Validation of a new approach for tau lepton identification in the ATLAS experiment

Sebastian Matberg

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

November 2012

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate kenntlich gemacht habe.

1. Gutachter: Prof. Dr. Klaus Desch
2. Gutachter: Prof. Dr. Jochen Dingfelder

Danksagungen

Contents

Chapter 1

Introduction

1.1 Motivation for the Analysis

When the idea for the PanTau algorithm first came up, its purpose was seen as a possible replacement of TauRec, the current τ lepton identification method in ATLAS. By making use of the particle flow concept, performance was expected to improve by exploiting the tracker's better energy resolution, compared to the calorimeter. A further advantage is the built-in opportunity to separate the τ lepton's hadronic decay modes from each other ('substructure' of the τ lepton), which is a natural consequence of PanTau's particle-oriented philosophy.

While all analyses benefit from a better energy resolution, the substructure is mainly important to spin-dependant studies. A particle's spin can be determined through a decay into a τ lepton and τ (anti-)neutrino, because neutrinos are virtually always left-handed (and anti-neutrinos right-handed).

Currently, PanTau is used as an optional addition to TauRec to identify decay modes, while PanTau's fake suppression is not performed by default. The Tau Working Group (***Tau WG***) of the ATLAS Collaboration will conduct further studies in the near future to test PanTau's performance, whose outcome will have great impact on the decision how to proceed with the tool.

Prior to these studies, PanTau's basic functionality had to be proven. This document presents an analysis of PanTau's handling of fake τ leptons, along with a technical test of its output mechanisms. These investigations have played a major part in demonstrating PanTau's physical abilities in order to qualify it as an adequate analysis tool.

1.2 The τ Lepton in Experimental Physics

1.2.1 The τ Lepton in the Standard Model of Particle Physics

The Standard Model of Particle Physics, often just referred to as the 'Standard Model', classifies particles as quarks, leptons and gauge bosons (see Fig. 1.1, where the Higgs boson is omitted for historical reasons).

The τ lepton (in short hand τ or τ^-) is the heaviest lepton in the SM, having a mass of 1.777 GeV/c^2. Its quantum numbers are: lepton number 1, electric charge -1 e, spin $\frac{1}{2}$, weak isospin $-\frac{1}{2}$ and τ-number 1. The anti-particle of the τ is the anti-τ ($\bar{\tau}$ or τ^+). The symbol τ is also often used as a generalisation of τ and $\bar{\tau}$.
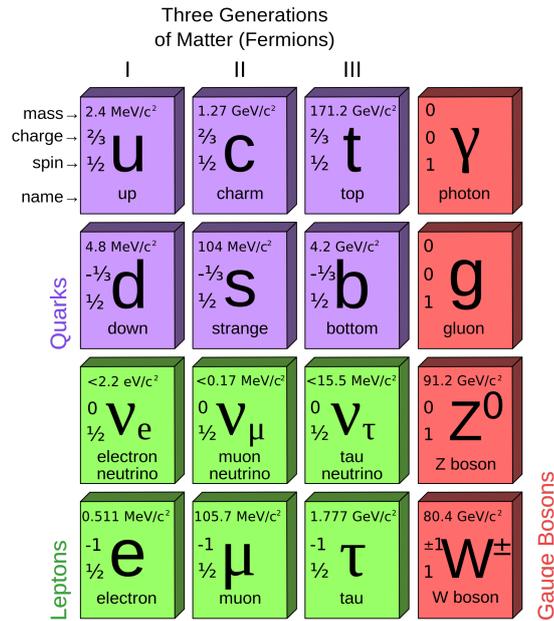
Figure 1.1: Particles of the Standard Model (omitting the postulated Higgs Boson) [1].

Due to its short lifetime of $T_\tau = 2.9 \times 10^{35}$s the τ lepton can only be measured indirectly by its decay products. This and the relatively high energy required to produce τ leptons in collisions are reasons for its late experimental discovery in the 1970s, as opposed to the discoveries of the other leptons in 1897 (electron) and 1936 (muon). It was the first lepton to be discovered at a particle collider (electron: cathode rays, muon: cosmic rays).

1.2.2 Scientific Significance

Today the scientifically most relevant applications of τ leptons are searches for undiscovered elementary particles. Among the most prominent examples are the search and measurement of the Higgs boson and particles predicted by models beyond SM, e. g. Supersymmetry (**SUSY**) models. Many of these undiscovered particles can decay into one or more τ leptons, thus making an excellent knowledge of their properties and decay channels as well as their accurate identification essential requirements to new physics discoveries.

1.2.3 Identification and Decay Modes

Unfortunately there are severe challenges in τ lepton identification. As mentioned above, the short lifetime induces the necessity of indirect measurement, which is complicated because of the variety of decay channels accessible to the τ lepton. Since the τ lepton's rest mass, unlike the other leptons', is high enough for it to decay even into hadrons, the resulting tracks of the newly formed particles make the decay look very similar to a hadronic jet induced by gluons or quarks (**QCD jet**).

Decay Mode	Short Hand	Fraction	Process
1 prong	1p0n	10.8 %	$\tau^- \to \nu_\tau + 1\pi^- + 0\pi^0$
1 prong 1 neutral	1p1n	25.5 %	$\tau^- \to \nu_\tau + 1\pi^- + 1\pi^0$
1 prong > 1 neutral	1pXn	10.4 %	$\tau^- \to \nu_\tau + 1\pi^- + [2..]\pi^0$
3 prong	3p0n	9.3 %	$\tau^- \to \nu_\tau + 2\pi^- + 1\pi^+ + 0\pi^0$
3 prong $\geq$ 1 neutral	3pXn	4.6 %	$\tau^- \to \nu_\tau + 2\pi^- + 1\pi^+ + [1..]\pi^0$
leptonical	lep	35.2 %	$\tau^- \to \nu_\tau + \begin{cases} 1\mu^- + \bar{\nu}_\mu \\ 1e^- + \bar{\nu}_e \end{cases}$

Table 1.1: Decay Modes

The τ^- decays into a ν_τ and a virtual W^-, which in turn decays either into a lepton and the corresponding anti-neutrino ($\mu^- + \bar{\nu}_\mu$ or $e^- + \bar{\nu}_e$) or into two quarks q, $\bar{q}'$ forming a meson (mostly a π^-; possibly more than one meson is formed via a resonance like ρ^- or a_1^-), as seen in Eq. 1.1. The decays for τ^+ are analogue.

$$\text{hadronical:} \quad \tau^- \to \nu_\tau + W^- \to \nu_\tau + q\bar{q}' \xrightarrow{(\rho^-)(a_1^-)} \begin{cases} \pi^- \\ \pi^- + \pi^0 \\ 2\pi^- + \pi^+ \\ K^- \\ \vdots \end{cases} \tag{1.1a}$$

$$\text{leptonical:} \quad \tau^- \to \nu_\tau + W^- \to \nu_\tau + \begin{cases} \mu^- + \bar{\nu}_\mu \\ e^- + \bar{\nu}_e \end{cases} \tag{1.1b}$$

These cases are further divided into a set of categories (in the following: decay ***modes***) according to their quantity of charged and neutral end products, where rare variants (like decays into more than three charged particles or those involving K^- for instance) are disregarded. In Table 1.1 these decay modes are defined, where 'p' stands for ***prong*** and 'n' for neutral particles. The letter 'X' stands for either '2 or more' (in the 1pXn case) or '1 or more' (in 3pXn). This ambiguous nomenclature was chosen because of the particular importance (i. e. relative high branching ratio) of the 1p1n mode compared to the 1-prong modes involving more than one neutral particle, as opposed to the relatively low branching ratio of 3pXn. Given in parenthesis are the respective branching ratios for each mode [2]. Even though π^0 mesons are short-lived particles and therefore arguably no end products in the narrower sense, they are listed directly due to their relatively straightforward reconstruction (they decay into 2γ with a branching fraction of 98.8 %). Again, decays are stated for τ^-, while τ^+ decays analogously.

Chapter 2

The ATLAS Experiment

The acronym *ATLAS* stands for A Toroidal LHC Apparatus and refers both to a particle detector and the corresponding physics collaboration. It is named after the particle accelerator Large Hadron Collider (*LHC*), which it is connected to.

The description of ATLAS and the LHC in this chapter are limited to a brief summary and focus on those facts which are referred to throughout this document. A more complete portrayal of the detector can be found in the given references [3, 4].

2.1 Large Hadron Collider

The LHC is a toroidal particle accelerator with a circumference of 27 km [5]. It was built approximately 100 m underground the Swiss-French border in the pre-existing tunnel of the previous accelerator named *LEP* for Large Electron-Positron Collider (see Fig. 2.1). It is based at the European Organization for Nuclear Research (*CERN*, short for the original name Conseil Européen pour la Recherche Nucléaire).

In the LHC, two beams of either protons or lead ions are accelerated and driven to a controlled collision at the experiments based at each of the four collision points. The building process of the collider and detectors took from 1998 to 2008, but due to a technical problem the LHC did not start operating till November 2009. Since March 2010 it is run at beam energies of 3.5 TeV (in 2010–2011) to 4 TeV (in 2012). In 2013–2014, the LHC will not be operational during a 20 month phase, planned for maintenance and upgrades. The new launch is planned for 2015 at 6.5 TeV beam energy.

2.2 ATLAS Detector

Among various experiments based at the LHC (most mentionable ALICE, CMS and LHCb), the largest collaboration – incorporating a total of about 3 000 scientists for the experiment – is associated with the ATLAS detector, which is a so-called general-purpose detector: It was designed in a way that a wide range of analyses in the context of searches for new physics as well as confirmation of known physics (e. g. limit settings in both situations) are possible. To achieve this, the machine had to include mechanisms to detect any known particles (except for neutrinos).
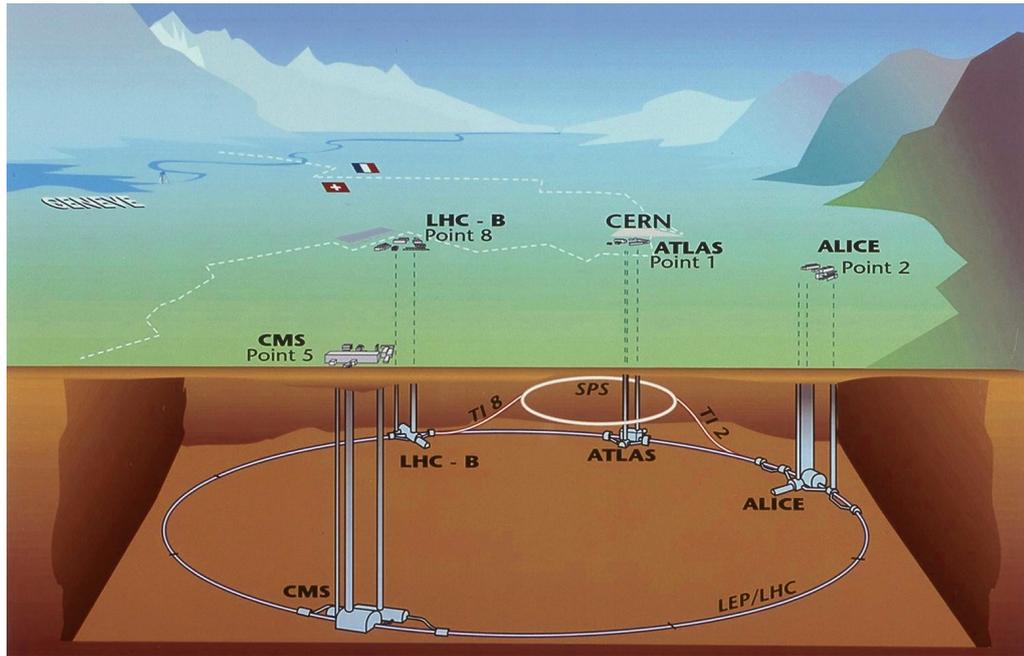
Figure 2.1: The LHC ring in the LEP/LHC tunnel at CERN near Geneva [6]. Illustrated are the four largest experiments at the LHC: ATLAS, ALICE, CMS and LHCb (here denoted as LHC-B). The former Super Proton Synchrotron (SPS), now functioning as a pre-accelerator for the LHC, is also depicted.

The ATLAS detector consists of multiple layers to achieve means of identification for diverse particle types (Fig. 2.2). It can approximately be described as a cylinder (this is also called barrel design) of 25 m diameter and 44 m length and weighs 7 000 t. The separate detecting layers are arranged in cylindrical shapes around the interaction point: the inner detector, the calorimeters and the muon chambers (outwards from the centre).

2.2.1 Inner Detector: Pixel, SCT and TRT

The inner detector is built as a cylinder with a 2.1 m diameter and a length of 6.2 m around the interaction point, at which the protons are lead to collision (Fig. 2.3a). It consists of pixel detectors, semiconductor trackers (***SCT***) and transition radiation trackers (***TRT***). Each of these detector types are arranged for one in barrel form around the detector's centre and, secondly, in form of disc-shaped end-caps (also called forward detectors or wheels), perpendicular to the beam axis. The central solenoid magnet surrounds the inner detector, providing it with a magnetic field of 2 T (cp. Ch. 2.2.4).

The main purpose of the inner detector is to achieve a good resolution for the vertices' positions and the particles' impact parameters and momenta.

The pixel detector has three layers of barrels with 5 cm, 9 cm and 12 cm radius, respectively, all located directly around the interaction point. This structure is bordered on each side by three end-cap pixel detector discs, each with 12 cm radius and an enclosed cavity of 9 cm radius.

The eight layers of the barrel SCT are fixed on four carbon-fibre cylinders and surround
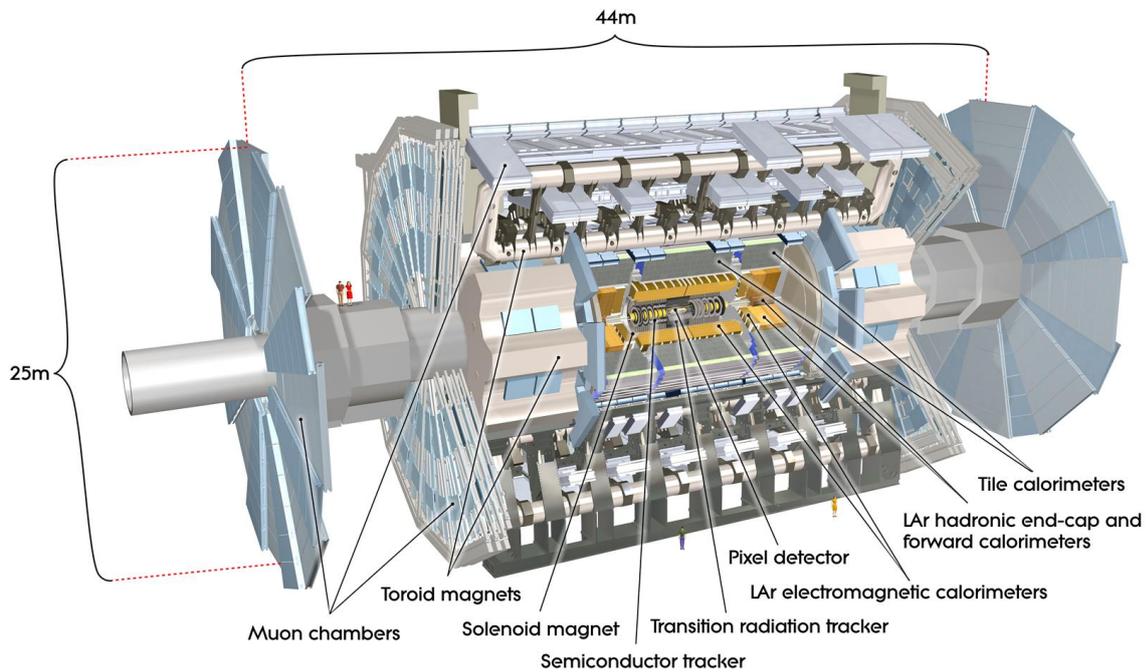
Figure 2.2: ATLAS detector [3]. The design is structured by means of several layers: inner detector, calorimeters and muon chambers. Magnets are located between the inner detector and the calorimeters, as well as inside the muon system.
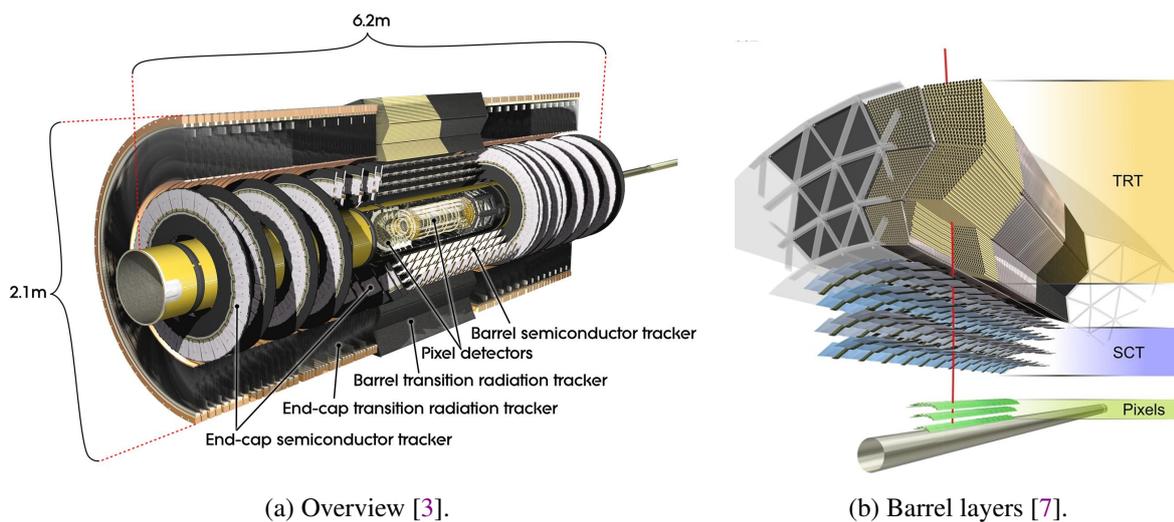
(a) Overview [3].

(b) Barrel layers [7].

Figure 2.3: Design of the inner detector. The pixel detectors are placed in the centre, enclosed by the SRT. Both are surrounded by the TRT. Each of these layers are structured as a barrel plus end-caps.

the pixel detector at radii between 30 cm and 52 cm. These are complemented by forward detectors on both sides of the barrel. Both parts of SCT are microstrip detectors using silicon as a semiconductor.

The outermost layer of the inner detector is the TRT and is based on straw chambers. Like the other parts of the inner detector, the TRT consists of a barrel and end-caps. The gaseous mixture used in the straws consists of 70 % Xe, 27 % CO_2 and 3 % O_2, with Xe being used to detect transition radiation induced from radiators located in between the straws. The wires inside the straws are made of gold-coated tungsten. The barrel TRT consists of 50 000 straws parallel to the beam axis (see Fig. 2.3b) and ranges from a radius of 56 cm to 107 cm, while the 36 end-cap discs are composed of 320 000 straws perpendicular to the beam.[4]

2.2.2 Calorimeters

Calorimeters are installed outside the inner detector to measure the particles' energies by letting them interact with dense matter, the absorber, thus producing electromagnetic or hadronic particle showers (depending on the given particle's physical properties). The showers in turn induce signals in a second material, the sensor.[1]

There are two kinds of calorimeters in the ATLAS detector (see Fig. 2.4). The inner part uses liquid argon (*LAr*) as a sensor, while the outer tile calorimeters (*TileCal*) use plastic scintillators. The particles emitted by the sensors are electrons in case of LAr and photons in case of plastic.

A more event-based approach of classifying the calorimeters is by nature of the analysed particle showers (electromagnetic or hadronic). In this manner, the LAr calorimeters can further be divided into the electromagnetic barrel (*EMB*), electromagnetic end-caps (*EMEC*) and hadronic end-caps (*HEC*). Additionally, there are two LAr forward calorimeters (*FCal*).

The tile calorimeter consists of the tile barrel and the tile extended barrel.

Used absorber materials are Pb in EMB and EMEC, Cu in HEC, W in FCal and steel in TileCal.

2.2.3 Muon Chambers

To detect muons, straw chambers are installed outside of the calorimeters (cp. Fig. 2.2). Mainly muons reach this part of the detector, since other types of particles are stopped in large part in the calorimeters – except for neutrinos, which pass through the muon system undetected as well. The muon chambers form the outermost barrel of ATLAS and in addition feature several end-cap wheels.

To induce the magnetic field needed to measure the muons' momenta by means of curved tracks, toroidal magnets are installed within the barrel (cp. Ch. 2.2.4).

---

[1] It is also possible to have one and the same material functioning as absorber and sensor (homogeneous calorimeter). However, this technique is not used in ATLAS.
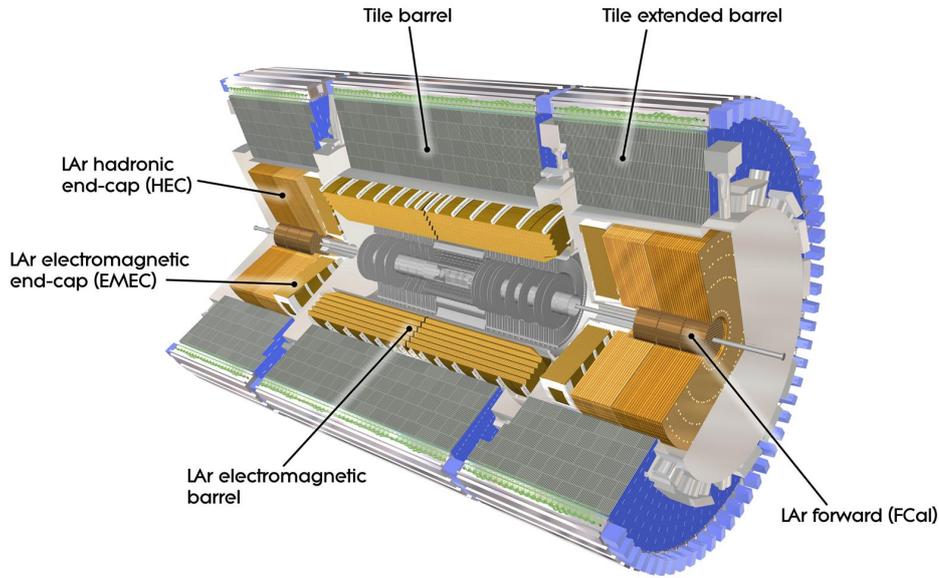
Figure 2.4: LAr and tile calorimeters [3].

2.2.4 Magnetic System

The muon chambers are pervaded by eight toroidal magnets, which are referred to as barrel toroid. Along with two end-cap magnets, each in turn consisting of eight toroidal magnets, and a central solenoid (Fig. 2.5a), they provide the magnetic field for the particle deflection in the detector (Fig. 2.5b). All magnets are superconducting and therefore need to be held in a cryostat. ATLAS uses liquid helium at a temperature of 4.8 K for cooling [8].
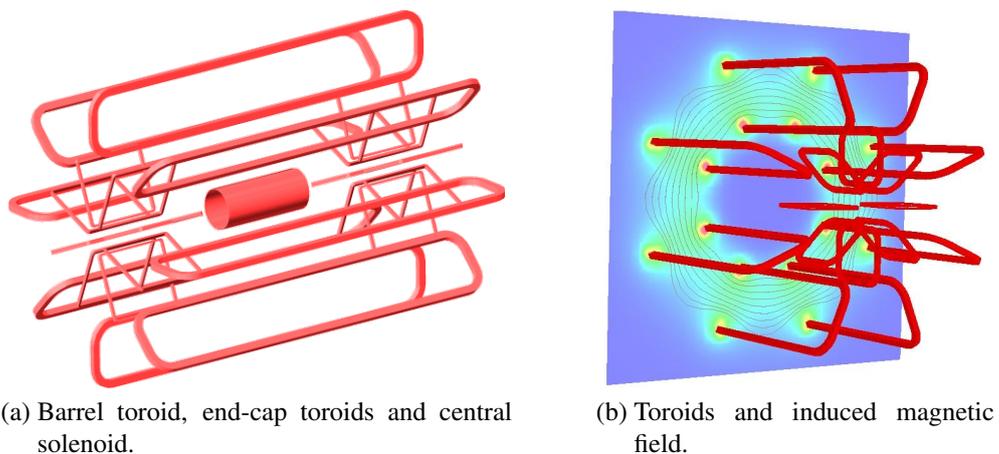
(a) Barrel toroid, end-cap toroids and central solenoid.

(b) Toroids and induced magnetic field.

Figure 2.5: Magnets and their induced field [8].

2.2.5 Pileup and Noise

One of the aims at the LHC and in ATLAS is to produce high luminosities in order to detect events with very low branching ratios in sufficient statistics to analyse them scientifically. This is realised by colliding bunches of 10^{11} protons at a rate of 40 MHz.

The such achieved luminosities reached values of up to 550 pb^{-1} per week in 2011 (see Fig. 2.6).
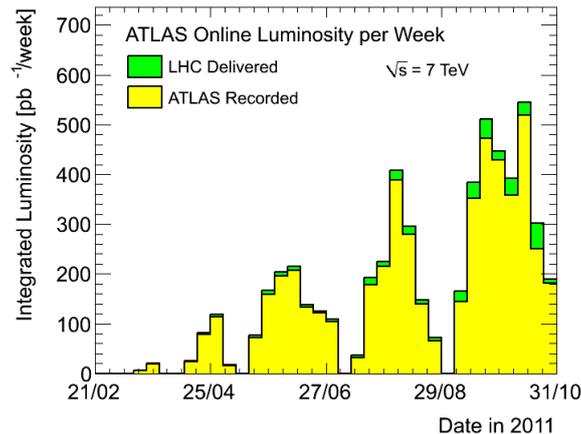
Figure 2.6: Luminosity in ATLAS for 2011, binned in weeks. [9]
Over time, higher luminosities can be reached due to optimisation of the machines.

One of the drawbacks of this practise is that several proton-proton collisions can occur during one bunch crossing, or processes from two consecutive crossings can follow too fast, that they overlap in the reconstructed event. These types of complications are called ***pileup***.

Most analysed processes have relatively low cross sections compared to QCD jets. Since it is unlikely for two or more processes with relatively low cross sections to happen in one event, most pileup processes are QCD jets with low energy (and thus low momentum).

Pileup can become a problem in physics studies, when an unexpected process influences the measurements of the analysed particles. Techniques to avoid this are often based on the expected low momentum of the jets induced by means of pileup events.

Further unwanted QCD jets originate in the proton remnants from the main collision. When two partons from the protons interact with each other, the proton remnants are not color-neutral anymore, which resolves in QCD jets. This is called the 'underlying event' and suppressed by similar means as pileup events, since the assumption of low momentum holds here as well.

Another nuisance is any form of electronic noise in the detector. While easy to understand, it is impossible to predict, and difficult to implement in simulations. In analyses it is usually handled by ignoring signals below a certain momentum threshold.

