncertainties that extend into unphysical regions above 1 or below 0. Five of the fitted quark fractions in the low $p_T(\ell\ell)$ region are exactly 1 (Table 7.3), all of which are in the high $p_T(\tau_{\text{had-vis}})$ bins. This is likely caused by a too low number of events in these bins where the quark fraction is expected to be very close to one.

Within the large uncertainties, the quark fraction from truth matching are compatible with the results obtained from the template fit method.

As expected, the fractions in the low $p_T(\ell\ell)$ region are smaller than the ones in the high $p_T(\ell\ell)$ region. $\tau_{\text{had-vis}}$ candidates with different numbers of associated tracks also exhibit different quark fractions.

#Tracks	$p_T(\tau_{\text{had-vis}})$ Region	Fitted Quark Fraction	Truth Matching
1	20 - 25 GeV	$0.64 \pm 0.039 \pm 0.15$	$0.40 \pm 0.00005 \pm 0.32$
1	25 - 30 GeV	$0.67 \pm 0.047 \pm 0.17$	$0.53 \pm 0.00011 \pm 0.26$
1	30 - 40 GeV	$0.822 \pm 0.037 \pm 0.073$	$0.67 \pm 0.00012 \pm 0.18$
1	40 - 60 GeV	$0.871 \pm 0.027 \pm 0.018$	$0.816 \pm 0.00017 \pm 0.095$
1	60 - 120 GeV	$0.893 \pm 0.017 \pm 0.024$	$0.916 \pm 0.00033 \pm 0.051$
3	20 - 25 GeV	$0.906 \pm 0.038 \pm 0.012$	$0.445 \pm 0.00005 \pm 0.219$
3	25 - 30 GeV	$0.902 \pm 0.037 \pm 0.022$	$0.551 \pm 0.00007 \pm 0.175$
3	30 - 40 GeV	$0.854 \pm 0.027 \pm 0.026$	$0.670 \pm 0.00005 \pm 0.118$
3	40 - 60 GeV	$0.929 \pm 0.018 \pm 0.008$	$0.802 \pm 0.00006 \pm 0.057$
3	60 - 120 GeV	$0.913 \pm 0.012 \pm 0.026$	$0.914 \pm 0.00007 \pm 0.016$

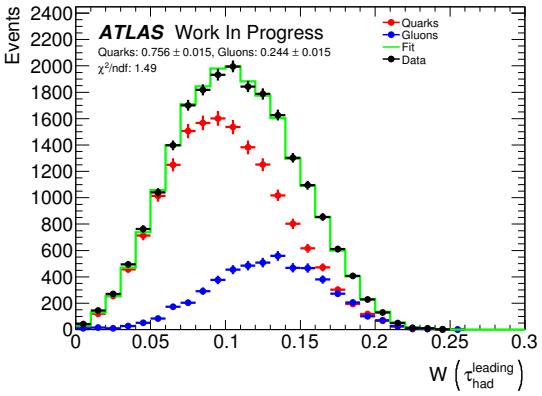
Table 7.2.: Quark fraction $q \pm$stat.$\pm$syst. in the full $p_T(\ell\ell)$ region as determined by a template fit on data or from truth matching on the $Z \to ee$ MC sample. The given systematic uncertainties origin from the unmatched $\tau_{\text{had-vis}}$ candidates (Section 7.1.3).

#Tracks	$p_T(\tau_{\text{had-vis}})$ Region	Fitted Quark Fraction	Truth Matching
1	20 - 25 GeV	$0.705 \pm 0.059 \pm 0.164$	$0.26 \pm 0.00008 \pm 0.46$
1	25 - 30 GeV	$0.523 \pm 0.061 \pm 0.108$	$0.28 \pm 0.00023 \pm 0.50$
1	30 - 40 GeV	$1.000 \pm 0.035 \pm 0.000$	$0.29 \pm 0.00043 \pm 0.48$
1	40 - 60 GeV	$1.000 \pm 0.059 \pm 1.000$	$0.25 \pm 0.0016 \pm 0.57$
1	60 - 120 GeV	$1.000 \pm 0.011 \pm 0.000$	$0.10 \pm 0.0075 \pm 0.53$
3	20 - 25 GeV	$0.801 \pm 0.040 \pm 0.016$	$0.32 \pm 0.00009 \pm 0.38$
3	25 - 30 GeV	$0.642 \pm 0.044 \pm 0.026$	$0.34 \pm 0.00017 \pm 0.39$
3	30 - 40 GeV	$0.721 \pm 0.052 \pm 0.056$	$0.36 \pm 0.00025 \pm 0.39$
3	40 - 60 GeV	$1.000 \pm 0.055 \pm 1.000$	$0.31 \pm 0.00078 \pm 0.43$
3	60 - 120 GeV	$1.000 \pm 0.022 \pm 0.580$	$0.24 \pm 0.0059 \pm 0.47$

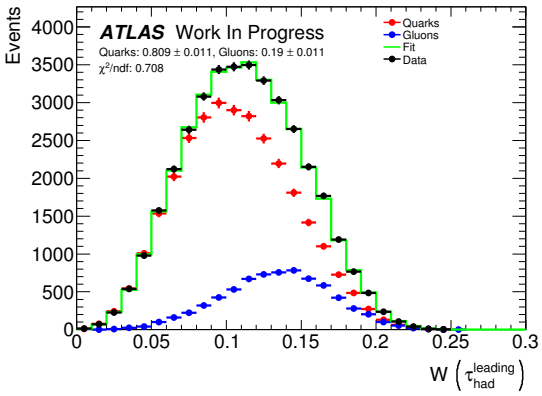
Table 7.3.: Quark fraction q±stat.±syst. in the $p_T(\ell\ell) < 25\,\text{GeV}$ region as determined by a template fit on data or from truth matching on the $Z \to ee$ MC sample. The given systematic uncertainties origin from the unmatched $\tau_{\text{had-vis}}$ candidates (Section 7.1.3).

