GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fakultät für
Physik

Master's Thesis

Untersuchung der Modellierung und Performance vom ATLAS Tau Trigger mit Hilfe von Tag & Probe Analysen mit Run-2 LHC Daten bei $\sqrt{s} = 13$ TeV

Investigation of the Modelling and Performance of the ATLAS Tau Trigger using Tag & Probe Analyses with Run-2 LHC data at $\sqrt{s} = 13$ TeV

prepared by

Vanessa Annabelle Grauer

from Witzenhausen

at the II. Physikalisches Institut

Thesis number:	II.Physik-UniGö-MSc-2022/06
Thesis period:	15th May 2022 until 15th November 2022
First referee:	Prof. Dr. Stan Lai
Second referee:	Prof. Dr. Arnulf Quadt

Zusammenfassung

Die Modellierung des τ Lepton HLT Rekonstruktionsalgorithmus wurde untersucht mit Hilfe einer $Z \rightarrow \tau_{lep}\tau_{had}$ Tag und Probe Analyse. Die genutzten Datensätze sind von LHC Run-2 Proton-Proton Kollisionen und wurden in den Jahren 2017 und 2018 mit dem ATLAS Detektor aufgenommen, was einer Datenmenge von 44.3 fb^{-1} beziehungsweise 58.5 fb^{-1} entspricht.

Diese Analyse betrachtet verschiedene Strategien um die Modellierung und Performance des HLT für zukünftige RNN Identifikation Trainings des τ Lepton Trigger für den bevorstehenden Run-3 des LHC zu verbessern. Die vorgeschlagene Korrektur für die Modellierung ist eine datengetriebene Umgewichtung der Signal MC Simulation anhand von ausgewählten Identifikation Variablen. Dafür wurden die Variablen nach ihrer Modellierung eingestuft. Dies wurde sowohl für den 1-Prong als auch für den 3-Prong Kanal durchgeführt.

Die in dieser Arbeit präsentierten eins- und zweidimensionalen Gewichtungen sind sehr rentable Strategien um die Übereinstimmung von Daten und MC Simulation mit wenig Nachteilen oder technischem Kostenaufwand zu verbessern. Die Gewichtungsstrategien, die sich als Beste erwiesen haben sind:

- eine zweidimensionale Gewichtungsstrategie basierend auf zwei verschiedenen Identifikations Variablen den für 1-Prong Kanal und

- eine eindimensionale Gewichtungsstrategie basierend auf einer Identifikations Variable für den 3-Prong Kanal.

Abstract

The modelling of the τ lepton HLT reconstruction algorithm with a $Z \rightarrow \tau_{lep}\tau_{had}$ tag and probe analysis has been investigated. Datasets from the LHC Run-2 proton-proton collisions recorded in the years 2017 and 2018 with the ATLAS detector have been utilised, corresponding to 44.3 fb^{-1} and 58.5 fb^{-1}, respectively.

This analysis probes different strategies to improve the modelling and performance of the HLT for future RNN identification trainings for τ lepton triggers in the upcoming Run-3 of the LHC. The suggested modelling improvement is to apply a data-driven re-weighting on the signal MC sample based on chosen identification variables. For this purpose, the HLT identification variables have been ranked according to the modelling of their distributions in both the 1-prong and 3-prong decay modes.

The presented 1-dimensional and 2-dimensional re-weightings are both viable strategies to improve the data and MC agreement with little disadvantages or computing cost if the correct variables are chosen to calculate the signal weight from. The re-weightings that have proven to perform the best are:

- a 2-dimensional re-weighting strategy based on two different identification variables for the 1-prong channel and

- a 1-dimensional re-weighting strategy based on a chosen identification variable for the 3-prong channel.

Therefore, these re-weightings of the signal MC sample are recommended to use when training an τ lepton ID algorithm Run-3.

Contents

1. Introduction

In particle physics, scientists strive to probe the nature of the constituents of matter and underlying fundamental interactions. The theoretical framework that has been developed to summarise this is called the Standard Model of particle physics (SM) [1–3]. This framework gives a description of all the known fundamental particles and interactions, except gravity. For very high energies around the *Planck scale*[1], gravitational effects start to dominate at particle level, hence quantum field theories describing the particles become inadequate. It is therefore known that the SM is not a complete theory of the nature of particles.

To explore the properties of the SM and even beyond, precise machinery for the measurement of particle collisions is needed. The ATLAS detector [4] at the Large Hadron Collider (LHC) [5] at CERN in Geneva is such a machine. It consists of a variety of the latest technological advancements to measure the collisions of highly energetic proton-proton beams. One of the most important parts of the ATLAS detector is the trigger system [6]. In order to measure the large amount of data coming in during the collisions, the triggers are required to make fast decisions during the data-taking in real time, while still having to be as precise as possible in determining the type and properties of the incoming particles. This is done in two steps with the hardware-based Level-1 (L1) trigger for pre-filtering of events combined with the software-based High-Level Trigger (HLT) for more precise real-time preprocessing [7].

This analysis investigates some properties of the HLT with regards to the modelling reconstructed τ leptons. The τ lepton triggers are vital for a lot of analyses, as an accurate reconstruction of these particles is often needed for signatures in and beyond the SM. The investigations in this thesis are performed using a tag and probe analysis with $Z \to \tau^+\tau^-$ events. This is a method to make use of the unique signature of the SR for a final state with one τ lepton decaying into hadrons ($\tau \to hadr. + \nu_\tau$) and other one decaying into a muon and its neutrino ($\tau \to \mu\nu_\mu\nu_\tau$).

In this thesis, a possible strategy to improve the performance of the HLT is presented through application of weights to the signal Monte Carlo simulation for a set of variables.

---

[1] The Planck scale is an energy region around the order of $\mathcal{O}(10^{19}$ GeV).

The data used to carry out this analysis has been recorded at the ATLAS detector in 2017 and 2018 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV corresponding to 44.3 fb^{-1} and 58.5 fb^{-1}, respectively.

This thesis is structured starting with Chapter 2, which introduces some theoretical aspects of high energy particle physics at hadron colliders. Chapter 3 discusses the experimental setup around the LHC and the ATLAS detector, as well as the trigger system that is used to detect particles. Chapter 4 gives an overview over the analysis strategy, including a description of the tag and probe analysis in the $Z \to \tau^+\tau^-$ channel. Following that is the main part of the analysis with Chapters 5 and 6. First, the modelling of a set of variables used for HLT particle identification trainings and correlations to the corresponding offline variables will be presented. These results will be later used to construct several one- and two-dimensional signal re-weighting strategies in Chapter 6. The different re-weightings have been applied to the $Z \to \tau^+\tau^-$ signal Monte Carlo sample and their effects are evaluated to see if an improvement of the agreement between data measurements and Monte Carlo distributions of identification variables can be seen. A better variable modelling overall would be advantageous for future HLT trainings used for particle reconstruction. A recommendation for the best performing configurations will be given. Lastly, a conclusion and an outlook will be drawn in Chapter 7.

2. Theoretical Background

This Chapter will give a brief overview over the foundation of particle physics, the Standard Model. Together with the underlying symmetries, the mechanism of electroweak symmetry breaking will be introduced in Section 2.1. In Section 2.2, the τ lepton is presented as its properties are important to understand in the context of this thesis. Different aspects of high-energy particle physics at hadron colliders will be considered as well, as detector signatures and experimental results of fundamental particles are included.

2.1. The Standard Model of Particle Physics

The Standard Model of particle physics is one of the greatest achievements of modern science. It is a quantum field theory that describes all currently known particles and fundamental interactions, except the gravitational force in one mathematical framework. It is a theory that was able to give many predictions about particle physics that are consistent with experimental results to very high precision. A summary of the particles contained in the SM can be found in Figure 2.1.

The SM is composed of twelve spin-$\frac{1}{2}$ particles, four spin-1 gauge bosons and one spin-0 scalar boson. The spin-$\frac{1}{2}$ particles are called fermions and are classified into six quark and six leptons. These particles are the building blocks of visible matter. Furthermore, the fermions are organised in generations. The quarks are ordered in three up-type (u, c, t) and three down-type quarks (d, s, b) with positive and negative electric charge respectively. Quarks have a particular property regarding their electric charge. The up-type particles hold a charge of +2/3 of the elementary charge, whereas down-type quarks have -1/3.

The six leptons are further categorised in a similar manner to the quarks. They are distinguished in three negatively charged fermions, the electron (e^-) and the heavier muons (μ^-) and τ^- leptons, as well as three electrically neutral, near massless neutrinos ν_e, ν_μ and ν_τ within each generation.

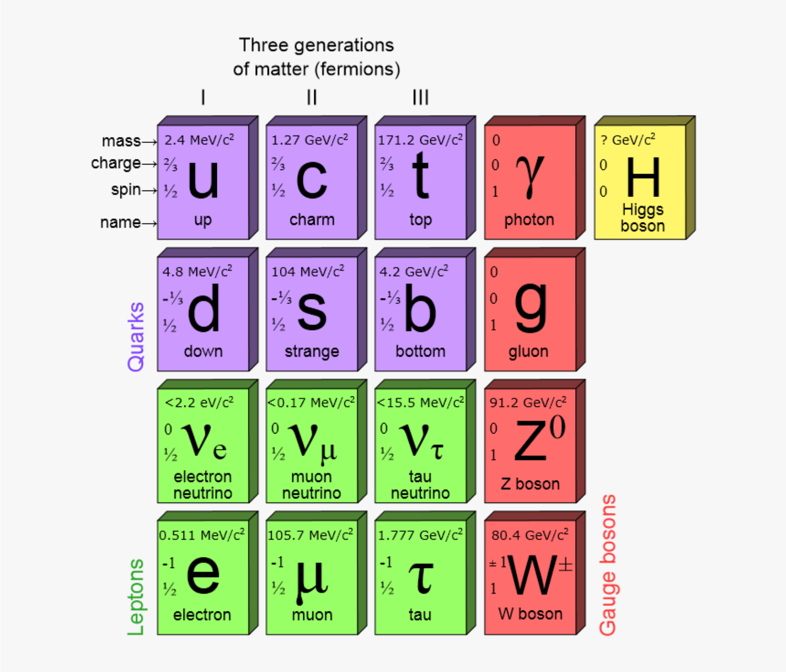


Figure 2.1.: Summary of the particles of the SM with some of their properties.

The generations of the SM are defined as the following [8]:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}.$$

The particles corresponding to each other between the generations have the same properties except for their masses, e.g. electrons and muons share the same properties, while the muon is significantly heavier with $m_\mu \approx 200 m_e$ [9]. Each particle has a counter-part, an anti-particle with the same mass but opposite quantum numbers, e.g. the electron's antiparticle is the positively charged positron.

The SM is a mathematical combination of different group theories. It obeys the

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

local gauge symmetry group, which describe the symmetry of the fundamental forces between all the known fundamental particles through field theories [8]. The gauge bosons mediate these fundamental forces.

The symmetry group $SU(3)_C$ defines the strong interaction. The theory describing the properties of this symmetry group is called *Quantum Chromodynamics*. The massless gluons (g) are the mediators of the strong force coupling to a colour charge. Only quarks and gluons carry such a charge and thus are the only particles that can interact strongly.

Quarks can interact via the electromagnetic and weak forces as well, due to also having an electric charge and weak isospin [10].

The electromagnetic force corresponds to the field theory of *Quantum Electrodynamics* and is described by an $U(1)_{EM}$ symmetry group. It is mediated by the massless photon (γ), which can only couple to electrically charged particles, namely quarks, charged leptons and the W boson.

The weak force is defined by the $SU(2)_L$ symmetry group. The mediators of the weak interaction, the W and the Z bosons, are the only massive gauge bosons in the standard model and are also the only particles that are able to couple to neutrinos [11, 12]. The W boson only couples to left-handed particles or right-handed anti-particles. In interactions with the W boson, a flavour changing charged current is involved, while the Z boson is neither flavour changing, nor does it transfer a charged current. The Z boson can couple to any particle with any handedness.

At hadron collider experiments, the electrically neutral Z boson is generally produced via quark-antiquark ($q\bar{q}$) annihilation. The large value of the mass of the Z boson is the reason that this particle decays quite fast, with a lifetime of 3×10^{-25} s and is therefore able to decay into most particles and its respective anti-particle except the heavier top quark [9]. Table 2.1 summarises the decay probabilities of the Z boson.

Z Decay Mode	Branching Fraction [%]
$\nu_\ell \bar{\nu}_\ell$	20.0
e^+e^-, $\mu^+\mu^-$, $\tau^+\tau^-$	3.4 (each)
$q\bar{q}$ (except t)	69.9

Table 2.1.: Branching fractions of all the possible Z boson decay processes. Lepton universality is assumed [9].

The combination of the $SU(2)_L$ and $U(1)_Y$ describes the unified electroweak sector [1]. In the process of the electroweak symmetry breaking (EWSB), the symmetry of $SU(2)_L \times U(1)_Y$ is spontaneously broken up to $U(1)_{EM}$ to produce the mediating particles of the electroweak sector [13]. This process yields a description for the masses of the massive gauge bosons. The basis of the mathematical description of EWSB is the Higgs potential, given by

$$V(\phi) = \mu^2 + (\phi^\dagger \phi) + \lambda(\phi^\dagger \phi)^2, \tag{2.1}$$

where $\phi = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}$ is a complex scalar doublet. The parameters μ^2 and λ are chosen to be smaller and greater than zero, respectively, in order to have a finite minimum in the potential [8]. This results in a "mexican-hat" potential, that is sketched in Figure 2.2. It can be seen that there is an infinite set of minima describing the vacuum states. Without loss of generality, $\langle\phi\rangle = \begin{pmatrix} 0 \\ v \end{pmatrix}$ can be chosen as the vacuum state. The constant v is called the *vacuum expectation value* (VEV), which can be calculated to $v \approx 246$ GeV [14]. This choice of the vacuum state spontaneously breaks the symmetry of the underlying Lagrangian.

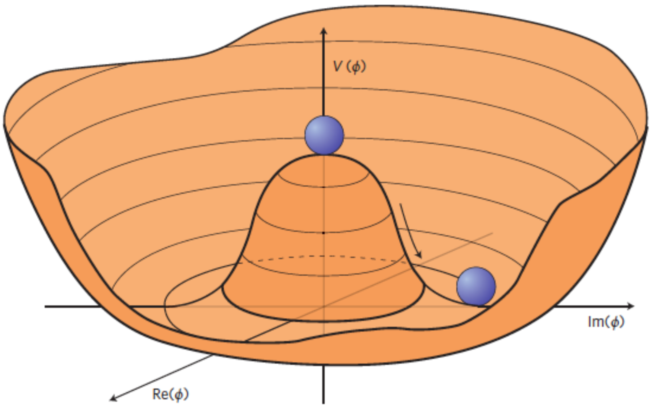


Figure 2.2.: The Higgs potential $V(\phi)$ with a sphere illustrating the process of symmetry breaking. The blue sphere demonstrates the arbitrary choice of a vacuum state occuring in spontaneous symmetry breaking [14].

The resulting mass term for the $W^{\pm}$ boson can be extracted as

$$m_W = \frac{1}{2} g_W v \approx 80.379 \pm 0.012 \text{ GeV [9]}.$$

The Z boson mass can be expressed as

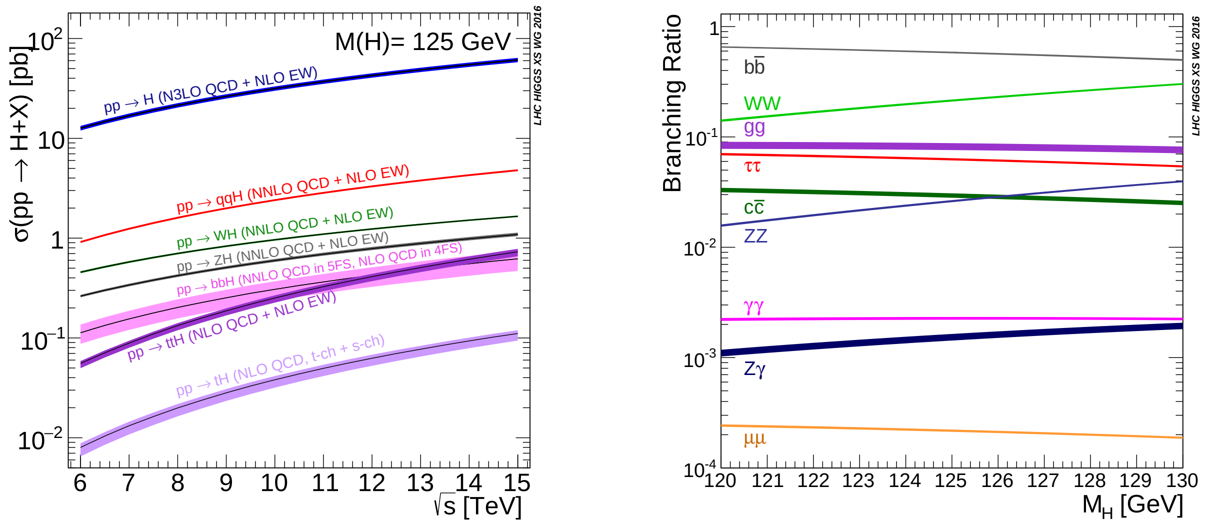
$$m_Z = \frac{1}{2} \frac{g_W}{\cos\theta_W} v \approx 91.1876 \pm 0.0021 \text{ GeV [9]}.$$

g_W is the weak coupling constant and $\theta_W \approx 28.74°$ is the empirically measured weak mixing angle [15].