2.3 Data Processing and Software

In a physics experiment of grand proportions such as ATLAS, a complex software machinery is crucial. It spans from the interface between hardware and software, the triggers, to the physics

analyses software.

The parts that are most important for the analysis at hand are sketched in the following chapters.

2.3.1 Triggers and Data Size Reduction

The large amount of observable events in ATLAS (40 MHz) is far too high to be stored in full. To reduce the unnecessary data flow as much as possible without losing significant numbers of interesting events, a combination of three trigger levels is used. At each level, multiple triggers exist to give physics analysts the opportunity to chose suitable conditions for their analysis, e. g. the presence of a particular kind of particle with a given energy in the event. No event is stored, which did not fulfil the conditions of any of the LVL1 and LVL2 triggers.

First level triggers (*LVL1*) are implemented purely electronically to make decisions within 2 μs possible. Only event properties that can be accessed within this time scale can be used by the LVL1 triggers, e. g. coarse information about calorimeter and muon chamber hits. The event rate is reduced to 75 kHz by these means.

The events passing at least one of the LVL1 triggers are analysed by the LVL2 triggers, which have access to the complete calorimeter and tracker information and scale the event rate down to 1 kHz.

The third level of triggers is represented by the 'event filters'. These perform detailed analyses on all events passing one or more of the LVL2 triggers to decide which events are finally saved. The resulting event rate is 0.2 kHz.

The triggers' task is not only to dispose of defective events. As already mentioned in Ch. 2.2.5, The LHC and ATLAS operate at high luminosities in order to be able to measure physics processes with very low branching ratios in sufficient quantities to obtain significant results in analyses. Hence, the triggers must also prevent common events from flooding the storage and at the same time take care to keep the more infrequent ones.

Therefore, triggers which are frequently released during a given run carry a 'prescale' P. This lets them only use every P-th event firing that trigger, while the rest are disregarded. The prescales for all triggers are stored, so they can be accounted for in physics analyses. However, triggers with high prescales are often avoided because the statistical error is larger and becomes complicated to calculate if several triggers with different prescales are used.

To further decrease the disk space for the data files needed in a particular analysis, three different techniques can be used that are tailored to the given purpose. Either one can remove more events according to an adjustable scheme. This process is called 'skimming'. Another possibility is to remove specific data containers from all events, if there are containers that are generally not needed in the analysis. This option is referred to as 'slimming'. Finally, the removal of only certain parts of a data container (i. e. a subset of variables or other objects) from all events is called 'thinning'. Naturally, these three methods can be combined.

2.3.2 Analysis Tools

Data Analysis is divided into two steps: reconstruction and physics analysis. Both can be performed within the software framework ATHENA, which is explained in some detail in

Ch. 3.1.1.

Some analyses, however, are executed externally on root ntuples created by ATHENA. At the University of Bonn, the analysis framework **SFM**, based on sframe (which itself is an analysis framework based on ROOT), has been developed to simplify and unite several distinctive physics analyses [10]. This tool has also been used extensively for the studies in Ch. 5.

In the course of this analysis, several extensions were contributed to SFM (e. g. the deviation plotter described in Ch. 5.1.3, and some more technical algorithms like a dumper for latex-formatted cut flow tables) and various external root tools were created to optimise interfaces and control over the program.

2.3.3 Coordinates and Related Variables

The common coordinate system at ATLAS uses the centre of the detector as origin. The x-axis points to the centre of the LHC, the y-axis to the top of the detector, and the z-axis in beam direction (towards LHCb). The projection of a particle's track into the xy-plane can be defined as $r(x, y)$. The angles ϕ and θ lie in the xy- and $zr(x, y)$-plane, respectively, and are measured with respect to the x- and z-axis. Rather than θ, the pseudorapidity η is often used, which is a transformation of θ (Eq. 2.1).

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

The advantage of η as a variable, rather than θ, is that it naturally provides a grid in which collision products are approximately equally distributed. This is due to the particles' relativistic boost in beam direction and the typical η-segmentation getting finer towards the (positive and negative) z-axis.

Distances are often indicated in terms of ΔR, which is the square root of the quadratic sum of differences in η and ϕ (Eq. 2.2). A cone spanned by these differences is a ΔR cone and, for a given value, referred to as e. g. 0.1-cone (for $\Delta R = 0.1$).

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.2}$$

The impact parameter in z-direction (longitudinal impact parameter) is denoted as z_0 and the longitudinal impact parameter, i. e. the closest approach to the origin in the xy-plane, as d_0.

Also the momentum p and energy E of a particle are separated into their longitudinal (p_L, E_L) and transverse (p_T, E_T) parts. The relative momenta of the partons within each proton before the collision are impossible to predict or measure. While the transverse components are negligible compared to the energies present in the collision, the longitudinal momenta are boosted because of the protons' relativistic speed. Therefore, the longitudinal components are normally avoided by grounding the physics analysis on transverse variables.

Since neutrinos are not detected in ATLAS, and neither are particles with $|\eta| > 4.9$, parts of the collision energy leave the detector unnoticed. This energy is called 'missing energy'. Its transverse component can be calculated easily from the detected particles' E_T's using transverse energy conservation, assuming a starting condition of $\sum E_T = 0$ (see above). The resulting missing transverse energy is abbreviated with **MET** or the symbol $\not{E}_T$. Often, it is

also referred to just as 'missing energy' – if not stated otherwise, usually the transverse part is meant.

2.3.4 Unit System

Calculations are performed in 'natural units' in ATLAS, which means the reduced Planck constant $\hbar$, the speed of light in vacuum c, and the Boltzmann constant k_B are set to 1 (Eq. 2.3).

$$\hbar = c = k_B = 1 \tag{2.3}$$

Energy, momentum and mass are measured in eV, length and time in eV^{-1} (Eq. 2.4). However, lengths and times are often used with SI units (mm, s) instead, and cross sections (luminosities) in the SI-related unit 'barn' b (inverse barn b^{-1}), with 1 b = $(10^{-28}\ m^2)$.

$$[\text{energy}] = [\text{momentum}] = [\text{mass}] = [\text{length}]^{-1} = [\text{time}]^{-1} = eV \tag{2.4}$$

Energy and momentum do not only share the same unit, but are also closely related (Eq. 2.5a). Since for relativistic particles, the mass can often be neglected, the terms energy and momentum are sometimes used equivalently (Eq. 2.5b).

$$E^2 = p^2 + m^2 \tag{2.5a}$$
$$\lim_{\beta \to 1} E^2 = p^2 \tag{2.5b}$$

Chapter 3

PanTau

This chapter covers the idea and scope of the τ lepton particle and decay mode reconstruction algorithm PanTau, as well as a summary of the used methods and their physical motivation. It also aims to explain how PanTau is embedded as a package within the ATHENA framework, which will help to point out the importance of validation tests (cp. Ch. 4) due to the complexity of its structure.

After a short guide about file formats used with respect to ATHENA, the reconstruction of τ leptons with TauRec, eflowRec and PanTau is addressed. The aim, however, is not to give a full description of these algorithms, which can be found in a more thorough form within the respective ATLAS papers (or other documents) on TauRec [11, 12], eflowRec [13, 14] and PanTau [15, 16, 17, 18]. The given explanations are rather intended as a sufficient preparation for the studies reported within this document.

3.1 Structure around PanTau

PanTau as a package of the analysis framework ATHENA depends on other packages' output, especially TauRec and eflowRec, which will be explained here.

3.1.1 ATHENA and Types of Data Files

ATHENA is a C++-based computing framework (configuration implemented in python) used to reconstruct and analyse collision events from ATLAS, based on real or simulated data.

Simulated data is produced by means of a Monte Carlo simulation tool (*MC*), including a simulation of the detector. Commonly used simulation tools are Alpgen and Pythia. Simulated as well as real data ('data' will mean measured data in the following where not stated otherwise, whereas 'MC' will stand for simulated data) are both stored in *RDO*s (Raw Data Objects), which is the input format for the ATHENA reconstruction tools. The output these tools provide are Analysis Data Objects (*AOD*s), Event Summary Data (*ESD*s) and/or root ntuples, which are – if they obey certain ATHENA-internal standards – referred to as *D3PD*s ('Derived Physics Data in its 3rd version'). The physics analysis tools usually run on AODs, ESDs or D3PDs, but the latter ones are more commonly used. There are, however, official tools ('D3PD Makers'), which create D3PDs from arbitrary AODs for a particular analysis. The tool `TauD3PDMaker`, for instance, builds D3PDs including information for τ lepton analysis.

While data is stored in an object oriented manner in the AODs and ESDs, ntuples are 'flat', i. e. the class information is lost. Compared to AODs, ESDs additionally contain more specific information, e. g. the deposited energy in the calorimeter cells.

There is a huge variety of core and optional packages for use within ATHENA, most of which can be divided into reconstruction or physics analysis packages. PanTau is one of the official reconstruction packages. It is, however, not run by default, but has to be activated manually.

3.1.2 Reconstruction Using Particle Flow

TauRec (short for Tau Reconstruction) is the current τ lepton identification tool in ATHENA. Since TauRec uses a rather traditional identification approach based on tracks in the inner detector and energy clusters in the calorimeter, PanTau was originally created as an alternative approach using so-called particle flow (or ***pflow***).

While traditional methods handle information according to its origin, i. e. (in ATLAS) depending on whether it comes from the tracker or the calorimeter, pflow algorithms try to use more particle-like structures from the beginning. In Fig. 3.1 an example is shown for both a traditional reconstruction and one based on pflow.[1] While the former stores and processes information depending on which part of the detector it comes from, the pflow algorithm aims for a more physics-inspired reconstruction concept by arranging data in particle-like structures from the beginning.
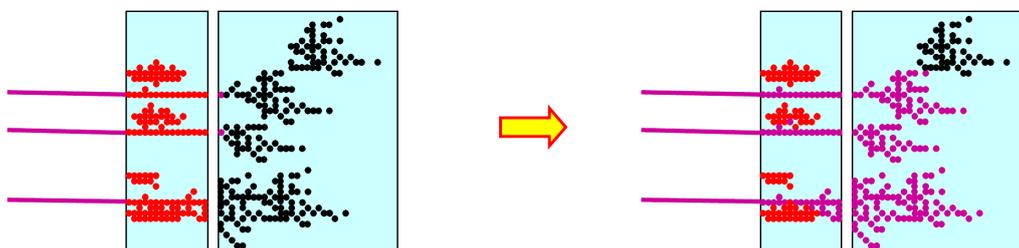
Figure 3.1: Concepts of particle reconstruction for traditional methods like TauRec (left) and pflow-based methods like PanTau (right) [19].
Charged mesons (pink) pass through the electromagnetic calorimeter (respective left blue box) and start hadronic showers in the hadronic calorimeter (respective right blue box). The traditional identification algorithm distinguishes between energies measured in the electromagnetic calorimeter (red dots) and those taken from the hadronic calorimeter (black dots). The pink dots in the hadronic calorimeter were matched to the pink tracks and help associating the black dots with the mesons' energy. The pflow algorithm on the other hand aims to distinguish the energy originating from the meson (pink dots) from the neutral energy induced by γ's (red dots) or neutral hadrons (black dots).

In ATLAS, the concept of particle flow is titled energy flow (or ***eflow***). Accordingly, a pflow object is called an eflow object (***EFO***).

---

[1] Fig. 3.1 is actually taken from a talk about pflow-based particle identification in the International Linear Collider project rather than ATLAS.

Currently, no eflow algorithm is being used by default for particle reconstruction in ATLAS. In 2007, M. Hodgkinson, D. Tovey and R. Duxfield developed **eflowRec** (short for eflow reconstruction), a program to find and store EFOs. Due to CPU constraints, it only runs on certain parts of each event (compare Ch. 3.1.3) [13, 14].

3.1.3 TauRec, EflowRec and PanTau

PanTau was originally developed as a new τ lepton identifier based on EFOs by S. Fleischmann, R. Prabhu and C. Limbach [15, 16, 17]. Currently, the main goal of PanTau is to reconstruct the decay mode of τ leptons found by TauRec. Although it would take too much time for eflowRec to run on whole events, using it to analyse regions around τ lepton candidates found by TauRec is a feasible approach. In this way, PanTau can run on EFOs in the important areas of an event (near possible τ leptons identified by TauRec) and examine these with respect to decay modes – and, in theory, also further fake suppression.

The important reconstruction steps of TauRec and eflowRec are described the following sections, while the PanTau algorithm is explained in more detail in Ch. 3.2.

TauRec

TauRec and eflowRec both make use of the `TopoCluster` algorithm, which is a program arranging calorimeter cells in small sets (clusters) that are likely to originate from the same particle or jet. To do so, it searches for cells with an absorbed absolute energy of more than four sigma above the expected value due to electromagnetic and pileup noise. Neighbouring cells with an absolute energy of more than two sigma above total noise expectation are recursively added to the cluster. Finally, all cells adjacent to the current outer border of the cluster are added as well, in case they have a finite energy value (in principle it is possible to have a sigma-relative threshold here also, but by default this is set to 0). If there is more than one local maximum of absolute energy in the cluster, it is split. [20]

While eflowRec uses the TopoClusters directly, TauRec takes advantage of an additional algorithm's output: The `TopoJet` finder groups TopoClusters together by searching 0.4-cones around clusters with energies of more than 1 GeV. All clusters in the cone are associated with the TopoJet. TopoJets with $E_T > 10$ GeV and $|\eta| < 2.5$ are used as seeds for tau candidates by TauRec. These are matched in an 0.3-cone to tracks which pass certain quality requirements (namely $p_T > 1$ GeV, $d_0 < 1.5$ mm and a minimum number of detector hits). [11, 12]

EflowRec

As stated above, this chapter will merely give a brief description of the elfowRec algorithm. A more thorough report is given in the referenced documents [13, 14].

The input for eflowRec are tracks and clusters in the vicinity of those TauRec tau candidates with a number of reconstructed tracks within a predefined range (currently $n_{\text{Tracks}} \in [1, 5]$).

Each track is extrapolated to the EM calorimeter's second layer, where it is matched to the TopoCluster closest to it in η and ϕ.[2] If the energy deposit E_{calo} in the cluster exceeds the

---

[2] In fact, the clusters' sizes in η and ϕ are also considered in the algorithm, effectively accounting for the standard

expected energy E_{exp} considering the track, minus its standard deviation σ_{exp}, weighted by an empirical factor k_2, this cluster is believed to contain the full energy corresponding to the track (Eq. 3.1). The TopoCluster is then removed, and the energy measurement by means of the calorimeter entries is replaced by the one from the tracker.

$$E_{\text{calo}} > E_{\text{exp}} - k_2 \sigma_{\text{exp}} \tag{3.1}$$

The expected energy E_{exp} is calculated by multiplying the track energy E_{track} with the average ratio of deposited energy E_{calo} to particle momentum p_{track} (Eq. 3.2). This ratio is obtained from single pions in MC and stored in reference tables, binned in p_{T} and η.

$$E_{\text{exp}} = E_{\text{track}} \frac{E_{\text{calo}}}{p_{\text{track}}} \tag{3.2}$$

If the condition in Eq. 3.1 is not satisfied, the charged particle's energy appears to have split between several clusters. EflowRec then considers calorimeter cells in an 0.2-cone around the track and removes them cell by cell, until the total subtracted energy equals the expected energy deposit E_{exp}.

The order in which cells are deleted is optimised for shower shapes in the respective p_{T} and η bin. Like in the first case, the calorimeter energy measurement is replaced by the tracker's.

The combination of a track and its matched TopoCluster is stored as a charged EFO and the remaining TopoClusters as neutral EFOs.

EflowRec adds an EM or HAD flag to neutral EFOs, which is set by the Local Hadron Calibration tools, depending on whether the largest amount of energy is deposited in the EM or HAD calorimeter. Currently, algorithms are being developed to distinguish π^0's from other neutral EFOs by means of Multivariate Analysis (*MVA*) within eflowRec. The outcome of this MVA is also stored with each neutral EFO.

All EFOs are stored in a container, which is the information PanTau uses to discriminate decay channels and (if desired) for fake suppression. Naturally, this structure implies that PanTau can only find τ leptons which were considered as tau candidates by TauRec in the first place.

3.2 PanTau Algorithm

The idea of PanTau is to run on EFOs around TauRec tau candidates. The sets of EFOs associated to a TauRec tau candidate are called seeds, and for each of those seeds a set of decay mode discriminating variables is calculated. These PanTau-specific variables are also called PanTau *features*.

---

deviations σ of the clusters' coordinates. Instead of choosing the cluster with the lowest ΔR to the track's extrapolation, the following term is minimised to find the matching cluster: [14]

$$\sqrt{\frac{(\eta_{\text{Track}} - \eta_{\text{Cluster}})^2}{\sigma_\eta^2} + \frac{(\phi_{\text{Track}} - \phi_{\text{Cluster}})^2}{\sigma_\phi^2}}.$$

The procedure is divided into four separate algorithms, which will be explained in the following sections. [15, 16, 17, 18]

3.2.1 Seed Building and Feature Extraction

The seed building takes place in the PanTau algorithm `TauSeedBuilderAlg`, which loops over all TauRec candidates with $n_{\text{Tracks}} \in [1, 5]$ and collects the EFOs within an 0.4-cone. These are associated with the tau candidate. All tau candidates with at least one associated charged EFO serve as seed for PanTau and are then 'PanTau tau candidates '.

For all seeds, each EFO's momentum is corrected by setting the EFO's mass to the pion mass (139.57 MeV for $\pi^{\pm}$ and 134.98 MeV for π^0). There are six different type flags which can be saved for each EFO, given it fulfils the respective conditions (see Table 3.1, for the conditions cp. Ch. 3.1.3: EflowRec). Every EFO has one of the two basic flags 'Neutral' or 'Charged', while the other flags are further specifications for one of the two basic types.

Flag	Conditions
Charged	nonzero charge
Neutral	zero charge
QualifiedTrack	Charged, track passing track selection[3]
EmNeutral	Neutral, tagged EM, $E_{\text{T}} > 1\,\text{GeV}$, $E > 1\,\text{GeV}$
HadNeutral	Neutral, tagged HAD, $E_{\text{T}} > 1\,\text{GeV}$, $E > 1\,\text{GeV}$
Pi0	Neutral, π^0 MVA result $> x$ (x is adjustable)

Table 3.1: Type flags assigned to EFOs by PanTau.
One of 'Neutral' or 'Charged' is always set, the other flags are optional. The requirements for E_{T} in EmNeutral and HadNeutral are redundant, because $E > 1\,\text{GeV}$ already implies $E_{\text{T}} > 1\,\text{GeV}$. The given expressions are still stated, to reflect the current implementation in the code.

Note that the currently implemented conditions on TauRec tracks are stronger than the PanTau conditions for a 'QualifiedTrack', which makes every charged EFO a charged qualified EFO. This is, however, subject to change.

EFOs within $\Delta R < 0.2$ around the tau candidate ('core region') are used to calculate the candidate's 4-momentum, while all EFOs within $\Delta R < 0.4$ ('sum region') are used for the calculation of PanTau features by the algorithm `JetFeatureExtractionTool`.

The `TauSeedBuilderAlg` then classifies the seeds as 'good' or 'bad'. Bad seeds are those with invalid TauRec information (like a null pointer to the tau candidate or n_{Tracks} outside the allowed range). An additional seed preselection with cuts on E_{T}, η, charge, and/or number of EFOs can also be applied to optimise this classification.

---

[3] These are the conditions for the track selection: $p_{\text{T}} \geq 1\,\text{GeV}$, $d_0 \leq 2\,\text{mm}$, $z_0 \sin(\theta) \leq 10\,\text{mm}$, $z_0 \leq 9999\,\text{mm}$, and at least 2 hits in the pixel detector and 7 in the SCT.

3.2.2 Tau Building

The algorithm `TauJetBuilderAlg` runs on each PanTau seed. For the bad seeds, it adds the variable `isPanTauCandidate=0` to the candidate, to indicate they failed the selection and are thus no valid PanTau candidates. For the good seeds, the five `TauDiscriminantTools` are called to get the likeliness for each decay mode (cp. Ch. 3.2.3).

One `TauDiscriminantTool` exists for every decay mode to be tested, which are currently 1p0n, 1p1n, 1pXn, 3p0n and 3pXn (cp. Table 1.1). Although in principle it is possible to discriminate any set of decay modes against any second set, by default each single of the above modes is discriminated against the rest.[4]

The likeliness values form a five-dimensional vector (in 'decay mode space'). This vector is normalised to 1, and its angle to each axis is computed to quantify how likely the corresponding decay mode is (the smaller the angle, the more probable the mode).

The mode whose axis spans the smallest angle to the candidate's vector is then assigned as the *loose* decay mode to the tau candidate. If this angle is below 45°, it is also assigned as *medium* mode[5], and for an angle smaller than 15° as *tight*. Hence, the loose flag has the lowest purity and highest efficiency, and vice versa for the tight flag. Note that this way, one particular tau candidate can never have different modes assigned for different purity flags. A loose mode is always assigned, while medium and tight can be set to 'none of the above'. Trivially, if the tight flag is set to one of the decay modes, the loose and medium flag are set as well (for the same mode).

3.2.3 Likeliness Calculation

To find each tau candidate's likeliness for each decay mode, the `TauDiscriminantTool` uses the root-based MVA program TMVA. Currently, the default MVA algorithm is BDT (Boosted Decision Trees).

BDT values are typically used to separate one signal against one background by placing a cut. They are not a feasible measure for comparisons of different decay modes, because the values can in principle be shifted with respect to each other. One and the same BDT cut can thus resolve in different efficiencies and purities for different modes, and will most likely not be optimal for more than one of them.

Instead, a reference table is created for each decay mode to link BDT values to the corresponding likeliness. In order to achieve this, a signal and background BDT histogram are created from MC for each decay mode. The signal sample contains only τ leptons of that mode, and the background sample all other ones, summed in their natural respective ratios,

---

[4] In principle, it is possible to also include 'fake' as a decay mode in this sense and use a `TauDiscriminant-Tool` to implement a fake suppression at this stage of PanTau. The main reason to separate these steps is that the fake suppression, if conducted at a later stage, can benefit from the information gained here, namely the classification in decay modes. The MVA to discriminate fakes from true τ leptons can thus be trained separately for performance on each of the above modes.

[5] The maximum angle of 45° was chosen for the medium flag because, in any number of dimensions, this criterion can never apply to more than one axis. The medium (and tight) flag therefore never applies to more than one decay mode, even without the additional criterion to take the axis with the lowest angle, which is necessary for the loose flag [18].

i. e. taking into account their branching fractions. Scanning through the BDT bins i, the signal efficiency e^i_{sig} is divided by the background efficiency e^i_{bkg} to get the likeliness l^i (Eq. 3.3a). The efficiencies are defined as the fraction of signal or background entries $B_{\text{sig/bkg}}$, respectively, above the current BDT value B^i (Eq. 3.3b, where n_{bins} is the number of bins in the histogram).

$$l^i = \frac{e^i_{\text{sig}}}{e^i_{\text{bkg}}} \tag{3.3a}$$

$$e^i_{\text{sig/bkg}} = \frac{\sum\limits_{j=i+1}^{n_{\text{bins}}} B^j_{\text{sig/bkg}}}{\sum\limits_{j=1}^{n_{\text{bins}}} B^j_{\text{sig/bkg}}} \tag{3.3b}$$

The mapping of BDT values B to the corresponding likeliness l is saved as a reference table for that mode. The value l, being a fraction of efficiencies, is comparable among different modes.

3.3 Calculating PanTau Features

The PanTau algorithm `JetFeatureExtractionTool` calculates the features for each tau candidate. There are two kinds of storage in ATHENA: Information that is stored in the 'persistent' part of the storage will be written to the AOD after execution of PanTau, whereas anything stored in the 'transient' part is lost by default. However, PanTau can be configured to create a D3PD during its execution, in which the transient information is saved.

For good seeds, the `JetFeatureExtractionTool` writes all features into the transient part of the storage, and all features that are used by the `TauDiscriminantTool` into the persistent part. For bad seeds, only the variable `isPanTauCandidate` is stored, in both the transient and the persistent storage.

Since there exist a variety of 121 PanTau features, it is impossible to present them all in suitable detail within the scope of this document.[6] The variables chosen to be described in this chapter have good separation power with respect to the different decay modes and are also subject matter of the studies in Ch. 5.

A more explicit list of PanTau features and their definitions is available online at the corresponding ATLAS TWiki page [21], and the algorithm itself can be enquired for more details [22].

One of the most intuitive PanTau features is NEFO, which is the number of EFOs in the tau candidate (cp. Ch. 3.2.1). There are a few other features that also correspond to numbers of EFOs, but only count a particular type of these, based on the classifications introduced in Ch. 3.2.1 (Table 3.1). These are NNeutralEFO (number of EFOs flagged 'Neutral'), NCharged-EFO (number of 'Charged' EFOs) and NChargedQualifiedEFO (EFOs with 'QualifiedTrack' flag),[7] NEMNeutralEFO (EFOs with 'EmNeutral' flag) and NHADNeutralEFO (EFOs with

---

[6] Recently, the design of the `JetFeatureExtractionTool` was changed, and the number of variables increased to over 600. These are, however, similar in structure and purpose to the ones described within this document.

[7] Note that, currently, NChargedQualifiedEFO is identical to NChargedEFO by construction (cp. Ch. 3.2.1).

'HadNeutral' flag).[8]

Neutral EFOs are often also referred to as EFO^0, charged EFOs as $EFO^\pm$ and charged quali-fied EFOs as $EFO^{*\pm}$.

Several variables in PanTau are fractions of E_T. A sub-category of these are features named EflowMeanEtXToSumEt, where X is either Chrg, Neut, HADNeut or All. The intuitive formula to calculate them is given in Eq. 3.4, where the mean $< E_T >_X$ only takes into account EFOs of type X.

$$\frac{< E_T >_X}{\sum E_T} \tag{3.4}$$

These features are a main focus in Ch. 5.3.

Since the purpose of the variables in PanTau is to separate the different decay modes against each other, and possibly to separate true from fake τ leptons, the separation power has to be evaluated, which is an ongoing process conducted by C. Limbach [23, 24, 25, 26, 27].

Some variables showing a good separation for at least one of the decay modes are given in Fig. 3.2. Note that the nomenclature for the modes was different at the time these graphs were produced: '1p' corresponds to 1p0n defined above, '1pn' to 1p1n and 1pXn combined, '3p' to 3p0n and '3pn' to 3pXn. In 'tau vs fake', all these modes are separated against fake tau candidates.

The variable EflowNeutralSumEt (Fig. 3.2a) describes the summed E_T of all neutral EFOs. It serves well to separate 1p0n against the other modes, since due to the τ lepton's branching fractions (cp. Table 1.1), the main background there is 1p1n and 1pXn. Since there are no neutral pions in 1p0n, the signal obtains far lower values for EflowNeutralSumEt than the rest of the modes. Only 3p0n is accountable for the non-negligible number of background entries at low values.