#Tracks	$p_T(\tau_{\text{had-vis}})$ Region	Fitted Quark Fraction	Truth Matching
1	20 - 25 GeV	$0.634 \pm 0.041 \pm 0.076$	$0.60 \pm 0.00012 \pm 0.13$
1	25 - 30 GeV	$0.748 \pm 0.064 \pm 0.103$	$0.717 \pm 0.00017 \pm 0.083$
1	30 - 40 GeV	$0.874 \pm 0.036 \pm 0.026$	$0.802 \pm 0.00013 \pm 0.056$
1	40 - 60 GeV	$0.834 \pm 0.022 \pm 0.013$	$0.891 \pm 0.00016 \pm 0.024$
1	60 - 120 GeV	$0.914 \pm 0.015 \pm 0.003$	$0.955 \pm 0.00025 \pm 0.014$
3	20 - 25 GeV	$0.928 \pm 0.029 \pm 0.218$	$0.574 \pm 0.00010 \pm 0.073$
3	25 - 30 GeV	$0.991 \pm 0.052 \pm 0.140$	$0.677 \pm 0.00010 \pm 0.049$
3	30 - 40 GeV	$0.907 \pm 0.026 \pm 0.007$	$0.759 \pm 0.00006 \pm 0.034$
3	40 - 60 GeV	$0.944 \pm 0.018 \pm 0.097$	$0.848 \pm 0.00006 \pm 0.016$
3	60 - 120 GeV	$0.877 \pm 0.017 \pm 0.040$	$0.928 \pm 0.00007 \pm 0.005$

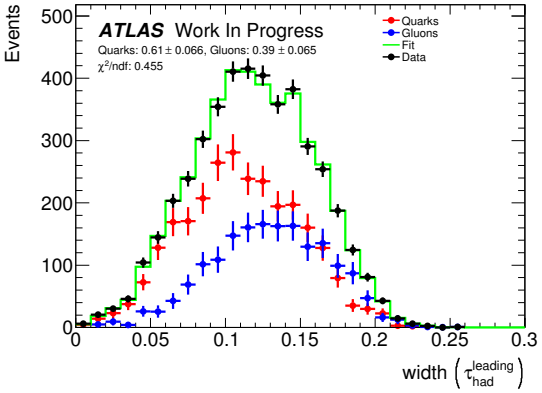
Table 7.4.: Quark fraction q±stat.±syst. in the $p_T(\ell\ell) > 25\,\text{GeV}$ region as determined by a template fit on data or from truth matching on the $Z \to ee$ MC sample. The given systematic uncertainties origin from the unmatched $\tau_{\text{had-vis}}$ candidates (Section 7.1.3).



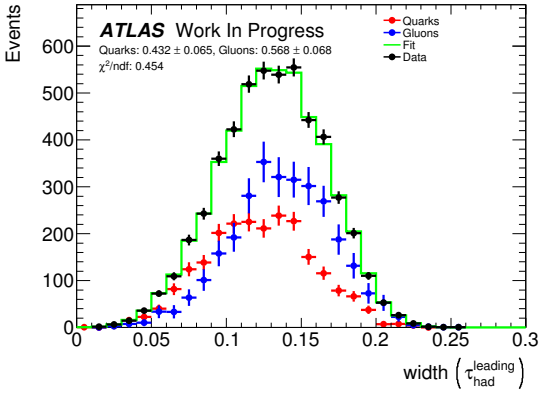
(a) Entire $p_T(\ell\ell)$ region, $\tau_{\text{had-vis}}$ candidates with one associated reconstructed track.



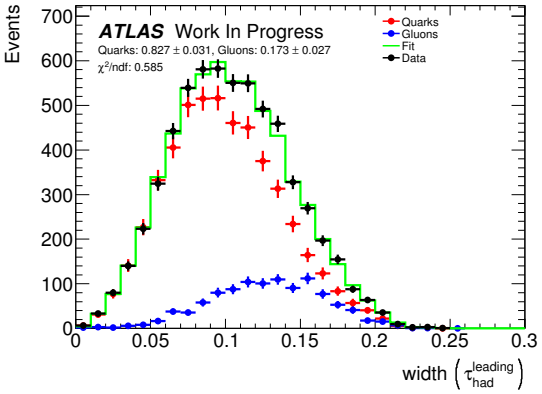
(b) Entire $p_T(\ell\ell)$ region, $\tau_{\text{had-vis}}$ candidates with three associated reconstructed track.



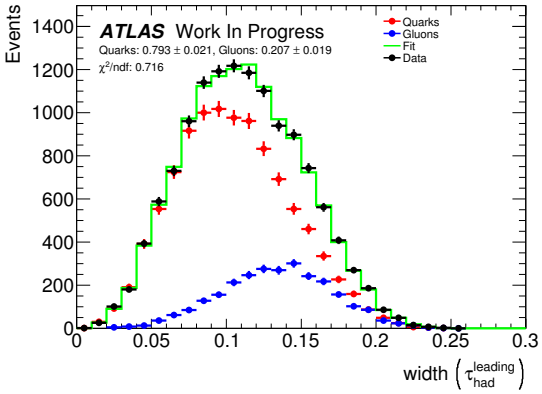
(c) Low $p_T(\ell\ell)$ region, $\tau_{\text{had-vis}}$ candidates with one associated reconstructed track.



(d) Low $p_T(\ell\ell)$ region, $\tau_{\text{had-vis}}$ candidates with three associated reconstructed track.



(e) High $p_T(\ell\ell)$ region, $\tau_{\text{had-vis}}$ candidates with one associated reconstructed track.



(f) High $p_T(\ell\ell)$ region, $\tau_{\text{had-vis}}$ candidates with three associated reconstructed track.

Figure 7.8.: Fit of the quark and gluon templates to data in different regions of $p_T(\ell\ell)$. The templates are obtained from the $Z \to ee$ MC sample using truth matching.