In addition, the Higgs boson [13] is postulated by the Higgs Lagrangian and is therefore seen as the evidence for the electroweak symmetry breaking. Its discovery in 2012 completes the theory of the SM [16, 17]. A parameterisation of the Higgs boson mass can theoretically be extracted from EWSB as well, with

$$m_H = \sqrt{\lambda} v.$$

It has been measured to be $m_H = 125.25 \pm 0.17$ GeV [9].



(a) Cross-sections for the different Higgs boson production channels. The centre-of-mass energy range that is used is the current most relevant range at the LHC.

(b) Branching ratios for all the possible Higgs boson decay modes for Higgs boson masses between 120 GeV and 130 GeV

Figure 2.3.: Properties of Higgs boson production and decay at the LHC [18].

Investigations of the Higgs boson and its properties is one of the main focuses of the LHC. Figure 2.3(a) shows the cross sections of the dominant Higgs boson production channels for different centre-of-mass energies for proton-proton collisions. The most probable process is gluon-gluon fusion (ggF) shown in dark blue. Vector boson fusion (VBH) is the next most probable production process to occur and can be seen in the red curve. At the current energies of the LHC of around $\sqrt{s} = 13$ TeV, the cross sections are expected to be $\sigma_{ggF} = 48.58^{(+2.22)}_{-3.27}(\text{theory}) \pm 1.56(\text{PDF}+\alpha_s)$ pb and $\sigma_{VBF} = 3.782 \pm 0.052$ pb [18].

In Figure 2.3(b), one can see the branching ratios of the various decay modes of the Higgs boson for different Higgs masses. This plot underlines that the Higgs boson can decay to every fundamental particle in the SM, except for the top quark and depends on the particle mass. The decay to massless particles is realized through massive particle loops. Fermion masses can be generated through the Yukawa coupling to the Higgs field, yielding masses of

$$m_f = \frac{1}{\sqrt{2}} g_f v,$$

where g_f is the Yukawa coupling constant [19].

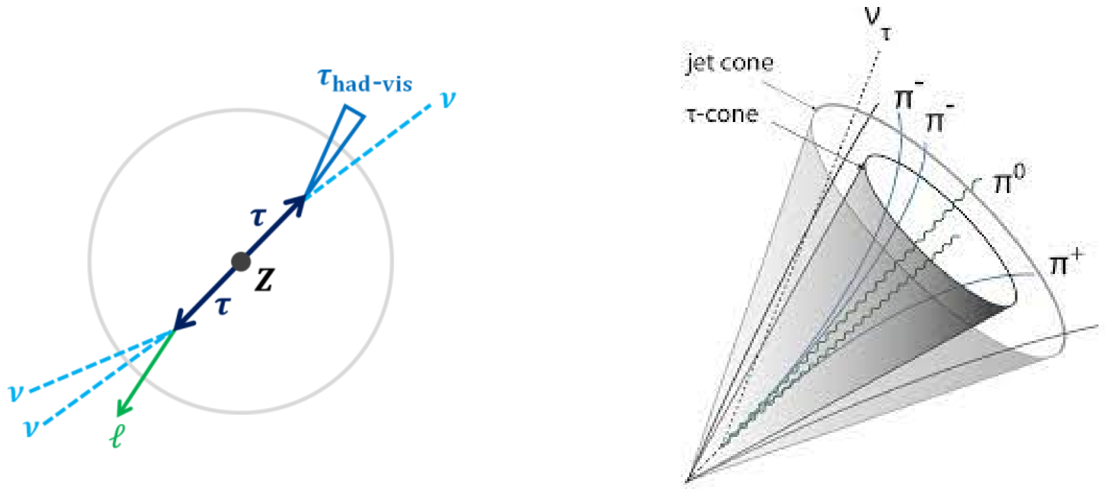
Although the SM offers a near complete picture of particle physics as of now, it is known that this theory has its shortcomings. For instance, it does not include gravity. Another problem of the SM is that it predicts the neutrino masses to be exactly zero, which is inconsistent with experimental results showing the observation of neutrino oscillations [20]. This is why particle physicists search for physics beyond the SM. One promising way to investigate physics beyond the SM is approached by investigating the Higgs sector. Ongoing investigations on $H \to \tau^+\tau^-$ decays or the searches for Higgs boson self-coupling in the $HH \to b\bar{b}\tau^+\tau^-$ channel are hoping to shed light on the shortcomings of the SM. To investigate and measure the Higgs boson properties in these channels, a good knowledge of τ leptons is crucial, as τ leptons couple more strongly to the Higgs boson than most other SM particles, with a branching ratio of ≈ 6 % (compare Figure 2.3(b)). Since one of the main backgrounds of these processes is the $Z \to \tau^+\tau^-$+heavy flavour channel, a good understanding of this background is needed in these investigations as well [21].

2.2. τ Leptons and their Properties

The τ lepton is the negatively charged lepton particle of the third generation. The mass of the τ lepton has been measured to be

$$m_\tau = 1776.86 \pm 0.12 \text{ MeV [9]},$$

and is the heaviest of the leptons. This relatively high mass makes this lepton interesting to investigate for Higgs boson studies, because the Higgs boson couples to fermions via the Yukawa-coupling. The τ lepton therefore exhibits the strongest coupling to the Higgs boson compared to most other particles in the SM. $Z \to \tau^+\tau^-$ decays are one of the main production channels for τ lepton pairs and is prominently used in performance studies for detector triggers. As already mentioned earlier, the probability for the $Z \to \tau^+\tau^-$ decay to occur is approximately 3.4 % (compare Table 2.1) [9]. The cross section for Z bosons decaying into leptons at $\sqrt{s} = 13$ TeV is $\sigma_{Z\to\ell\ell} \approx 699 \pm 5$ (syst) ± 17 (lumi) pb [22].
The τ lepton can decay leptonically or hadronically. If a τ lepton decays hadronically, a hadron is in the decay products and the signature in the particle detector is similar to that of hadronic jets from QCD processes. A leptonic decay implies, that the τ lepton decays into a lepton of a lower generation and respective neutrinos. The probability for a τ lepton to decay to either an electron or a muon is at approximately 36 % in total. A depiction, where the τ leptons decay either hadronically or leptonically is drawn in Figure 2.4(a). For the hadronic decay mode, the signatures are further differentiated into decays with one charged hadron or three charged hadrons in the final state. These final states are called 1-prong and 3-prong and have decay probabilities of 49 % and 15 %, respectively.



(a) $Z \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ decay mode

(b) Comparison of a τ lepton jet vs. a hadron induced jet

Figure 2.4.: The relevant τ lepton decay mode and a visualisation of a τ jet.

Further classification of the 1-prong and 3-prong final states can be retrieved from Table 2.2, as well as a summary of the possible τ lepton decays can be seen in Figure 2.5 [9]. Due to its short decay time of $\tau = 2.9 \times 10^{-13}$ s, the τ lepton can only be investigated through its decay products in particle experiments. Additionally, the τ lepton has a unique signature in the detector. The jets arising from the τ lepton start at a specific primary vertex. Furthermore, the jets widths are smaller than the ones of jets coming from QCD [9]. These properties are visualised in Figure 2.4 (b), where a comparison of a hadronic jet and a τ lepton jet signature can be seen.

One method that makes use of the properties of the τ lepton decay products is called the tag and probe method, which focuses on the signature that can be seen in Figure 2.4 (b) in order to investigate the reconstruction of τ lepton trigger in particle detectors. The tag and probe method will be introduced in more detail in Chapter 4.

	Decay Mode	Branching Fraction [%]
1-prong	$\tau^\pm \to \pi^\pm \nu_\tau$	10.8
	$\tau^\pm \to \pi^\pm \pi^0 \nu_\tau$	25.5
	$\tau^\pm \to \pi^\pm 2\pi^0 \nu_\tau$	9.3
	$\tau^\pm \to \pi^\pm 3\pi^0 \nu_\tau$	1.0
	$\tau^\pm \to K^\pm \nu_\tau +$ neutral particles	1.5
3-prong	$\tau^\pm \to 2\pi^\pm \pi^\mp \nu_\tau$	9.0
	$\tau^\pm \to 2\pi^\pm \pi^\mp \pi^0 \nu_\tau$	2.7
	$\tau^\pm \to K^\pm \pi^\pm \pi^\mp \nu_\tau +$ neutral particles	0.3

Table 2.2.: Summary of the decay probabilities of the τ lepton decaying hadronically for the 1-prong and 3-prong final states [9].

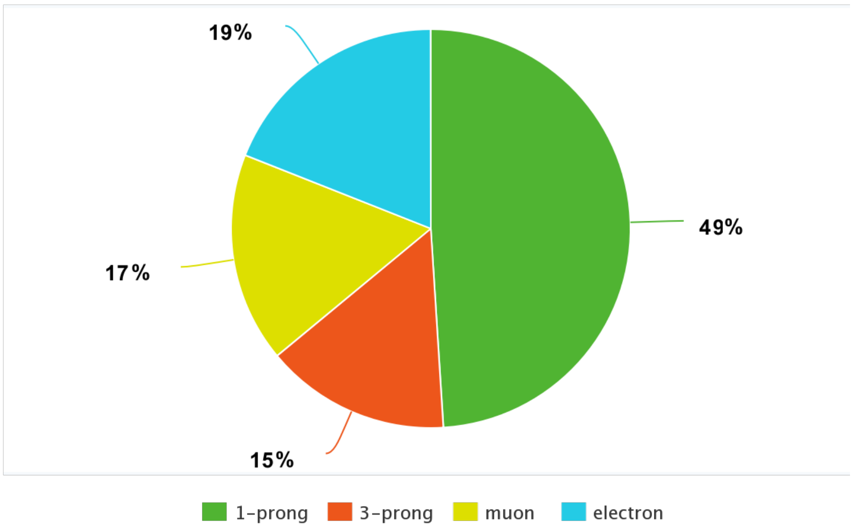


Figure 2.5.: Pie chart of possible τ Lepton decays. The decay modes of 1-prong and 3-prong are all the channels mentioned in Table 2.2. Muon and electron include the corresponding neutrinos needed for these processes [9]. The numbers in this chart are rounded and higher-pronged decay modes than the 3-prong decay, e.g. the 5-prong decay mode, are not depicted.

3. The Large Hadron Collider and The ATLAS Experiment

In the following Chapter, the experimental setup for particle detection and identification is presented. The data analysed in this thesis stems from proton-proton collisions at the LHC at CERN in Geneva that has been recorded with the ATLAS experiment, one of the several detectors installed at the LHC. Throughout Sections 3.1 and 3.2, a summary of the technical components will be given for the LHC and the ATLAS experiment, respectively. Lastly in Section 3.3, the trigger system of the ATLAS detector will be explained in more detail as the High-Level Trigger (HLT) will be a core part of the studies presented in this thesis.

3.1. The LHC

The LHC is the largest accelerator complex in the world; with 27 km in circumference, it is located approximately 100 m underground at CERN in Geneva, Switzerland and collides proton-proton beams with a centre-of-mass energy of up to $\sqrt{s} = 14$ TeV. A ring of superconducting dipole and quadrupole magnets is responsible to accelerate and focus the particle beams. There are four major experiments at the LHC: ATLAS [4], CMS [23], LHCb [24] and ALICE [25]. All of these experiments entail large collaborations of a various number of scientists. The location of the detectors in the accelerator ring and the whole accelerator complex at CERN can be seen in Figure 3.1.

To accelerate the proton beams, the particles are sent through Linac 2, which accelerates the beams up to 50 MeV. Afterwards, the bunches pass the Proton Synchrotron Booster and the Proton Synchrotron, with the energies increasing up to 25 GeV. In the final steps the particles are accelerated by the Super Proton Synchroton up to 450 GeV and are injected in the LHC to reach the goal energy of 6.8 TeV per particle beam and bunch crossings at a rate of 25 ns.

One of the other technical goals for the LHC is a high luminosity in order to have large

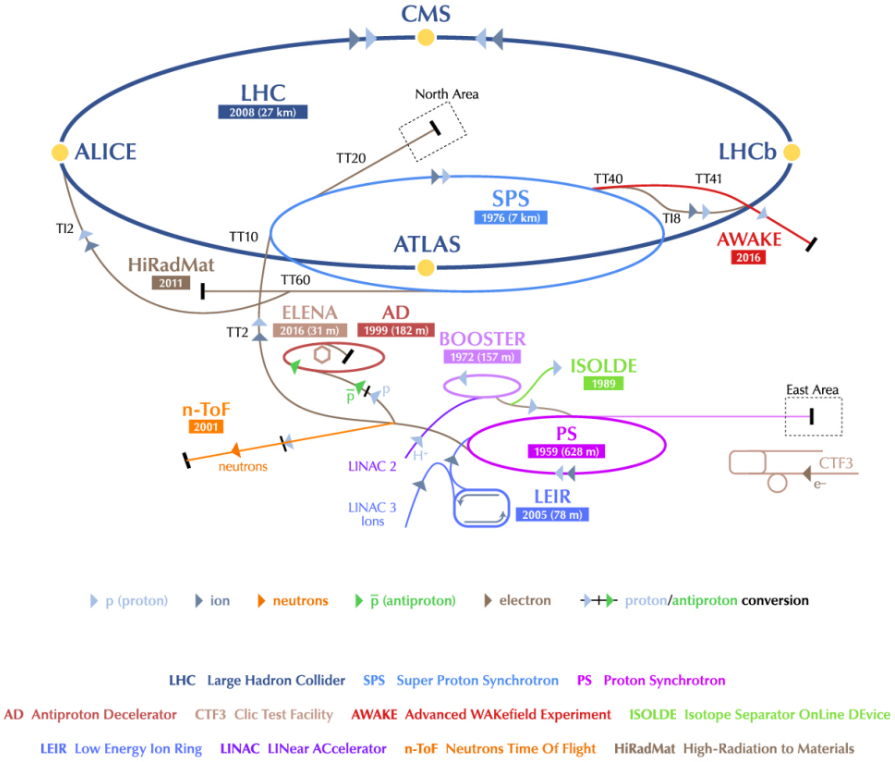


Figure 3.1.: The Large Hadron Collider accelerator complex.

datasets to evaluate. The luminosity is the ratio of detected events to a cross section σ,

$$L = \frac{1}{\sigma} \cdot \frac{dN}{dt}.$$

Usually, the integrated luminosity, defined as $L_{int} = \int L dt$, is considered for particle accelerator performance. Throughout this thesis, all luminosities used, denoted with L will refer to the time integrated luminosity.

3.2. The ATLAS Detector at the LHC

The ATLAS (A Toroidal LHC ApparatuS) detector is a general-purpose detector. The cylindrical multi-layer detector system has a 4π solid angle coverage around the interaction point to observe as many particle interactions as possible in a short amount of time during the collisions. The basic composition of the detector can be seen in Figure 3.2 [4].

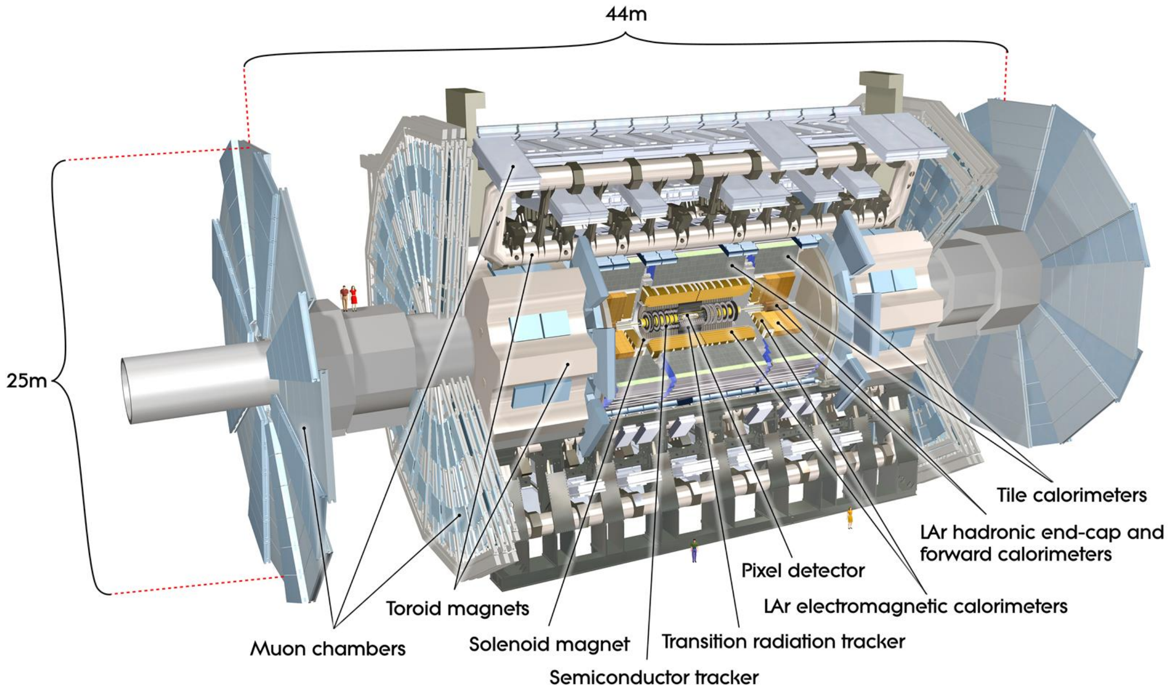


Figure 3.2.: A 3D model of the ATLAS detector system [26].