EflowChargedSumEt (Fig. 3.2b) is defined analogously, i. e. as the sum of the charged EFOs' E_T. Since for 3p0n, the main background is 1p0n, 1p1n and 1pXn (cp. the higher branching fractions for 1-prong decays), the magnitude of the charged EFOs' E_T is obviously a good measure for the separation. The 3p0n τ leptons have on average higher values than the 1-prong ones – and even than 3pXn, since less energy is left for the three charged pions if neutral pions are emitted as well.

For the same reasons as just stated, the 1-prong modes are also the main background for 3pXn. One variable with reasonable separation power is SumCharge (Fig. 3.2c), the sum of all EFOs' charges, which can only have ±1 as true values. However, it can become any other integer value if the number of tracks is not reconstructed correctly. In the 3-prong cases, the probability of counting too many or too few tracks is higher than for 1-prong, because there are three times as many tracks that could in principle get lost. If one of the tracks is lost, SumCharge becomes 0 or ±2, depending on which of the pions is the lost one. These values have more entries for 3pXn than for 1-prong. The same holds, for the same reasons, also for 3p0n, as can be seen in the respective reference [25]. This is not a major problem, due to the lower branching fraction, but is one of the factors reducing the separation power.

EflowStdDevEt is a variable suitable to separate true τ leptons from fakes (Fig. 3.2d). It

---

[8] There is not yet a variable corresponding to the 'Pi0' flag, which was added only recently.
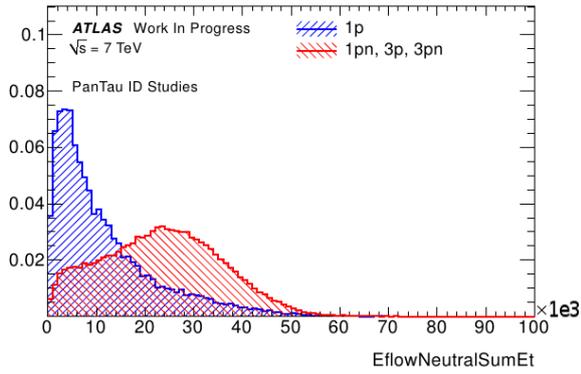
(a) EflowNeutralSumEt, 1p vs rest [23].
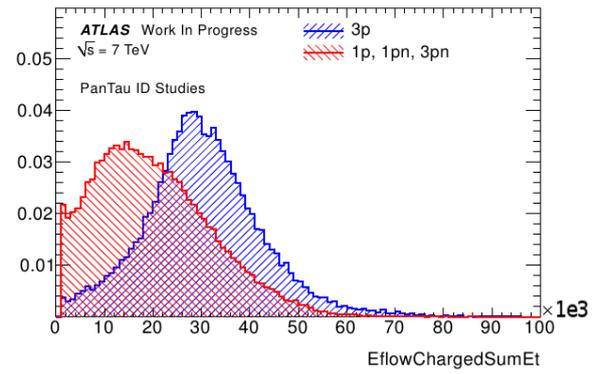
(b) EflowChargedSumEt, 3p vs rest [25].
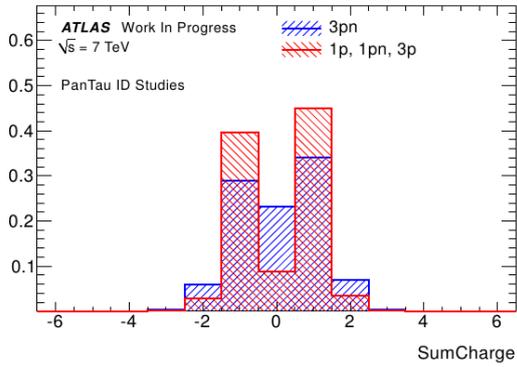
(c) SumCharge, 3pn vs rest [26].
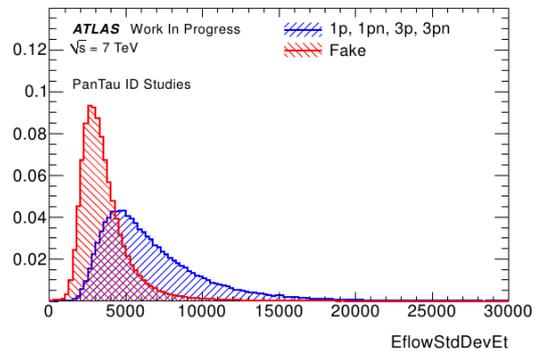
(d) EflowStdDevEt, tau vs fake [27].

Figure 3.2: Separation power of different PanTau features for different decay modes.

calculates the standard deviation $\sigma(E_{\mathrm{T}})$ of all EFOs' transverse energies (Eq. 3.5, where N is the number of EFOs).

$$\sigma(E_{\mathrm{T}}) = \sqrt{\frac{1}{N} \sum_i \left(E_{\mathrm{T}}^i - < E_{\mathrm{T}} >\right)^2} \qquad (3.5)$$

The standard deviation is smaller for fakes than for true τ leptons. One of the reasons for this effect could be the higher multiplicity in jets, which results in a lower average for E_{T} and thus lower absolute deviations from the average. Furthermore, QCD jets have relatively high branching fractions for emitting low E_{T} gluons, which reduces the value of $\sigma(E_{\mathrm{T}})$.

Chapter 4

Consistency Checks

PanTau is a complex algorithm, and before it can be used by physics analysts, it has to be made sure that it works as expected and gives appropriate results. These consistency checks are the scope of this chapter.

Both of the described tests use Pythia MC samples for the processes $Z \rightarrow \tau\tau$, $W \rightarrow \tau\nu$, $Z'_{\mathrm{SSM\,250}} \rightarrow \tau\tau$, $Z'_{\mathrm{SSM\,500}} \rightarrow \tau\tau$ and several jet processes (Pythia J0-J5). The exact sample names are stated in Appendix A (Table A.3).

4.1 Container Storage

As mentioned before, PanTau uses EFOs as seeds that eflowRec finds in the vicinity of TauRec tau candidates. TauRec writes information for each of its tau candidates into a 'DetailsContainer'. The names of the variables inside this container are preceded by the prefix `TauDetails`.

The variables created by PanTau have the prefix `PanTauDetails` and by default are only stored transiently, i. e. not written into the AOD. However, PanTau generates a second, persistent container, in which the variables have the prefix `TauPanTauDetails`, and which is linked in the original DetailsContainer TauRec created (see Fig. 4.1).
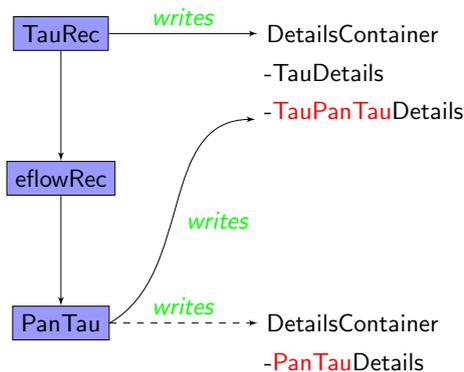
Figure 4.1: Details Containers from TauRec and PanTau

Trying to confirm PanTau's consistency, the first step was to cross-check if the internal PanTau DetailsContainer (`PanTauDetails`) and the transient one, linked inside the TauRec

DetailsContainer (`TauPanTauDetails`), indeed share the same information. This check was successfully made by comparing each variable's histogram for a variety of ten different MC samples (see above) with 1000 events each (for examples see Fig. 4.2).

(a) Jet Sphericity

(b) E_{T} in $0.1\Delta R$ over E_{T} in $0.4\Delta R$
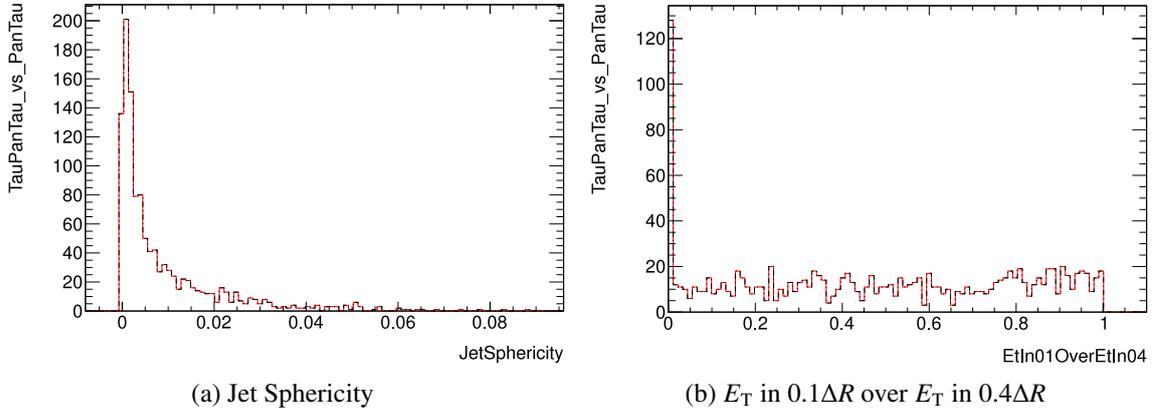
Figure 4.2: Plot of one and the same variable from `TauPanTauDetails` (red) and `PanTauDetails` (dashed black) for a $W \to \tau\nu$ sample with 1000 events.
Histograms should – and do – match exactly.

The only visual mismatches were due to technical necessary casting of integer valued variables to floating point numbers. This, combined with an imperfect binning routine, resulted in histograms not coinciding as obviously (Fig. 4.3). The reason is that only integer valued bins are filled with `PanTauDetails`, while `TauPanTauDetails` have an 'integer-friendly' binning and are thus easier to read. To prove the equality of histograms, only bins with non-zero entries were compared. This test was successfully as well for these cases.

(a) absolute value of sum of charges

(b) number of neutral EFOs
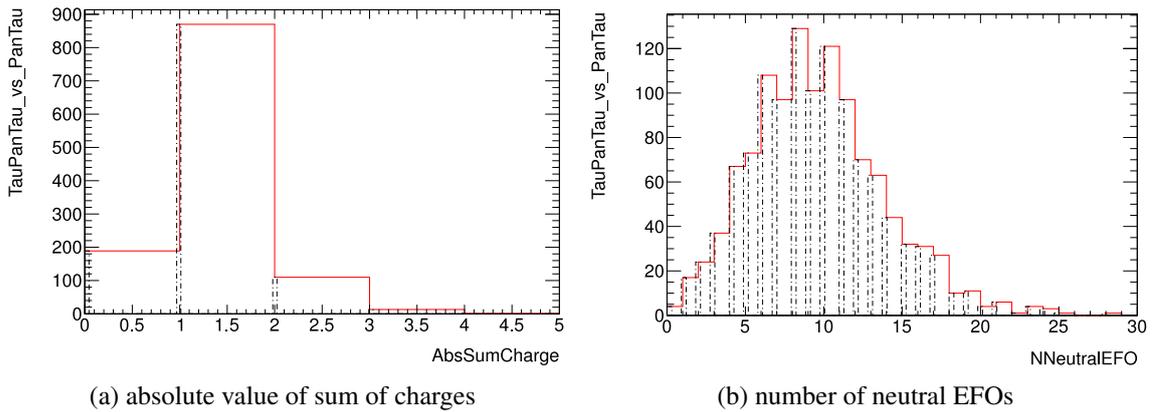
Figure 4.3: Plot of one and the same variable from `TauPanTauDetails` (red) and `PanTauDetails` (dashed black) for a $W \to \tau\nu$ sample with 1000 events.
Histograms match, disregarding the different binning.

26

4.2 Output Channel Supervision

There are three stages of information in PanTau (see Fig. 4.4): When PanTau evaluates the input, the information is first stored internally (in form of the above mentioned DetailsContainers). This can optionally be written to a D3PD (cp. Ch. 3.3), the 'internal D3PD'. Then the persistent part of the TauRec DetailsContainer (also containing the `TauPanTau` variables) is written to the AOD, from where one can access this information and manually write it to a second D3PD for validation purposes, the 'validation D3PD'. Finally, the official ATHENA D3PDMaker is the package most people use to create D3PDs from AODs, so this 'official D3PD' is the one that will most commonly be utilised for further analysis.
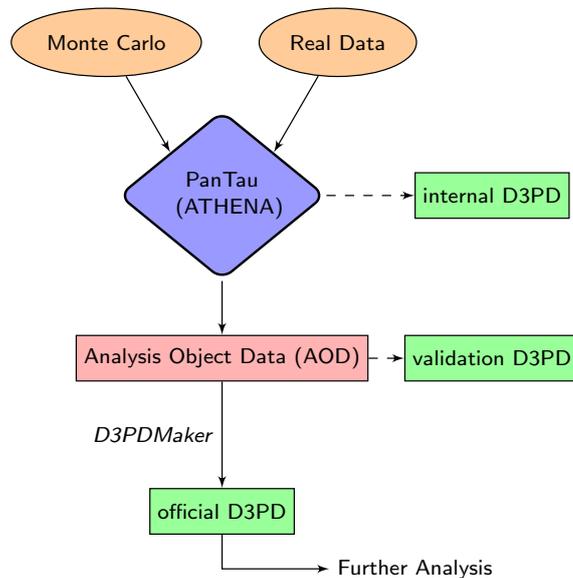
Figure 4.4: Creation of D3PDs with PanTau information

To control the PanTau output, it had to be made sure that the information in all three layers is absolutely identical. This implies a comparison of the internal, validation and official D3PDs, leading to similar kinds of plots as for the `TauPanTau`/`PanTau` DetailsContainer comparison – and unsurprising results, since the expectation for this validation was to find nothing unreckoned. Again, this was performed for ten different MC samples (see above), each containing 1000 events (for examples see Fig. 4.5).

(a) invariant mass M of all EFOs, for official (red) and internal (dashed black) D3PD for a $J0$ sample.

(b) E_T in $0.1\Delta R$ over E_T in $0.2\Delta R$ (inverse sigmoid transformed), for internal (red) and validation (dashed black) D3PD for a $J3$ sample

(c) mean angle between leading charged qualified EFO components, for official (red) and validation (dashed black) D3PD for a $Z' \to \tau\tau$ sample

(d) number of charged EFOs, for official (red) and internal (dashed black) D3PD for a $Z \to \tau\tau$ sample

Figure 4.5: Plot of some variables from two different D3PDs for different samples with 1000 events each.
Histograms evidently match exactly.

Chapter 5

Background Validation

Since usually a large fraction of analysed tau candidates are fakes (mostly QCD jets), it is essential that these jets are well-described by PanTau. It is arguably even as important as the performance on real τ leptons: If for instance 80 % of the tau candidates in a sample are fakes and 90 % of these can be identified as such, the purity of the remaining 28 % of candidates has become higher significantly. However, if real τ leptons are not described well at all, even a good understanding of fakes cannot help, e. g. if 98 % of real τ leptons are rejected due to misbehaviour of the program. Thus, the validation of PanTau's behaviour on background is only the first step and should be followed up by an analysis for true τ leptons.

5.1 Matching of Measured against Simulated Data

The outcome of data analysis has to be compared against MC simulations; it has to be checked if assumptions being made about the physics in the algorithm are in fact valid. Therefore, the data results have to be described by MC as well as possible.

However, QCD jets are impossible to calculate precisely and can thus only be modelled by means of approximation algorithms, whose optimisation is still a subject of current research. Imprecisions therefore arise whenever QCD plays a great role. Since the scope of this study are solely QCD jets, discrepancies between data and MC are in principle unavoidable. The aim is rather to find the weakest spots, and to estimate the quality of the agreement in general in order to judge if the differences are explicable by QCD approximations alone.

Should a considerable part of PanTau features indeed be described poorly by MC, a possible work-around would be to train the background BDTs on real data rather than on MC. The huge disadvantage in that case would be that it is virtually impossible to control exactly which processes a data sample consists of. Training on MC, if possible, is therefore more trustworthy, which makes for another reason that a good accordance of data and MC (within the aforementioned limitations) is desirable.

5.1.1 Choosing the Physics Process for the Analysis

The background analysis can in principle be performed on any kind of process without large fractions of events including τ leptons. One aim is to find a process which produces 'τ-like' objects, i. e. such objects that could, in a physics analysis, easily be mistaken for actual τ leptons. This is mostly the case with QCD jets. Another intention is to have an easily accessible

process, i. e. one that is straightforward to select by applying the right cuts (compare Ch. 5.1.2). Furthermore, high data statistics due to large cross sections are a desirable feature.

The process that was chosen as basis for this study is a Z boson decaying into two muons while recoiling against a QCD jet. The quantum processes leading to such an event are shown as Feynman diagrams in Fig. 5.1.

(a) t-channel with gluon jet

(b) t-channel with quark jet

(c) s-channel with quark jet
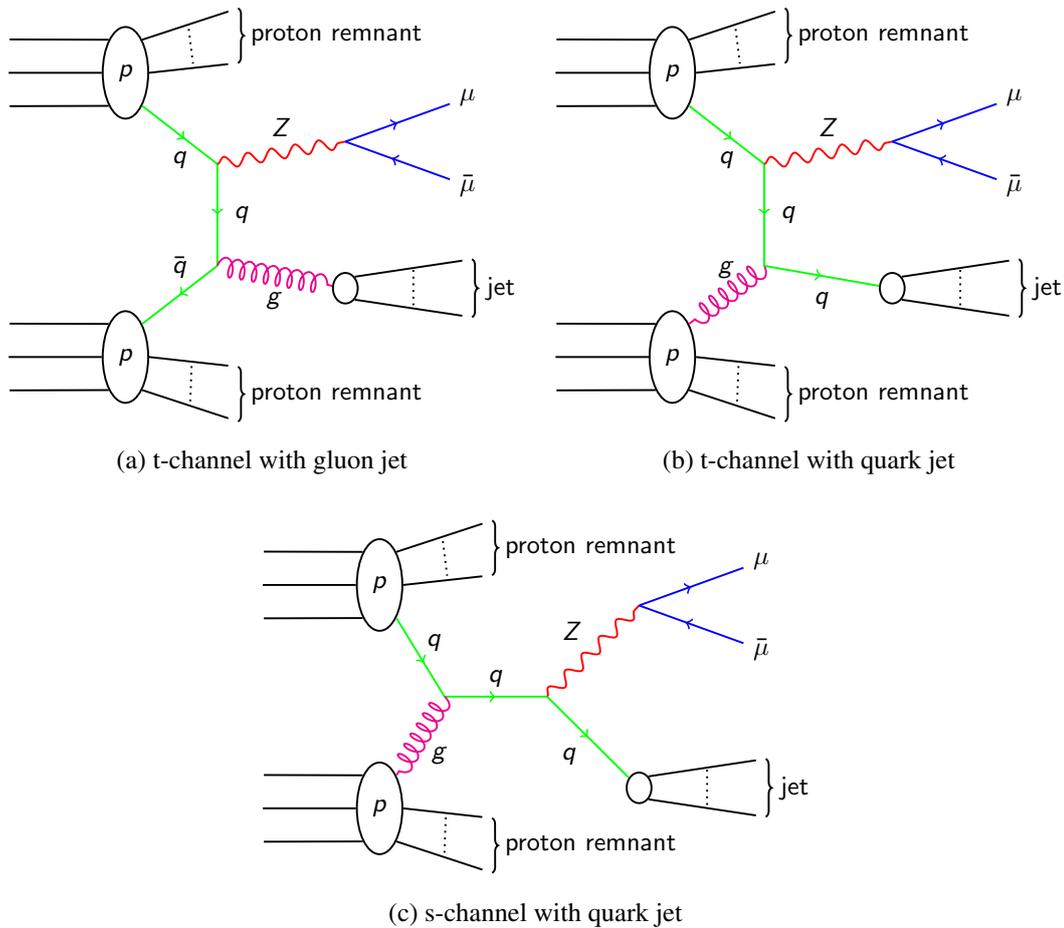
Figure 5.1: Feynman diagrams of $Z \to \mu\mu$ decays involving a recoil jet.
Five more can be produced by reverting the time direction of the (anti-)quark paths and/or swapping the two protons. Note that for the t-channel with gluon jet (Fig. 5.1a), these changes are equivalent. In the s-channel, no gluon jet can be produced.

In the proton-proton collision, a quark q and an antiquark $\bar{q}$ interact to produce a Z boson, which in turn decays into a muon μ and an antimuon $\bar{\mu}$. One of the (anti-)quarks emits a gluon g (initial state radiation) before the Z boson is formed. This gluon in turn produces a QCD jet, which is likely to be reconstructed as a tau candidate (if certain requirements are met, e. g. for it not to have too many tracks). The Feynman diagram for this part of the process is depicted in Fig. 5.1a. When the paths of the initial quark and the gluon it emits are switched in the graph, another part evolves, resulting in a quark producing the QCD jet rather than the gluon (Fig. 5.1b). Although physically independent processes, there is no way of controlling if the

interacting partons from the protons are two quarks or a quark and a gluon. Since the aim of this study is the analysis of QCD jets regardless their nature (quarkonic or gluonic), this difference will be neglected and the processes are analysed by means of their combination.

Trivially, the paths of the (anti-)quarks can be reversed in time in both of the Feynman graphs, resulting in two more first order contributions to the process. In Fig. 5.1b, yet another contribution is obtained by swapping the two protons, while in Fig. 5.1a this procedure is equivalent to reverting the quark edges.

Finally, three more contributions come from the s-channel of the quark-gluon-process and its alterations with reversed quark direction and/or swapped protons (Fig. 5.1c).

Another Feynman diagram that might play a role consists of two gluons (from the protons), interacting by means of a quark loop (box diagram) to form the Z and a gluon, which emerges as a gluonic jet. In spite of being a higher order diagram, this might still have a significant impact, because the probability is higher for gluons to be the interacting partons than for quarks.

Assuming negligible p_T of the colliding partons (cp. Ch. 2.3.3), the Z boson's and the QCD jet's transverse momenta have to add up to zero. The jet is then called a 'recoil jet' of the Z, because it recoils against the boson. If the Z boson then decays into two muons, it is relatively easy to reconstruct, since muons are the only particles detected in the muon chambers, and the QCD jet has to be opposite of their combined direction in the xy-plane (Fig. 5.2).
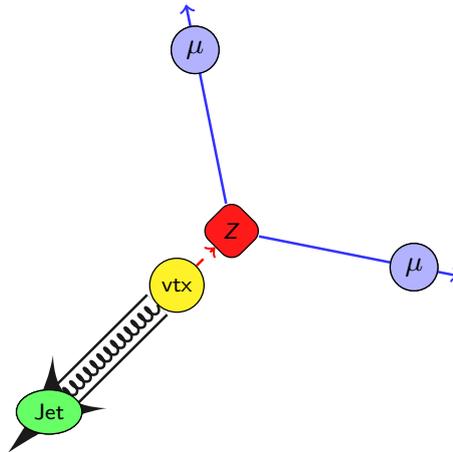
Figure 5.2: $Z \rightarrow \mu\mu$ process with a recoil jet, projected onto the xy-plane. At the vertex (yellow) a QCD jet (black/green) and a Z boson (red) are formed. The Z boson continues to decay into two muons (blue).

The $Z \rightarrow \mu\mu$ process has a relatively high cross section and is easily selectable by requiring two muons with an invariant mass $M_{\mu\mu}$ near the Z-mass. Identification of muons is far less complicated than that of τ leptons, since they can be recognised in the muon chamber due to their relatively long lifetime of 2.2×10^{-6}s [2].

In the chosen process, most of the recoil jets are gluonic, i. e. produced by a gluon rather than a quark. While in this study, the jets' origins are not differentiated, it might be preferable in later studies to distinguish them in order to gain performance by taking advantage of their different shower shapes and multiplicities.

5.1.2 Selecting Events and Tau Candidates

To filter out the desired events from the magnitude of recorded collisions at ATLAS, a selection of several cuts is needed. These cuts are applied to real data as well as to MC in order to achieve a fair comparison between both. In the following, cuts on events will be referred to as ***event cuts*** and cuts on tau candidates as ***tau cuts***. The first few event cuts, being necessary regardless the aim of the analysis, are called ***cleaning cuts***.

The used sequence of cuts (***cut flow***) is based on one stated in an article by the Muon WG (ATLAS Muon Working Group) [28]. Some of their cuts had to be adapted to be suitable for 2011 data (as opposed to 2010 data in the original paper), and a few additional cuts were added concerning tau candidates. The resulting cut flow is presented in the following paragraphs and – in short – in Table 5.1. The corresponding histograms can be found in Fig. 5.3 (lumi-weighted, cp. Ch. 5.1.6) and Fig. 5.4 (event cuts, in percentage of the number of events before cuts).

All corrections and weightings that are explained in the chapters to come (muon corrections, luminosity and pileup weights, global MC scale) are already implied in all histograms, unless stated otherwise. This way, conclusions can be drawn from any graphs without worrying about effects that are explained in subsequent parts of the document.

Used Samples

The used MC samples were produced by the Alpgen generator. For processes of type $Z \rightarrow \mu\mu +$ Jets, these are divided into Np0, Np1, ..., Np5. The respective number indicates how many jets emerge in the main process ('main jets'). Nevertheless, in each of these samples more QCD jets can be evident, which come from the underlying event and/or pileup collisions. These usually carry far lower energies (cp. Ch. 2.2.5) and can thus be easily suppressed by cutting on a minimum p_T (cp. Fig. 5.8). The samples that were generated for this study also require the main jet(s) to have a p_T of 20 GeV or more.

The data samples in this study are from data periods M6 to M10 (without M7, for which no eflowRec containers were available) of 2011. The collision energy is $7\,TeV$, and the integrated luminosity of the used runs totals $332.02\,\mathrm{pb}^{-1}$.[1]

All used samples are listed in Appendix A (MC samples in Table A.4, data samples in Table A.5). The D3PDs were produced by means of ATHENA 17.2.3.7.4 with development versions of PanTau and the TauD3PDMaker-01-08-15.

Cleaning Cuts

The cleaning cuts' purpose is to get a reliable set of events for the analysis.

Only runs mentioned in the Good Run List (***GRLs***) can be used for analysis. These are sets of usable runs for each period of data taking. Runs not included in these lists should not be

---

[1] Note that no run from period M9 is included in the good run list (the good run list is defined within this chapter), and the runs from M6 only account for $28.26\,\mathrm{pb}^{-1}$. Thus, the used data effectively consists of periods M8 and M10 (cp. Table A.5).