7.4. Quark and Gluon Fake Rate Extraction

As discussed in Section 6.1.1, the fake rate of a selection i with fractions q_i of quark initiated and g_i of gluon initiated jets can be assumed to be given by:

$$FR_i = q_i \cdot FR_q + g_i \cdot FR_g, \tag{7.3}$$

where FR_q and FR_g are the fake rates on pure samples of quark or gluon initiated jets. Since all jets originate either from a quark or a gluon, the sum of the quark and gluon fractions is required to be one ($q_i + g_i = 1$). This allows for the removal of the gluon fraction g_i from Equation 7.3:

$$FR_i = q_i \cdot FR_q + (1 - q_i) \cdot FR_g, \tag{7.4}$$

If Equation 7.4 holds true, the measurement of two fake rates FR_1 and FR_2 in regions with different known quark fractions q_1 and q_2, allows the calculation of the pure fake rates FR_q and FR_g:

$$FR_q = \frac{(1-q_2)\cdot FR_1 - (1-q_1)\cdot FR_2}{q_1 - q_2} \quad \text{and} \quad FR_g = \frac{q_2 \cdot FR_1 - q_1 \cdot FR_2}{q_2 - q_1} \tag{7.5}$$

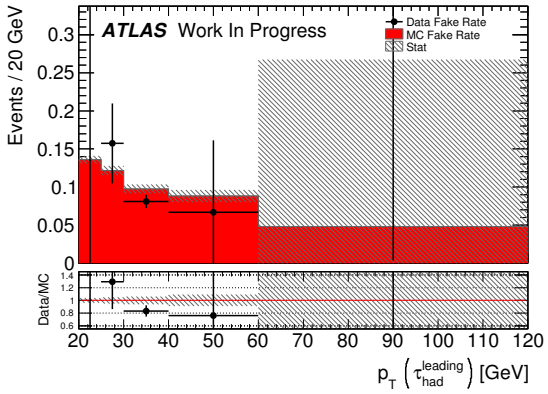
The estimation of uncertainties on these fake rates is discussed in Section A.3.3.

7.4.1. Extracted Fake Rates

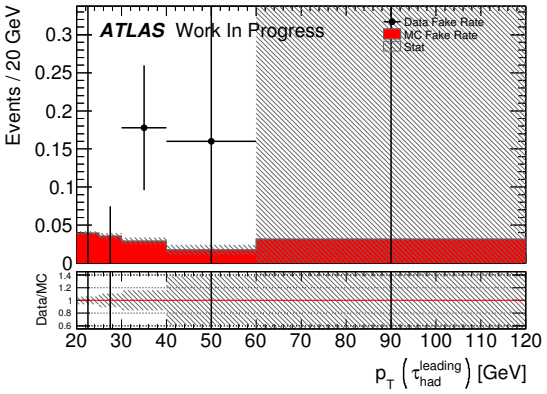
$p_T(\tau_{\text{had-vis}})$-binned fake rate have been measured on data using the tag and probe method separately for $\tau_{\text{had-vis}}$ candidates with one and three associated tracks and in the two regions defined in Section 7.2. Template fits as described in Section 7.1 have been performed in the same $p_T(\tau_{\text{had-vis}})$-binned selections to obtain quark fractions. With these measurements, Equation 7.5 has been applied to extract $p_T(\tau_{\text{had-vis}})$-binned quark and gluon fake rates.

These fake rates for the medium working point of the tau identification algorithm are shown in Figure 7.9 (fake rates for the loose and tight working points can be seen in the Figures A.3 and A.4). Due to a non-converging template fit, the extracted fake rates for $\tau_{\text{had-vis}}$ candidates with three tracks is exactly zero in the highest p_T bin. As a comparison, the figures also show the quark and gluon fake rate as extracted from the $Z \to ee$ MC

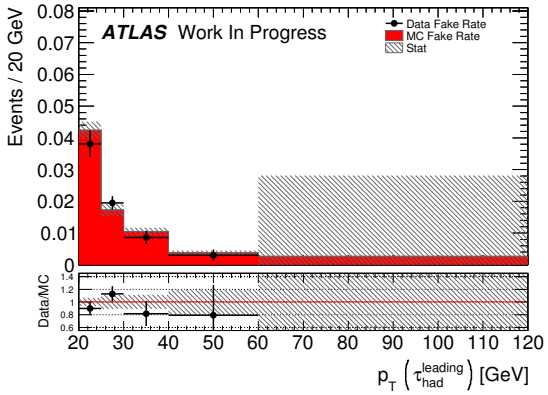
sample using truth matching information.



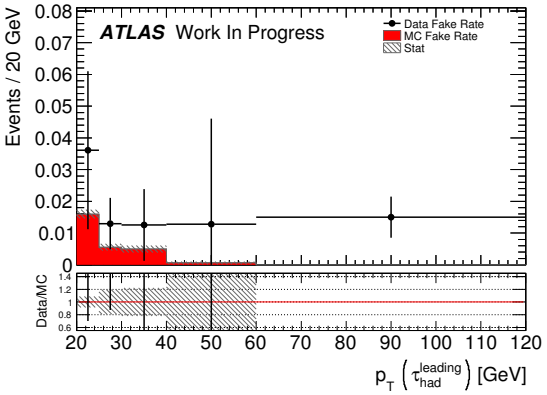
(a) Quark jets with one associated track.



(b) Gluon jets with one associated track.



(c) Quark jets with three associated tracks.



(d) Gluon jets with three associated tracks.

Figure 7.9.: Extracted quark and gluon jet fake rate from data in comparison to quark and gluon fake rates obtained from the $Z \rightarrow ee$ MC sample through truth matching. The shown hatchings display the statistical uncertainty on the MC fake rates (see Appendix A.3.1). The fake rates in this figure are calculated at the medium working point of the tau identification algorithm, for the loose and tight working point see Figures A.3 and A.4.

Some uncertainties shown in the Figures 7.9, A.3 and A.4 extend into the unphysical negative region. This is caused by the propagation of quark fraction uncertainties that extend below 0 or above 1. These are unphysical cases, where the approximative handling of the errors as Gaussian uncertainties breaks down. However, since these artifacts do not influence the interpretation of the given result, these uncertainty estimations are deemed sufficient.

Within the given uncertainties, the extracted fake rates from data seem to agree with the fake rates obtained trough truth matching. However, especially the extracted gluon fake rates in particular have very large uncertainties associated to them. The main source

of these uncertainties is the template fit method. In Section 7.3 it was shown, that the statistical uncertainties dominate the template fit when applied to the quark and gluon enriched regions. This effect is enhanced, when an additional binning in $p_T(\tau_{\text{had-vis}})$ is applied.

It is worth noting, that whenever the extracted quark fake rate is smaller than the MC prediction, the gluon fake rate is higher than the prediction and vice versa. This is an effect of the solution of the linear system: Since the mixture of both extracted fake rates needs to result in the measured rates, the divergences of the extracted fake rates from their "true value" need to balance each other out. In this sense, this effect increases the confidence in the compatibility of the extracted fake rates with the MC predictions.