The Atlas experiment employs a right-handed coordinate system that matches the detector geometry, with the nominal interaction point as the origin of this coordinate system. The x-axis is defined as pointing from the interaction point to the centre of the accelerator ring of the Lhc and the positive y-axis points upwards. The z-axis is located in the same direction as the particle beam. The azimuthal angle ϕ extends around the beam axis and the polar angle θ is defined as the angle from the beam axis. Usually, instead of θ, the pseudorapidity is used. The pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$, because it is suitable for the detector geometry at hand [4].
The main instruments of the Atlas detector can be categorised as the following:

- **The inner detector**: The charged particles after the proton-proton-collisions are recorded in this part of the detector as tracks. A 2 T central magnetic field parallel to the beam axis is applied by using a large solenoid magnet, completely surrounding the detector. The inner detector has a length of 6.2m and a diameter of 2.1m. Furthermore, the inner detector consists of a pixel detector, a silicon strip detector and a transition radiation tracker (TRT). The pixel part of the detector is placed close to the initial beam interaction point to provide a high resolution measurement for track and vertex reconstruction in a very dense environment. It consists of approximately 92 million silicon pixels. The strip detector that is build around the pixel detector is made of approximately 6 million silicon readout strips. Together, these two detectors cover a range of $|\eta| < 2.5$ Finally, the TRT makes use of the fact that light charged particles emit transition radiation when passing through matter to record more particle information. This property is especially useful for electron identification. The tracker consists of > 300000 straw tubes filled with Xenon, Carbon Dioxide and Oxygen. The inner detector is able to handle the collider requirements such as bunch spacing and pile-up in a short amount of time [4].

- **The Calorimeter System**: All SM particles except neutrinos and the muon lepton shower by energy depletion in the calorimeter system. Using this information together with the information acquired by the inner detector, it is possible to reconstruct the 4-vector for the incident particles quite well. The calorimeter system is split up into an electromagnetic and a hadronic calorimeter, the ECAL and the HCAL. Both of the ECAL and the HCAL are sampling calorimeters. The ECAL consists of liquid argon (LAr) active and passive material made of lead to cover a thickness of > 22 radiation lengths (X_0) in the barrel and $> 24X_0$ in the end-caps. The barrel and end-caps cover a region of $|\eta| < 1.475$ and $1.35 < |\eta| < 3.2$, respectively. It is suitable to detect energy showers coming from electromagnetic particles, giving a precise energy measurement of electrons and photons. To detect energy showers from hadronic particles on the other hand, the HCAL is needed. The HCAL is a LAr calorimeter that is placed outside of the ECAL and its thickness corresponds to approximately ten interaction lengths λ. The HCAL is similar to the ECAL, split up into barrels, end-caps and forward calorimeter. In the barrels and extended barrels, a scintillating tile calorimeter is installed, covering $|\eta| < 1.7$. The passive material used here is steel. The end-caps and forward calorimeter are made of LAr with a spatial coverage of $1.5 < |\eta| < 3.2$ and $3.1 < |\eta| < 4.9$, respectively [4].
- **The Muon System**: A tracking detector, used to detect muons, which pass through the calorimeter system undetected. Since muons are minimal ionising particles, the process of Bremsstrahlung is suppressed. The magnetic system is made of three magnets to create toroidal electromagnetic fields with field strengths ranging from 0.5 to 1 T. The tracking is done by cathode strip chambers and resistive plate chambers for triggering in the inner layer. In the outer layer, there are cathode strip chambers (a kind of multiwire proportional chambers) to cover high pseudorapidities and thin gap chambers to obtain trigger information [4].

A summary of the resolutions and η coverages of the different parts of the ATLAS detector can be found in Table 3.1.

Detector Component	Required resolution	η coverage Measurement
Tracker	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	±2.5
EM calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	±3.2
Hadronic calorimeter (jets)		
barrel and end-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	±3.2
forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \lvert\eta\rvert < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\%$ at $p_T = 1$ TeV	±2.7

Table 3.1.: Resolution and η coverage of the different parts of the ATLAS detector. The units for E and p_T are in GeV [4].

3.3. The ATLAS Trigger

The particle collisions deliver a large amount of data in a short amount of time. As saving all of this data is neither possible nor feasible, a trigger system is needed in order to reduce the rate at which the data is saved. Due to unpredictable QCD processes, soft scatterings and other processes that are not of much interest, there is much data that needs to be filtered out of the collision rates that exceed 40 MHz. The uninteresting data occurring in the detector makes up a large portion of the particle collisions. The trigger system implemented in the ATLAS detector is a multi-level trigger system which splits up into a hardware-based Level-1 (L1) trigger and a software-based high-level trigger (HLT). The combination of these systems makes it possible to identify a variety of particles and jets, ranging from muons and τ leptons to b-jets and other hadronic energy depositions. Jets are the signature that occurs in a detector, when particles are depleting their kinetic energy when they are slowed down by the detector material. Jets build in scatter experiments starting from particles with colour charge. The jets collimate from surrounding hadronic objects in order to obey colour confinement. A schematic setup of the trigger system can be seen in Figure 3.3. This thesis will focus on the trigger settings that are used for τ lepton reconstruction.

The L1 trigger is responsible for the first rate reduction from 40 MHz down to 100 kHz with a latency of 2.5 μs and is derived from the energy detected in ECAL, the HCAL and the muon system. The L1 trigger system is split up into a calorimeter part and a muon part for the initial data acquisition, consisting of multi-chip modules. After that, the data is processed further in the central trigger and finally the accepted data is sent to the detector read-out. This setup can be understood in Figure 3.3 in the Level-1 block.

For the τ lepton trigger, the L1 provides a geometry for Regions of Interest (RoI) around the tau candidate. From Figure 3.4, it can be seen that these RoIs are defined by $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ trigger tower blocks forming a 2×2 core region and a 4×4 isolation region. The information about the τ candidate from the L1 is then sent to the HLT. These RoIs are particularly important for the identification of the visible part of the hadronically decaying τ leptons, the $\tau_{had,vis}$ candidates [6], [7]. A more detailed discussion about L1 RoIs will be given in Chapter 4.

The HLT trigger reduces the data coming from the L1 trigger further to a rate of 1 kHz with a latency of 200 ms for permanent storage with the help of > 28000 processors.

In the HLT τ lepton trigger, the candidate processing is split into a *calo-only* preselection, a *track* preselection and an *offline-like* selection. Both calo-only and track preselection use information from the RoIs and the Fast TracKer (FTK), respectively. In these selection steps, the energies are calibrated and additional requirements on the candidates are applied via algorithms. The offline-like selection utilises algorithms that are as close to an offline analysis as possible to achieve good efficiencies given the time restrictions. In high energy particle physics, *online* calculations are done in real-time by e.g. the HLT, while *offline* calculations are performed with local analyses with more time and computing resources available. For online τ lepton identification (ID), usually a boosted decision tree (BDT) is applied for the trigger decision. A decision tree is an algorithm used for particle classification, that constructs value prediction based on a set of input variables. If the trees are incrementally build and dependent of earlier trees, the decision tree is *boosted.* Recurrent neural networks (RNN) on the other hand, have proven to exceed the signal efficiency of decision trees for τ leptons (see e.g. [27]), hence why efforts are being made to switch to RNNs for the trigger decision. RNNs are a class of artificial neural networks that are suitable to predict dynamic systems. More details about this in the context of the τ lepton trigger will be given in Chapter 4.

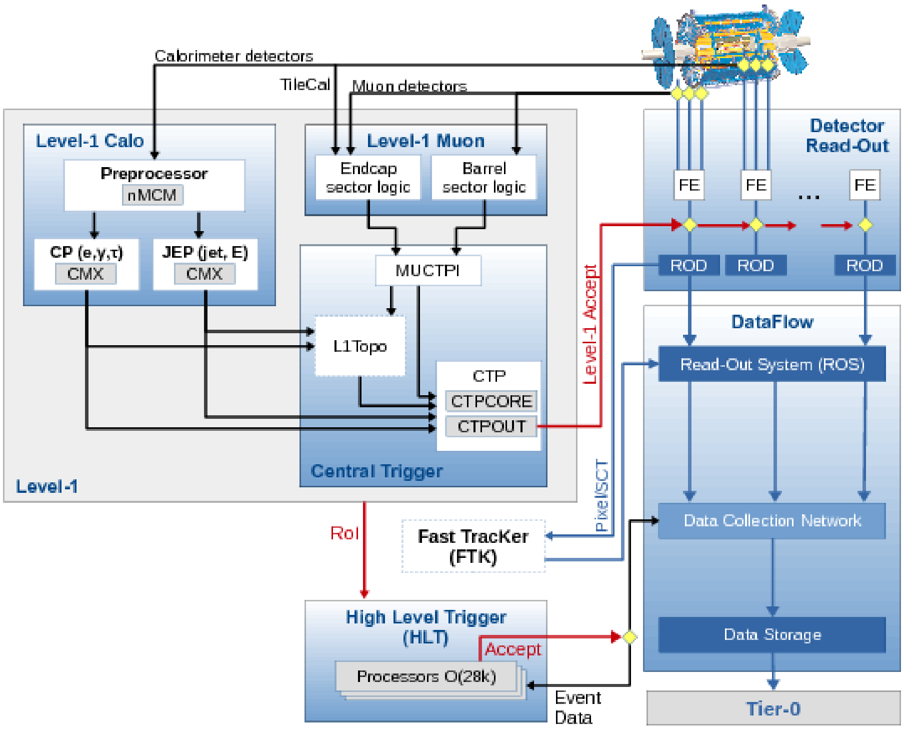


Figure 3.3.: The trigger system at the ATLAS experiment [6]. It can be seen that the trigger system is split up into the L1 and HLT. The L1 is a hardware based trigger, that collects data from all the detector parts and selects RoIs to send to the HLT trigger for further event filtering, which are then saved into the data storage.

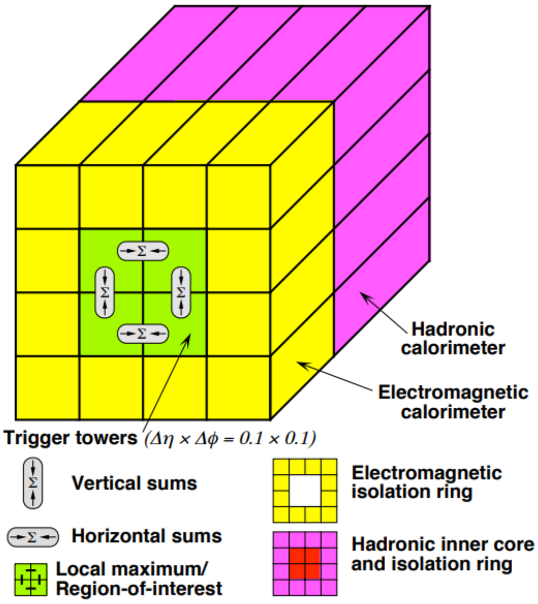


Figure 3.4.: Schematic setup of the EM calorimeter tower used for the identification of RoIs in the tau trigger algorithm [7]. The parts of the hadronic calorimeter trigger tower can also be seen. The processed data is then sent to the HLT.

4. $Z \rightarrow \tau\tau$ Tag and Probe Analysis in the $\tau_{\text{lep}}\tau_{\text{had}}$ Final State

In this Chapter, a more technical description of the analysis about an investigation of the HLT τ lepton trigger modelling is presented as well as a motivation for this analysis is given. Beginning in Section 4.1, the object reconstruction together with the specific τ trigger are explained. Additionally, the method of identifying the τ lepton candidates with a BDT and the variables used to train this decision tree are defined. In Section 4.2, the tag and probe method to reconstruct the $\tau_{lep}\tau_{had}$ signal region that is used in this thesis is motivated and introduced. Finally in Section 4.3, the offline event selection and background estimations of the tag and probe analysis are presented.

4.1. Object Reconstruction and Identification

A large part of the analysis in this thesis relies on an accurate and efficient reconstruction of hadronically decaying τ leptons. To achieve this, specific τ trigger algorithms have been developed. In general, these triggers function similar to other particle triggers, i.e. split into a two layered subsystem. The main goal of this thesis is to improve the second, software-based part of the trigger chain, the HLT for τ lepton identification.

A summary of the τ lepton trigger process with respect to the processing time can be understood in Figure 4.1. During data taking, the measured particle information coming from the calorimeters first gets sent to the L1 trigger. As already mentioned in Section 3.3, the L1 selects RoIs for the hadronically decaying τ leptons with hardware based trigger towers. These towers have the dimensions of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ with an η-coverage of $|\eta| < 2.5$. The τ-triggers identify $\tau_{had,vis}$ candidates by requiring certain p_T or energy thresholds within the core region and an isolation region. The core region is a trigger tower cluster with $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$. The isolation region is defined as the $\Delta\eta \times \Delta\phi = 0.4 \times 0.4$ ring of trigger tower clusters around the core region [28].

Based on the η and ϕ coordinates, the distance parameter ΔR can be defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{4.1}$$

This parameter is used in trigger algorithms for $\tau_{had,vis}$ candidate selection. Due to the short computing times of the L1, this trigger is not precise in reconstructing energies. Without sufficient time for decision making and specifically calibrated clustering algorithms, this trigger mainly filters the large amount of data [28]. After passing the L1, the HLT can use the information from Level 1 to apply reconstruction algorithms for further processing.

For the proceeding reconstruction of the $\tau_{had,vis}$ candidate, the HLT applies algorithms in three consecutive steps with rising requirements on the CPU. In the first step, the *calo-only preselection*, the τ lepton four vectors are reconstructed with the information gained from the *TopoClusters*, which are energy calibrated calorimeter cells in a $\Delta R = 0.2$ cone around the τ lepton direction. In this step, a calibration for the τ energy scale as well as pile-up corrections are applied [28, 29].

Afterwards, track information is added to the candidates in the *track preselection* step. This is done by applying the *two-stage fast tracker*, which is a pattern recognition algorithm. The two stages refer to the ΔR requirements on the p_T tracks. The first stage is to identify the highest p_T track in the cone $\Delta R < 0.1$ around the τ lepton direction, while the second stage identifies the tracks coming from the same incident point as the leading p_T track in the $\Delta R < 0.4$ cone [30].

For the final step, the HLT τ lepton trigger applies a BDT cut along with other preselection cuts to the $\tau_{had,vis}$ lepton candidates that are explained in Section 4.2.

In order for the HLT to be able to select the τ leptons as close to an offline analysis as possible in terms of performance, a BDT is trained on a set of different object identification variables. Consequently, each τ lepton candidate is assigned with a BDT score, which is used to create baseline working points that are set at different identification efficiency criteria [29]. A detailed description of a BDT training within a similar context can be found in Ref. [31]. The variables used in the training and within the scope of this analysis are defined as the following [32]:

- p_T: Transverse momentum of the leading τ lepton.
- ΔR_{max}: The maximum distance ΔR between the $\tau_{had,vis}$ direction and other tracks associated to the $\tau_{had,vis}$ candidate in the core region.
- f_{cent}: The centrality fraction. The fraction of energy deposited in the region

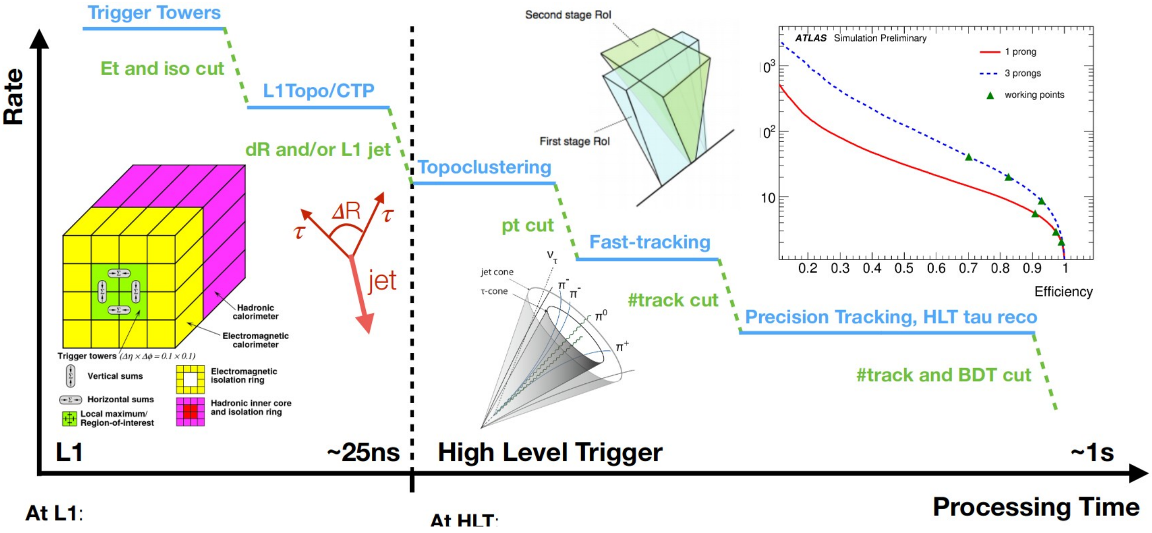


Figure 4.1.: A sketch of the τ-trigger candidate selection chain with respect to the rates and the processing time [28].