Cleaning Cuts
Good Run List
Trigger EF_mu18_MG_medium
Jet Cleaning

Cuts on Events
exactly 2 μ's with opposing charge
both μ's: $p_T^{\text{muon}} > 20\,\text{GeV}$
both μ's: $|\eta^{\text{muon}}| < 2.4$
at least 1 τ_{cand}: $p_T > 10\,\text{GeV}$
μ's: medium quality
μ's: combined
$S(d_0) < 3$
$z_0 < 10\,\text{mm}$
both μ's: $p_T^{\text{muon}}(\Delta R < 0.2) < 0.1 p_T$
$M_{\mu\mu} = (91 \pm 5)$ GeV
$\angle_\phi(\mu_1, \mu_2) \leq 2.8$

Cuts on τ_{cand}'s
$p_T > 15\,\text{GeV}$
opposite μ's within $D\phi = 0.3$
$|\eta| < 2.4$
$|\eta| \notin [1.37, 1.52]$

Table 5.1: Event and tau cut flow.
Event cuts regard properties of events and are thus applied once per event (note that cleaning cuts, per definition, are event cuts as well). Tau cuts apply to each tau candidate in every event.
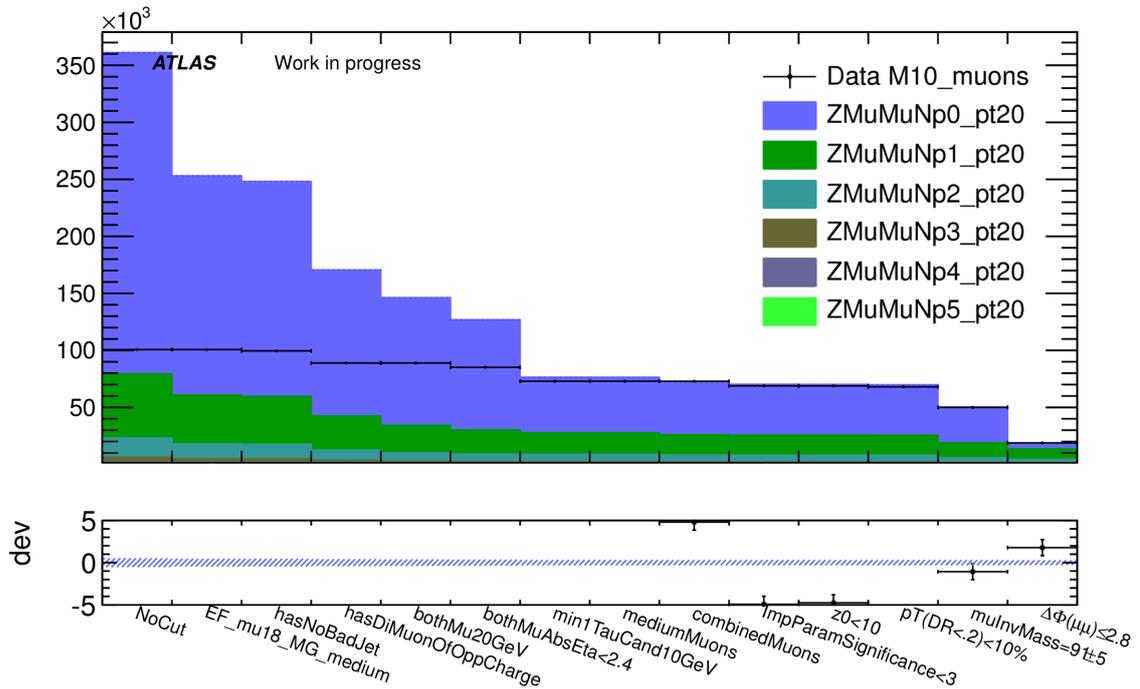
(a) Number of events after each cut. Note that data is skimmed (cp. Ch. 5.1.4).
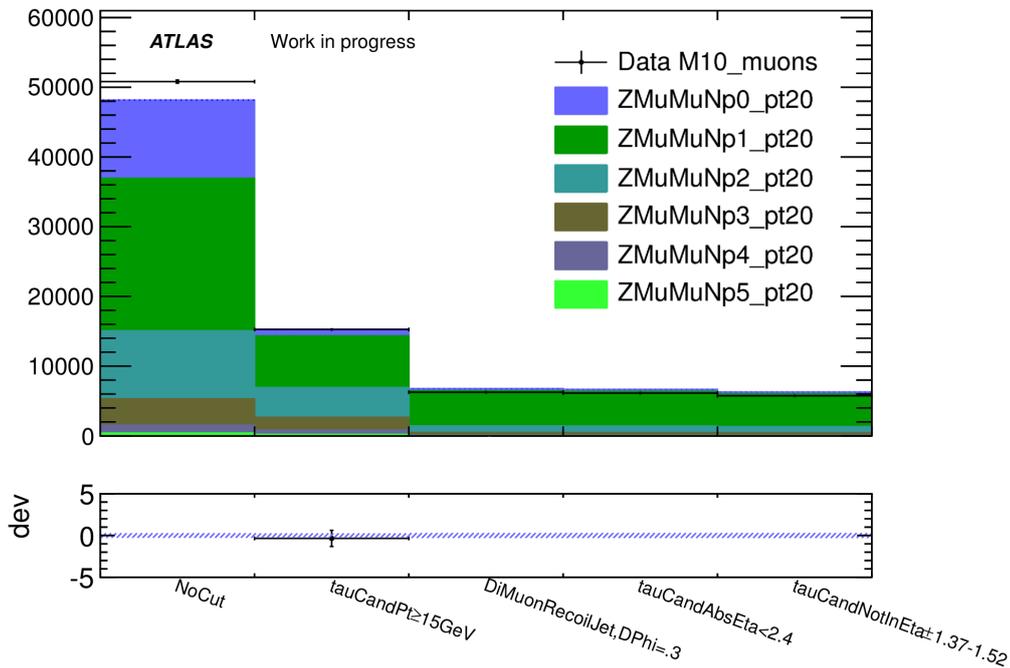
(b) Number of tau candidates after each cut. All event cuts are already applied before 'NoCut'.

Figure 5.3: Cut flow histograms (curves lumi-weighted, cp. Ch. 5.1.6).
Data and MC match almost perfectly after the event cut on $M_{\mu\mu}$ (by construction, see Ch. 5.1.8). After the tau cut requiring a recoil jet (i.e. for the tau candidate to be opposite the Z), there are more candidates in MC than in data.
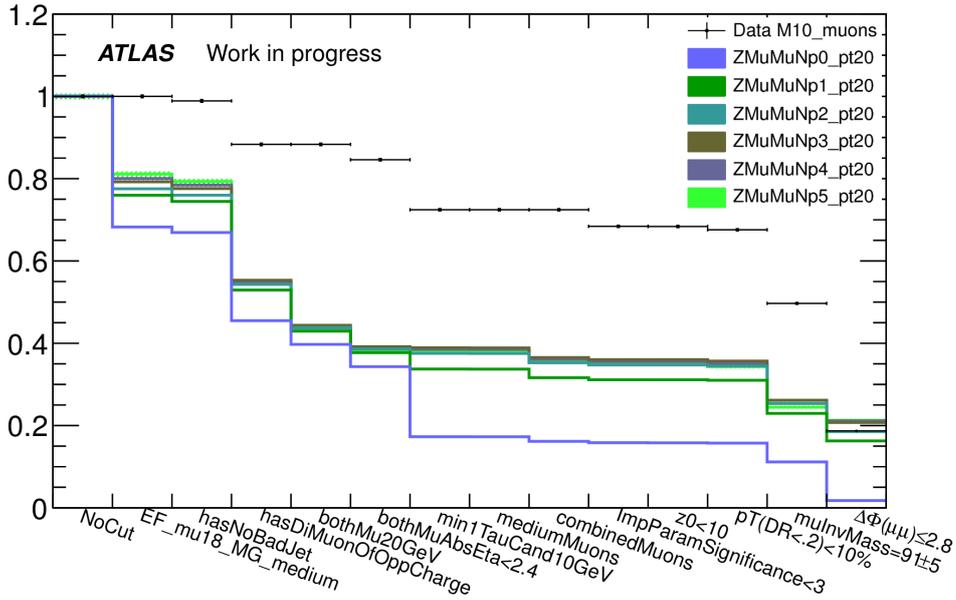
Figure 5.4: Event cut flow (numbers relative to first bin). Note that data is skimmed (cp. Ch. 5.1.4).

used for analyses, as they might contain physically wrong information. According flags are set after run time according to the present conditions of the detector.

The trigger ('EF_mu18_MG_medium') filters all events without muons having at least a momentum of 18 GeV. The Muon WG used triggers requiring 10 GeV and 13 GeV, depending on the respective data taking period of 2010, but for 2011 the 18 GeV trigger was the lowest working trigger without a prescale (with exception of one single run, having a prescale of 1.015). The subsequent ***jet cleaning*** removes every event from the selection which contains any 'bad jets'.

Event Cuts

For obvious reasons, the first physics cut requires two muons of opposing charge in the event (sometimes also referred to as a ***di-muon***), and the first condition involving tau candidates is to have at least one of reasonable transverse momentum (here: $p_T > 10$ GeV) in the event. From Fig. 5.4, it is obvious that the Np0 sample loses about half its events in the latter cut, which is expected because it is the sample with the lowest p_T jets.

Furthermore, both muons should fulfil certain quality conditions such that their reconstruction by means of the muon identification algorithm (in this case 'STACO') can be trusted. These requirements are a minimum transverse momentum (here: $p_T^{\text{muon}} > 20$ GeV), a maximum absolute η-angle ($|\eta^{\text{muon}}| < 2.4$), the STACO tag ***medium*** (which in turn requires a minimum of detector hits), and a positive ***combined*** flag (meaning that the muon could be matched to a track). The Muon WG did not cut on STACO medium muons. They did, however, demand a minimum number of detector hits in each of the pixel, SCT and TRT detectors. This constraint is in turn part of the STACO medium flag. In addition, the muons have to be isolated, which is to say the transverse momentum measured within a ΔR-cone around each muon is not to

exceed a percentage of its own transverse momentum (here: $p_T^{muon}(\Delta R < 0.2) < 0.1 p_T^{muon}$).

The p_T histogram for muons is shown in Fig. 5.5. The Np0 sample peaks just below 45 GeV, which is compatible to half the mass of the Z boson the muons originate from (which is (91.1876 ± 0.0021) GeV [2]). In the other samples, this energy is shared with one or more (high p_T) jets. The more main jets are present in the sample, the broader the curve becomes, since more combinatorial possibilities arise to distribute the momentum in the event.

Evidently, only a small fraction of muons have a p_T of less than 20 GeV (partly due to the trigger already requiring at least one muon with 20 GeV for MC and 18 GeV for data), so the respective cut does not alter the distributions significantly. The only visible effect is that it slightly favours the Np0 sample, which is not a desirable feature, since the sample of interest is Np1. However, the effect is not critical due to its proportions. Another important impact of this cut is that it neutralises the different trigger thresholds for MC and data.
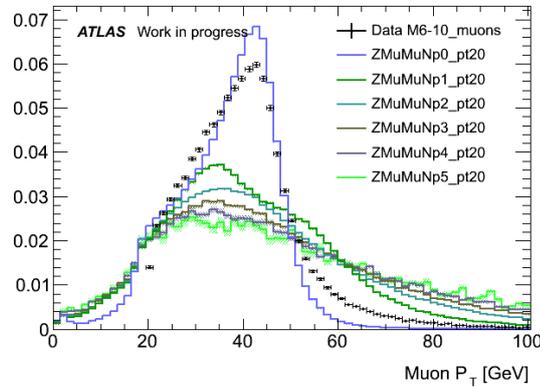
Figure 5.5: p_T^{muon} before cut on $p_T^{muon} > 20$ GeV (curves normalised to unit area).
The fraction of muons cut away is small. The sample with the least fraction in the region of $p_T^{muon} < 20$ GeV is Np0.

Two cuts are applied on the muons' impact parameters to ensure they actually originated from the primary vertex (as opposed to the majority of cosmic rays). One is for the longitudinal impact parameter z_0 not to exceed 10 mm, while the other prohibits a transverse impact parameter significance of 3 and higher. For a transverse impact parameter d_0 with uncertainty $\sigma(d_0)$, the definition of the impact parameter significance $S(d_0)$ is stated in Eq. 5.1.

$$S(d_0) = \frac{d_0}{\sigma(d_0)} \tag{5.1}$$

To assert the actual origin of the muons from a Z boson, their invariant mass has to lie in a certain range around the Z mass. The cut used here is $M_{\mu\mu} = (91 \pm 5)$ GeV. In Fig. 5.6, the invariant mass is shown before the cut. The histograms' shapes are all consistent with the sum of a Gaussian, peaking at 91 GeV, and a polynomial of degree 1 with a small intercept and a minor negative slope. Since the polynomial part can be used to estimate the data's background and to find the correct MC scale in one step, this histogram will be discussed in more detail in Ch. 5.1.8.
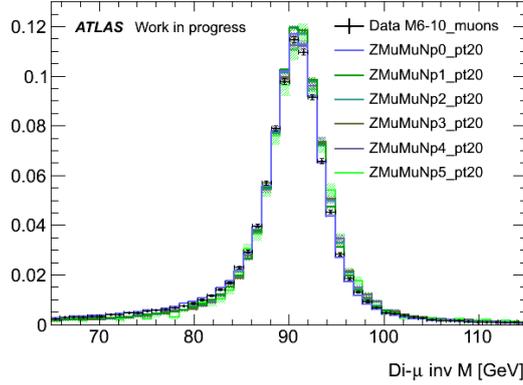
Figure 5.6: $M_{\mu\mu}$ before cut on $M_{\mu\mu} = (91 \pm 5)$ GeV (curves normalised to unit area).
All curves can be described as a sum of a Gaussian, peaking at 91 GeV, and a small background with a slightly negative slope.

In case a recoil jet occurred with a finite transverse momentum (or more than one jet, with a finite sum of transverse momenta), the Z boson must balance it with an equal p_T in the opposite direction. The two muons, emerging back-to-back in the Z boson's rest frame, therefore cannot be back-to-back in the laboratory rest frame, but include a ϕ-angle of $\angle_\phi(\mu_1, \mu_2) \neq \pi$. The greater the muon's p_T gets, the smaller is the angle $\angle_\phi(\mu_1, \mu_2)$. Thus, a cut can be applied on this included ϕ-angle to select the events with a recoil jet (here: $\angle_\phi(\mu_1, \mu_2) \leq 2.8$). An illustration is given in Fig. 5.7a, and the histogram of $\angle_\phi(\mu_1, \mu_2)$ before the cut is shown in Fig. 5.7b.
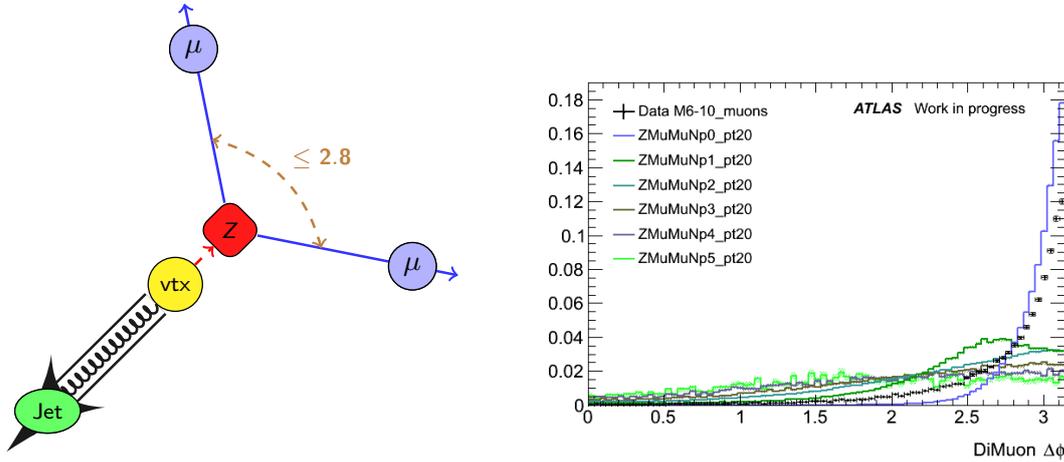
(a) Illustration of the angle $\angle_\phi(\mu_1, \mu_2)$.

(b) $\angle_\phi(\mu_1, \mu_2)$ before cut (curves normalised to unit area).

Figure 5.7: Event cut on $\angle_\phi(\mu_1, \mu_2) \leq 2.8$ (graphical illustration and variable histogram).
The Np0 sample peaks at π and will be affected by the cut to the most extent. Of the other samples, Np1 and Np2 have the largest fractions in the region above 2.8.

As expected, Np0 peaks at π due to the absence of a jet to recoil against, while in Np1 some

transverse momentum is given to the jet, which results in a smaller ϕ-angle between the muons. In Np2, smaller angles are made possible again by the second jet (if they point roughly in the same direction), as are greater angles (if they point in opposite directions). For even higher jet multiplicities, there are more and more possibilities to combine the jet vectors, which lets the angle obtain nearly arbitrary values (except that very small angles are suppressed, since the muons still emerge back-to-back in the *Z* boson's rest frame).

While the cut clearly reduces the share of the unwanted Np0 events considerably, it also cuts away about a third of Np1 events (also cp. Fig. 5.4, which substantiates these estimations). Though the decision on the exact value for the optimal cut is in the eye of the beholder, the main argument for such a low cut is that the magnitude of accessible statistics permits this sacrifice of signal for the sake of a higher background suppression.

Tau Cuts

So far, the only requirement regarding tau candidates is to have at least one with $p_T > 10\,\mathrm{GeV}$ in the whole event. The next cut is for each candidate to have $p_T > 15\,\mathrm{GeV}$. The p_T spectrum before this cut is presented in Fig. 5.8. Despite its main purpose to increase the particles' reconstruction quality, the cut also diminishes the Np0 fraction once again due to its high presence in the p_T bins below 15 GeV.
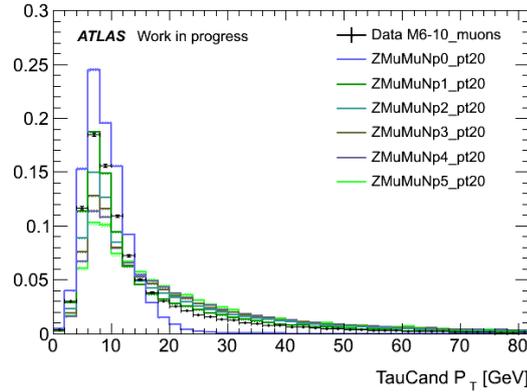
Figure 5.8: $p_T^{\tau_{\text{cand}}}$ before cut on $p_T^{\tau_{\text{cand}}} > 15\,\mathrm{GeV}$ (curves normalised to unit area).
The distributions have very similar shapes, but Np0 has the highest ratio of $p_T^{\tau_{\text{cand}}}$ values below 15 GeV.

Another condition on the tau candidates is for them to be opposite in ϕ to the supposed *Z* boson (cp. Fig. 5.9a), which insures a candidate to be consistent with a *Z*'s recoil jet. To achieve this, the bisection of the muons' ϕ-angles is calculated. The angle $\Delta\phi_{Zj}$ included by the bisection and the tau candidate has to be close to π (here: $\Delta\phi_{Zj} \in [\pi - 0.3, \pi]$). Note that $\Delta\phi_{Zj}$ is defined here as the absolute difference in ϕ, so $\Delta\phi_{Zj} \in [0, \pi]$. The tolerance level will be called $D\phi$, thus $D\phi = 0.3$ here (cp. Table 5.1).

From the $\Delta\phi_{Zj}$ histogram in Fig. 5.9b it is evident that the the Np1 sample indeed peaks at π and is favoured by the cut as a consequence. The cut flow histogram in Fig. 5.3b supports this fact, but also shows that in MC, more tau candidates survive this cut than in data. A possible

reason is that too many candidates from Np2 and Np3 are kept. However, the effect is relatively small.

(a) Tau cut: opposite μ's within $D\phi = 0.3$.   (b) $\Delta\phi_{Zj}$ before cut (curves normalised to unit area).
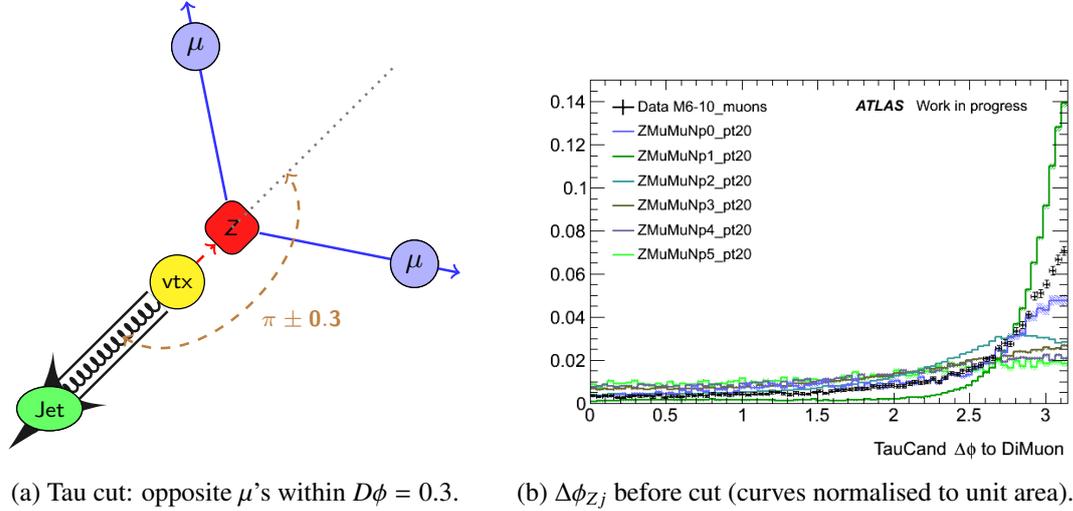
Figure 5.9: Tau cut on $\Delta\phi_{Zj} \in [\pi - 0.3, \pi]$ (graphical illustration and variable histogram). The main sample in the remaining $\Delta\phi_{Zj}$ region is the desired Np1.

Lastly, two cuts on the tau candidates' η were added to certify their detection in a detector regime where a reliable reconstruction is possible ($|\eta| < 2.4$ and $|\eta| \notin [1.37, 1.52]$).

5.1.3 Error Bars and Deviation Histograms

For statistical data, uncertainties are Poisson-distributed. Since only one of the two parameters for this distribution is usually known, the Gaussian approximation is used to compute errors simply as $\sqrt{N}$ for N bin entries, even though, strictly speaking, this is only valid for large values of N.

In many analyses, as in the one at hand, this does not have a big impact on the results. Although methods exist to give better approximations for Poisson errors at low multiplicities (J. Heinrich proposed several approaches in 2003 [29]), the exact quantity is impossible to determine.

For the qualitative purpose of this study, it was sufficient to assume Gaussian errors even for bins with low multiplicity. For plots highlighting data/MC differences, like the deviation plots introduced within this chapter, a nonzero error has to be defined to avoid division by 0. The choice was made to set the error to 1, though this is not optimal and should be replaced by a mathematically more consistent approach (e. g. one if those given in the aforementioned document by J. Heinrich).

Up to this point, all histograms were area-normalised in order to compare their shapes. In fact, for decay mode reconstruction, only the shapes matter, since PanTau uses efficiencies to train the MVA (cp. Ch. 3.2.3). However, it makes sense to have a look at plots normalised by luminosity (cp. Ch. 5.1.6) to be able to compare histograms intuitively. The sum of MC samples then has to match the data histogram, if the modelling was accurate.

To quantify how well the histograms match, several techniques can be applied, some of which are explained in detail in Ch. 5.2. For the moment, it is sufficient to visualise the divergence in the plots and its proportion to the errors. A commonly used way to achieve this is by adding a second graph below the histograms, showing the data/MC ratio ('ratio plot'), with a straight line at 1 indicating accordance. This has the disadvantage of being asymmetric: If, for example, there are double as many data events as MC in one bin, the graph has an entry at 2, while for the reverse case, the value is 0.5. This implicates that, when looking at such a plot, one has to view downward divergences as more crucial than upwards ones, which is not intuitive. One way to solve this is to apply a logarithmic scale, which still is not helpful for fast intuitive interpretation.

An alternative approach is to plot the difference of the two histograms (number of data events N_{data}^i minus number for MC N_{MC}^i), scaled by their combined errors σ_{comb}^i for each bin i. The result is denoted as 'deviation' (dev^i) in this document (Eq. 5.2a). The combined error is calculated by adding the squared errors for data (σ_{data}^i) and MC (σ_{MC}^i) and taking the square root (Eq. 5.2b).

$$\mathrm{dev}^i = \frac{N_{\mathrm{data}}^i - N_{\mathrm{MC}}^i}{\sigma_{\mathrm{comb}}^i} \qquad\qquad i = \text{bin number} \qquad\qquad (5.2\mathrm{a})$$

$$\sigma_{\mathrm{comb}}^i = \left(\sigma_{\mathrm{data}}^i\right)^2 + \left(\sigma_{\mathrm{MC}}^i\right)^2 \qquad\qquad (5.2\mathrm{b})$$

While the data error (in the Gaussian approximation, see above) is obtained by taking the square root of the number of bin entries N_{data}^i (Eq. 5.3a), for MC all applied scaling factors, i. e. s_{muon}, s_{lumi}, s_{pileup} and s_{global}, have to be included (Eq. 5.3b). The factor s_{muon} will be introduced in the course of muon corrections (Ch. 5.1.5), s_{lumi} is needed to scale the N_{gen}^i generated MC events to the data's integrated luminosity (Ch. 5.1.6), s_{pileup} accounts for different pileup conditions in data and MC (same chapter) and depends on the average interactions per bunch crossing μ, and finally, s_{global} is a global weighting factor obtained by a fit (Ch. 5.1.8).

$$\sigma_{\mathrm{data}}^i = \sqrt{N_{\mathrm{data}}^i} \qquad\qquad (5.3\mathrm{a})$$

$$\sigma_{\mathrm{MC}}^i = s_{\mathrm{muon}} s_{\mathrm{lumi}} s_{\mathrm{pileup}} s_{\mathrm{global}} \sqrt{N_{\mathrm{gen}}^i} = \sqrt{s_{\mathrm{muon}} s_{\mathrm{lumi}} s_{\mathrm{pileup}} s_{\mathrm{global}} N_{\mathrm{MC}}^i} \qquad\qquad (5.3\mathrm{b})$$

Analogously to the line at 1 in ratio plots, a line at 0 is shown in deviation histograms to indicate the position of MC. To this line, the MC error, also scaled by the combined error, can be appended, as the data's error (scaled in the same way) is drawn at the deviation values. This makes it possible to identify the separate errors' impact on the combined error.