8. Conclusion

Fake rate measurements on 2015 ATLAS data and a $Z \to ee$ MC sample have been performed using the tag and probe method. These fake rates are provided as functions of the transverse momentum and the pseudorapidity of the $\tau_{\text{had-vis}}$ candidates as well as the amount of pileup in the event. A direct comparison revealed an imperfect modelling of the fake rates by the MC in low $p_T(\tau_{\text{had-vis}})$-regions. Scale factors are provided to correct the calculated MC fake rates.

In accordance with a 2011 study [31] it is assumed, that the mismodelling of the fake rates is the effect of a mismodelling of the fraction of quark induced jets in the events. A template fit method has been successfully applied to the data to extract the quark fraction. This measured fraction was compared to the fraction obtained from MC using a truth matching algorithm. Within the uncertainties of the measurement, both predictions are compatible.

A separation of the data into two $p_T(\ell\ell)$ regions has been optimised for a separate measurement of fake rates and quark fractions in both regions. The results of these measurements have been unfolded to obtain pure quark and gluon jet fake rates. A comparison of the obtained fake rates with predictions obtained from MC with the truth matching algorithm hint at an agreement within the limits of the available statistics.

The 2015 ATLAS dataset used in this thesis corresponded to about $3.2\,\text{fb}^{-1}$. The 2016 ATLAS dataset already contains roughly ten times as much data. With this increased statistics it will be possible to gain a more robust conclusion. In addition it would be possible to also use other, similar processes like $Z \to \mu\mu$ events for the tag and probe method to further increase the available statistics.

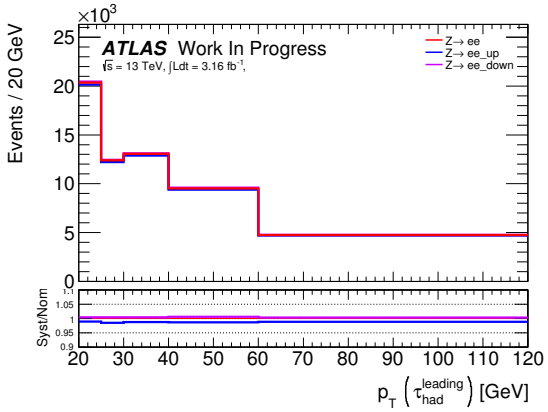
Another limiting factor of these results is the systematic uncertainty from the unmatched $\tau_{\text{had-vis}}$ candidates in the used MC samples. A better understanding of the origin of these candidates could drastically reduce the systematic uncertainties on the extracted quark fraction, both on data and in the MC sample.

The use of two independent physics processes for the fake rate measurements, that have different quark fractions, could further reduce the uncertainties on the extracted quark and gluon fake rates.

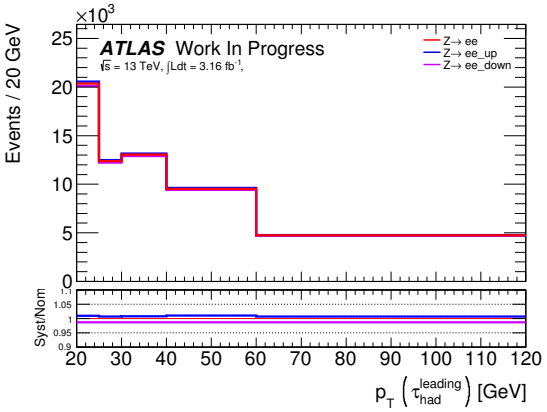
If the hypothesis holds true, that the fake rate distribution in a physics process depends solely on the quark fraction in this process, it will be possible to distribute the $p_T(\tau_{\text{had-vis}})$-binned pure quark and gluon fake rates along with a tool to measure the quark fraction in a given selection. The pure fake rates could then be mixed with the quark fractions into the expected total fake rate distribution for any physics process. This would remove the need to estimate the fake contribution in an individual way for every analysis involving hadronically decaying tau leptons.

A. Appendix

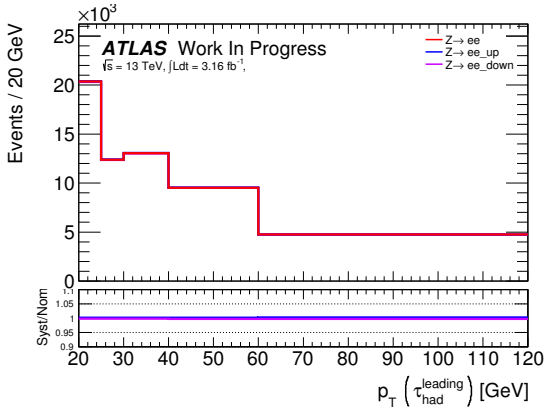
A.1. Additional Figures



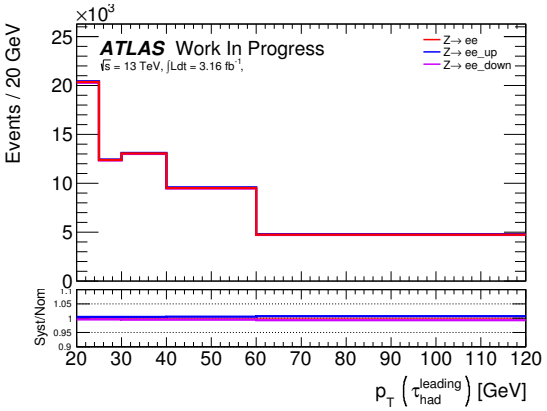
(a) EGamma resolution.



(b) EGamma scale.

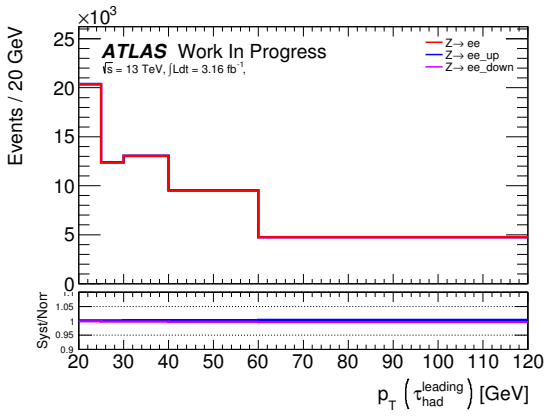


(c) Electron isolation SF.