$\Delta R < 0.1$ compared to the energy in the region $\Delta R < 0.2$ associated to the $\tau_{had,vis}$ candidate.

- $\bar{d}_{innerTrack}$: Average distance in the transverse plane of the inner track to the $\tau_{had,vis}$ direction.
- m_{track}: Invariant mass of the sum of the four momenta of the tracks of all tracks in the core and isolation regions, assuming a pion mass for each track.
- $f_{EM}^{track-HAD}$: Fraction of the $\pi^{\pm}$ energy, defined as the electromagnetic energy deposited in the EM calorimeter in the core region of the associated $\tau_{had,vis}$ candidate with respect to the total electromagnetic energy.
- f_{track}^{EM}: Ratio of the electromagnetic energy from the $\tau_{had,vis}$ candidate to the sum of the transverse track momenta.
- $\frac{E_T}{p_T^{lead}}$: Ratio of the sum of transverse energy deposited in the core region of the $\tau_{had,vis}$ candidate and momentum of the leading track in the core region.
- $|S_{lead}|$: Significance of the transverse impact parameter b of the leading track. Usually, the absolute value of the significance is used. This variable is only relevant for the 1-prong $\tau_{had,vis}$ decay. In this case, the significance is the measured impact parameter, divided by its uncertainty.
- S_T^{flight}: Significance of the transverse flight path. This variable describes the decay length in the transverse plane of the $\tau_{had,vis}$ that has more than one tracks associated

with it. Therefore, this variable is only relevant for multi-prong $\tau_{had,vis}$ candidates. In this case, the significance is the measured flight path distance in the transverse plane, divided by its uncertainty.

- p_T^{ratio}: Ratio of a p_T estimation of the $\tau_{had,vis}$ using the vector sum of track momenta and cluster energy in the core region, to the calorimeter measurement of the $\tau_{had,vis}$ p_T.

The distributions of all listed variables for both the 1-prong and 3-prong channels can be found in the Appendix A.

For future runs, the τ-trigger tuning will switch over to recurrent neural networks (RNN), as these algorithms have proven in the past to perform with higher efficiencies [27]. Most offline analyses already use an RNN for τ lepton ID. The content of Chapter 6 will evaluate if there is a recommendation about the variables used for the training that can be made to improve the RNN considered for the HLT even further.

4.2. Tag and Probe Method

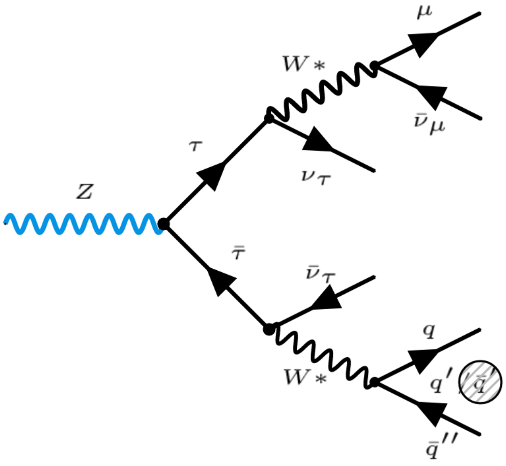


Figure 4.2.: The $Z \rightarrow \tau_{lep}\tau_{had}$ channel, used in the tag and probe analysis for the τ-trigger studies.

The tag and probe method with the process $Z \rightarrow \tau_{lep}\tau_{had}$ is suitable for τ-trigger studies, as the $Z \rightarrow \tau_{lep}\tau_{had}$ channel provides high statistics and purity at lower $\tau_{had,vis}$ p_T regions between 20 GeV and 60 GeV[1].

---

[1] There are less events for higher p_T regions. A $t\bar{t}$ tag and probe analysis can be used to cover these regions. See more about the $t\bar{t}$ tag and probe analysis in Appendix B

The tag and probe analysis of the $Z \to \tau_{lep}\tau_{had}$ channel uses the kinematics of one leptonically decaying and one hadronically decaying τ lepton coming from a Z resonance. This can be seen in Figure 4.2. In the process of the τ lepton decay, the first τ lepton decays into a muon and missing energy. This muon is identified (tag) through a single muon trigger with a loose BDT ID, where every candidate is required to have $p_T > 27.3$ GeV and $|\eta| < 2.5$ to pass the trigger requirements [33]. The tracks are reconstructed with the information from the inner detector and the muon spectrometer. Furthermore, events are vetoed if more than one lepton is present, as specifically the region with only one muon is of interest.

To probe the $\tau_{had,vis}$ candidate, a τ lepton is reconstructed that has an opposite electric charge to the previously tagged muon, without any other leptons in the vicinity. The τ leptons are required by the HLT to have $p_T > 25$ GeV and $|\eta| < 2.47$ in order to trigger. The region between $1.37 < |\eta| < 1.52$ is excluded from this cut to account for the detector geometry. A differentiation of the τ candidates for either one or three core tracks in the region is also made. These candidates then correspond to 1-prong or 3-prong τ leptons. The associated jets have to pass a medium BDT working point imposed by the HLT [32]. In general, the final state is required to have zero jets generated by bottom quarks to ensure orthogonality to possible $t\bar{t}$ tag and probe analyses. Jets originating from bottom quarks have properties differing from other hadronic jets, therefore they are tagged by a DL1r identifier algorithm [34]. These differing properties are e.g. a larger track impact parameter.

This analysis employs an event selection that allows a tag and probe analysis in $\sqrt{s} = 13$ TeV proton-proton collisions recorded in Run-2 from 2015-2018. The datasets of the years 2017 and 2018 are the main focus within this analysis. These years correspond to a total amount of data of 44.3 fb^{-1} and 58.5 fb^{-1}, respectively.

In order to investigate the HLT, an offline analysis has been applied in order to simulate possible modelling improvement strategies. The event selection for this offline analysis and processes to estimate the backgrounds will be discussed in detail in Section 4.3.

4.3. Offline Event Selection and Background Estimation

The signal and the background samples of the considered $Z \to \tau_{lep}\tau_{had}$ channel have been generated by Monte Carlo algorithms (MC), in order to be able to compare the measurement with SM predictions. The background samples that need to be simulated are the processes of Z+jets, W+jets, single top quark and $t\bar{t}$ quark pairs. All MC samples used are normalised to the next-to-next-to-leading-order (NNLO) cross sections. For a detailed description of the individual MC generators, see Ref. [32]. A summary of the used MC algorithms can be seen in Table 4.1.

Process	ME	Non-perturbative	Miscellaneous
W/Z+jets	Powheg-Box v2+CT10 [35–40]	Pythia8+AZNLO [41–43]	Photos++[44, 45]
$t\bar{t}$, single top Wt- and s-channel	Powheg-Box v2+CT10 [46]	Pythia6+P2012 [47, 48]	MadSpin[49], Photos++
single top t-channel	Powheg-Box v1+CT10f4	Pythia6+P2012	MadSpin

Table 4.1.: Summary of MC generators used for the NNLO simulation of the signal and background samples. ME stands for the matrix element, non-perturbative to include NLO or NNLO effects and miscellaneous sets more generators to simulate the top quark decay in the case of MadSpin and to include photon radiation from charged leptons in the case of Photos++ [28].

For the tag and probe signal region (SR) and background offline selection, more kinematic variables are considered. The first variable is the transverse mass m_T between the muon and the missing transverse energy:

$$m_T(\mu, E_T^{miss}) = \sqrt{2p_T(\mu)E_T^{miss}(1 - \cos(\Delta\phi(\mu, E_T^{miss})))}. \tag{4.2}$$

Another variable is the sum of the azimuthal angles of the muon and the $\tau_{had,vis}$ candidate together with their respective missing energies. It is defined as

$$\sum \cos(\Delta\phi) = \cos(\Delta\phi(\mu, E_T^{miss})) + \cos(\Delta\phi(\tau_{had,vis}, E_T^{miss})). \tag{4.3}$$

These variables, together with the invariant mass between the muon and the visible τ_{had} candidate $m_{vis}(\mu, \tau_{had-vis})$ offer a good separation between the $Z \to \tau\tau$ signal and the Z/W+jets and same sign backgrounds. As can be seen in Figure 4.3, the requirements for events to pass the SR selection that have been chosen based on these variables to filter out background events are $m_T(\mu, E_T^{miss}) < 50$ GeV, $45 < m_{vis}(\mu, \tau_{had-vis}) < 90$ GeV and $\sum \cos(\Delta\phi) > -0.15$. Figure 4.4 shows the p_T distributions of the leading τ lepton before (a) and after (b) applying the SR cuts. It can be seen that a significant amount of all background events have been cut out to ensure a higher signal purity.

For the offline analysis, the possibility of applying an additional RNN ID is given. There are three different RNN ID working points based on a cut on the RNN score considered: loose, medium and tight. Each of these working points has a set of τ ID measurement scale factors applied to the selection in order to improve the agreement between the data and MC samples. The loose working point has a minimal cut on the RNN score. This makes this selection similar to no RNN ID cut, except for the additional weights. The

Variable	SR	W CR	QCD CR	$t\bar{t}$ CR	High-m_T CR
Muon isolation	yes	yes	inverted	yes	yes
Opposite sign	-1	+1	-1	-1	-1
$m_T(\mu, E_T^{miss})$ [GeV]	< 50	> 60	< 50	> 40	> 60
$\sum \cos(\Delta\phi)$	> -0.15	< 0.0	> -0.15	> -0.15	-
$m_{vis}(\mu, \tau_{had-vis})$ [GeV]	[45,90]	[45,90]	[45,90]	-	-
E_T^{miss} [GeV]	-	> 30	-	-	> 20

Table 4.2.: Summary of cut requirements for each SR and the CRs. This is the selection used for the offline analysis. The charge of the tagged muon and the probed τ lepton as well as the isolation of the muon are being considered. Non-isolated muons in jets tend to be from QCD processes. The high-m_T CR is a dedicated CR to calculate the $k^{OS/SS}$ scale factors.

medium and tight working points cut out events below an RNN score of 0.9 and 0.95, respectively. The RNN scores and cuts based on the different working points can be seen in Figure 4.5. The baseline selection for this analysis was a loose RNN ID, changing to medium or tight for systematic considerations as discussed in Section 6.3.

Events that have been reconstructed in the offline analysis are matched to online counterparts in the process. If a tagged τ lepton can be geometrically matched to a probed τ lepton in $\Delta R < 0.2$, then the candidate is saved as a "true" τ lepton event. Otherwise, it is added to the "fake" τ lepton background contribution.

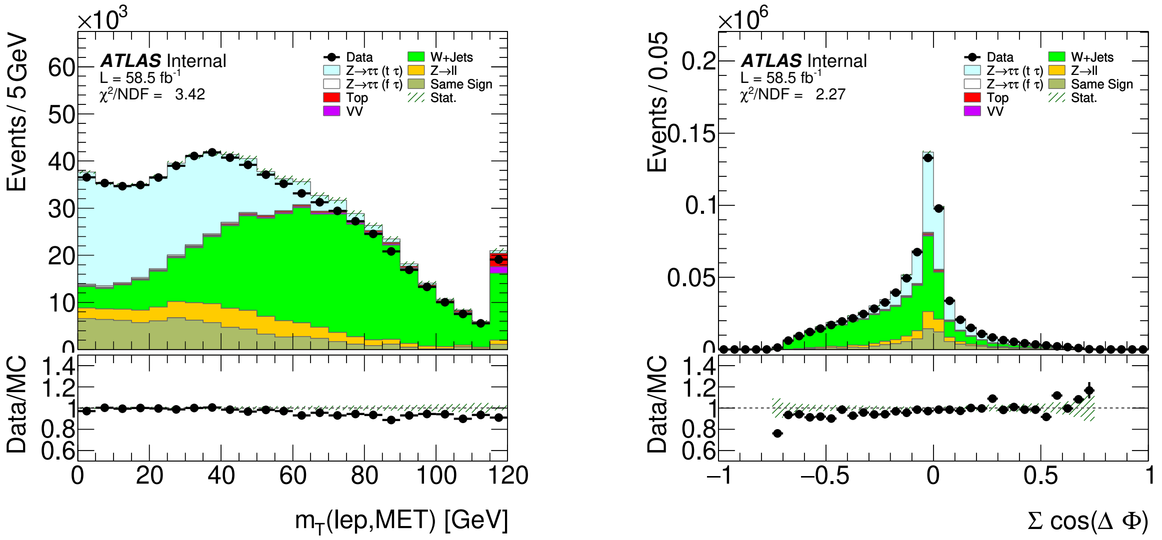
Using earlier defined SR cut variables, respective control regions (CR) for the main background contributions from QCD multi-jet events and $W \to \mu\nu$+jets as well as a $t\bar{t}$ CR can be defined as well. The CRs are used for data-driven scale factor calculations. The necessary cuts to change between the respective regions are summarised in Table 4.2.

An important step for the background estimation is to take advantage of the required charge properties of the τ_{lep} and the $\tau_{had,vis}$ candidates. In this analysis, a background estimation is used, that takes out the events of the μ-τ pairs with the "same-sign" (SS) and singles out the pairs with the "opposite-sign" (OS) for the SR. This data-driven process is called the OS-SS estimation and the full estimate of the OS part of the event yields for the SR can be calculated to

$$OS_{data} = OS_{QCD} + OS^{MC}_{Z\to\tau\tau} + k^{OS} \cdot OS^{MC}_{W} - k^{OS} \cdot OS^{MC}_{Top}, \tag{4.4}$$

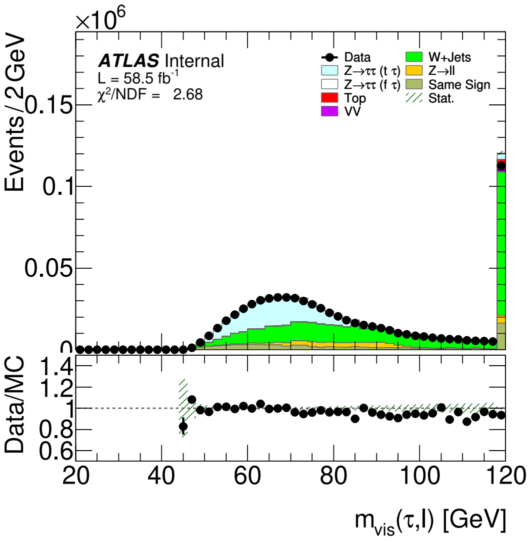
$$OS_{QCD} = r_{QCD} \cdot (SS_{data} - SS^{MC}_{Z\to\tau\tau} - k^{SS} \cdot SS^{MC}_{W} - k^{SS} \cdot SS^{MC}_{Top}), \tag{4.5}$$

where for Equation 4.4, the main background contributions of the channel are considered. In Equation 4.5 the QCD background is estimated from the SS data. Combining the terms together with $Z \to \tau\tau = OS^{MC}_{Z\to\tau\tau} - r_{QCD} \cdot SS^{MC}_{Z\to\tau\tau}$ and $(W/Top) = k^{OS} \cdot OS^{MC}_{W/Top} -$



(a) $m_T(\mu, E_T^{miss})$, the SR cut on this variable is for events above 50 GeV.

(b) $\sum \cos(\Delta\phi)$, the SR cut on this variable is for events below -0.15.

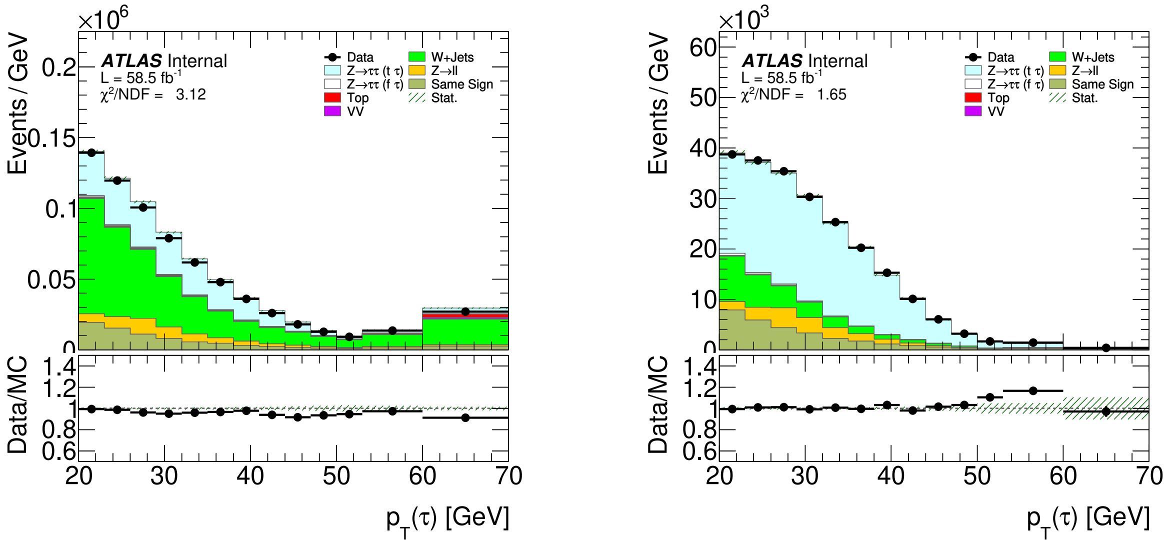


(c) $m_{vis}(\mu, E_T^{miss})$,the SR cut on this variable is for events below 45 GeV and above 90 GeV.