When used to ratio plots, one has to read deviation histograms with care, since (in this comparison) they tend to overemphasise differences in certain N-ranges. To illustrate this, both values were computed for five different numbers of N_{data}^i and arbitrary N_{MC}^i between 0 and $3N_{\mathrm{data}}^i$ (Fig. 5.10).[2]

The deviation gives the data/MC difference in direct proportion to the combined error, but the ratio overemphasises differences for small numbers in bins, and understates those at high

---

[2] Note that N_{MC}^i, unlike N_{data}^i, can take non-integer values due to the scaling factor s_{lumi}.

numbers. In ratio plots, it is therefore always important to analyse the error bars as well, while in deviation histograms these are optional and rather give additional information on the errors' main origin.
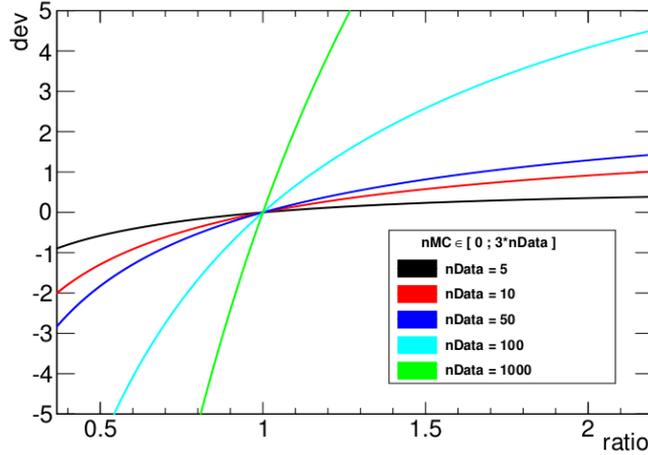
Figure 5.10: Deviation and data/MC ratio for simulated bin entries. Errors were taken as square root of the respective values, for data and MC.

For low numbers of N, the deviation suffers strongly from the Gaussian error approximation, as visible in the black line for $N_{\text{data}}^i = 5$ in Fig. 5.10.

As already mentioned, the error was set to 1 for $N = 0$ to avoid technical problems. This mathematical inaccuracy will not have a great impact on the analysis. However, it should be kept in mind when reading the deviation histograms: Black dots with nonzero error bars do not imply a finite number of data entries in the corresponding bin. The error could be set to 1 because there were no data entries, in which case the point is still drawn in the deviation histogram due to technical imperfections of the implementation.

These minor technical flaws will soon be corrected in the SFM framework, but the mentioned implications still hold for the graphs in this document.

5.1.4 Skimming

Working at an integrated luminosity of $332.02\,\text{pb}^{-1}$, file sizes can reach unmanageable magnitudes. To reduce disk usage as well as download and computing time, the used D3PDs were skimmed, i. e. some pre-cuts were applied before even running the cycle (cp. Ch. 2.3.1). These cuts clearly have to be looser than the ones in the cycle, but tighter than the used trigger(s). They are stated in Table 5.2.

Also, it has to be confirmed that the skimming does not cut off more than it is supposed to. One simple method to achieve this is by comparing the skimmed and the original D3PD of one of the samples. If the test is successful, the skimming algorithm is assumed to work correctly and the other samples' skimmed versions can be trusted to include all desired events.

The result is shown in Fig. 5.11: As expected, the normal and the skimmed D3PD are a match after the cut on the muons' invariant mass, which corresponds to the last of the skimming cuts. Henceforward, the skimmed version is an equivalent replacement of the full D3PD.

Trigger EF_mu18_MG_medium
at least 2 μ's: $p_\text{T} > 20\,\text{GeV}$
at least 1 τ_cand: $p_\text{T} > 10\,\text{GeV}$
at least 2 μ's: loose quality
at least 2 μ's: combined
at least 2 μ's: $M_{\mu\mu} = (90 \pm 30)\,\text{GeV}$

Table 5.2: Skimming cuts.
Before skimmed data can be compared to MC, these represent the minimum cuts that have to be applied beforehand.
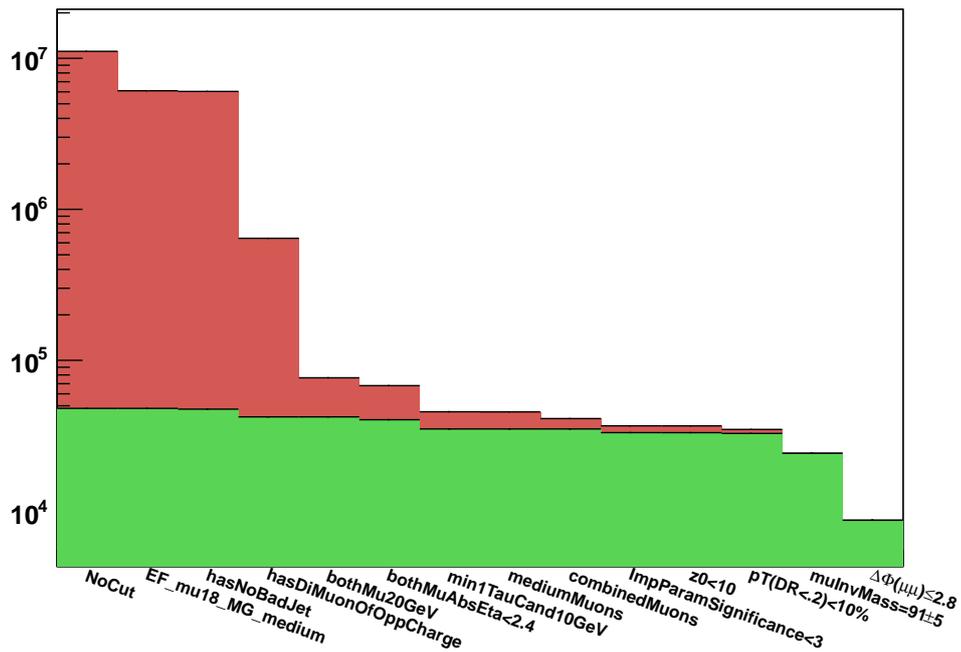
Figure 5.11: Event- and tau-wise cut flow: Number of events after each event cut for full (red) and skimmed (green) D3PD, tested on data period M10/2011.
Their difference decreases with every cut. Starting from the $M_{\mu\mu}$ cut, the D3PDs are identical.

5.1.5 Necessary Corrections

In addition to the cleaning cuts, some further corrections are necessary in order to achieve sensible results in the analysis. Some proposals for these can be retrieved from the Muon WG.

When dealing with muons, particular care has to be taken regarding the reconstruction due to general misbehaviour of the MC with reference to the muon momenta. Because the assumed reconstruction resolution in MC is too optimistic, the muon momentum is smeared over with a Gaussian distribution to correct this (***momentum smearing***). The Muon WG provides assigned tools for this, as well as for the ***muon efficiency corrections***, which are required for similar reasons and add another weighting factor to MC events.

The effect these corrections have on the reconstructed muons' momentum is shown in Fig. 5.12. Evidently, the overestimation in the peak region (Fig. 5.12a) is fixed after their application (Fig. 5.12b). The weighting factor s_{muon} introduced by the muon corrections is 0.9755. Note that the altered $p_{\mathrm{T}}^{\mathrm{muon}}$ resolution also has an impact on the behaviour of some of the cuts: Due to the corrections, the number of MC events decrease by about 2.34 % (in addition to the scaling factor).

(a) $p_{\mathrm{T}}^{\mathrm{muon}}$ before corrections.   (b) $p_{\mathrm{T}}^{\mathrm{muon}}$ after corrections.
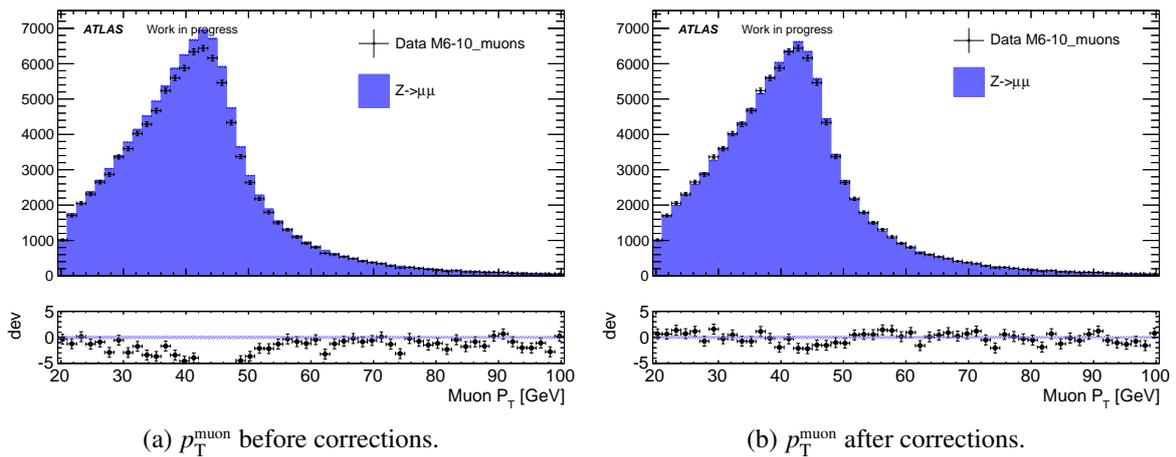
Figure 5.12: $p_{\mathrm{T}}^{\mathrm{muon}}$ before and after muon corrections are applied (curves lumi-weighted, cp. Ch. 5.1.6). The corrections mainly serve to artificially deteriorate the $p_{\mathrm{T}}^{\mathrm{muon}}$ resolution to compensate for a bug in the MC. As a result, the data is described far better by MC than without the corrections.

5.1.6 Luminosity and Pileup Weights

Two kinds of preparations have to be made before the actual analysis even starts: The luminosity (short hand ***lumi***) and pileup weights have to be calculated (compare Fig. 5.13).

To be able to compare data and MC, the integrated luminosity L (cp. Eq. 5.4a) which the generated MC samples correspond to has to be calculated, and an according luminosity weight s_{lumi} has to be applied to all MC events. The generated luminosity L_{gen} equals the number N_{gen} of generated MC events divided by the total cross section σ of the process (Eq. 5.4b). Trivially,
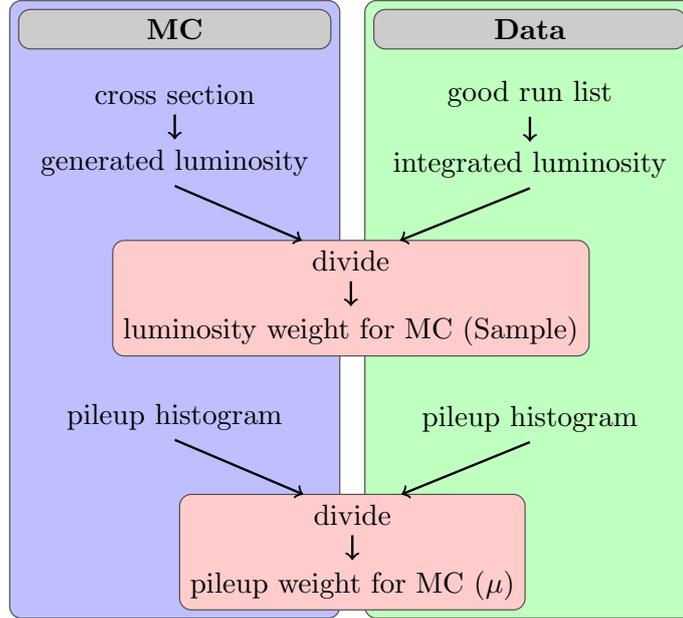
Figure 5.13: Necessary preparations regarding MC and data samples (dependencies of weights are stated in brackets).

the luminosity weight s_{lumi} is the ratio of data luminosity L_{data} and L_{gen} (Eq. 5.4c).

$$L := \int \mathcal{L} \, \mathrm{d}t \qquad (5.4a)$$

$$L_{\text{gen}} = \frac{N_{\text{gen}}}{\sigma} \qquad (5.4b)$$

$$s_{\text{lumi}} = \frac{L_{\text{data}}}{L_{\text{gen}}} \qquad (5.4c)$$

For the luminosity calculation of data, only runs from the GRL are considered, since other runs are not allowed to be used for analyses.

During the running of the experiment, pileup conditions vary critically over time. During a run, the luminosity decreases naturally, so μ is lower for late events in a run than for early ones. During a year, luminosity typically increases due to optimisation of the machine.

To account for these effects in MC samples, ***pileup reweighting*** has to be applied. This means that each MC event is weighted by a factor dependent on its pileup conditions (i. e. the average number of interactions per bunch crossing μ). This weighting factor s_{pileup} is determined by simply dividing the μ-histogram for the used MC sample by the one for the applicable data periods. In other words, each MC event is weighted by the quotient of the number $N_{\text{data}}(\mu_e)$ of data events with the same μ as the current event e and the analogue number $N_{\text{MC}}(\mu_e)$ of MC

events with the same μ (Eq. 5.5).

$$s_{\text{pileup}} = \frac{N_{\text{data}}(\mu_e)}{N_{\text{MC}}(\mu_e)} \tag{5.5}$$

In this manner, events differing strongly from real data with respect to pileup conditions will be weighted by a small factor, whereas more 'data-like' MC events will be multiplied by a larger weighting factor. The resulting μ-histogram for MC is therefore be a considerably better match to the one for data (Fig. 5.14).

(a) μ before pileup reweighting. (b) μ after pileup reweighting.
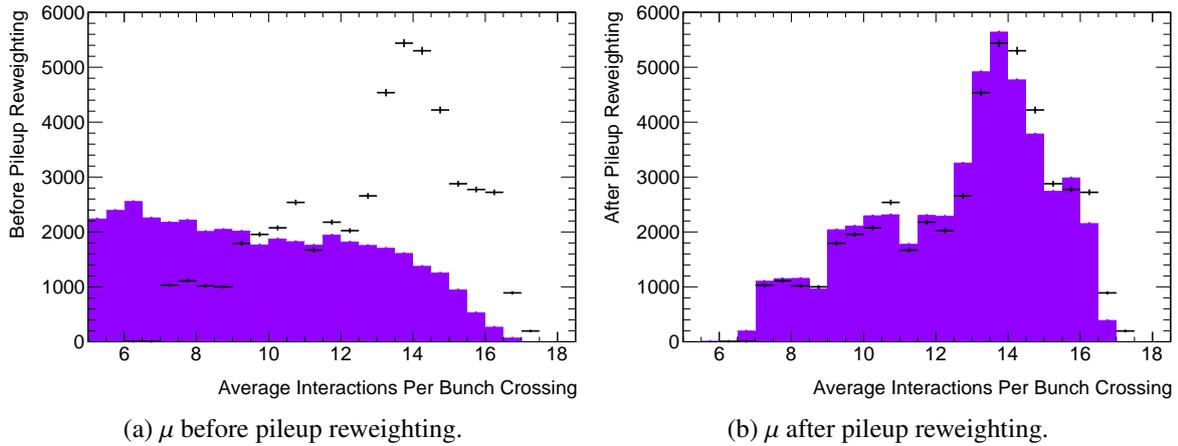
Figure 5.14: Average interactions per bunch crossing μ before and after pileup reweighting of MC (curves lumi-weighted).

Before the reweighting, pileup conditions of data and MC are completely different. The reweighting remodels the MC histogram's shape to fit the data.

This can have a huge impact on pileup-sensitive variables, such as the ΔR to the closest jet (Fig. 5.15). From Fig. 5.15a it is evident that the MC hugely overestimates the values before pileup reweighting is applied: In the low-ΔR region, MC has far smaller bin entries than data, and for high ΔR the MC entries are too high (although the effect is less visible there, due to the curve's lower gradient). In the 0-bin, there are more than double the entries for MC compared to data, which shows a lower jet multiplicity per event for MC.

All of these effects are a direct consequence of the lower μ in MC, as seen in Fig. 5.14a: With low pileup, the overall jet multiplicity is lower and therefore the ΔR between two jets is higher on average. The number of single jets in an event (0-bin) is also higher for lower values of μ.

After the pileup reweighting, the data/MC differences are reduced immensely (Fig. 5.15b), which proves the distinctive pileup conditions as their origin.

5.1.7 Analysis Cycle

Due to the complex structure of the cycle, particular caution has to be applied when designing the arrangement of loops, especially the ones embedded within each other. Therefore, the used

(a) ΔR to closest jet before pileup reweighting.

(b) ΔR to closest jet after pileup reweighting.
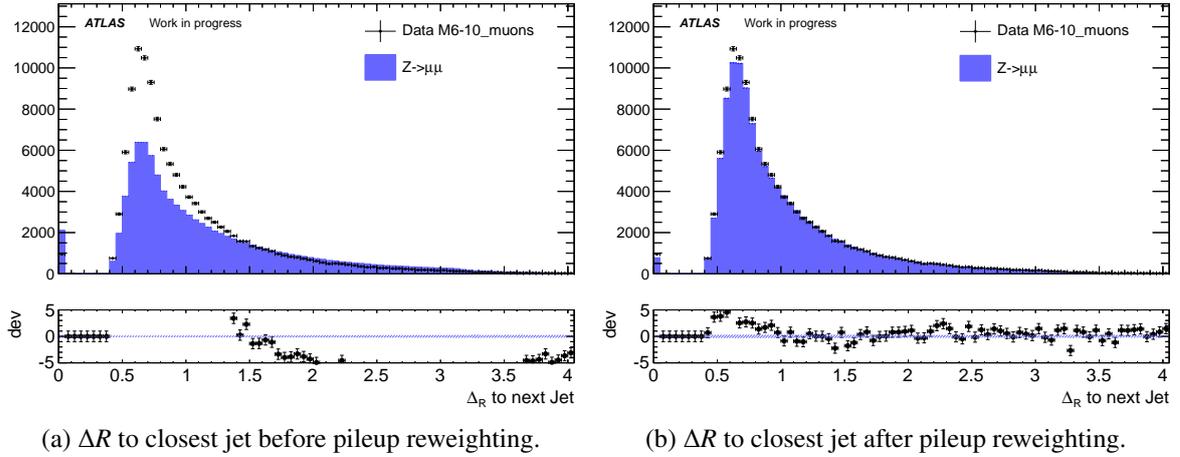
Figure 5.15: ΔR to closest jet before and after pileup reweighting (curves lumi-weighted). Values below 0.4 are not allowed, to avoid counting the ΔR between the tau candidate and its own reconstructed jet (cp. Ch. 3.2). Values at 0 are filled whenever no second jet could be found outside the 0.4-cone.

Before the reweighting, huge deviations are evident, in particular in the lower regions and the 0-bin. The reweighting eliminates these differences almost completely.

cycle for the background analysis deserves some explaining. The basic design of the cycle is illustrated in Fig. 5.16.

The main loop is executed per event and begins with some preparations, which, in the MC case, are dependent on the current sample and pileup conditions. The last preparation step is a loop over the muons in the event and depends on the different measured p_{T} values of the respective muon.

Subsequently, the fundamental operations are arranged in two double loops: First all event cuts are scanned. Whenever a cut criterion is not met, the rest of the steps for the current event are skipped and the next event is examined. As long as the criterion for the present event cut is fulfilled, firstly histograms are filled for variables regarding the whole event (as opposed to τ lepton related variables). In Fig. 5.16, these are referred to as ***event histograms*** (and ***tau histograms***, respectively). Secondly, tau histograms are filled for all tau candidates (inner loop). Thirdly, the second double loop is executed if all event cuts have been passed. For each tau cut, tau histograms are filled in a loop over all tau candidates, given that the particular tau cut is satisfied by the active tau candidate (otherwise the algorithm continues with the next tau candidate in the current event). Also, for each tau cut the number of tau candidates passing it is stored. For each tau candidate passing the last demanded tau cut, the binned histograms (as described in Ch. 5.1.10) are filled as well.

A final loop over all tau cuts fills the event histograms for each cut which was passed by one or more tau candidates in the previous loop.
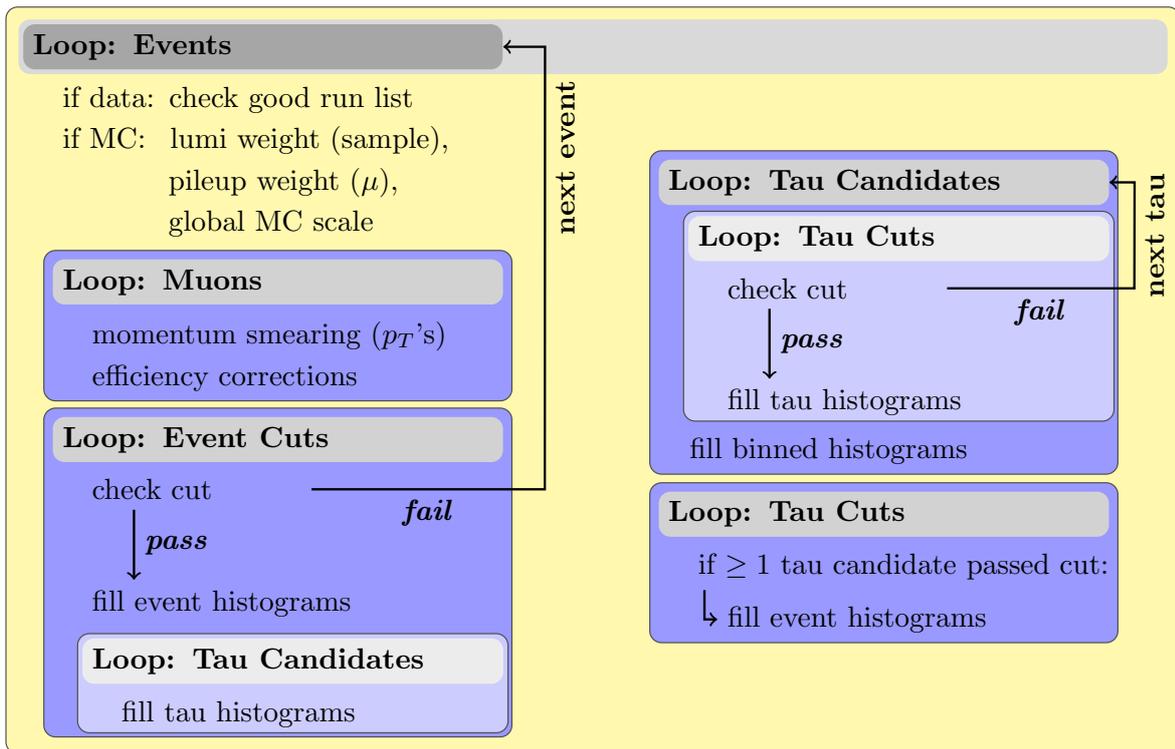
Figure 5.16: Graphical presentation of the analysis cycle (dependencies of actions are stated in brackets).
After some initial preparations for data and MC events, including a loop over all muons to apply basic corrections (cp. Ch. 5.1.5), the cycle consists of two double loops over tau candidates and either event cuts or tau cuts, as well as a final loop over tau cuts.

5.1.8 Background Estimation

For the analysis, data samples are taken from the muon stream. The choice of MC samples is based on the desired $Z \to \mu\mu$ processes. Since no further MC samples are added for completeness, it might well be that even after all physics cuts some data events remain, which have other origins and can therefore not be expected to be described by the used MC samples.

Note that in τ lepton studies all fake taus are background, while in this background study fake taus are signal, so the term 'background' here only refers to those tau candidates which do not represent $Z \to \mu\mu$ recoil jets.

Since for the above reasons, these background events cannot be described by MC, it has to be shown that their magnitude is small compared to the signal. To achieve this, the invariant mass $M_{\mu\mu}$ of the two muons (before the cut on the Z mass) is fitted with a discontinuous function f, which is composed of the bin entries of the MC histogram $h(M_{\mu\mu})$, multiplied by a scaling factor s_{global}, and a linear part $aM_{\mu\mu} + b$, where the slope a and the intercept b are further fit parameters (Eq. 5.6).

$$f(M_{\mu\mu}) = s_{\text{global}} \cdot h(M_{\mu\mu}) + aM_{\mu\mu} + b \tag{5.6}$$

The plot of the fit is shown in Fig. 5.17.
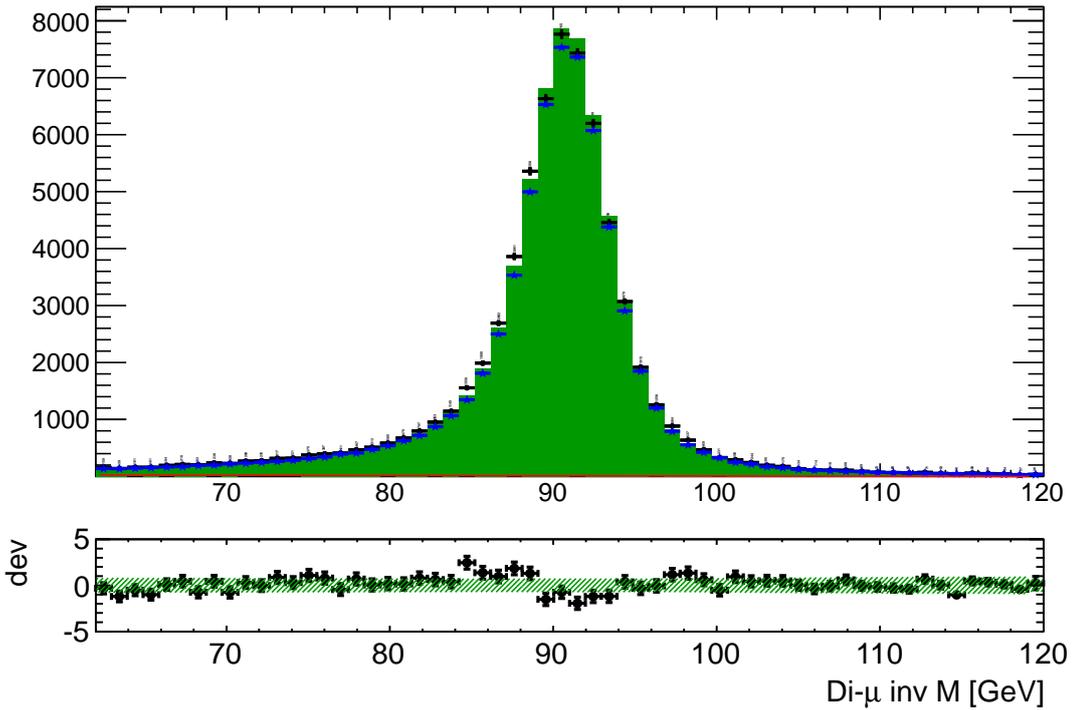
Figure 5.17: Invariant mass $M_{\mu\mu}$ of the di-muon. The data histogram (black) is fitted with the sum of a linear function (red) and a multiple of the MC histogram (blue). The green area represents the optimised fit function.
Results for the fit parameters are presented in the first row of Table 5.3.

The value of the scaling factor s_{global} obtained from the fit result can be used as an additional

MC weighting factor (in fact, all lumi-weighted histograms that appeared so far already include this weighting). The physical interpretation of this global MC scale is a correction of the branching ratio that was used to calculate the samples' generated luminosity (cp. Ch. 5.1.6).

The linear part of the fit, integrated over the later-used region of $[86, 96]$ GeV, gives the estimation of background events in the analysis. Dividing it by the data integral over the same region produces the ratio of background to total events of 0.23 %. This is small enough for it to be be neglected in the further analysis.