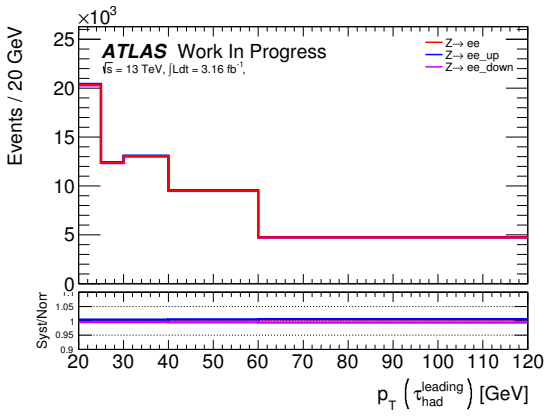


(d) Electron mediumLLH SF.

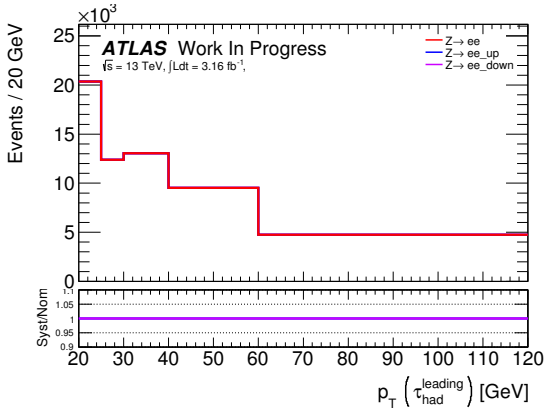
Figure A.1.: Direct effect of the different systematic variations on the $Z \to ee$ MC events. Shown is the p_T distribution for leading $\tau_{\text{had-vis}}$ candidates in events passing the selection defined in Section 5.1.



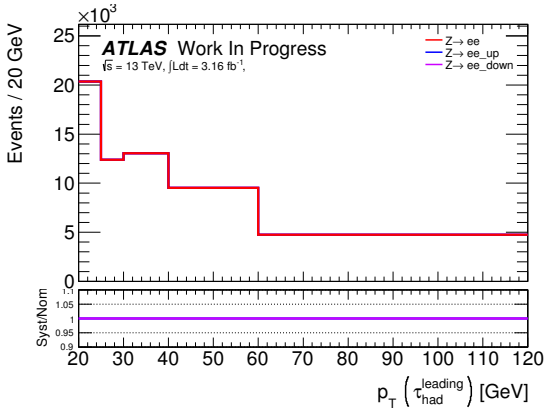
(a) Electron reconstruction SF.



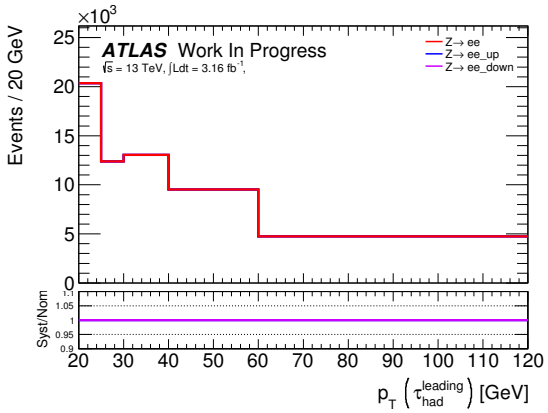
(b) Electron trigger SF.



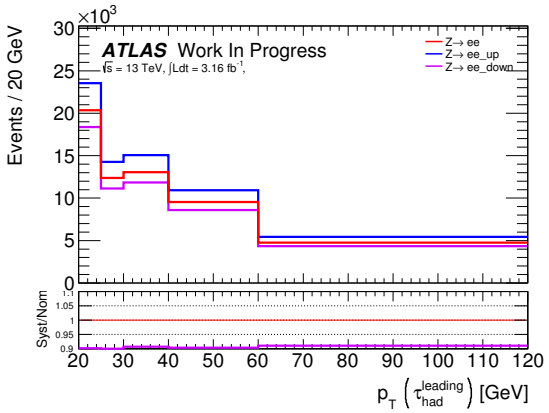
(c) Tau energy scale detector.



(d) Tau energy scale insitu.

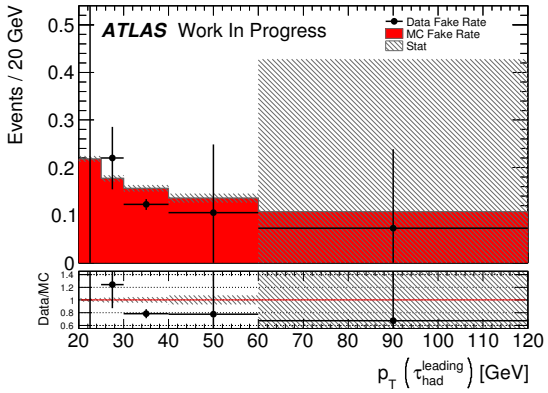


(e) Tau energy scale model.

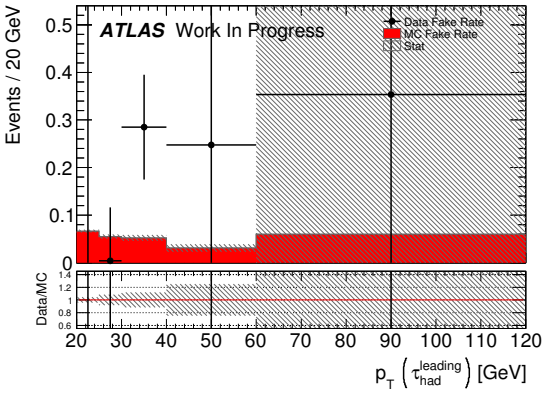


(f) Width of the Z^0 mass window.