Figure 4.3.: Important kinematic variables that are considered in the SR selection. The three distributions all correspond to the 1-prong decay and loose offline ID. These distributions include overflow bins, which explain the high count bin at the end of the distribution.

$r_{QCD} \cdot k^{SS} \cdot SS^{MC}_{W/Top}$, the estimation of the full background results in

$$OS - SS_{total} = r_{QCD} \cdot SS_{data} + Z \rightarrow \tau\tau + W + Top. \tag{4.6}$$



(a) before

(b) after

Figure 4.4.: p_T distributions for the 1-prong channel before and after applying three SR preselection cuts.

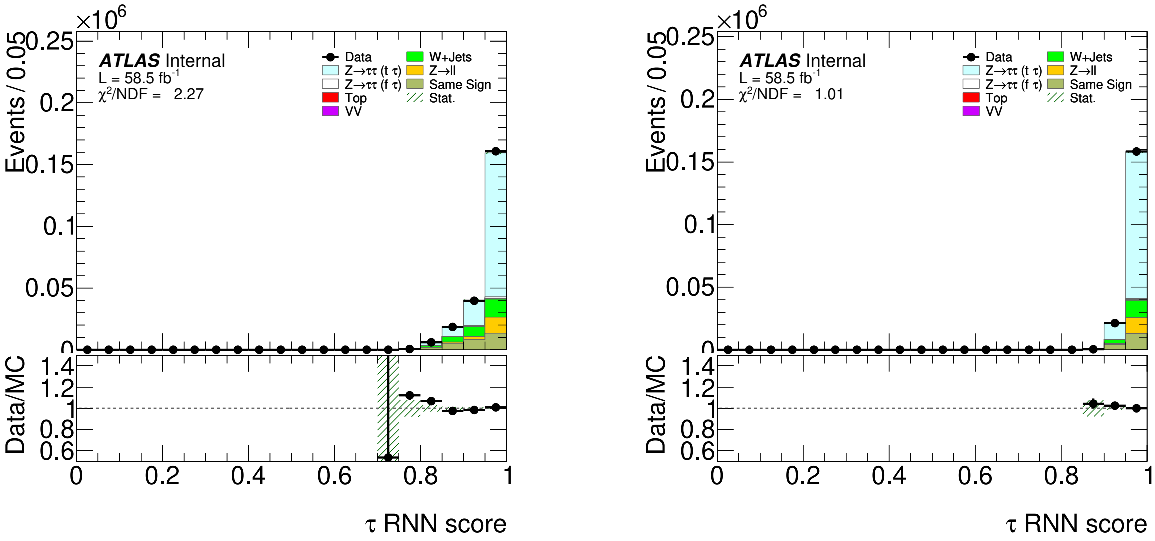
The factors r_{QCD}, k^{SS} and k^{OS} are data-driven correction factors calculated within the framework used in this analysis. r_{QCD} is a scale factor extracted from the QCD CR. It is calculated from

$$r_{QCD} = \frac{OS_{data} - OS_{MC}}{SS_{data} - SS_{MC}}. \tag{4.7}$$

Figure 4.6 shows the distributions of the OS and SS part of the QCD CR, respectively. The scale factor can be calculated from the ratio of the differences between the data and MC samples from Figures 4.6(a) and (b) according to Equation 4.7.
The scale factors k^{SS} and k^{OS} are used to normalise W+jets and top quark background contributions to data and are calculated from a high-m_T CR. They are extracted according to

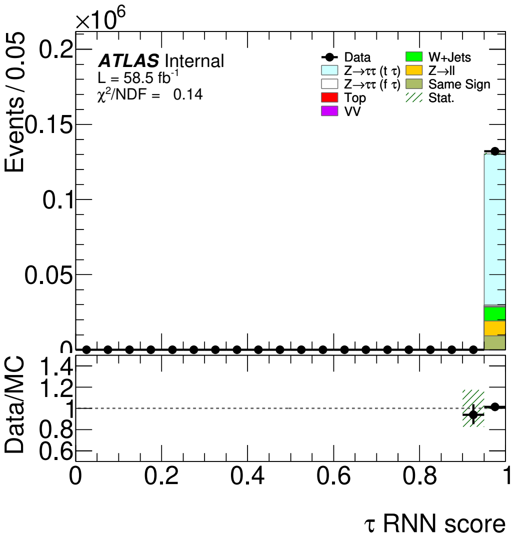
$$k_X^{OS/SS} = \frac{OS/SS_{data} - OS/SS_{non-XMC}}{OS/SS_{XMC}}, \tag{4.8}$$

where $X = W,\ top$. An example of the calculated scale factors for a 2018 selection with loose RNN ID and without further τ trigger requirements is given in Table 4.3. As these scale factors are calculated with a data-driven method, they have to be recalculated for every dataset used. Table 4.4 shows an example of the yields for the different samples in the signal region after applying the scale factors for the 1-prong channel with a loose offline RNN ID applied.



(a) loose offline RNN ID

(b) medium offline RNN ID

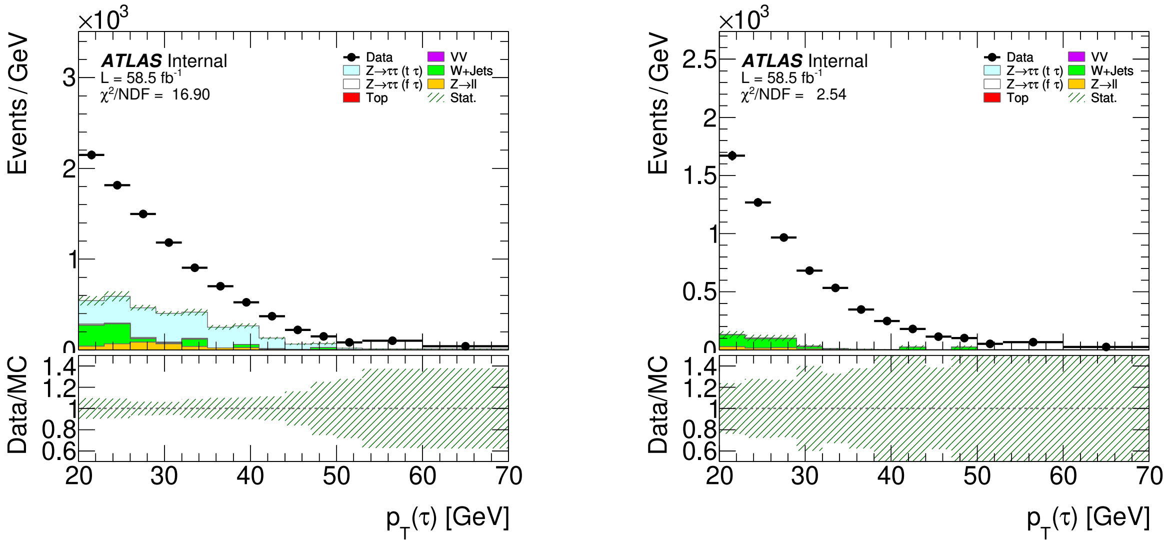


(c) tight offline RNN ID

Figure 4.5.: RNN scores with different ID working points applied. The loose working point only differs to the no ID working point in an additional efficiency measurement weighting that has been applied in order to improve the agreement between data and MC samples.

Scale factor	1-prong	3-prong
r_{QCD}	1.167 ± 0.028	1.386 ± 0.049
k_W^{OS}	1.070 ± 0.008	1.216 ± 0.017
k_W^{SS}	1.161 ± 0.013	1.280 ± 0.026
k_{Top}^{OS}	1.036 ± 0.008	1.069 ± 0.014
k_{Top}^{SS}	1.716 ± 0.039	1.595 ± 0.055

Table 4.3.: Scale factors calculated by the OS-SS estimation from 2018 data with a loose offline RNN ID in the 1-prong and 3-prong selections. There is no further τ trigger requirement applied.



(a) Opposite sign QCD control region

(b) Same sign QCD control region

Figure 4.6.: p_T distribution for the OS and SS regions of the QCD CR. These regions are used in the calculation of the r_{QCD} scale factor. The distributions both correspond to the 1-prong decay and loose offline ID.

Sample	Event Yield
Data	121548 ± 350
$Z \to \tau\tau$	92111 ± 420
Same Sign	11932 ± 290
W+jets	10257 ± 345
$Z \to \ell\ell$	4140 ± 32
VV	247 ± 7
Top	200 ± 6

Table 4.4.: Event Yield for a 1-prong selection from 2018 data and loose offline RNN ID applied.

5. Modelling of the High Level Trigger for τ Identification

Over the course of this Chapter, a quantitative analysis of the variables used by the τ trigger at the HLT level that were introduced in Section 4.2 is presented. In Section 5.1, the agreement between the data sample and the combined MC samples for all the ID variables is reviewed. In Section 5.2, 2D correlations between the same variables for different data-taking years for data and the signal MC sample will be investigated in order to quantify the performance of the HLT with respect to the offline reconstruction. Both data sets from 2017 and 2018 will be discussed.

5.1. Data/Monte Carlo Agreement

The agreement between the data sample and the combined signal and background MC samples is presented. In order to obtain a sensible comparison between the HLT and offline distributions of the ID variables introduced in Section 4.2, an additional "tau25_medium1_tracktwo"[1] trigger is added to the offline variables, to mimic the selection of the HLT with a medium BDT working point, which is currently used for the HLT variables. This trigger is applied for all studies detailed in Chapter 5. It adds an additional cut on the leading τ lepton p_T for candidates with $p_T < 25$ GeV. One objective to improve the modelling of the HLT is to follow the offline reconstruction as well as possible in the limited computing time available. Hence, why comparisons of HLT variables to the offline distributions are made throughout this chapter.

A statistical χ^2 test is used to quantify the agreement between measured data and the simulated MC samples. The χ^2 factor for a comparison between an unweighted and a

---

[1] This trigger follows the naming scheme for the τ trigger chains that are available at HLT level. *medium1* denotes the algorithm that has been used for τ ID, in this case a medium BDT working point. *tracktwo* describes the track algorithm that is applied at trigger level.

weighted dataset is calculated from

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - Np_i)^2}{Np_i} + \sum_{i=1}^{r} \frac{(w_i - Wp_i)^2}{\sigma_i^2} \tag{5.1}$$

$$= \sum_{i=1}^{r} \frac{(Wn_i - Nw_i)^2}{W^2 n_i + N^2 \sigma_i^2}. \tag{5.2}$$

The variables n_i and w_i are the events per bin i for unweighted and weighted histograms, respectively. N and W are the total number of events of the unweighted and weighted histogram. σ_i is the uncertainty of the weighted histogram per bin. The unweighted and weighted histograms correspond to the data and simulation in this analysis, as the data sample comes from a measurement and the background MC samples have been normalised with the OS-SS estimation explained in Section 4.3. The signal sample has an additional set of weights from the offline RNN ID applied.

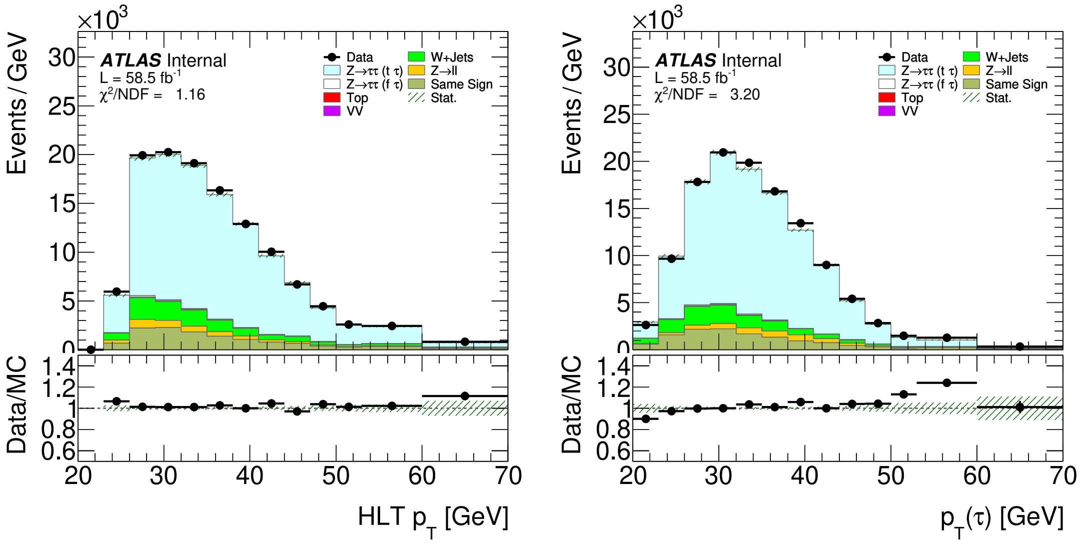
	HLT		Offline	
Variables	1p	3p	1p	3p
p_T	1.16	1.23	3.20	2.03
ΔR_{max}	3.56	0.45	2.33	1.72
f_{cent}	1.46	7.65	5.39	3.12
$\bar{d}_{innerTrack}$	3.67	2.03	4.19	2.75
m_{track}	-	2.70	-	3.54
$f_{EM}^{track-HAD}$	7.40	4.99	8.09	9.51
f_{track}^{EM}	3.86	8.79	3.64	5.71
$\frac{E_T}{p_T^{lead}}$	0.59	1.83	1.00	2.22
$\lvert S_{lead}\rvert$	16.04	-	5.99	-
S_T^{flight}	-	1.24	-	1.30
p_T^{ratio}	1.36	1.25	2.32	1.74

Table 5.1.: Reduced χ^2 values for the variables in the 1-prong and 3-prong channels for the 58.5 fb^{-1} dataset from 2018. For the 1-prong decay, $|S_{lead}|$, $f_{EM}^{track-HAD}$, and f_{track}^{EM} have the highest χ^2/NDF values for HLT. For 3-prong, this applies to f_{track}^{EM}, f_{cent} and $f_{EM}^{track-HAD}$.

For this thesis, the reduced χ^2 rather than the χ^2 values from Equation 5.1 are used for the discussion. The reduced χ^2, or χ^2/NDF is calculated dividing the χ^2 calculated earlier by the degrees of freedom. In the case of histograms, the degrees of freedom are equal to the number of bins. A Histogram shows a good agreement of data and MC for low χ^2 or if $\chi^2/NDF \to 1$. For $\chi^2/NDF \to 0$, a perfect agreement between data and MC is given, though undesirable, as the simulation of data for probability dependent processes will always vary by a small amount from measurements. The final assessment of variable modelling is based on both the reduced χ^2, as well as a comparison of the distributions.

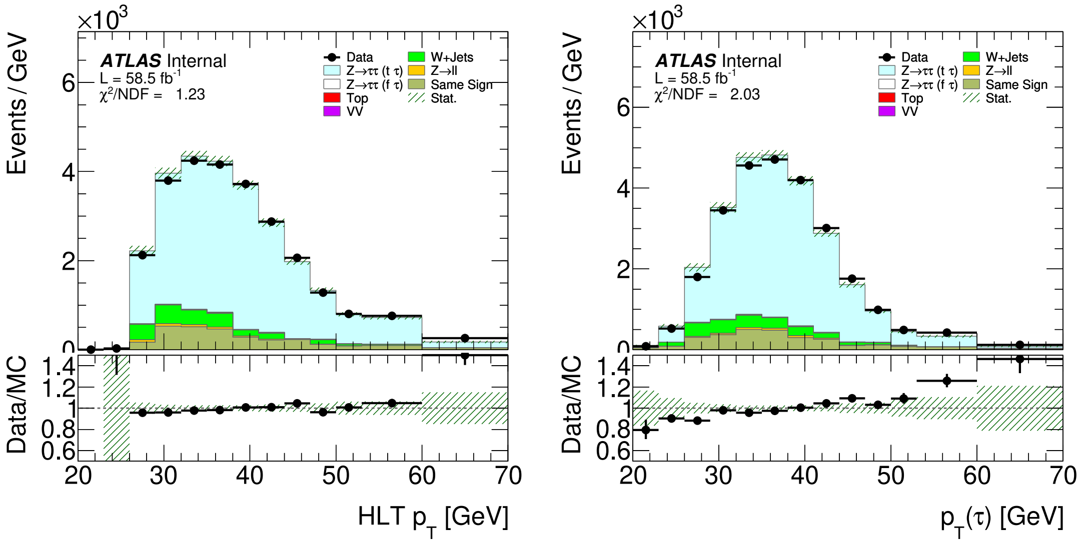
The results of the χ^2/NDF calculations can be retrieved from Table 5.1 for the HLT and offline variables for both 1-prong and 3-prong channels.