As a sensitivity check, the fit is re-evaluated for smaller $M_{\mu\mu}$ intervals to find out how severe the change in the background ratio would be (see Table 5.3 and Fig. 5.18). With values varying between 0.23 % and 0.61 %, the increase indeed is not severe enough to doubt the original estimate.

Fit Region	Scale s_{global}	Slope a	Offset b	$\frac{BG}{Data}$ in $[86, 96]$ GeV
$[62, 120]$ GeV	1.042 ± 0.004	-0.40 ± 0.09	48.3 ± 9.7	0.23 %
$[67, 115]$ GeV	1.039 ± 0.005	-0.72 ± 0.14	82.5 ± 14.5	0.33 %
$[72, 110]$ GeV	1.035 ± 0.005	-0.95 ± 0.23	111.4 ± 23.6	0.49 %
$[77, 105]$ GeV	1.033 ± 0.005	-0.90 ± 0.45	113.1 ± 44.4	0.61 %

Table 5.3: Fit results for background estimation and MC scaling.

After applying the global scaling to MC, the difference between data and MC regarding the number of events is only due to the background events. By cutting on the Z-mass in the next step, a large fraction of these is removed, which leaves a very good data/MC match at this point, with slightly too many MC events (cp. Fig. 5.3a).

5.1.9 First Results

After confirming that no background events must be expected to corrupt the analysis, the study of PanTau features can finally begin.

As described in Ch. 3.3, a variety of 121 variables are calculated by the `JetFeatureExtractionTool`. Obviously, the entirety of their histograms cannot be included within this chapter, so only some examples are shown in Fig. 5.19.

The presented variables are the ones that were argued to have good separation power in Ch. 3.3 (EflowNeutralSumEt works well to separate 1p0n against other modes, the same holds for EflowChargedSumEt and 3p0n, SumCharge and 3pXn, as well as EflowStdDevEt and 3pXn).

The variables all show very good data/MC agreement, especially EflowNeutralSumEt. Consequently, they can be utilised for their respective separation purposes in PanTau's MVA (cp. Ch. 3.2.3).

Most PanTau features display a similarly good accordance of data and MC, but are not discussed in detail due to the proportions of this document.

---

[3] Note that Fig. 5.18a is identical to Fig. 5.17 and is only included here for sake of comparison.

(a) $M_{\mu\mu} \in [62, 120]\,\text{GeV}$ [3]

(b) $M_{\mu\mu} \in [67, 115]\,\text{GeV}$

(c) $M_{\mu\mu} \in [72, 110]\,\text{GeV}$

(d) $M_{\mu\mu} \in [77, 105]\,\text{GeV}$
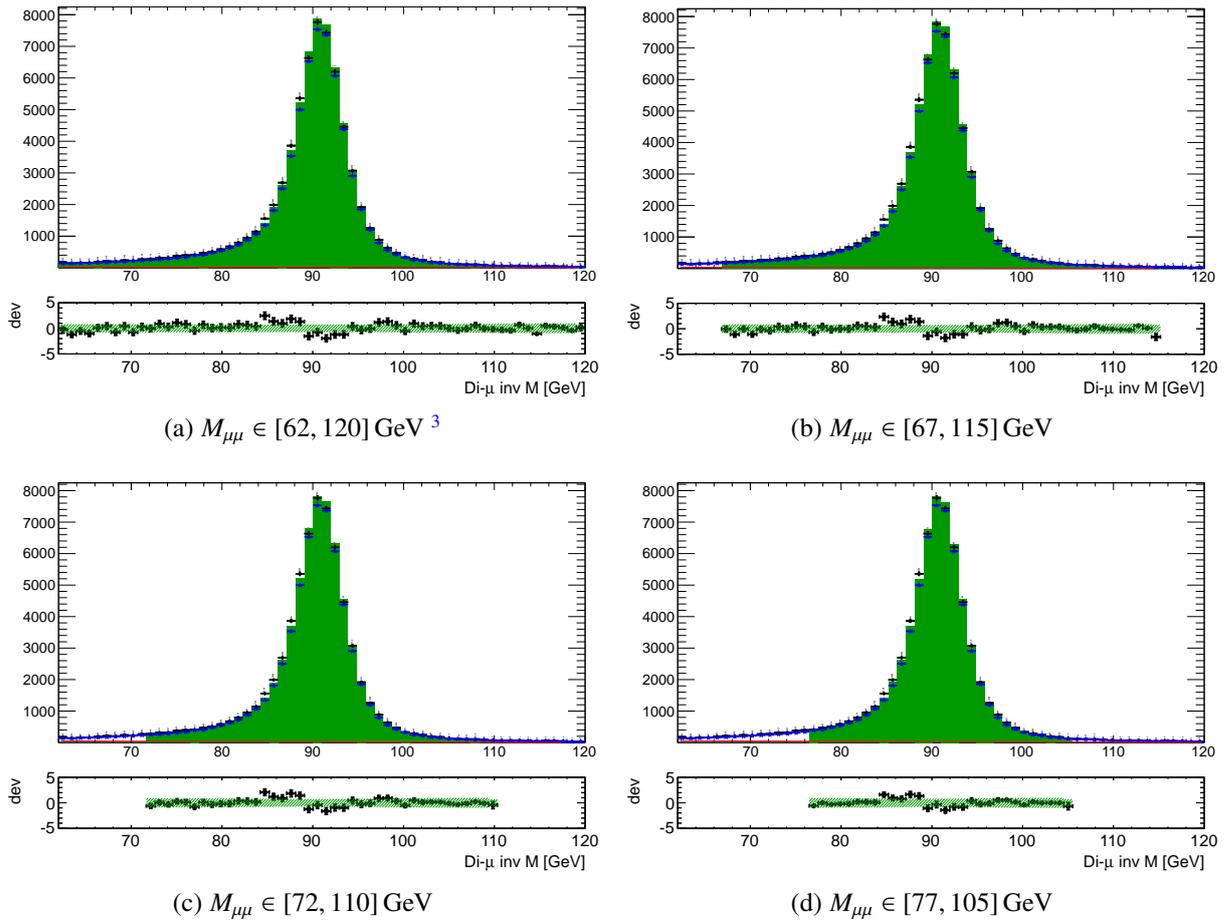
Figure 5.18: Background fits for different $M_{\mu\mu}$ ranges. As before, the data histogram (black) is fitted with the sum of a linear function (red) and a multiple of the MC histogram (blue). The green area represents the respectively optimised fit function (cp. Fig. 5.17).
Evidently, the fit parameters are not sensitive on the length of the interval.

(a) EflowNeutralSumEt.

(b) EflowChargedSumEt.

(c) SumCharge.

(d) EflowStdDevEt
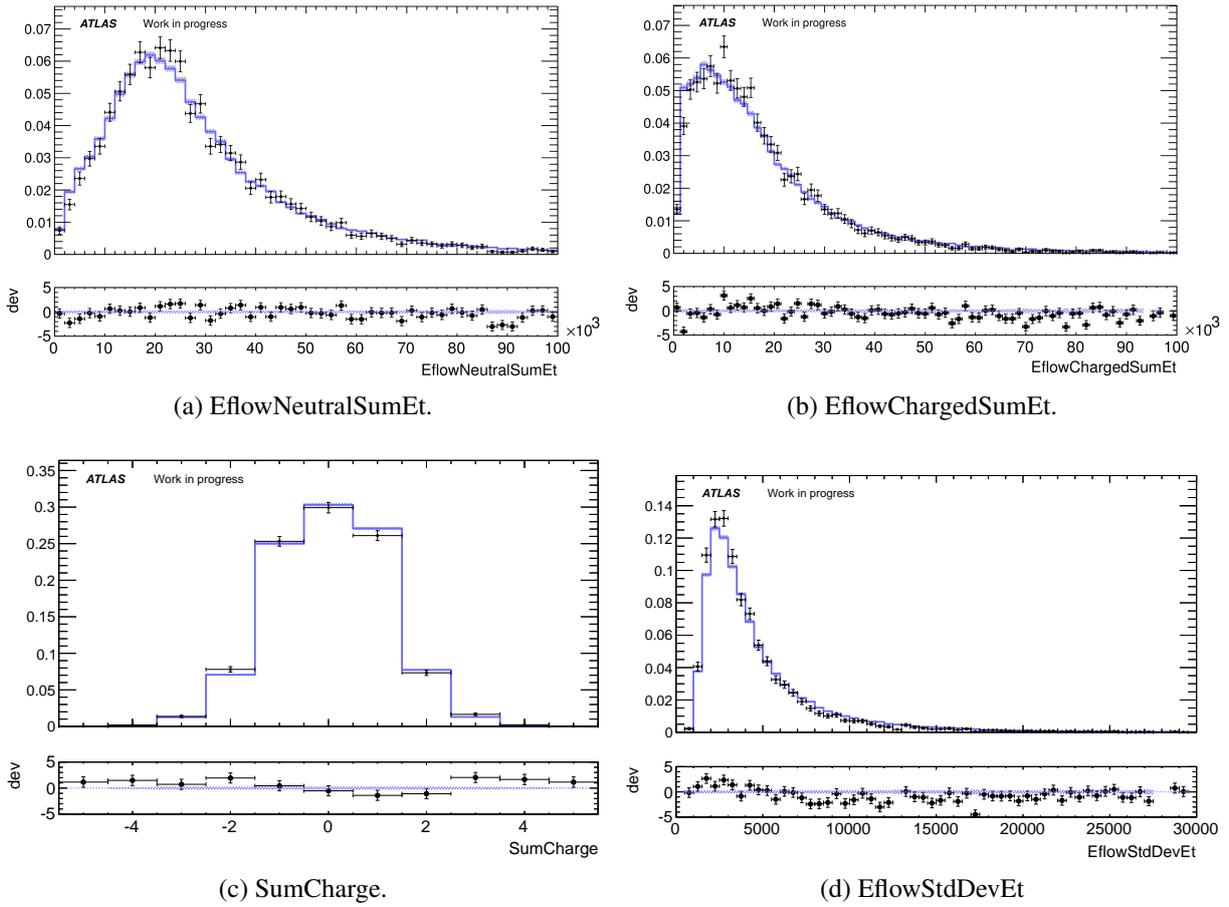
Figure 5.19: Histograms for different PanTau features (curves normalised to unit area).
These are variables with good separation power for at least one of the decay modes, or against fakes (cp. Ch. 3.3).

5.1.10 Physics Binnings

Apart from analysing the PanTau features' liabilities, another reason to perform the data/MC comparison is to spot the physical circumstances under which the variables perform particularly well and rather poorly, respectively.

Therefore, all histograms were divided into different binnings. To avoid confusion, this kind of binning will also be referred to as 'physics binnings' and the binning in each histogram's x-axis as 'variable binnings'. The physics binnings will enable the opportunity to not only inspect each variable under different circumstances (e. g. a particular p_T region), but also to study the general performance of PanTau variables in that region itself (see Ch. 5.2.2).

The used binnings are listed in Table 5.4, where $N(\text{EFO}^{*\pm})$ denotes the number of charged qualified EFOs.[4]

Binning Variable	Ranges
μ	$[10; 13)$, $[13; 14)$, $[14; 15)$, $[15; 16)$, $[16; 18]$
E_T	$[15; 18)$, $[18; 26)$, $[26; 43)$, $[43; 200)$
η	$[-2.8; -1.2)$, $[-1.2; -0.4)$, $[-0.4; 0.4)$, $[0.4; 1.2)$, $[1.2; 2.8)$
ϕ	$[-\pi; -0.5\pi)$, $[-0.5\pi; 0)$, $[0; 0.5\pi)$, $[0.5\pi; \pi)$
$N(\text{EFO}^{*\pm})$	$[1]$, $[2]$, $[3]$, $[4]$, $[5]$, $[6; 7]$
ΔR to closest jet	$[0.4; 0.55)$, $[0.55; 0.65)$, $[0.65; 0.725)$, $[0.725; 0.975)$, $[0.975; 1.475)$, $[1.475; 5)$

Table 5.4: List of physical binnings and their respective bin ranges. The intervals were chosen to create approximately equal numbers of tau candidates in each bin (except for the highest $N(\text{EFO}^{*\pm})$ bin) and virtually no tau candidate in the overflow and underflow bins. Energies are given in GeV.

For ΔR to the closest jet, values below 0.4 are not allowed to avoid counting the ΔR between the tau candidate and its own reconstructed jet (cp. Ch. 3.2). Values at 0 are filled whenever no second jet could be found outside the 0.4-cone.

To exemplify the resulting histograms, the variable SumCharge is given in all E_T bins in Fig. 5.20. Overall agreement of data and MC is very good. The 0-bin for $E_T \in [15; 18)\,\text{GeV}$ is overestimated by MC a bit, but the data value lies almost within the error bars. The deviation plot shows that the difference is less than $2\sigma_{\text{comb}}$.

Most PanTau variables share the positive data/MC comparison, as well as the indifference with regards to E_T and other binnings. However, the total number of histograms to be analysed amounts to $121 \times 30 = 3\,630$ at this point, which makes it difficult to keep an overview over general trends.

In order to prioritise the variables and physics bins to be inspected first, an attempt was made to quantify the data/MC agreement mathematically. This is described in the following chapter.

---

[4] Note that the inclusive upper border of the last μ bin in Table 5.4 is not a typo.

(a) $E_T \in [15; 18)$.

(b) $E_T \in [18; 26)$.

(c) $E_T \in [26; 43)$.

(d) $E_T \in [43; 200)$.
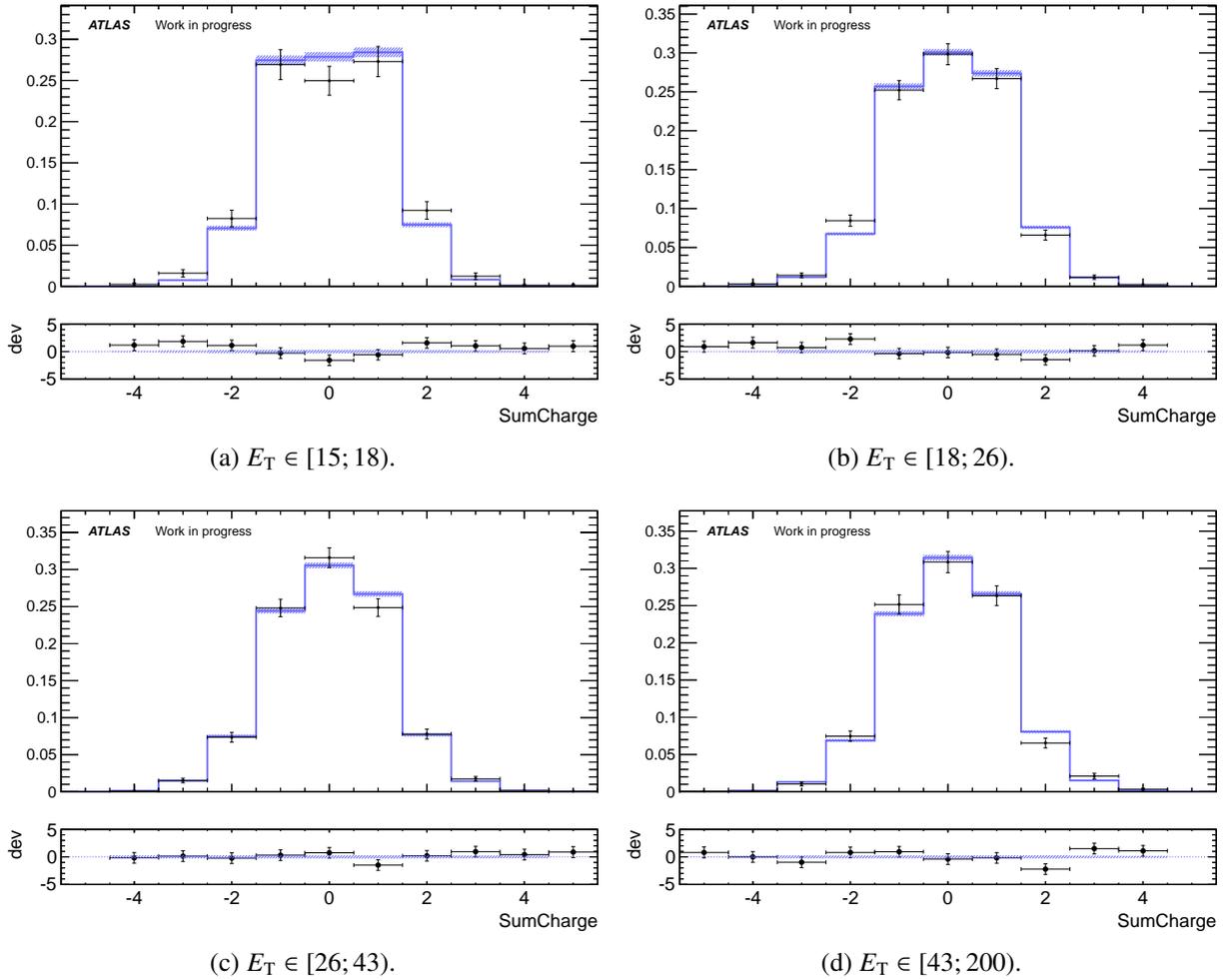
Figure 5.20: SumCharge in different E_T bins (curves normalised to unit area). Energies are given in GeV.

5.2 Quantifiers for the Degree of Accordance

The actual data/MC comparison is a process which is hard to be quantified in objective metrics. For each PanTau variable and each physics binning, the histograms for data and MC have to be checked for regions (or whole histograms) with bad agreement, which then can be investigated further. To assist with the inspecting of the vast quantities of graphs, and also to substantiate the otherwise subjective classification in well and poorly agreeing histograms, it is reasonable to develop an algorithm to statistically analyse the graphs, thus providing a first classification.

The aim, however, is not to find a substitute for the rational investigation of graphs with the naked eye, which no computational method can provide (some of the different methods' shortcomings are elaborated in the following sections). It is still practical to have one or two quantifiers at hand in order to get a first overview over the huge amount of histograms that are to be considered and help prioritise further studies.

5.2.1 χ^2 and Kolmogorov-Smirnov Tests

A very common statistical test to estimate the probability that two histograms originate from the same distribution is the χ^2 test. The χ^2 value is calculated by summing the squared difference of the bin entries of the data (N_{data}^i) and MC histogram (N_{MC}^i), divided by the square of their combined error σ_{comb}^i (cp. Eq. 5.2b), over all bins i in the region of interest (Eq. 5.7).

$$\chi^2 = \sum_i \frac{\left(N_{\text{data}}^i - N_{\text{MC}}^i\right)^2}{\left(\sigma_{\text{comb}}^i\right)^2} \qquad i = \text{bin number} \qquad (5.7)$$

Under consideration of the number of degrees of freedom n_{df}, the resolving χ^2 value can be transformed into a p-value by means of the inverse distribution function. The number of bins with finite entries corresponds to n_{df} and the p-value represents the probability to achieve a data/MC agreement of the present quality or worse, given that they originate from the same distribution. The p-value thus equals the integral over the χ^2 probability density function $f(x; n_{\text{df}})$ from the realised value of χ^2 to infinity (Eq. 5.8).

$$p = \int_{\chi^2}^{\infty} f(x; n_{\text{df}}) \, \mathrm{d}x \qquad (5.8)$$

The formula for the probability density function $f(x; n_{\text{df}})$ is stated in Eq. 5.9a, where $\Gamma(z)$ is the Gamma function (Eq. 5.9b). Most computer programs, however, equivalently use the

inverse distribution function $F^{-1}(x; n_{\text{df}})$ for the computation of p.

$$f(x; n_{\text{df}}) = \frac{x^{(n_{\text{df}}-2)/2}\,e^{-x/2}}{2^{n_{\text{df}}/2}\,\Gamma(n_{\text{df}}/2)} \tag{5.9a}$$

$$\Gamma(z) = \int\limits_{0}^{\infty} e^{-t}t^{z-1}\,dt \tag{5.9b}$$

The main drawback of the χ^2 test is that is only valid for large enough statistics, which is one of the reasons the Kolmogorov-Smirnov test (***KS test***) is often used as an alternative approach. This is despite the fact that the KS test in principle is only valid when comparing a histogram to a known, continuous function (as opposed to an unknown function approximated by a histogram, as is the case here). For the sake of comparison, it is still legitimate to perform both tests.

In the KS method, the test value d is the supremum of the differences between cumulative data (S_{data}^{i}) and MC (S_{MC}^{i}) histograms up to bin i (Eq. 5.10a, with S^i defined in Eq. 5.10b).[5]

$$d = \sup_{i}\left|S_{\text{data}}^{i} - S_{\text{MC}}^{i}\right| \tag{5.10a}$$

$$S^{i} = \sum_{j=1}^{i} N^{j} \tag{5.10b}$$

To find the corresponding p-value, the distribution function for d has to be found, which is not trivial if both histograms follow an unknown distribution. It can be achieved by sampling data points while using the given histograms as approximation for the distribution functions.

A consequence of using the cumulative functions S^i in the KS test is that the histograms' shapes are accounted for, while the χ^2 test is a pure bin-by-bin comparison. An advantage of the χ^2 test is that the bin errors are taken into account, as opposed to the KS test. On the other hand, the common approximations often used to calculate the errors yield incorrect results for bin entries (cp. Ch. 5.1.3), which could weaken this advantage.

For the purpose of this analysis, these differences do not play a huge role, since the aim is merely to pick the variables with the worst agreement, rather than to categorise each single variable. The expectation is that the two algorithms do not differ to a great extent in this task, which will be proven to be a valid assumption in Ch. 5.2.2 (also see Fig. 5.21).

---

[5] If the KS test is used to compare a histogram to a known continuous function $f(x)$, its values x^i have to be compared to both the respective upper (i) and lower ($i-1$) bin borders of the histogram. This corresponds to changing Eq. 5.10a to

$$d = \sup_{i}\left(\left|F(x^i) - S(x^i)\right|, \left|F(x^i) - S(x^{i-1})\right|\right).$$

This is not necessary here, because all histograms in the analysis have the same variable binning for data and MC.

5.2.2 Optimising the Tests' Reasonability

One of the main problems with both described mathematical methods is that they produce only one (statistically independent) number.[6]

This suffices in mathematically clearer case studies, e. g. when testing if a given histogram originates from a particular distribution function or not, or proving that an event cannot be explained by classical physics. In the analysis at hand, this is not the case. The compared histograms are known to deviate from one another to some extent (due to QCD approximations in MC), and slight, but interesting, mismatches beyond this effect might not result in hugely different p-values, regardless the test method.

For two histograms originating from the same distribution, the p-values are uniformly distributed. Thus, a small p is not a strong indicator to a great mismatch of data/MC – even though the probability for a mismatch is obviously higher than if p is large. To get a better idea of the distribution of p-values, one needs to perform the test multiple times on separate data sets.

If for example the test is performed five times, each time on one fifth of the present statistics, and the five resolving p-values are all small, they are highly unlikely to be uniformly distributed. Then again, each of them is less reliable than the one value was before, due to the smaller samples. The presented method aims at a trade-off between these effects, while at the same time using the subdivision itself to obtain more meaningful results.

This is realised by using the five pileup bins as the means of subdivision (cp. Ch. 5.1.10). By construction, these bins share roughly the same number of entries. To find the variables with the worst data/MC agreement, the resolving five p-values are categorised into the three bins shown in Table 5.5, which will be referred to as 'base' binning.

Binning Variable	Ranges
p	A: [0 %; 1 %], B: [1 %; 50 %], C: [50 %; 100 %]

Table 5.5: Bins and their labels for histogram of p-values.
Being a probability, p can only take values in [0 %; 100 %], so the underflow and overflow bins by definition are always empty.

When sorting the list of variables by decreasing entries in the above p-bins, the first variables are likely to have a bad data/MC agreement.[7] Since the five p-values correspond to different pileup bins, this method uncovers PanTau features with bad agreement in all pileup regions, rather than pileup-sensitive variables that are otherwise well-described.

To test how different the p-values from the two presented test algorithms are, a scatter plot of the variables' rankings was created (Fig. 5.21). It shows a very slight correlation between the two methods, which proves the calculated p-values' low significance as indicators for good or bad data/MC agreement. However, for the purpose of quickly spotting the histograms with the worst agreement among a large quantity of plots, it works well enough: Variables with extremely high (or low) rankings in one method tend to score bad (or good, respectively) with

---

[6] The p-value is derived from the χ^2 or d variable, respectively, which technically is a second number. However, this number yields no different statistical information, since they are connected by a bijective transformation.
[7] Note that the p-bin entries are uniformly distributed for excellent data/MC agreement, i. e. expected to be around $(0, 2.5, 2.5)$.

the other one. One particular type of variable that is highlighted in the correlation plot is the 'N-EFO type'. These variables are the number of EFOs or of a subset of EFOs (e. g. charged EFOs) and are described very poorly by MC, for reasons discussed in Ch. 5.3.1. It is evident that for this variable type, the correlation between χ^2 and KS test is quite high.
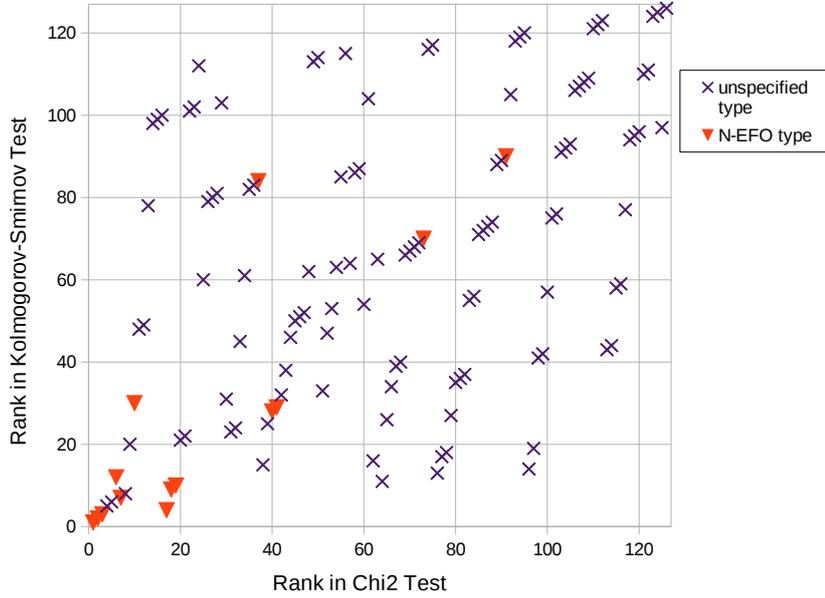
Figure 5.21: Ranking correlation for χ^2 and KS method (used ob variables). Low numbers stand for high rankings, i. e. bad data/MC agreement according to the respective test algorithm. The rankings are fairly uncorrelated, yet a slight correlation seems to apply, especially for extreme values. Correspondingly, the entries for N-EFO-like variables (red triangles) have a much higher correlation than the average of the unspecified variables (violet crosses), as the former tend to very low rankings.