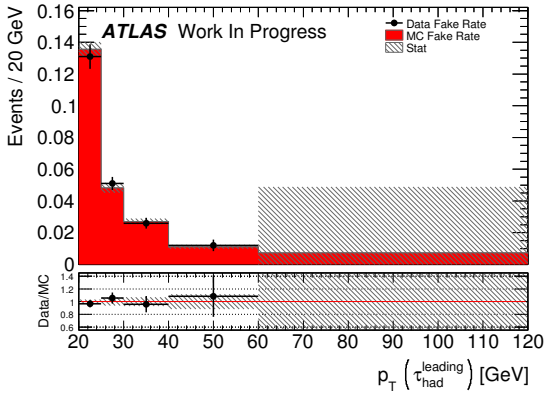
Figure A.2.: Direct effect of the different systematic variations on the $Z \to ee$ MC events. Shown is the p_T distribution for leading $\tau_{\text{had-vis}}$ candidates in events passing the selection defined in Section 5.1.



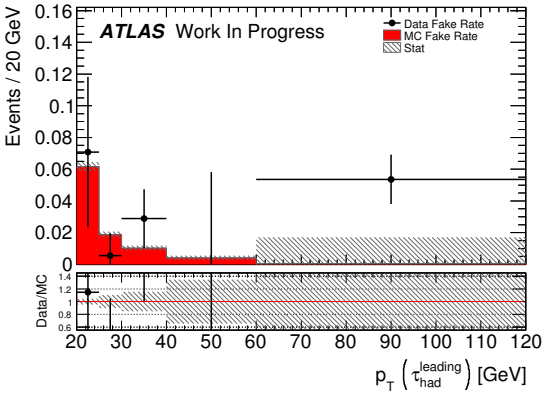
(a) Quark jets with one associated track.



(b) Gluon jets with one associated track.

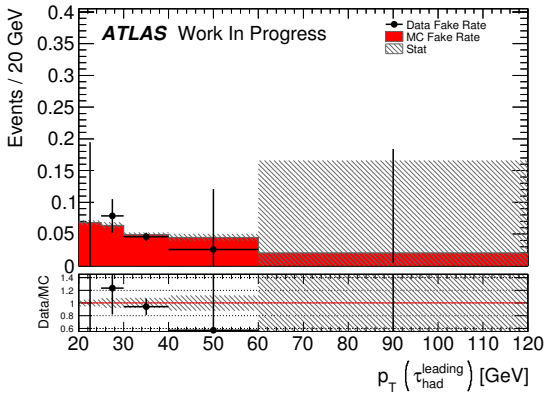


(c) Quark jets with three associated tracks.

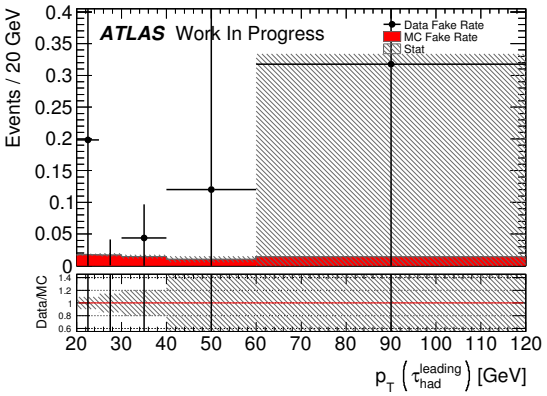


(d) Gluon jets with three associated tracks.

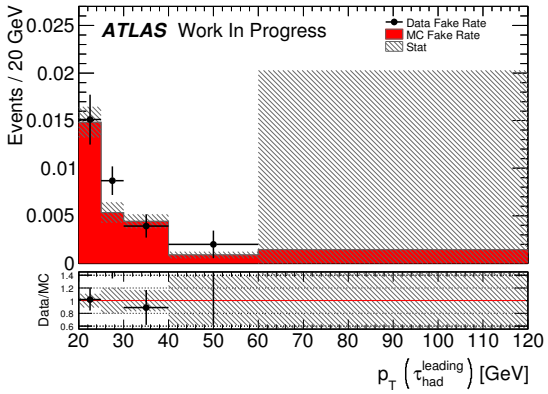
Figure A.3.: Extracted quark and gluon jet fake rate from data in comparison to quark and gluon fake rates obtained from the $Z \rightarrow ee$ MC sample using a truth-matched $\tau_{\text{had-vis}}$ candidate. The shown fake rates are calculated at the loose working point of the tau identification algorithm.



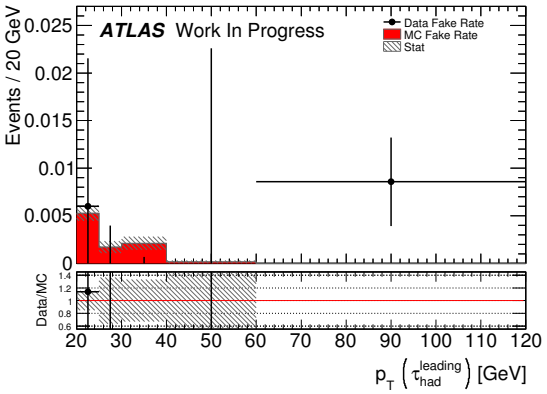
(a) Quark jets with one associated track.



(b) Gluon jets with one associated track.



(c) Quark jets with three associated tracks.



(d) Gluon jets with three associated tracks.

Figure A.4.: Extracted quark and gluon jet fake rate from data in comparison to quark and gluon fake rates obtained from the $Z \rightarrow ee$ MC sample using a truth-matched $\tau_{\text{had-vis}}$ candidate. The shown fake rates are calculated at the tight working point of the tau identification algorithm.