Generally, the modelling of the variables is acceptable, although there are some variables such as $|S_{lead}|$ or $f_{EM}^{track-HAD}$ in the 1-prong channel that stand out with high χ^2/NDF values. Figure 5.1 shows the distributions of the τ lepton p_T for both the 1-prong and 3-prong decay. Figures 5.1(a) and (c) show the p_T variable as reconstructed by the HLT



(a) HLT, 1p

(b) Offline, 1p



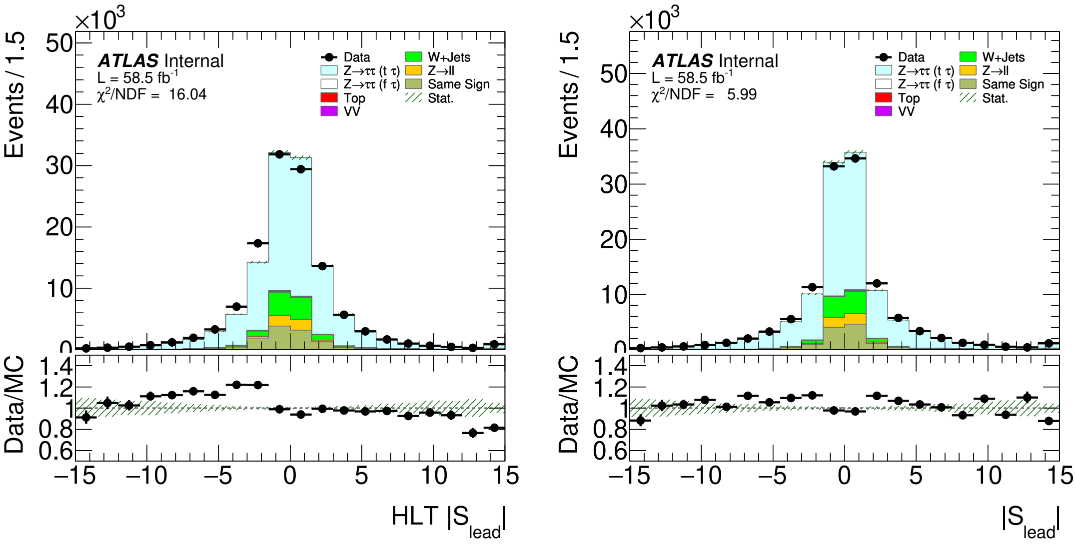
(c) HLT, 3p

(d) Offline, 3p

Figure 5.1.: Different distributions of p_T for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

and Figures 5.1(b) and (d) show the same distributions corresponding to the offline reconstruction.

The agreement between data and MC here is good throughout all the distributions, although in the high p_T tails, the deviations of the MC samples from data and uncertainties of the reconstructions rise because of small number of events. To account for that, the last bins have been combined into one large bin for all p_T distributions. For the 1-prong τ lepton decay channel, the highest χ^2/NDF value for the HLT stems from the $|S_{lead}|$ with a reduced χ^2 of 16.04. This variable is shown in Figure 5.2(a). The distribution shows mismodelling especially on the left side of the peak for HLT and in the tails for offline. The peak of the distribution from the offline plot in Figure 5.2(b) exhibits a better agreement between signal and MC, which is confirmed by $\chi^2/NDF = 5.04$.



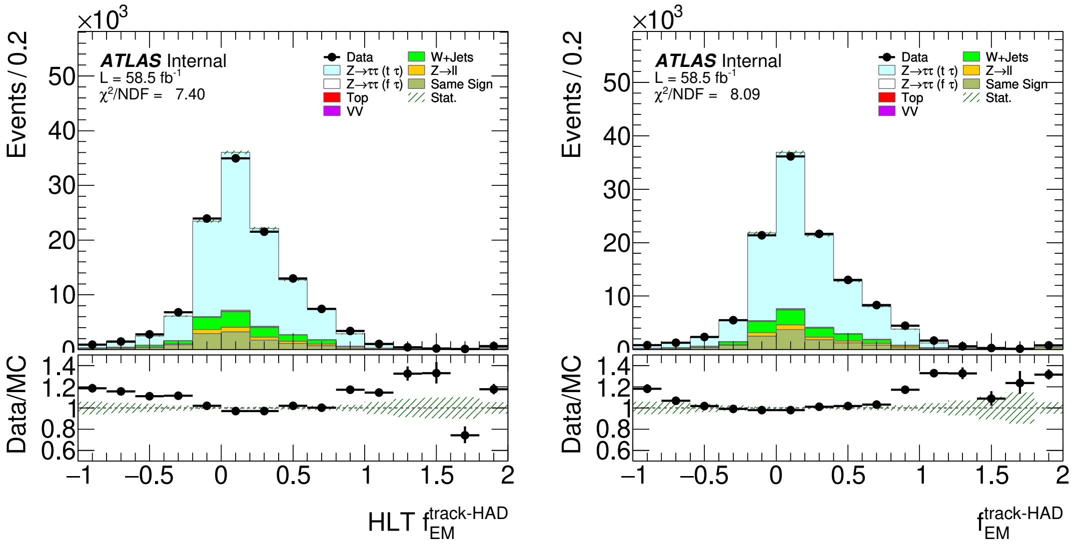
(a) HLT, 1p

(b) Offline, 1p

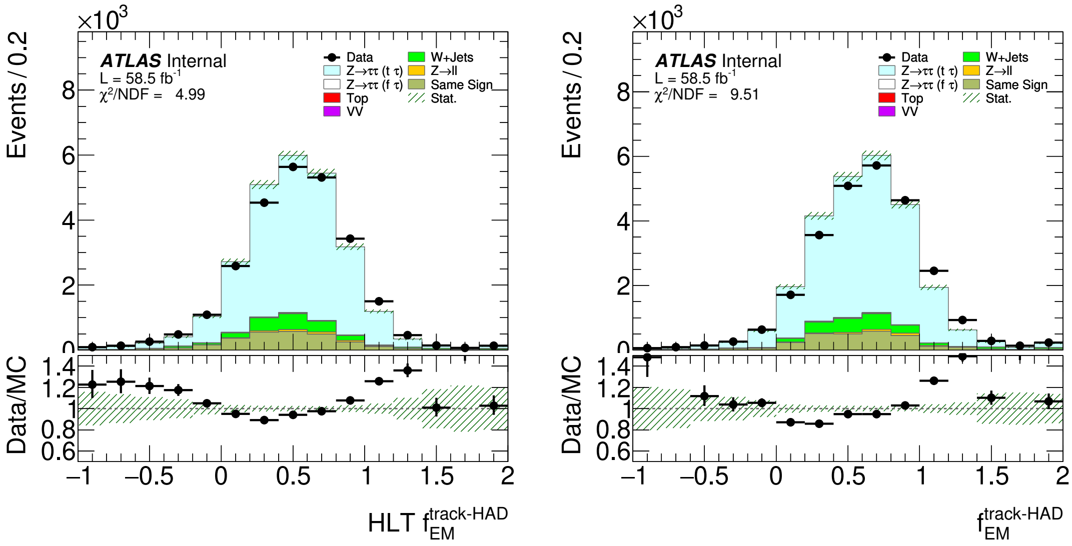
Figure 5.2.: Different distributions of $|S_{lead}|$ for HLT and offline reconstruction in the 1-prong channel. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

The variable with the second highest reduced χ^2 value for the HLT distributions, is $f_{EM}^{track-HAD}$ with a $\chi^2/NDF = 7.40$. It is visible from the distributions for $f_{EM}^{track-HAD}$ in Figure 5.3, that the agreement in the peak and tail regions can be improved upon for both the HLT in 5.3(a) and 5.3(c) and the offline reconstructions in 5.3(b) and 5.3(d). This variable is modelled overall similarly for online and offline. In general, the HLT reconstructs the τ leptons quite well for 1-prong τ candidates overall.

For the 3-prong τ lepton decay channel, the highest χ^2/NDF values for the HLT variables stem from f_{track}^{EM}, f_{cent} and $f_{EM}^{track-HAD}$. The distributions of f_{track}^{EM} for the HLT and offline



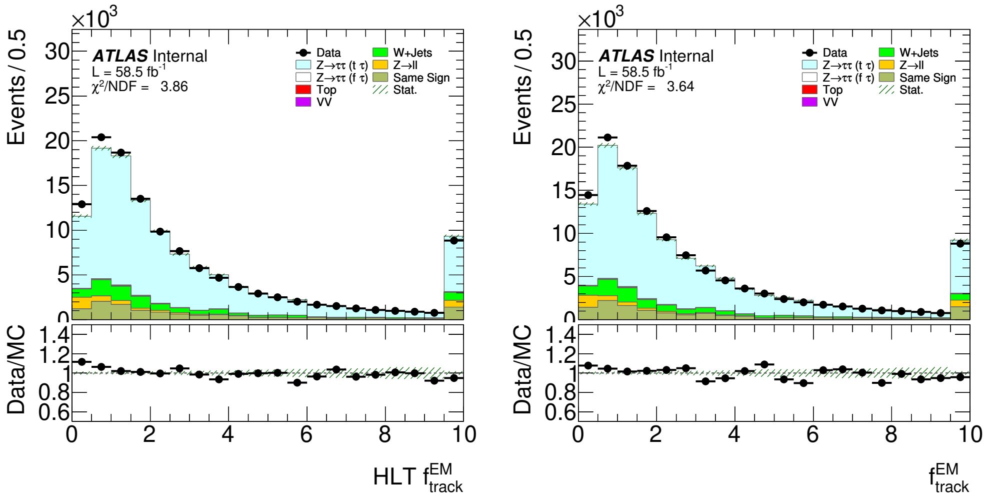
(a) HLT, 1p (b) Offline, 1p



(c) HLT, 3p (d) Offline, 3p

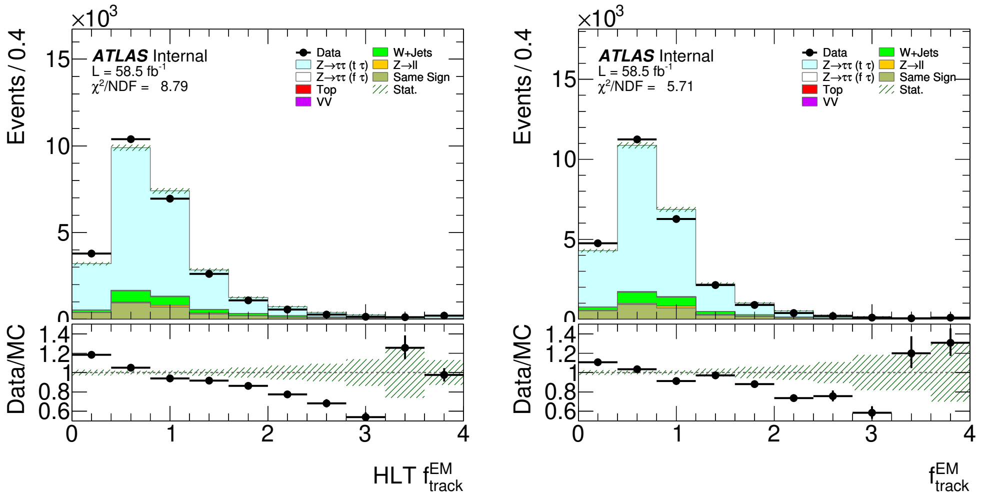
Figure 5.3.: Different distributions of $f_{EM}^{track-HAD}$ for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

reconstructions can be seen in Figures 5.4(a), (c) and Figures 5.4(b), (d), respectively. Especially for the tail regions, the data and MC agreement suffers from a low number of events in the bins. This results in the high value of e.g. $\chi^2/NDF = 8.79$ for the distribution of Figure 5.4(c). This effect is only apparent for 3-prong for this variable.



(a) HLT, 1p

(b) Offline, 1p

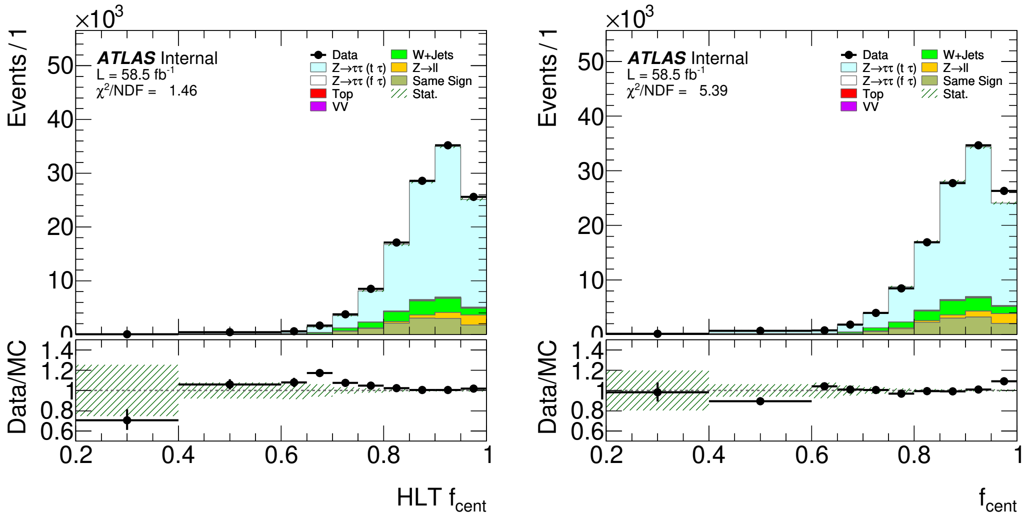


(c) HLT, 3p

(d) Offline, 3p

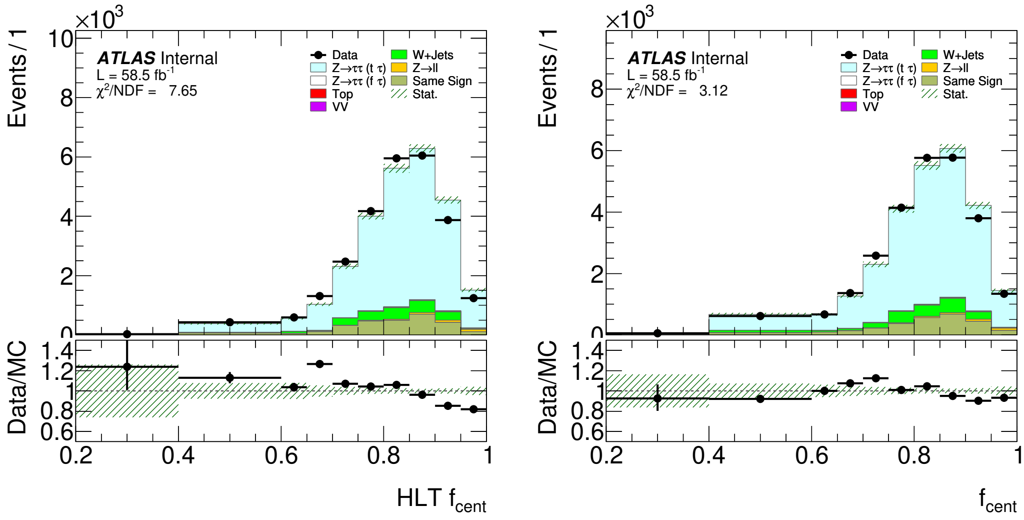
Figure 5.4.: Different distributions of f^{EM}_{track} for HLT and offline reconstruction, 1-prong and 3-prong. These distributions correspond to the 2018 58.5 fb^{-1} dataset.

The variable with the second highest χ^2/NDF in the 3-prong channel with a value of 7.61 for HLT is f_{cent}. It is expected that the centrality fraction is more shifted towards 1, because the leading τ lepton candidate deposits most of its energy in the core region of $\Delta R < 0.1$. Thus, for the bins between 0 and 0.4 the number of events is low, which results in a high χ^2/NDF value, as the reconstruction and therefore the data and MC agreement suffers. To counteract this effect, the bins in the lower statistics regions have been combined. Overall, the 3-prong channel distribution modelling is worse than for the 1-prong channel. Using these results of the variables with the highest χ^2/NDF values, a strategy to improve the modelling for future RNN trainings for the HLT will be discussed in the Chapter 6.



(a) HLT, 1p

(b) Offline, 1p



(c) HLT, 3p

(d) Offline, 3p

Figure 5.5.: Different distributions of f_{cent} for HLT and offline reconstruction, 1-prong and 3-prong. These distributions correspond to the 2018 58.5 fb^{-1} dataset.