To make sure that the ranking exposes the most important PanTau features (i. e. the ones with the worst agreement) at the top of the list, the p-values from both methods were filled into one histogram, because the low correlation of χ^2 and KS test shows their almost complementary nature. To leave both methods' results accessible, bin entries from both separate methods were kept in the tables, although the sorting algorithm only took into account the summed bin entries from both.

The upper part of the such produced list is given in Table 5.6, while the complete list can be found in Appendix A (Table A.1).

For each of the physics binnings, p-values can be calculated from every PanTau feature's histogram, so there is no need to artificially subdivide the samples here. Because of the greater magnitude of data points, plotting the p histograms (in linear or logarithmic binning) is more sensible for physics binnings than for single variables. An example for such a plot is given in Fig. 5.22 in all three kinds of mentioned p-binnings. For the most part of the analysis, however, the base binning is used.

While for the features' p histograms only the base binning is reasonable due to the few entries, its main advantage also for the physics binnings' p histograms is that the number of

(a) Standard histogram.

(b) Logarithmic binning.

(c) Base binning.
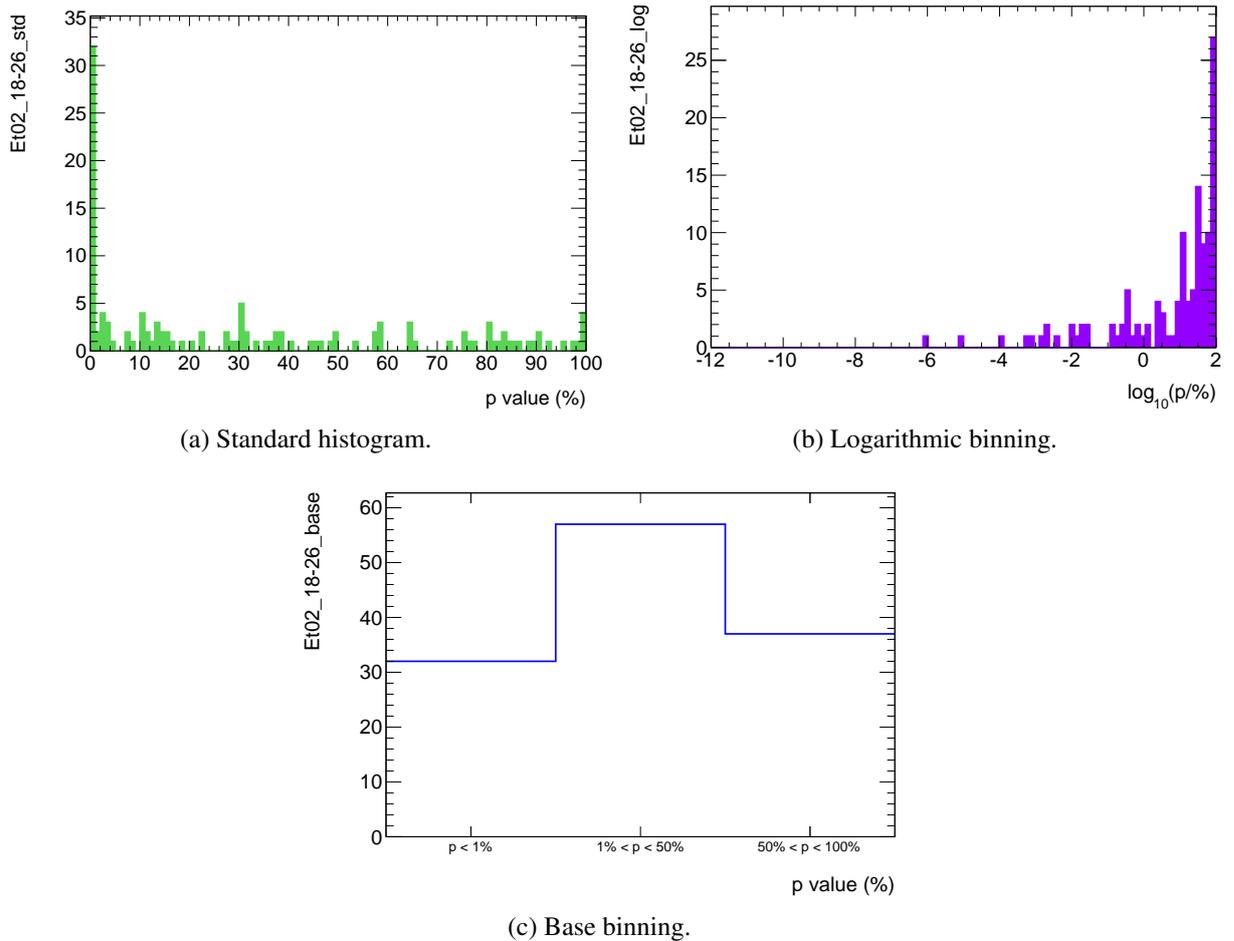
Figure 5.22: Indicative *p* histogram for the E_T bin [18;26] GeV.

In a standard uniform binning (Fig. 5.22a), the histogram is not well readable due to the first bin's magnitude. A logarithmic binning (Fog. 5.22b), on the other hand, makes the important comparison between lower and upper half of the axis unintuitive to see. The plot in base binning (Fig. 5.22c) makes it easy to spot the number of especially low *p*-values, as well as the relation of values below and above 50 %.

Sum			χ^2			KS			Variable
A	B	C	A	B	C	A	B	C	
10	0	0	5	0	0	5	0	0	NEFO
10	0	0	5	0	0	5	0	0	NNeutralEFO
9	1	0	4	1	0	5	0	0	NEFOOut02
8	2	0	4	1	0	4	1	0	EflowMeanEtAllToSumEt
8	2	0	4	1	0	4	1	0	EflowMeanEtNeutToSumEt
8	1	1	4	0	1	4	1	0	NEFOOut03
7	3	0	4	1	0	3	2	0	NHADNeutralEFO
7	3	0	3	2	0	4	1	0	RatioChargedToTotalNumber
7	2	1	2	2	1	5	0	0	NEFOIn02To04
6	3	1	2	2	1	4	1	0	NEFOIn03
6	3	1	2	2	1	4	1	0	NEFOOut01
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 5.6: Quantifier outputs for PanTau features, sorted by their sum (top 11 entries). The subcolumns represent the bins of the methods' p-values (cp. Table 5.5). The first three columns give the sum of the bin entries from the χ^2 and the Kolmogorov-Smirnov method, the other columns represent the methods' separate outcomes. For the full list, see Table A.1.
Clearly, many N-EFO-like variables reach low *p*-values in both test methods.

values below 1 % can be spotted immediately, while the comparison to a uniform distribution is still relatively straightforward (by comparing the third bin to the sum of the first two).

5.3 Analysis of Discrepancies

The discrepancies in PanTau features between data and MC are analysed with respect to the distinct variables and different physics binnings. The prioritisation mainly follows Table 5.6.

The huge number of histograms could not be included within this document, but will be available online for a limited time at the referenced address [30].

5.3.1 N-EFO-like Variables

As already indicated in Ch. 5.2.2, many of the N-EFO-like features show bad data/MC agreement. Histograms for the basic variables of this type (namely, the ones not depending on particular ΔR-cones) are depicted in Fig. 5.23. These are among the best separating PanTau features (cp. Ch. 3.3) and are therefore very desirable to perform well. Hence, it is particularly worth analysing where discrepancies come from and if they can be fixed.

The two variables obtaining only *p*-values below 1 % in both the χ^2 and KS test are NEFO and NNeutralEFO. The corresponding histograms are shown in Fig. 5.23a and Fig. 5.23b, respectively. In both cases, MC overestimates the multiplicity of EFOs. In contrast to this, NChargedEFO is underestimated by MC (Fig. 5.23c), but only to an extent consistent with
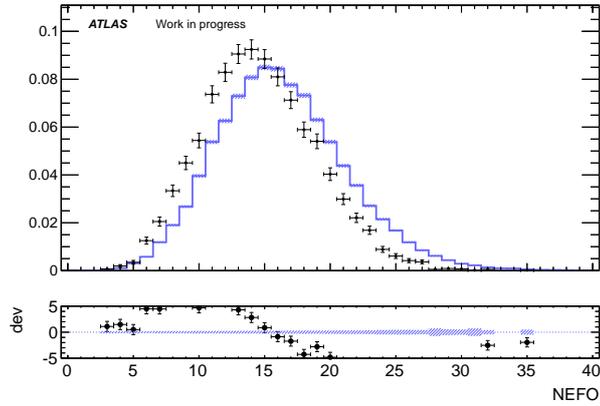
(a) NEFO.
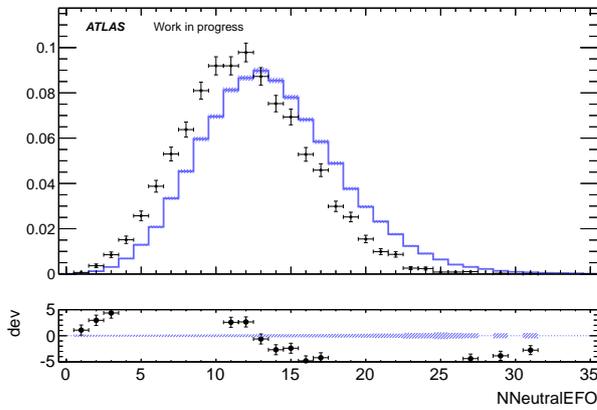
(b) NNeutralEFO.
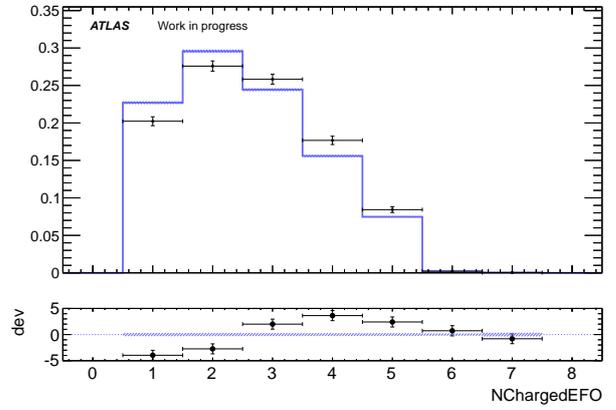
(c) NChargedEFO.
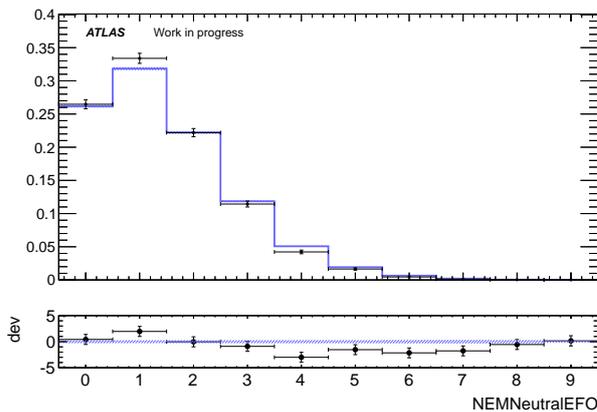
(d) NEMNeutralEFO.
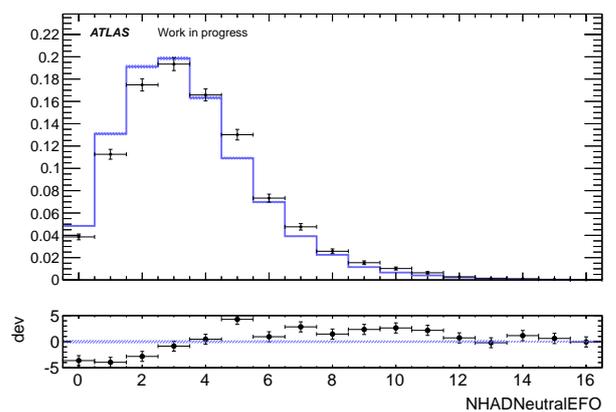
(e) NHADNeutralEFO.

Figure 5.23: Histograms for five basic N-EFO-like variables (curves normalised to unit area). MC overestimates NEFO and NNeutralEFO by 1–2 EFOs. In NChargedEFO and NHAD-NeutralEFO, the shift is directed in the opposite direction, and less critical. The differences between data and MC are lowest in NEMNeutralEFO.

known MC misbehaviour on QCD jets. This claim can be supported through comparison to the number of Tracks for each tau candidate, calculated by TauRec, which shows the same properties in data/MC deviations (Fig. 5.24). It can hence not be an effect caused by PanTau.

Note that, even though eflowRec only creates EFOs near TauRec candidates with 1–5 tracks (cp. Ch. 3.1.3), a small fraction of PanTau candidates can still have more than 5 charged EFOs, as seen in Fig. 5.23c. This happens when EFOs are assigned to more than one TauRec candidates (if they are close to each other). The effect is small enough to be neglected and will be fixed in later revisions of PanTau, by checking that EFOs are only allocated to one tau candidate.

NEFO is the sum of NNeutralEFO and NChargedEFO, but clearly dominated by the higher values of NNeutralEFO and thus shares its shift to higher values (in MC).

Peculiarly, the two subsets of NNeutralEFO behave differently. In NHADNeutralEFO, only a slight shift can be seen, and it points into the other direction (Fig. 5.23e). The data/MC differences here are comparable in magnitude to the ones in NChargedEFO. In NEMNeutral-EFO the shapes match very well (Fig. 5.23d).

The most obvious interpretation of this fact regards the additional energy cuts included in the definition of the EM and HAD flags: Contrary to the other basic N-EFO-like variables, NEM-NeutralEFO and NHADNeutralEFO require $E > 1\,\text{GeV}$ for each EFO (cp. Tab. 3.1), which could improve the features' reliabilities drastically due to reduction of pileup and calorimeter noise (which is not unlikely to be described poorly by MC).

In fact, an energy cut on neutral calorimeter clusters is applied in other decay mode reconstruction algorithms for τ leptons as well, e. g. in the π^0-tagger by B. Winter, where it was proven to reduce electromagnetic noise, as well as pileup deposits, by large amounts [31].

M. Hodgkinson and C. Limbach presently investigate if this is the reason for the different results of NNeutralEFO and its EM-/HAD-flagged subsets. Furthermore, adding energy cuts to all used EFOs could turn out not only to improve data/MC agreement, but also to have a positive effect on decay mode reconstruction by PanTau due to the enhanced background suppression.

In this case, the optimal cut values have to be tested. The background analysis hints at a lower threshold for NEMNeutralEFO, since the data/MC agreement is very good and purity could be reduced unnecessarily, and a higher one for NHADNeutralEFO, because a slight shift between data and MC is still visible. However, these values will have to be identified empirically by computing the efficiency and purity for different scenarios, while monitoring the data/MC accordance.

For the mentioned features it was demonstrated that the main data/MC differences are likely to share the same origin. Similar assumptions can be made for further variables based on numbers of EFOs (or subsets of EFOs) also achieving high rankings in Table 5.6.

5.3.2 Variables Based on E_T

Two further PanTau features obtaining poor results in χ^2 and KS tests are EflowMeanEtAllTo-SumEt and EflowMeanEtNeutToSumEt (Fig. 5.25, where EflowMeanEtChrgToSumEt is added for comparison).
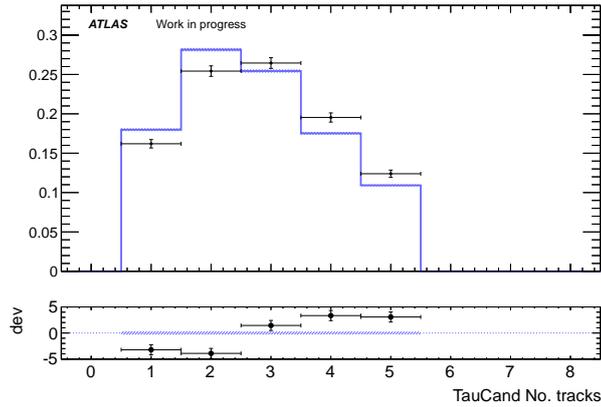
Figure 5.24: Histogram for the number of Tracks for each tau candidate (normalised to unit area).
The data/MC agreement is analogue to the PanTau feature NChargedEFO (cp. Fig. 5.23c).
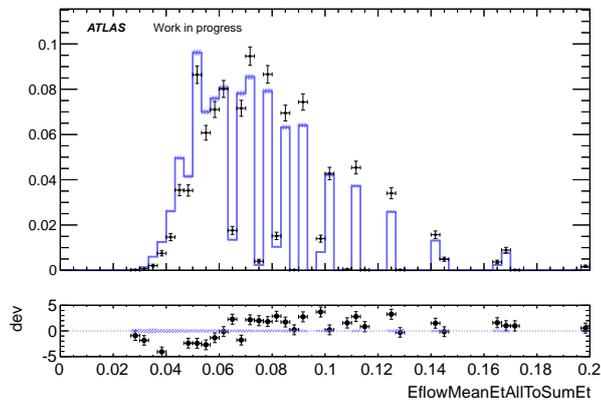
(a) EflowMeanEtAllToSumEt.
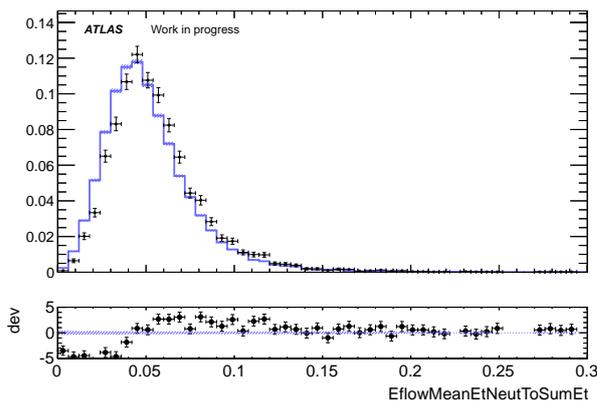
(b) EflowMeanEtNeutToSumEt.
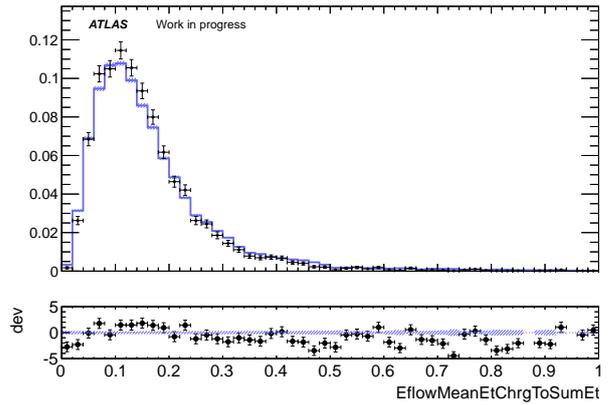
(c) EflowMeanEtChrgToSumEt.

Figure 5.25: Histograms for three variables based on E_{T}-fractions (curves normalised to unit area).
In EflowMeanEtAllToSumEt and EflowMeanEtNeutToSumEt, data/MC discrepancies are
large for low values of the respective feature, but very small for high values. Eflow-
MeanEtChrgToSumEt shows an overall good agreement.

Note that the fragmented structure of the EflowMeanEtAllToSumEt histogram in Fig. 5.25a is a direct consequence of its formula (cp. Ch. 3.3):

$$\frac{<E_\mathrm{T}>}{\sum E_\mathrm{T}} = \frac{\frac{\sum E_\mathrm{T}}{N}}{\sum E_\mathrm{T}} = \frac{1}{N}.$$

At low values, the intervals between obtainable values are smaller than the bin size and the histogram looks continuous, while at higher values the intervals become larger.

In EflowMeanEtAllToSumEt, data and MC are a good match for values greater than 0.07, and in EflowMeanEtNeutToSumEt (Fig. 5.25b), for values greater than 0.05. The lower parts of the histograms, however, are overestimated by MC. The variable EflowMeanEtChrgToSumEt (Fig. 5.25c) does not share this effect, but has a good agreement of data and MC over the whole range.

These effects are perfectly consistent with the findings from Ch. 5.3.1: In MC, more neutral EFOs are counted than in data, and the additional ones in MC seem to be the ones with low energy. Thus, the mean E_T of neutral EFOs is shifted down in MC, and so is EflowMeanEtNeutToSumEt. Since this variable is proportional to the inverse number of neutral EFOs, low values are accentuated naturally (cp. structure of EflowMeanEtAllToSumEt, where this effect is not smoothed by multiplication with an arbitrary number between 0 and 1). Hence, the shift is most visible in the lower bins, while for high values, it is blurred.

This also holds for the relation between NEFO and EflowMeanEtAllToSumEt. The shift in NEFO results in data/MC deviations in EflowMeanEtAllToSumEt which are more distinctive for lower values than for high ones. In this case, the latter variable is the first one's inverse, which proves that it is a purely mathematical effect.

Since NChargedEFO is shifted in the other direction (compared to NEFO) and less strongly, one would expect that EflowMeanEtChrgToSumEt behaves contrarily to EflowMeanEtAllToSumEt, namely by having too much MC in the first few bins and a little bit too little in the higher region. This is exactly what can be seen in the histogram, taking into account the general presence of more MC than data (also cp. Fig. 5.23c).

5.3.3 Different Regions

To evaluate and compare data/MC agreement in different physics bins, χ^2 and KS tests are performed for all feature histograms in each bin. The full results are shown in Appendix A (Table A.2).

Analogously to the variables analysis, a correlation plot for the two tests is inspected as a first step in order to get an overview of how reliable the results seem to be (Fig. 5.26).

Even though the test statistics per p-histogram are higher here than for the feature analysis, the obtained values should still be taken with care and only used to find the most interesting regions. The correlation between the test methods is still not high enough to prove their feasibility.

The tests agree about the two bins with the most occurrences of small p-values, $E_\mathrm{T}^{\tau_{\mathrm{cand}}} \in [43; 200)$ and $\phi \in [-0.5\pi; 0)$ (cp. Table 5.7). However, the results are more ambiguous than for the variables' p-values, as most entries are in the (50 %–100 %) bin.
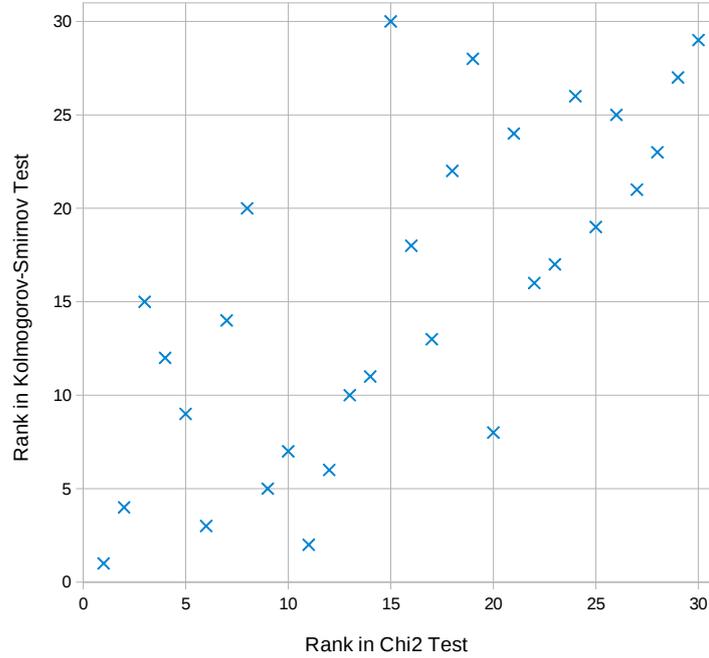
Figure 5.26: Ranking correlation for χ^2 and KS method (used on physics bins). Low numbers stand for high rankings, i. e. bad data/MC agreement according to the respective test algorithm. The correlation is slightly higher than for the feature histograms (Fig. 5.21).

Sum			χ^2			KS			Region
A	B	C	A	B	C	A	B	C	
97	117	38	40	67	19	57	50	19	$E_{\mathrm{T}}^{\tau_{\mathrm{cand}}} \in [43; 200)$
74	114	64	36	57	33	38	57	31	$\phi \in [-0.5\pi; 0)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 5.7: Quantifier outputs for physics bins, sorted by their sum (top 2 entries). The subcolumns represent the bins of the methods' p-values (cp. Table 5.5). The first three columns give the sum of the bin entries from the χ^2 and the Kolmogorov-Smirnov method, the other columns represent the methods' separate outcomes. For the full list, see Table A.2.

The respective p histograms are shown in Fig. 5.27 and Fig. 5.28. In both cases, large fractions (29 %–45 %) are in the first bin, but the rest is distributed plainly to some extent.

(a) χ^2 test.
(b) KS test.
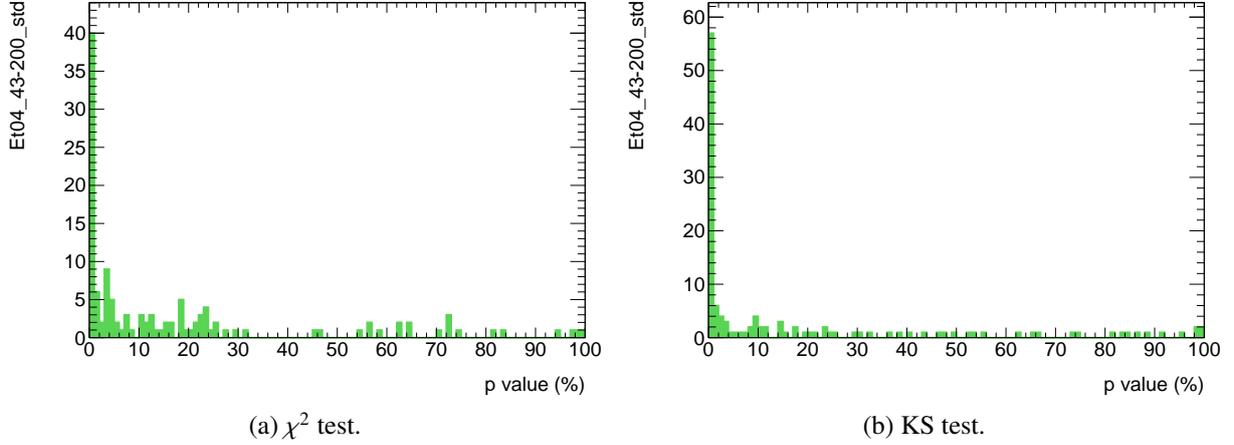
Figure 5.27: p histograms for $E_{\mathrm{T}}^{\tau_{\mathrm{cand}}} \in [43; 200)$.