A.2. ATLAS Tau ID BDT Input Variables

The following section is taken from [24]:

Central energy fraction ($f_{\mathbf{cent}}$): Fraction of the calorimeter transverse energy deposited in the region $\Delta R < 0.1$ with respect to all energy deposited in the region $\Delta R < 0.2$ around the $\tau_{\text{had-vis}}$ candidate. It is calculated by summing the energy deposited in all cells belonging to TopoClusters with a barycentre in these regions, calibrated at the EM energy scale. [Figure 4.4(a)]

Leading track momentum fraction ($f^{-1}_{\mathbf{leadtrack}}$): The transverse energy sum, calibrated at the EM energy scale, deposited in all cells belonging to TopoClusters in the core region of the $\tau_{\text{had-vis}}$ candidate, divided by the transverse momentum of the highest-p_T charged particle in the core region. [Figure 4.4(b)]

Track radius ($R^{0.2}_{\mathbf{track}}$): p_T-weighted ΔR distance of the associated tracks to the $\tau_{\text{had-vis}}$ direction, using track only tracks in the core region. [Figure 4.4(c)]

Leading track IP significance ($|S_{\mathbf{leadtrack}}|$): Absolute value of transverse impact parameter of the highest-p_T track in the core region, calculated with respect to the TV, divided by its estimated uncertainty. [Figure 4.4(d)]

Fraction of tracks p_T in the isolation region ($f^{\mathbf{track}}_{\mathbf{iso}}$): Scalar sum of the p_T of tracks associated with the $\tau_{\text{had-vis}}$ candidate in the region $0.2 < \Delta R < 0.4$ divided by the sum of the p_T of all tracks associated with the $\tau_{\text{had-vis}}$ candidate. [Figure 4.4(e)]

Maximum ΔR ($\Delta R_{\mathbf{Max}}$): The maximum ΔR between a track associated with the $\tau_{\text{had-vis}}$ candidate and the $\tau_{\text{had-vis}}$ direction. Only tracks in the core region are considered. [Figure 4.4(f)]

Transverse flight path significance ($S^{\mathbf{flight}}_{\mathbf{T}}$): The decay length of the secondary vertex (vertex reconstructed from the tracks associated with the core region of the $\tau_{\text{had-vis}}$ candidate) in the transverse plane, calculated with respect to the TV, divided by its estimated uncertainty. It is defined only for multi-track $\tau_{\text{had-vis}}$ candidates. [Figure 4.5(a)]

Track mass ($m_{\mathbf{track}}$): Invariant mass calculated from the sum of the four-momentum of all tracks in the core and isolation regions, assuming a pion mass for each track. [Figure 4.5(b)]

Fraction of EM energy from charged pions ($f_{\mathbf{EM}}^{\mathbf{track\text{-}HAD}}$): Fraction of the electromagnetic energy of tracks associated with the $\tau_{\text{had-vis}}$ candidate in the core region. The numerator is defined as difference between the sum of the momentum of tracks in the core region and the sum of cluster energy deposited in the hadronic part of each TopoCluster (including the third layer of the EM calorimeter) associated with the $\tau_{\text{had-vis}}$ candidate. The denominator is the sum of cluster energy deposited in the electromagnetic part of each TopoCluster (presampler and first two layers of the EM calorimeter) associated with the $\tau_{\text{had-vis}}$ candidate. All clusters are calibrated at the LC energy scale. [Figure 4.5(c)]

Ratio of EM energy to track momentum ($f_{\mathbf{track}}^{\mathbf{EM}}$): Ratio of the sum of cluster energy deposited in the electromagnetic part of each TopoCluster associated with the $\tau_{\text{had-vis}}$ candidate to the sum of the momentum of tracks in the core region. All clusters are calibrated at the LC energy scale. [Figure 4.5(d)]

Track-plus-EM-system mass ($m_{\mathbf{EM+track}}$): Invariant mass of the system composed of the tracks and up to two most energetic EM clusters in the core region, where EM cluster energy is the part of TopoCluster energy deposited in the presampler and first two layers of the EM calorimeter, and the four-momentum of an EM cluster is calculated assuming zero mass and using TopoCluster seed direction. [Figure 4.5(e)]

Ratio of track-plus-EM-system to p_T ($p_{\mathbf{T}}^{\mathbf{EM+track}}/p_{\mathbf{T}}$): Ratio of the $\tau_{\text{had-vis}}$ p_T , estimated using the vector sum of track momenta and up to two most energetic EM clusters in the core region to the calorimeter-only measurement of $\tau_{\text{had-vis}}$ p_T . [Figure 4.5(f)]

A correction depending linearly on μ, the average number of pileup interactions per bunch crossing computed from the instantaneous luminosity [], is applied to each discriminating variable. The usage of μ, instead of the number of reconstructed interaction vertices per event, N_{PV} , provides compatibility with the High Level Trigger, that cannot run a full event primary vertex fit.

A.3. Error Estimations

A.3.1. Binomial Errors for Fake Rates

In general, the error propagation for a ratio $r = a/b$, where both a and b are fluctuating, is given as:

$$\sigma_r^2 = \left(\frac{\partial r}{\partial b}\right)^2 \sigma_b^2 + \left(\frac{\partial r}{\partial a}\right)^2 \sigma_a^2 + 2\frac{\partial r}{\partial b}\frac{\partial r}{\partial a}\mathrm{cov}(a,b). \tag{A.1}$$

For a fake rate $FR = k/n$, where k is the number of events in a subset of the set of n events, the covariance term becomes $\mathrm{cov}(n,k) = FR \cdot n = \sigma_k^2$ [35]. With this, the following formula can be obtained:

$$\sigma_{FR}^2 = \frac{(1 - 2 \cdot \frac{k}{n}) \cdot \sigma_k^2 + (\frac{k}{n})^2 \cdot \sigma_n^2}{n^2}, \tag{A.2}$$

which is the formula used in the ROOT [20] function `TH1::Divide()`, when the option `"B"` (for binomial errors) is specified.

In the case of Poisson uncertainties $\sigma_k = \sqrt{k}$ and $\sigma_n = \sqrt{n}$, Equation A.2 can be simplified to the usual form of a binomial error estimation:

$$\sigma_{FR}^2 = \frac{\frac{k}{n}(1 - \frac{k}{n})}{n^2} \tag{A.3}$$

A.3.2. Error Propagation for Scale Factors

For the scale factor $s = FR^{Data}/FR^{MC}$ between the fake rate in data and the MC fake rate, the usual error propagation (Equation A.1) can be applied. Since the two fake rate measurements are independent from each other, the covariance term vanishes and the uncertainty of the scale factor is given by:

$$\sigma_s^2 = \left(\frac{1}{FR^{MC}} \cdot \sigma_{FR^{MC}}\right)^2 + \left(\frac{FR^{Data}}{(FR^{MC})^2} \cdot \sigma_{FR^{MC}}\right)^2 \tag{A.4}$$