5.2. Online/Offline Correlations

The distributions of each offline and online variable are compared against each other in 2D histograms to calculate a correlation factor to further investigate the differences between HLT and offline distributions to search for grounds of improvements of the HLT modelling. The correlation factor is calculated to quantify the similarity between the HLT and offline variables. The correlation factor for two variables x and y is defined as

$$r_{xy} = \frac{\mathrm{cov}[x, y]}{\sigma_x \cdot \sigma_y}, \tag{5.3}$$

with the co-variance between x and y and σ_x, σ_y being the uncertainties of x and y. Two variables are highly (anti-)correlated for $r \to 1$ ($r \to -1$). There is no correlation between two distributions if and only if the correlation factor is equal to zero. A high correlation between the HLT and offline distributions is desirable, so that the HLT yields more offline-like results in real time with less resources available.

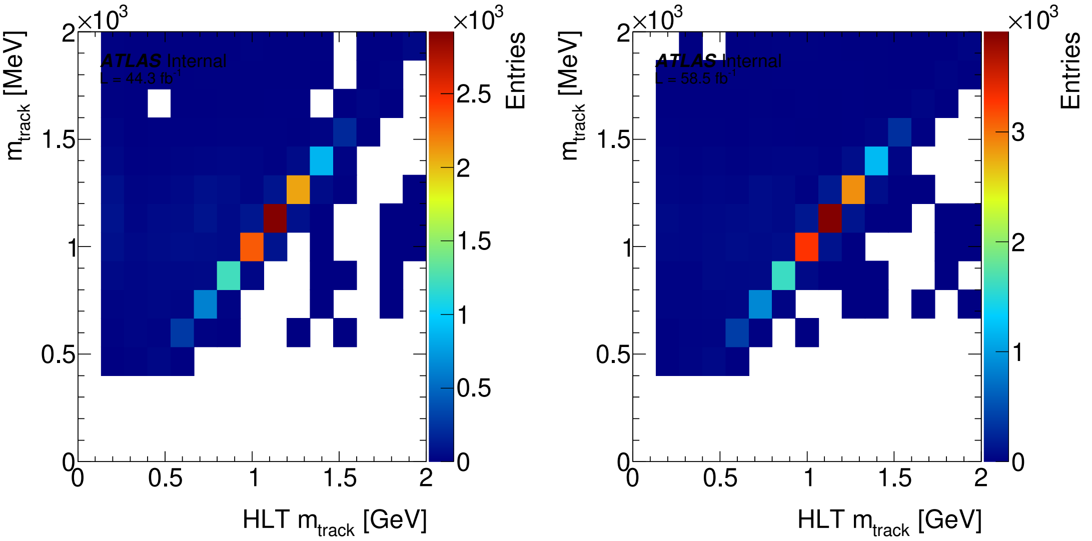
A fraction of the Run-2 dataset recorded between 2015 and 2018 is being used within this analysis. The 44.3 fb^{-1} and 58.5 fb^{-1} datasets from 2017 and 2018, respectively, are compared for the correlation factors between online and offline variable distributions. The results of the determination of all relevant correlation factors can be found in Table 5.2.

	2017		2018	
Variables	1p	3p	1p	3p
p_T	0.831	0.754	0.837	0.760
f_{cent}	0.836	0.827	0.846	0.838
ΔR_{max}	0.591	0.695	0.607	0.719
$\bar{d}_{innerTrack}$	0.622	0.591	0.648	0.630
m_{track}	-	0.603	-	0.689

Table 5.2.: Summary of correlation coefficients between offline and online for the datasets of the years 2017 and 2018. Only a fraction of the identification variables have been considered here. The correlation coefficients for 2018 data are summarised in Tables 5.3 and 5.4.

From Table 5.2, it can be seen that most of the variables are quite correlated to the respective offline variables. Looking at the correlation factor of ΔR_{max}, a lower correlation can be noticed compared to most of the ID variables. The value of $r = 0.603$ for m_{track} in 2017 in the 3-prong channel is one of the least correlated variable overall. The variables that are less correlated to the offline reconstruction seem to be the ones that are based on track reconstruction: ΔR_{max}, $\bar{d}_{innerTrack}$ and m_{track}. Between 2017 and 2018 in data-taking, the track reconstruction algorithm went through several changes to improve its performance.

A slight improvement of the correlation factors for the worst performing variables can be observed in the 2018 dataset, with an approximately $r = 0.006$ to $r = 0.09$ rise in correlation. This can be in particular seen in Figure 5.6, where the 2-dimensional distributions of m_{track} are depicted for 2017 and 2018 in 5.6(a) and (b), respectively.



(a) 2017, r=0.603

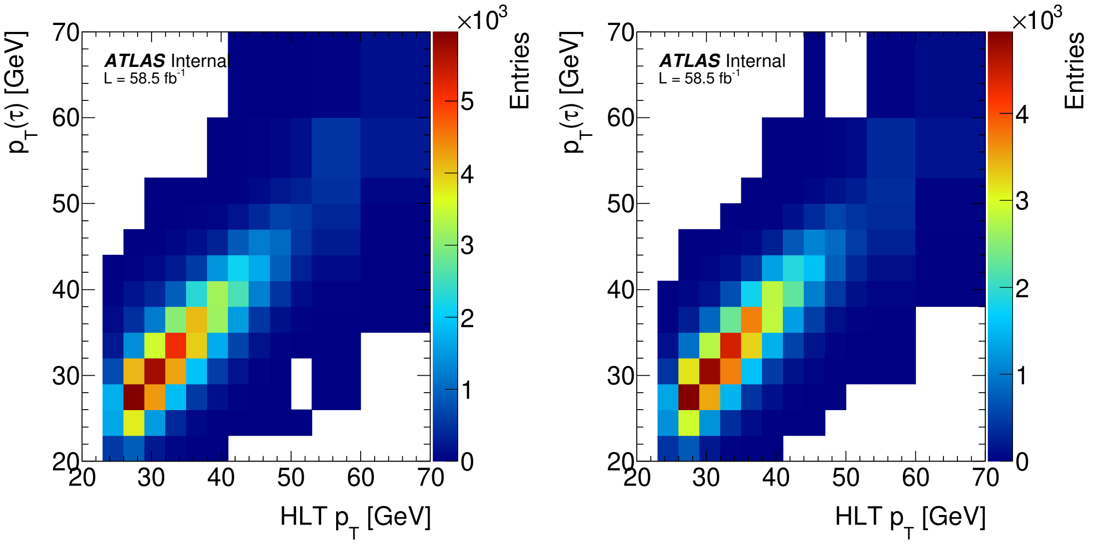
(b) 2018, r=0.689

Figure 5.6.: 2D distributions of m_{track} for the (a) 2017 dataset and (b) 2018 dataset. These distributions are in the 3-prong decay mode. An improvement of the correlation factor calculated according to Formula 5.3 between the years can be seen. The correlation factor rises by $\approx 9\ \%$ from $r = 0.603$ to $r = 0.689$ between the 2017 and 2018 datasets, respectively

Although the performance difference is overall quite small, the 2018 dataset will be focussed on going forward in this analysis. It has a larger amount of data available and improves correlations to the offline reconstruction. Furthermore, the latest track reconstruction algorithm that will also be used for future data-taking in Run-3 is applied, which makes the results of this analysis more useful for comparisons between different data-taking periods. Therefore, only 2018 data will be considered for the rest of this analysis.

Generally, the results presented throughout this analysis will be relevant for the upcoming Run-3 of data-taking at the ATLAS detector. The knowledge gained from Run-2 data can be used for future τ trigger reconstruction algorithms.

To conclude the investigations of the 2D distributions, all the correlations for the data and signal MC have been calculated separately. This can be seen in Figures 5.7 and 5.8 with the distributions for the p_T of the τ lepton for the 1-prong and 3-prong decays. The amount of events available for the background MC histograms was not sufficient to make a clear statement regarding correlations.



(a) Data, r=0.831

(b) Signal, r=0.842

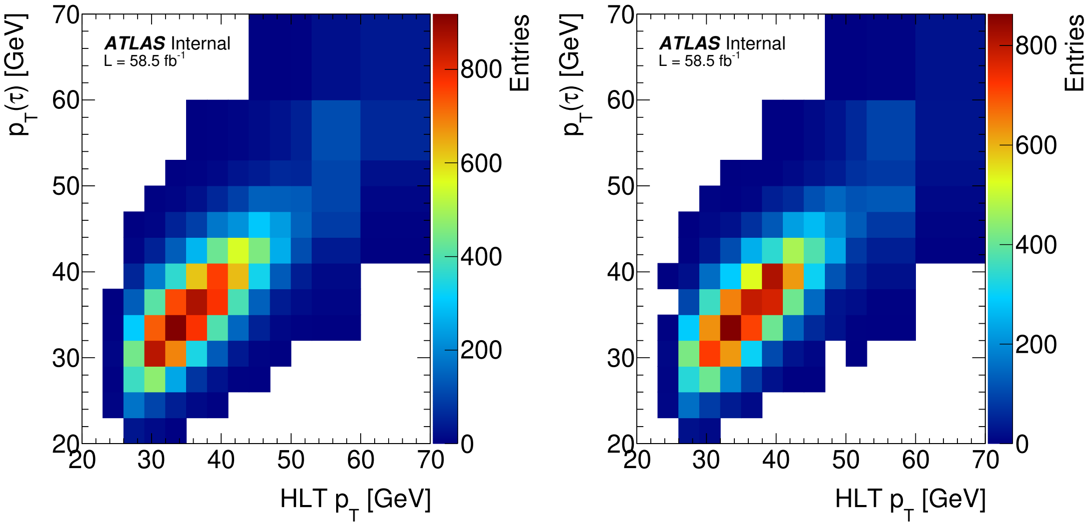
Figure 5.7.: 2D distributions of p_T for 2018 and the 1-prong channel. The plots are split up for (a) the data sample and (b) the $Z \rightarrow \tau\tau$ signal MC sample. The correlation factors calculated according to Formula 5.3 are quite similar to each other with correlation factors of $r = 0.831$ for data and $r = 0.842$ for signal.

The signal MC histogram in Figure 5.7 for instance, agrees quite well with the data sample. Thus, the same consequences can be drawn for signal as well as data variables corresponding to track calculation that were improved for 2018 data. The results of the correlation calculations are summarised in Table 5.3. The two variables that are also less correlated to the offline distributions for the 1-prong channel are the p_T and f_{cent}. The latter variable is sensitive to pile-up, so it could be possible to improve this variable by investigating the performance of the pile-up suppression.

The 3-prong decay channel yields similar results as the 1-prong decay channel. Looking at Figure 5.8, the correlations for the τ lepton p_T between offline and HLT seems to worsen in the 3-prong channel, compared to the respective 1-prong decay in Figure 5.7. The results of the calculation of the correlation factors are noted in Table 5.4. While the correlation factors become lower for the variables p_T or $f_{EM}^{track-HAD}$, some correlations just change slightly, e.g. for f_{cent} and some even seem to be more highly correlated than for the 1-prong channel, e.g. ΔR_{max}.

Variables	Data	Signal MC
p_T	0.83	0.84
ΔR_{max}	0.61	0.56
f_{cent}	0.83	0.80
$\bar{d}_{innerTrack}$	0.65	0.58
$f_{EM}^{track-HAD}$	0.91	0.92
f_{track}^{EM}	0.96	0.96
$\frac{E_T}{p_T^{lead}}$	0.99	0.99
$\lvert S_{lead}\rvert$	0.89	0.92
p_T^{ratio}	0.94	0.93

Table 5.3.: Correlation coefficients for the 1-prong channel, split up into the data and $Z \to \tau\tau$ signal MC samples.



(a) Data, r=0.755

(b) Signal, r=0.757

Figure 5.8.: 2D distributions of p_T for 2018 in the 3-prong channel. The plots are split up for (a) data and (b) the $Z \to \tau\tau$ signal MC sample. The correlation factors calculated according to Formula 5.3 are quite similar to each other with correlation factors of $r = 0.755$ for data and $r = 0.757$ for signal.

Variables	Data	Signal MC
p_T	0.76	0.76
ΔR_{max}	0.72	0.67
f_{cent}	0.82	0.80
$\bar{d}_{innerTrack}$	0.63	0.58
m_{track}	0.69	0.67
$f_{EM}^{track-HAD}$	0.86	0.85
f_{track}^{EM}	0.91	0.92
$\frac{E_T}{p_T^{lead}}$	0.98	0.97
$\lvert S_{lead}\rvert$	0.89	0.90
p_T^{ratio}	0.91	0.91

Table 5.4.: Correlation coefficients for the 3-prong channel, split up into the data and $Z \to \tau\tau$ signal MC samples.

6. Re-Weighting Studies for the $Z \to \tau^+\tau^-$ Process

In this Chapter, the main results of the signal re-weighting procedure will be presented. An advantage of this signal re-weighting is that this method extracts data-driven scale factors to provide a more accurate prediction of HLT ID variables with low additional computing costs. The goal of these investigations is an overall improvement of the HLT variable modelling which could help to improve the HLT modelling to make ID algorithm processes, e.g. the training of the RNN identification algorithm more efficient in the upcoming Run-3 of data-taking.

Beginning in Section 6.1, the strategy to re-weight the signal MC to the data sample based on ID variable distributions will be introduced. Throughout Section 6.2, the results of a 1-dimensional re-weighting for both the 1-prong and 3-prong channels are summarised. In Section 6.3, the 1-dimensional re-weighting is expanded into an analysis with 2-dimensional weight sets and the results are presented. Cross-checks with varying offline RNN requirements to evaluate the quality of the calculated weights and systematic uncertainties to the 1-dimensional weight sets from Section 6.2 are considered in Section 6.4. To conclude, recommendations on the best re-weightings for both the 1-prong and 3-prong decays are given in Section 6.5. Throughout this chapter, the 2018 dataset has been used.

6.1. Re-Weighting Strategy

After investigating the agreement between data and simulation, the results can be used to derive a strategy of re-weighting $Z \to \tau\tau$ events based on the considered ID variable distributions. The goal is to achieve a better modelling across all variables, without losing any quality of the well-modelled variables. The idea is to normalise the signal sample of the variables with the highest χ^2/NDF values, i.e. the worst modelled variables to agree exactly with the data. The set of weights can be calculated from event yields with

$$w_i = \frac{n_i^{data} - n_i^{MC}}{n_i^{signal}}. \tag{6.1}$$

The weights w_i are calculated for each bin i, where n_i^{data}, n_i^{MC} and n_i^{signal} are the number of data, background MC and signal MC events in bin i, respectively. The statistical uncertainty for each weight is computed using Gaussian error propagation on Equation 6.1, resulting in

$$\sigma_{w_i}^{stat} = \frac{1}{n_i^{signal}} \sqrt{\sigma^2_{n_i^{data}} + \sigma^2_{n_i^{MC}} + w_i^2 \cdot \sigma^2_{n_i^{signal}}}, \tag{6.2}$$

where $\sigma_{n_i^{data}}$, $\sigma_{n_i^{MC}}$ and $\sigma_{n_i^{signal}}$ are the statistical errors per bin i for data, background and signal, respectively.

The re-weighting process is visualised in Figure 6.1: The distribution of $\bar{d}_{innerTrack}$ in Figure 6.1(a) has been used to calculate a weight for each bin, which are applied to the signal MC sample in Figure 6.1(b). It can be seen throughout the whole distribution, that the ratio of data over MC samples is consistently one, resulting in a reduced χ^2 value of exactly zero.

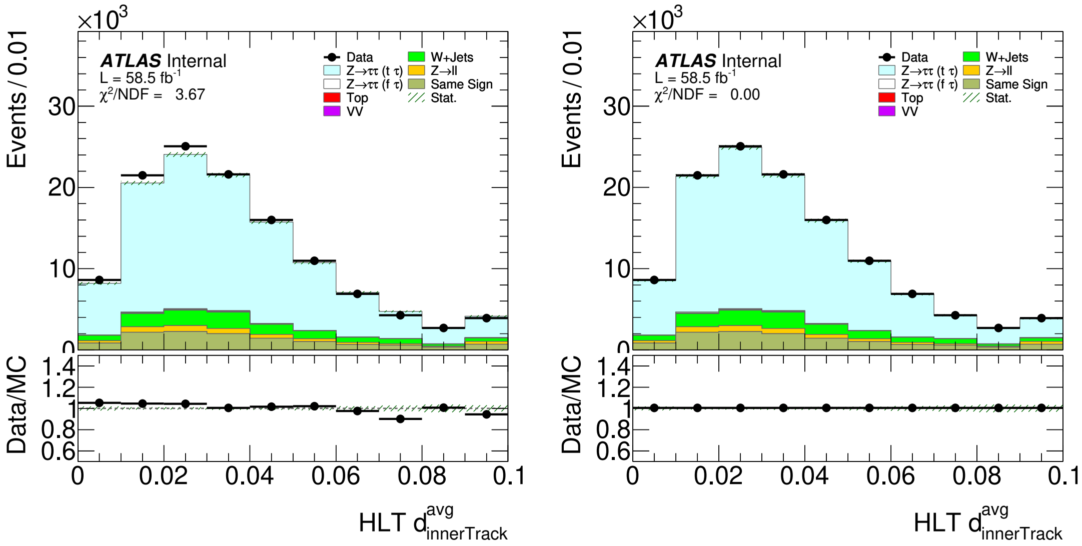
Subsequently, the weights calculated from Equation 6.1 are applied to the remaining variables of the selection to check the changes in data and MC agreement. An improvement of the modelling will be assessed based on both the results of χ^2 tests and the distributions. The selected variables and results of the re-weightings will be presented in Section 6.2.