(a) χ^2 test.
(b) KS test.
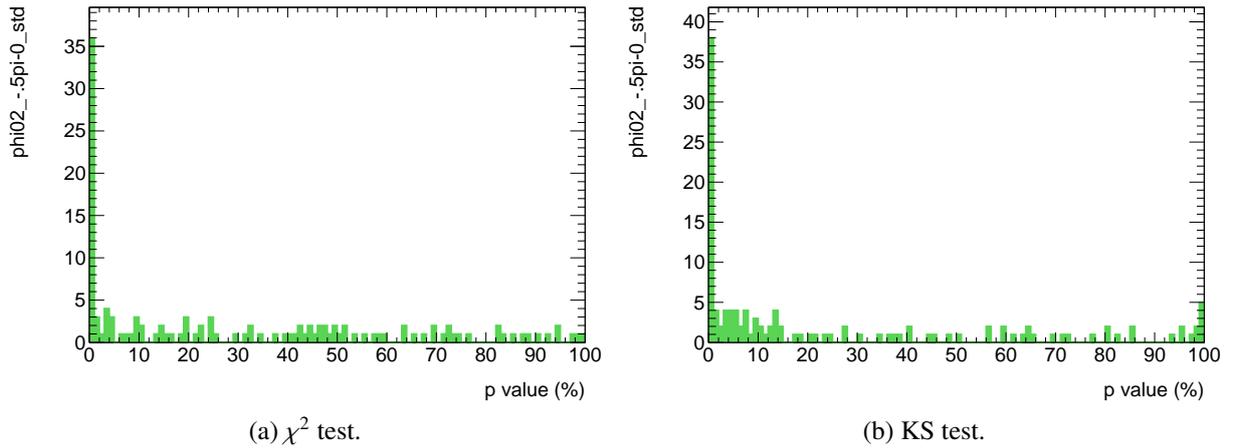
Figure 5.28: p histograms for $\phi \in [-0.5\pi; 0)$.

The poor results for the E_{T} bin indicate a negative energy dependency of the algorithm (worse data/MC accordance for higher E_{T}). On the other hand, the fact that one of the ϕ bins scores a result just as bad, even though physics are symmetric in ϕ, hints at the low significance of the test results.

Indeed, browsing through the graphs of the binned features [30], the effect of huge deviations for high E_{T} (or medium ϕ) could not be reproduced. In view of the fact that no other obvious effects were found in terms of dependency on any one of the physics binnings, not even other E_{T} regions or pileup, PanTau can be acknowledged as a stable algorithm regarding the tested parameters for the time being.

To authenticate this claim, the study has to be re-evaluated as soon as the cuts on minimum E_{T} for neutral EFOs (cp. Ch. 5.3.1) have been implemented and optimised.

Furthermore, a study on true τ leptons is essential to substantiate this hypothesis.

5.3.4 Overall Performance

As already argued in the beginning of Ch. 5.2, the χ^2 and KS tests are only meant to find a good starting point for the analysis. Browsing through all the PanTau features' inclusive and binned histograms, it is obvious that PanTau achieves very good results and most variables are well described by MC in all regions (at least when the statistics are sufficient).

Apart from the problems with N-EFO-like variables described in Ch. 5.3.1, that also have a large impact on many of the other variables, and, trivially, on the binning in $N(\mathrm{EFO}^{*\pm})$, the vast majority of the plots show a very good data/MC agreement. Even for the mentioned problems, a promising approach to fixing them is currently being implemented in eflowRec.

Chapter 6

Summary and Outlook

6.1 Conclusions from the Background Validation

The analysis of data/MC agreement in PanTau variables can be considered a success for Pan-Tau.

Most variables perform very well (see Ch. 5.1.9). For the ones showing large discrepancies, the conditions of the differences are well-understood, which mainly boil down to bad MC description of noise and pileup in the calorimeter, i. e. calorimeter clusters with low energy. In future revisions of eflowRec and PanTau, these problems will be attended to.

6.2 Sustainability of Developed Analysis Tools

Most of the tools that were created in order to perform these studies were integrated into the SFM framework (cp. Ch. 2.3.2) and are available for further application in ATLAS related analyses.

They can, for instance, be used to repeat the background analysis after changes in eflowRec and PanTau have been implemented. A large part of the algorithm could in principle also be re-used for similar studies on true τ leptons. However, these will be conducted by the ATLAS Tau WG by means of more complex tools.

6.3 Future Tasks and Challenges

The next step that has to be taken from here, as explained in Ch. 5.3, is to study the impact of energy cuts on neutral EFOs in eflowRec. Background studies have to be re-evaluated afterwards.

Naturally, studies on true τ leptons have to follow, which will hopefully reconfirm the good results of the background analysis.

This will provide sufficient evidence that PanTau is a reliable program and its results can be trusted. Subsequently, PanTau's performance with regards to decay mode separation (and, possibly, fake suppression) will be one of the main focuses.

Appendices

Appendix A

Tables

Sum			χ^2			KS			Variable
A	B	C	A	B	C	A	B	C	
10	0	0	5	0	0	5	0	0	NEFO
10	0	0	5	0	0	5	0	0	NNeutralEFO
9	1	0	4	1	0	5	0	0	NEFOOut02
8	2	0	4	1	0	4	1	0	EflowMeanEtAllToSumEt
8	2	0	4	1	0	4	1	0	EflowMeanEtNeutToSumEt
8	1	1	4	0	1	4	1	0	NEFOOut03
7	3	0	4	1	0	3	2	0	NHADNeutralEFO
7	3	0	3	2	0	4	1	0	RatioChargedToTotalNumber
7	2	1	2	2	1	5	0	0	NEFOIn02To04
6	3	1	2	2	1	4	1	0	NEFOIn03
6	3	1	2	2	1	4	1	0	NEFOOut01
5	2	3	3	0	2	2	2	1	EflowJetMomentDRChrg
4	5	1	0	4	1	4	1	0	EflowInvariantMassAllAboveEtCut
4	4	2	1	3	1	3	1	1	EflowJetMomentDRAll
4	4	2	2	2	1	2	2	1	EflowMaxDeltaPtLeadingCharged-QualifiedComponents
4	4	2	2	2	1	2	2	1	EflowMeanDeltaPtLeadingCharged-QualifiedComponents
3	7	0	2	3	0	1	4	0	NEFOIn01To04
3	6	1	1	4	0	2	2	1	EtIn01OverEtIn04InvSigm
3	6	1	1	4	0	2	2	1	JetAplanarity
3	5	2	1	3	1	2	2	1	EtIn01OverEtIn02
3	5	2	0	3	2	3	2	0	EtIn01To02OverTotalEtInvSigm
3	4	3	2	0	3	1	4	0	EflowInvariantMassChargedQualified
3	4	3	0	2	3	3	2	0	EflowInvariantMassNeutral
3	4	3	1	3	1	2	1	2	NChargedEFO
3	4	3	1	3	1	2	1	2	NChargedQualifiedEFO
2	8	0	0	5	0	2	3	0	EtIn01To02OverTotalEt
2	7	1	2	3	0	0	4	1	EflowMaxDRToJetAxis

Table A.1: Quantifier outputs for PanTau features. *(continued on next page)*

(continued from previous page)

Sum			χ^2			KS			Variable
2	7	1	1	3	1	1	4	0	EtIn01To02ToJetAxis
2	7	1	2	3	0	0	4	1	JetThrustMinor
2	6	2	0	3	2	2	3	0	EflowLeadingHADNeutralEt-ToSumHADNeutralEt
2	6	2	1	2	2	1	4	0	EflowMeanChrgDRToLeadingChrgEFO
2	6	2	1	4	0	1	2	2	EflowStdDevNeutDRToJetAxis
2	6	2	0	3	2	2	3	0	EtIn01OverEtIn02InvSigm
2	6	2	0	4	1	2	2	1	EtIn01OverEtIn04
2	6	2	1	3	1	1	3	1	JetSphericity
2	5	3	0	2	3	2	3	0	EflowHADNeutralEtToSumNeutralEt
2	5	3	1	3	1	1	2	2	EflowMeanChrgDR
2	5	3	0	3	2	2	2	1	EflowStdDevEtToSumEt
2	5	3	2	3	0	0	2	3	SeedRecoType
2	4	4	2	3	0	0	1	4	AbsSumCharge
2	4	4	1	2	2	1	2	2	EflowJetMomentDR2All
2	4	4	2	1	2	0	3	2	EtIn02To04OverTotalEtInvSigm
2	4	4	2	3	0	0	1	4	EtIn04To05OverTotalEt
2	4	4	2	3	0	0	1	4	EtIn04To05ToJetAxis
2	3	5	2	2	1	0	1	4	EflowAngleFirstNeutralSecondNeutral
2	3	5	2	1	2	0	2	3	EtIn02OverEtIn04InvSigm
2	3	5	2	1	2	0	2	3	EtIn02To03ToJetAxis
2	3	5	2	2	1	0	1	4	EtIn03To04ToJetAxis
2	3	5	2	1	2	0	2	3	SumCharge
2	2	6	2	1	2	0	1	4	EflowEMNeutralSumEt
2	2	6	2	2	1	0	0	5	JetFoxWolfram4
1	8	1	0	4	1	1	4	0	EflowLeadingHADNeutralEtToSumNeutralEt
1	7	2	0	3	2	1	4	0	EflowChargedToSumEt
1	7	2	0	4	1	1	3	1	EflowDRLeadingChargedLeadingNeutral
1	7	2	0	3	2	1	4	0	EflowHADNeutralEtToSumChargedEt
1	7	2	0	3	2	1	4	0	EflowHADNeutralEtToTotalEt
1	7	2	0	4	1	1	3	1	EflowInvariantMassCharged
1	7	2	1	3	1	0	4	1	EflowInvariantMassSignificantEFOs
1	7	2	1	3	1	0	4	1	EflowRatioDeltaPt12To13LeadingCharged-QualifiedComponents
1	7	2	1	3	1	0	4	1	EflowRatioDeltaPt23To13LeadingCharged-QualifiedComponents
1	7	2	1	4	0	0	3	2	JetThrustMajor
1	6	3	1	2	2	0	4	1	EflowEtLeadingChrgQualifiedOverTotalEt
1	6	3	1	3	1	0	3	2	EflowMeanEtHADNeutToSumEt

Table A.1: Quantifier outputs for PanTau features. *(continued on next page)*

Sum			χ^2			KS			Variable
1	6	3	1	4	0	0	2	3	EtIn02To04ToJetAxis
1	6	3	1	4	0	0	2	3	EtIn03To04OverEtIn04InvSigm
1	6	3	1	4	0	0	2	3	NEFOIn01
1	5	4	0	2	3	1	3	1	EflowStdDevE
1	5	4	1	2	2	0	3	2	EflowStdDevEtLeadingCharged-QualifiedComponents
1	5	4	0	2	3	1	3	1	EtIn01ToJetAxis
1	4	5	1	1	3	0	3	2	EflowAngleJetAxisLeadingChargedQualified
1	4	5	0	1	4	1	3	1	EflowChargedToNeutralEt
1	4	5	1	0	4	0	4	1	EflowHADNeutralSumEt
1	4	5	0	1	4	1	3	1	EflowStdDevEt
1	4	5	1	2	2	0	2	3	JetFoxWolframRatioFW2OverFW1
1	3	6	1	1	3	0	2	3	EflowAngleChargedAxisEMNeutralAxis
1	3	6	1	1	3	0	2	3	EflowAngleEMNeutralAxisJetAxis
1	3	6	1	3	1	0	0	5	JetFoxWolfram2
1	3	6	1	3	1	0	0	5	JetFoxWolframRatioFW234OverFW1pow4
1	2	7	1	2	2	0	0	5	JetFoxWolfram3
1	1	8	1	0	4	0	1	4	EflowAngleFirstChargedSecondNeutral
0	8	2	0	5	0	0	3	2	EflowChargedSumEt
0	7	3	0	4	1	0	3	2	EflowInvariantMassAll
0	7	3	0	3	2	0	4	1	EflowJetMomentDRNeut
0	7	3	0	3	2	0	4	1	EflowMeanEtChrgToSumEt
0	7	3	0	4	1	0	3	2	EflowStdDevDRToJetAxis
0	7	3	0	4	1	0	3	2	EflowStdDevHADNeutDRToJetAxis
0	7	3	0	4	1	0	3	2	EtIn02To04OverTotalEt
0	7	3	0	4	1	0	3	2	NEFOIn02
0	6	4	0	3	2	0	3	2	EflowMaxDRLeadingCharged-QualifiedComponents
0	6	4	0	3	2	0	3	2	EflowMeanHADNeutDR
0	6	4	0	3	2	0	3	2	EflowNeutralToSumEtInvSigm
0	6	4	0	2	3	0	4	1	EflowStdDevChrgE
0	6	4	0	3	2	0	3	2	JetPlanarity
0	5	5	0	3	2	0	2	3	EflowAngleChargedAxisJetAxis
0	5	5	0	3	2	0	2	3	EflowAngleChargedAxisNeutralAxis
0	5	5	0	2	3	0	3	2	EflowAnglePlaneFirstCharged-FirstNeutralSecondNeutral
0	5	5	0	1	4	0	4	1	EflowChargedToSumEtInvSigm
0	5	5	0	1	4	0	4	1	EflowStdDevChrgEt
0	5	5	0	2	3	0	3	2	EflowStdDevHADNeutEt

Table A.1: Quantifier outputs for PanTau features. *(continued on next page)*

(continued from previous page)

Sum			χ^2			KS			Variable
0	5	5	0	3	2	0	2	3	NEMNeutralEFO
0	4	6	0	2	3	0	2	3	EflowEMNeutralEtToTotalEt
0	4	6	0	2	3	0	2	3	EflowLeadingHADNeutralEt
0	4	6	0	4	1	0	0	5	EflowMeanEMNeutDR
0	4	6	0	2	3	0	2	3	EflowStdDevChrgDRToJetAxis
0	4	6	0	1	4	0	3	2	EflowStdDevHADNeutE
0	4	6	0	3	2	0	1	4	JetFoxWolframRatioFW1PlusFW2OverFW4
0	4	6	0	4	1	0	0	5	JetThrust
0	3	7	0	1	4	0	2	3	EflowAngleLeadingChrgQualified-2ndLeadingChrgQualified
0	3	7	0	2	3	0	1	4	EflowAngleNeutralAxisJetAxis
0	3	7	0	1	4	0	2	3	EflowInvariantMassEMNeutral
0	3	7	0	2	3	0	1	4	EflowMeanAngleBtwLeadingCharged-QualifiedComponents
0	3	7	0	3	2	0	0	5	EflowMeanHADNeutEt
0	3	7	0	2	3	0	1	4	EflowStdDevEMNeutDRToJetAxis
0	3	7	0	2	3	0	1	4	EflowStdDevNeutEt
0	3	7	0	3	2	0	0	5	JetFoxWolfram1
0	3	7	0	3	2	0	0	5	JetFoxWolframRatioFW4pow4OverFW1
0	3	7	0	1	4	0	2	3	JetOblateness
0	2	8	0	0	5	0	2	3	EflowEMNeutralEtToTotalEtInvSigm
0	2	8	0	1	4	0	1	4	EflowMaxAngleLeadingCharged-QualifiedComponents
0	2	8	0	2	3	0	0	5	EflowMeanHADNeutE
0	2	8	0	1	4	0	1	4	EflowNeutralSumEt
0	2	8	0	2	3	0	0	5	EflowStdDevNeutE
0	2	8	0	2	3	0	0	5	RatioChargedToNeutral
0	1	9	0	1	4	0	0	5	EflowDihedralAngleLeading-ChargedQualified
0	1	9	0	1	4	0	0	5	EflowSecondNeutralEt
0	0	10	0	0	5	0	0	5	EflowSumEAll

Table A.1: Quantifier outputs for PanTau features.
The subcolumns represent the bins of the methods' p-values (cp. Table 5.5). The first three columns give the sum of the bin entries from the χ^2 and the Kolmogorov-Smirnov method, the other columns represent the methods' separate outcomes.

Sum			χ^2			KS			Region
A	B	C	A	B	C	A	B	C	
97	117	38	40	67	19	57	50	19	$E_\mathrm{T}^{\tau\mathrm{cand}} \in [43; 200)$
74	114	64	36	57	33	38	57	31	$\phi \in [-0.5\pi; 0)$
70	111	71	26	69	31	44	42	40	ΔR to closest jet $\in [1.475; 5)$
69	105	48	30	59	22	39	46	26	$N(\mathrm{EFO}^{*\pm}) = 1$
66	112	74	32	46	48	34	66	26	$\eta \in [1.2; 2.8)$
65	109	78	32	57	37	33	52	41	$E_\mathrm{T}^{\tau\mathrm{cand}} \in [18; 26)$
64	114	74	27	60	39	37	54	35	$\phi \in [-\pi; -0.5\pi)$
64	107	81	34	49	43	30	58	38	$\mu \in [14; 15)$
63	109	66	26	57	36	37	52	30	$N(\mathrm{EFO}^{*\pm}) = 2$
63	100	89	27	57	42	36	43	47	$\eta \in [-2.8; -1.2)$
60	105	87	29	57	40	31	48	47	$N(\mathrm{EFO}^{*\pm}) = 3$
60	99	93	26	56	44	34	43	49	$\phi \in [0.5\pi; \pi)$
58	124	70	25	67	34	33	57	36	$\eta \in [0.4; 1.2)$
55	126	71	20	73	33	35	53	38	$\mu \in [10; 13)$
55	107	90	23	61	42	32	46	48	$E_\mathrm{T}^{\tau\mathrm{cand}} \in [26; 43)$
52	117	83	28	57	41	24	60	42	$\eta \in [-1.2; -0.4)$
50	111	91	24	61	41	26	50	50	$\mu \in [13; 14)$
49	104	99	19	57	50	30	47	49	ΔR to closest jet $\in [0.975; 1.475)$
45	127	80	18	69	39	27	58	41	$\eta \in [-0.4; 0.4)$
44	105	103	18	57	51	26	48	52	$\phi \in [0; 0.5\pi)$
43	122	87	22	64	40	21	58	47	$E_\mathrm{T}^{\tau\mathrm{cand}} \in [15; 18)$
39	112	101	16	64	46	23	48	55	ΔR to closest jet $\in [0.55; 0.65)$
34	121	97	16	60	50	18	61	47	ΔR to closest jet $\in [0.4; 0.55)$
34	118	100	20	53	53	14	65	47	$N(\mathrm{EFO}^{*\pm}) = 4$
31	113	108	17	60	49	14	53	59	ΔR to closest jet $\in [0.725; 0.975)$
29	117	106	18	63	45	11	54	61	$\mu \in [15; 16)$
28	102	122	21	42	63	7	60	59	$N(\mathrm{EFO}^{*\pm}) = 5$
25	121	106	15	58	53	10	63	53	ΔR to closest jet $\in [0.65; 0.725)$
25	74	151	25	44	56	0	30	95	$N(\mathrm{EFO}^{*\pm}) \in [6; 7]$
17	92	143	13	48	65	4	44	78	$\mu \in [16; 18]$

Table A.2: Quantifier outputs for physics bins.
The subcolumns represent the bins of the methods' p-values (cp. Table 5.5). The first three columns give the sum of the bin entries from the χ^2 and the Kolmogorov-Smirnov method, the other columns represent the methods' separate outcomes.

Sample Name	Number of Events
mc11_7TeV.105009.J0_pythia_jetjet.recon.AOD.e815_s1273_s1274_r3043_tid607212_00	1000
mc11_7TeV.105010.J1_pythia_jetjet.recon.AOD.e815_s1273_s1274_r3043_tid607213_00	1000
mc11_7TeV.105011.J2_pythia_jetjet.recon.AOD.e815_s1273_s1274_r3043_tid607214_00	1000
mc11_7TeV.105012.J3_pythia_jetjet.recon.AOD.e815_s1273_s1274_r3043_tid607215_00	1000
mc11_7TeV.105013.J4_pythia_jetjet.recon.AOD.e815_s1273_s1274_r3043_tid607216_00	1000
mc11_7TeV.105014.J5_pythia_jetjet.recon.AOD.e815_s1273_s1274_r3043_tid607217_00	1000
mc11_7TeV.106052.PythiaZtautau.recon.AOD.e825_s1356_s1353_r3043_tid644919_00	1000
mc11_7TeV.107054.PythiaWtaunu_incl.recon.AOD.e825_s1349_s1300_r3043_tid619639_00	1000
mc11_7TeV.107381.Pythia_Zprime_tautau_SSM250.recon.AOD.e916_s1351_s1300_r3097_tid644952_00	1000
mc11_7TeV.107382.Pythia_Zprime_tautau_SSM500.recon.AOD.e916_s1351_s1300_r3097_tid644953_00	1000

Table A.3: Used MC samples for consistency checks (cp. Ch. 4).

Sample Name	Cross Section
mc11_7TeV.107660.AlpgenJimmyZmumuNp0_pt20.merge.AOD.e835_s1299_s1300_r3043_r2993	835.85 pb
mc11_7TeV.107661.AlpgenJimmyZmumuNp1_pt20.merge.AOD.e835_s1299_s1300_r3043_r2993	167.675 pb
mc11_7TeV.107662.AlpgenJimmyZmumuNp2_pt20.merge.AOD.e835_s1299_s1300_r3043_r2993	50.4125 pb
mc11_7TeV.107663.AlpgenJimmyZmumuNp3_pt20.merge.AOD.e835_s1299_s1300_r3043_r2993	13.9875 pb
mc11_7TeV.107664.AlpgenJimmyZmumuNp4_pt20.merge.AOD.e835_s1299_s1300_r3043_r2993	3.4375 pb
mc11_7TeV.107665.AlpgenJimmyZmumuNp5_pt20.merge.AOD.e835_s1299_s1300_r3043_r2993	0.9625 pb

Table A.4: Used MC samples for background validation (cp. Ch. 5.1.2). NpX stands for $Z \to \mu\mu + X$ Jets.

Sample Name	Integrated Luminosity
data11_7TeV.periodM10.physics_Muons.PhysCont.AOD.t0pro09_v01	151.579 pb^{-1}
data11_7TeV.periodM9.physics_Muons.PhysCont.AOD.t0pro09_v01	0 pb^{-1}
data11_7TeV.periodM8.physics_Muons.PhysCont.AOD.t0pro09_v01	152.181 pb^{-1}
data11_7TeV.periodM6.physics_Muons.PhysCont.AOD.t0pro09_v01	28.26 pb^{-1}

Table A.5: Used data samples for background validation (cp. Ch. 5.1.2). In period M9, no run is included in the good run list.

Bibliography

[1] Commons Wikimedia. File:Standard Model of Elementary Particles.svg. http://commons.wikimedia.org, June 2006.

[2] Beringer J et al. (Particle Data Group). tau PR D86 010001. http://pdg.lbl.gov, January 2012.

[3] The ATLAS Collaboration et al. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008.

[4] ATLAS Experiment. ATLAS Webpage. http://www.atlas.ch, 2012.

[5] Lyndon Evans and Philip Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008.

[6] Deutsches Elektronen-Synchrotron DESY. LHC und Experimente. http://www.weltmaschine.de.

[7] Joao Pequenao. Computer generated image of the whole ATLAS detector. March 2008.

[8] ATLAS Magnet Group. ATLAS Magnet System. http://atlas-ma.web.cern.ch, 2002-2010.

[9] ATLAS Experiment. CERN Webpage. http://atlas.web.cern.ch/Atlas/Collaboration/, 2012.

[10] Stilling, J. SFM Introduction. https://wiki.physik.uni-bonn.de/atlas/public/index.php/SFM, 2012.

[11] Stan Lai and Anna Kaczmarska. Reconstruction and Identification of Hadronic Tau Decays with ATLAS. 2008.

[12] Tau WG. TauRec Twiki. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TauRec, 2009.

[13] Hodgkinson M C, Tovey D, and Duxfield R. Energy Flow Reconstruction with eflowRec Combined Reconstruction Software in Athena 12.0.6. Internal Note ATL-COM-PHYS-2007-082, (CERN), October 2007.

[14] Hodgkinson M, Tovey D, and Duxfield R. Energy Flow Reconstruction with the eflowRec Combined Reconstruction Software in Athena 15.6.9.8. Technical Report ATL-PHYS-INT-2011-031, (CERN), Geneva, Apr 2011.

[15] Fleischmann S. *Tau lepton reconstruction with energy flow and the search for R-parity violating supersymmetry at the ATLAS experiment*. PhD Thesis, (Universität Bonn), 2011.

[16] Prabhu R. *Studies into tau reconstruction, missing transverse energy and photon induced processes with the ATLAS detector at the LHC*. PhD Thesis, (Universität Bonn), September 2011.

[17] Limbach C. Leptonische Tau-Zerfälle in supersymmetrischen Prozessen im ATLAS-Detektor. Diplom Thesis, (Universität Bonn), October 2009.

[18] Limbach C. PhD Thesis (work in progress, Universität Bonn).

[19] Thomson M. Particle Flow Calorimetry and ILC Detector Design. Plenary Talk ALCPG 2009, (Albuquerque US), September 2009.

[20] Menke, S. TopologicalClustering Twiki. https://twiki.cern.ch/twiki/bin/viewauth/Atlas/TopologicalClustering, 2008.

[21] Limbach, C. PanTauListOfVariables. https://twiki.cern.ch/twiki/bin/viewauth/Atlas/PanTauListOfVariables, 2011.

[22] Fleischmann, S. TauJetFeatureExtractionTool: Calculating PanTau Features. https://svnweb.cern.ch/trac/atlasoff/browser/Reconstruction/PanTau/PanTauAlgs/tags/PanTauAlgs-00-01-02/src/TauJetFeatureExtractionTool.cxx#L89, 2012.

[23] Limbach, C. 1p 25-50 GeV Optimisation. https://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=198714, 2012.

[24] Limbach, C. 1pn 25-50 GeV Optimisation. https://indico.cern.ch/getFile.py/access?contribId=2&resId=1&materialId=slides&confId=198714, 2012.

[25] Limbach, C. 3p 25-50 GeV Optimisation. https://indico.cern.ch/getFile.py/access?contribId=2&resId=2&materialId=slides&confId=198714, 2012.

[26] Limbach, C. 3pn 25-50 Gev Optimisation. https://indico.cern.ch/getFile.py/access?contribId=2&resId=3&materialId=slides&confId=198714, 2012.

[27] Limbach, C. Tau vs Fake 25-50 Gev Optimisation. https://indico.cern.ch/getFile.py/access?contribId=2&resId=4&materialId=slides&confId=198714, 2012.

[28] Georges Aad et al. Measurement of the production cross section for Z/gamma* in association with jets in pp collisions at sqrt(s) = 7 TeV with the ATLAS detector. *Phys.Rev.*, D85:032009, 2012.

[29] Heinrich, Joel. Coverage of Error Bars for Poisson Data. http://www-cdf.fnal.gov/physics/statistics/notes/cdf6438_coverage.pdf, May 2003.

[30] Matberg, Sebastian. Histograms for PanTau Variables in Several Binnings. 736p., 94.02MB. https://www.dropbox.com/s/vpxonr7vst0e1pe/binnedGraphs_11-18.pdf, November 2012.

[31] Winter B T. Reconstruction of neutral pions in hadronic tau lepton decays in the ATLAS detector. MSc Thesis, (Universität Bonn), 2012.