A.3.3. Error Propagation for Extracted Quark-/Gluon Fake Rates

As discussed in Section 7.4, the fake rate of quark or gluon initiated jets FR_q or FR_g can be estimated from two fake rates FR_i ($i = 1,2$) measured in selections with different fractions of quark initiated jets q_i:

$$FR_q = \frac{(1-q_2)\cdot FR_1 - (1-q_1)\cdot FR_2}{q_1 - q_2} \qquad \text{and} \qquad FR_g = \frac{q_2\cdot FR_1 - q_1\cdot FR_2}{q_2 - q_1} \tag{A.5}$$

The uncertainty on this fake rate can be estimated by propagating the errors of the measured fractions and fake rates:

$$\sigma^2_{FR_x} = \sum_i \left(\frac{\partial FR_x}{\partial FR_i}\cdot\sigma_{FR_i}\right)^2 + \left(\frac{\partial FR_x}{\partial q_i}\cdot\sigma_{q_i}\right)^2, \tag{A.6}$$

where $x = q, g$. The necessary differential derivations are given by:

$$\frac{\partial FR_q}{\partial FR_1} = \frac{1-q_2}{q_1-q_2}, \qquad \frac{\partial FR_q}{\partial FR_2} = \frac{q_1-1}{q_1-q_2}, \tag{A.7}$$

$$\frac{\partial FR_q}{\partial q_1} = (q_2-1)\cdot\frac{FR_1-FR_2}{(q_1-q_2)^2}, \qquad \frac{\partial FR_q}{\partial q_2} = (q_1-1)\cdot\frac{FR_1-FR_2}{(q_1-q_2)^2}, \tag{A.8}$$

$$\frac{\partial FR_g}{\partial FR_1} = \frac{q_2}{q_2-q_1}, \qquad \frac{\partial FR_g}{\partial FR_2} = \frac{-q_1}{q_2-q_1}, \tag{A.9}$$

$$\frac{\partial FR_g}{\partial q_1} = q_2\cdot\frac{FR_1-FR_2}{(q_2-q_1)^2}, \qquad \frac{\partial FR_g}{\partial q_2} = -q_1\cdot\frac{FR_1-FR_2}{(q_2-q_1)^2}, \tag{A.10}$$

Bibliography

[1] The ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at* $\sqrt{s}$ *= 7 and 8 TeV*, ATLAS-CONF-2015-044, CMS-PAS-HIG-15-002

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

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

[4] A. Salam, *Elementary Particle Physics: Relativistic Groups and Analyticity*, in N. Svartholm, editor, *8-th Nobel Symposium*, page 367, Almquvist and Wiksell, Stockholm (1968)

[5] K. A. Olive, et al. (Particle Data Group), *2014 Review of Particle Physics*, Chin. Phys. **C38**, 090001 (2014)

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

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

[8] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145**, 1156 (1966)

[9] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716**, 1 (2012)

[10] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys.Lett. **B716**, 30 (2012)

[11] S. Heinemeyer, et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, arXiv:1307.1347

[12] S. Abachi, et al. (D0), *Observation of the top quark*, Phys.Rev.Lett. **74**, 2632 (1995)

[13] M. Thomson, *Modern Particle Phyiscs*, Cambridge University Press, Cambridge (2013), First Edition

[14] D. Griffiths, *Introduction to Elementary Particles*, Wiley-VCH, Weinheim (2008), Second, Revised Edition

[15] P. Langacker, *The Standard Model and Beyond*, Taylor & Francis, Boca Raton (2010)

[16] O. S. Brüning, et al., *LHC Design Report*, CERN, Geneva (2004)

[17] ATLAS Collaboration, *Luminosity Public Results Run 2*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2

[18] S. Chatrchyan, et al. (CMS Collaboration), *Measurement of the weak mixing angle with the Drell-Yan process in proton-proton collisions at the LHC*, Phys.Rev. **D84**, 112002 (2011)

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

[20] R. Brun, F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl.Instrum.Meth. **A389**, 81 (1997)

[21] A. Ruiz Martínez, A. Collaboration, *The Run-2 ATLAS Trigger System* (2016), ATL-DAQ-PROC-2016-003

[22] W. Lampl, et al., *Calorimeter clustering algorithms: Description and performance* (2008), ATL-LARG-PUB-2008-002

[23] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **0804**, 063 (2008)

[24] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATLAS-PHYS-PUB-2015-045

[25] P. N. S. Frixione, C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **0711**, 070 (2007)

[26] S. M. T. Sjostrand, P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178**, 852 (2008)

[27] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, JHEP **1409**, 145 (2014)

[28] J. Pumplin, et al., *New Generation of Parton Distributions with Uncertainties from Global QCD Analysis*, JHEP **07**, 012 (2002)

[29] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. **C74**, 2941 (2014)

[30] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. **C76**, 292 (2016)

[31] ATLAS Collaboration, *Measurement of the Mis-identification Probability of τ Leptons from Hadronic Jets and from Electrons*, ATLAS-CONF-2011-113

[32] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV* (2016), ATL-PHYS-PUB-2016-015

[33] ATLAS Collaboration, *Determination of the tau energy scale and the associated systematic uncertainty in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC in 2012* (2013), ATLAS-CONF-2013-044

[34] R. Barlow, C. Beeston, *Fitting using finite Monte Carlo samples*, Comp. Phys. Comm. **77**, 219 (1993)

[35] G. Ranucci, *Binomial and ratio-of-Poisson-means frequentist confidence intervals applied to the error evaluation of cut efficiencies*, arXiv:0901.4845

Acknowledgements

I want to thank Prof Stan Lai for giving me the opportunity to work on this very interesting topic in his research group. My thanks also go to PD Dr Jörn Große-Knetter for volunteering to be the second referee for my thesis.

I am deeply thankful for the help Prof Lai and Dr Michel Janus offered me during the last year and their patience. Furthermore, my thanks go to other members of the institute, including Antonio De Maria, Eric Drechsler and Dr Zinonas Zinonos. I am looking forward to working with you all in the future.

Thank you!

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

Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.
Darüberhinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht auszugsweise, im Rahmen einer nichtbestandenen Prüfung an dieser oder einer anderen Hochschule eingereicht wurde.

Göttingen, den 8. Januar 2017

(Timo Dreyer)