6.2. Determination of Weights

6.2.1. Variable Selection

The first step is to select a set of variables for the 1-prong and 3-prong τ lepton decay channels to base the weight calculation on. The variables have been selected based on the results from Section 5.1. For the 1-prong channel, the four variables with the highest χ^2/NDF values have been selected:

1. $|S_{lead}|$ with a $\chi^2/NDF = 16.0$ (Figure 5.2(a)),
2. $f_{EM}^{track-HAD}$ with a $\chi^2/NDF = 7.1$ (Figure 5.3(a)),
3. f_{track}^{EM} with a $\chi^2/NDF = 3.9$ (Figure 5.4(a)),
4. $\bar{d}_{innerTrack}$ with a $\chi^2/NDF = 3.6$ (Figure 6.1(a)).



(a) no weight

(b) Signal MC weighted to data

Figure 6.1.: Distributions of $\bar{d}_{innerTrack}$ in the 1-prong channel, where either (a) no weights or (b) the signal MC weights calculated based on this variable have been applied.

The reason that four variables are considered in the 1-prong channel is the similarity between the χ^2/NDF values of the variables f^{EM}_{track} and $\bar{d}_{innerTrack}$. For the 3-prong channel on the other hand, three variables with the highest χ^2/NDF values were chosen for the re-weightings:

1. f^{EM}_{track} with a $\chi^2/NDF = 8.7$ (Figure 5.4(c)),
2. f_{cent} with a $\chi^2/NDF = 7.6$ (Figure 5.5(c)),
3. $f^{track-HAD}_{EM}$ with a $\chi^2/NDF = 5.0$ (Figure 5.3(c)).

All of these variables have been used to calculate a set of weights and the effects of applying these weights on the other ID variables are summarised in the following Sections 6.2.2 and 6.2.3 for the 1-prong and 3-prong channels, respectively.

6.2.2. Results for the 1-prong Channel

Revisiting the results from Section 5.1 for the HLT variables and 1-prong channel, it can be seen that the MC predictions are relatively close to the data samples for most variables. The distributions that have been chosen for the re-weighting together with ΔR_{max} are the only ones with $\chi^2/NDF > 1.50$. For each of the variables for the 1-prong channel presented in Section 6.2.1, weights have been calculated and applied to the signal MC sample. A summary of the updated χ^2/NDF values for the ID variables without weights and normalised with the variables $|S_{lead}|$, f_{track}^{EM}, $f_{EM}^{track-HAD}$ and $\bar{d}_{innerTrack}$ can be found in Table 6.1.

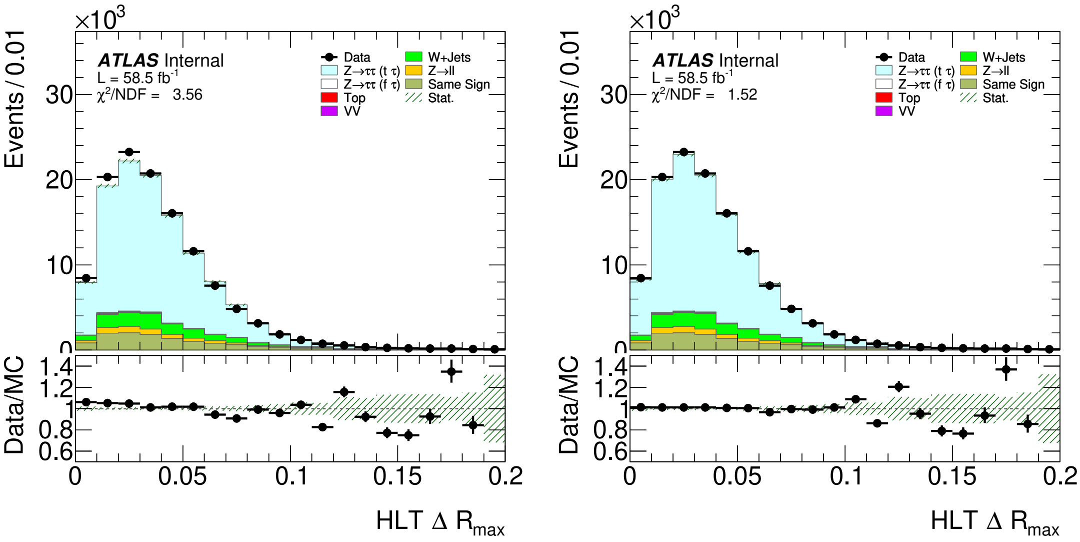
It is evident, that the weight set based on $|S_{lead}|$ hardly affects the other ID variables given the results for the χ^2/NDF for the different distributions compared to the non-weighted values. This is due to a low correlation between $|S_{lead}|$ and all the other variables (See Table 6.3 in Section 6.3). Therefore, this re-weighting will not be considered further for the 1-dimensional re-weighing studies.

A good example of the different effects of the other weight sets can be seen in Figure 6.2, where the distributions of ΔR_{max} have been plotted without weights in 6.2(a) and with the weight sets calculated from $\bar{d}_{innerTrack}$, f_{track}^{EM} and $f_{EM}^{track-HAD}$ applied in Figures 6.2(b), (c) and (d), respectively. For this variable, an improvement of the agreement between the data and MC samples after applying the re-weightings can be observed. In the peak region around $\Delta R_{max} = 0.03$, the $\bar{d}_{innerTrack}$ weights and the f_{track}^{EM} weights especially yield good results for the data and MC agreement, while the $f_{EM}^{track-HAD}$ weights do not seem to affect the ΔR_{max} in a significant way.

	no weight	$\lvert S_{lead}\rvert$	f_{track}^{EM}	$f_{EM}^{track-HAD}$	$\bar{d}_{innerTrack}$
p_T	1.2	1.3	1.2	1.3	1.3
ΔR_{max}	3.6	3.6	2.2	3.1	1.5
f_{cent}	1.5	1.4	1.4	1.5	1.7
$\bar{d}_{innerTrack}$	3.7	3.7	1.4	2.9	0.0
$f_{EM}^{track-HAD}$	7.5	7.7	4.2	0.0	6.8
f_{track}^{EM}	3.9	4.1	0.0	3.9	2.5
$\frac{E_T}{p_T^{lead}}$	0.6	0.7	2.7	0.9	0.9
$\lvert S_{lead}\rvert$	16.0	0.0	16.3	16.1	16.3
p_T^{ratio}	1.4	1.2	1.6	6.3	1.4

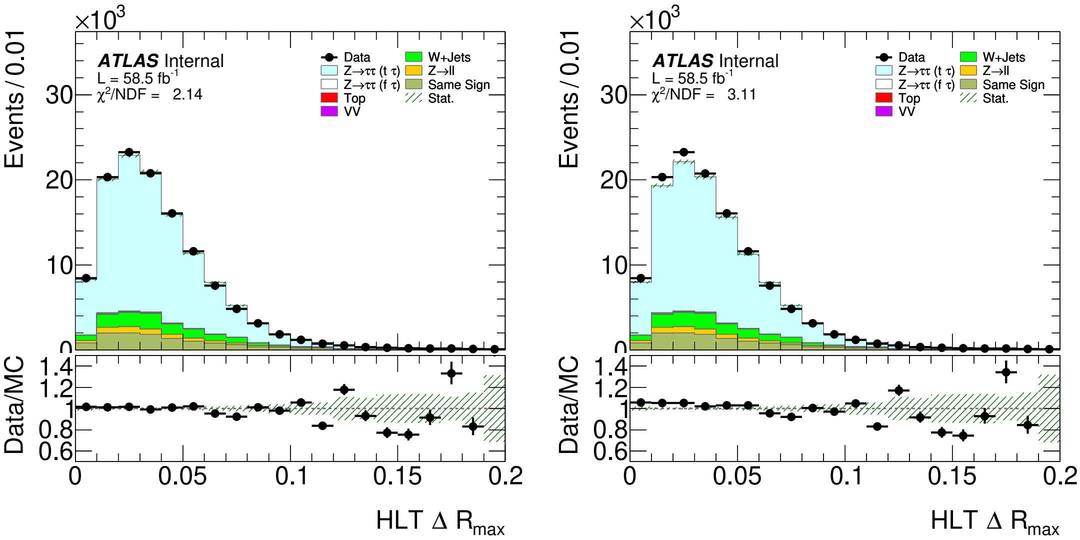
Table 6.1.: Results of the reduced χ^2 value for different re-weightings in the 1-prong channel. Based on these results, the $\bar{d}_{innerTrack}$ re-weighting is recommended.

Given the p_T^{ratio} distribution in Figure 6.3, the re-weightings yield similar results for the $\bar{d}_{innerTrack}$ and the f_{track}^{EM} weights compared to the unweighted distribution. The $f_{EM}^{track-HAD}$ re-weighting in Figure 6.3(d) increases the χ^2/NDF value by almost a factor of five. Regarding the modelling, large deviations through low statistics effects in the lower and upper tails of the distributions at $[0, 0.5]$ and $[1.7, 2.5]$, respectively, can be observed. In the peak region of the distribution, a slight decrease of the data and MC agreement can be observed compared to Figure 6.3(a). This is a strong argument to discard the f_{track}^{EM} re-weighting, as a worsening of the χ^2/NDF value compared to the non-weighted distributions was used as an exclusion criterion.



(a) no weight

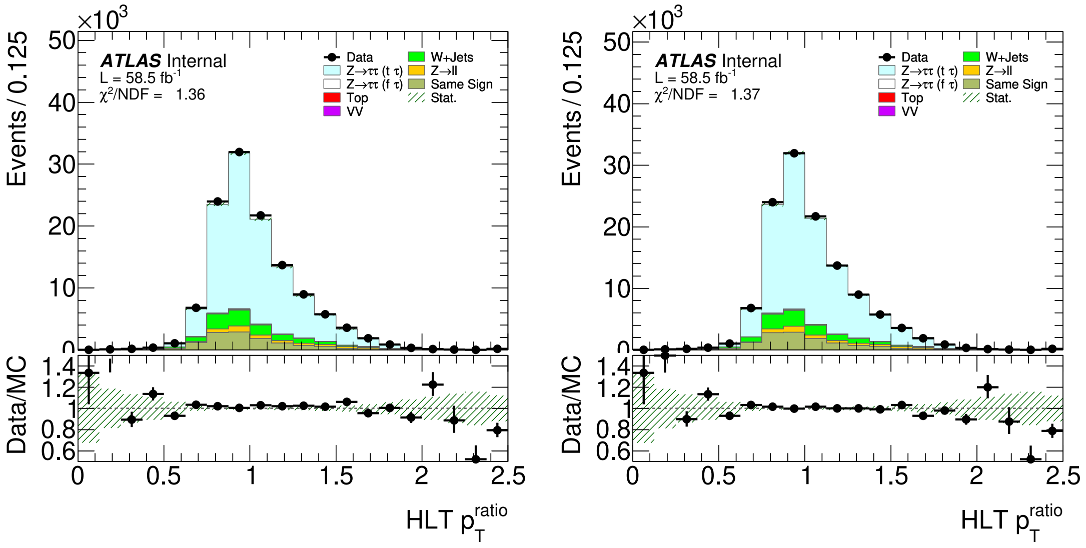
(b) $\bar{d}_{innerTrack}$



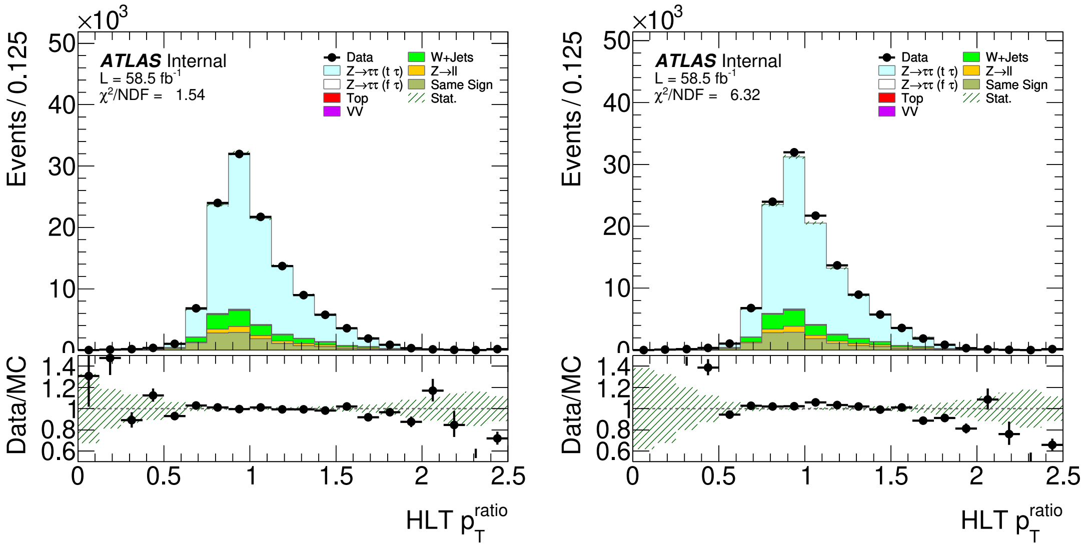
(c) f^{EM}_{track}

(d) $f^{track-HAD}_{EM}$

Figure 6.2.: Distributions of ΔR_{max} for the 1-prong channel (a) without weights and with three weight sets calculated from (b) $\bar{d}_{innerTrack}$, (c) f^{EM}_{track} and (d) $f^{track-HAD}_{EM}$ applied.



(a) no weight (b) $\bar{d}_{innerTrack}$



(c) f^{EM}_{track} (d) $f^{track-HAD}_{EM}$

Figure 6.3.: Distributions of p_T^{ratio} for the 1-prong channel (a) without weights and with three weight sets calculated from (b) $\bar{d}_{innerTrack}$, (c) f^{EM}_{track} and (d) $f^{track-HAD}_{EM}$ applied.

The best results stem from the $\bar{d}_{innerTrack}$ and f^{EM}_{track} re-weighting. These two weight sets perform similarly well, although the improvement of the modelling for the variable ΔR_{max} from the $\bar{d}_{innerTrack}$ weights is taken more strongly into account, as ΔR_{max} is one of the more important variables for RNN ID. Looking at the χ^2/NDF of the $\frac{E_T}{p_T^{lead}}$ distribution in Table 6.1, the value worsens for the f^{EM}_{track} weights by a small margin. This gives reason to prefer the $\bar{d}_{innerTrack}$ weights over the f^{EM}_{track} weights.

When considering all the aspects presented together with the χ^2/NDF results from Table 6.1, the best performing re-weighting for the 1-prong channel is with the weights calculated with the $\bar{d}_{innerTrack}$ variable. The improvement that the $f_{EM}^{track-HAD}$ re-weighting yields does not outweigh the worsening in data and MC agreement in other variables. The $\bar{d}_{innerTrack}$ re-weighting on the other hand does not impose any negative effects on the other variables and also improves on the variables that did have a comparably worse modelling before applying weights, e.g. ΔR_{max} or $f_{EM}^{track-HAD}$. These positive effects also weigh in favour of the $\bar{d}_{innerTrack}$ re-weighting rather than the similarly well performing f_{track}^{EM} re-weighting.

To summarise, the best 1-dimensional re-weighting for the 1-prong channel can be calculated from $\bar{d}_{innerTrack}$. These weights are shown in Figure 6.4. For the bins with relatively high event counts in the interval of $[0, 0.07]$, the weights deviate from one by a few per cent. The rest of the weights stem from bins with low event counts, hence the higher value of the statistical errors and deviations that can be found in these bins. Although the weight values only deviate from one by a small amount, the improvements on the data and MC agreement are quite significant. All ID variable distributions with the $\bar{d}_{innerTrack}$ re-weighting applied can be found in Figures 6.5(a)-(i).

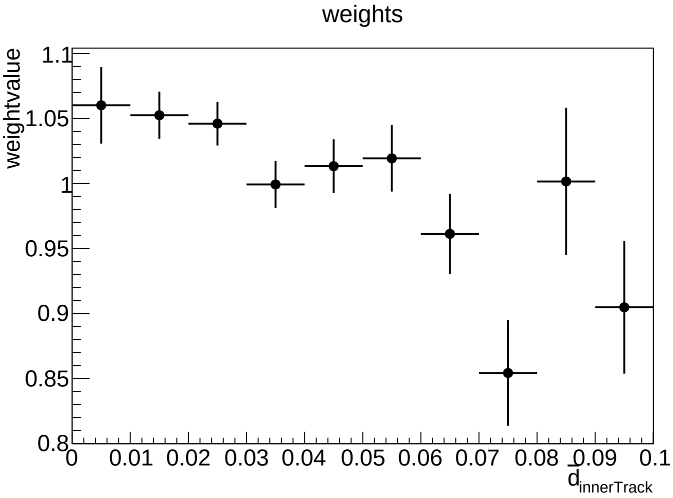
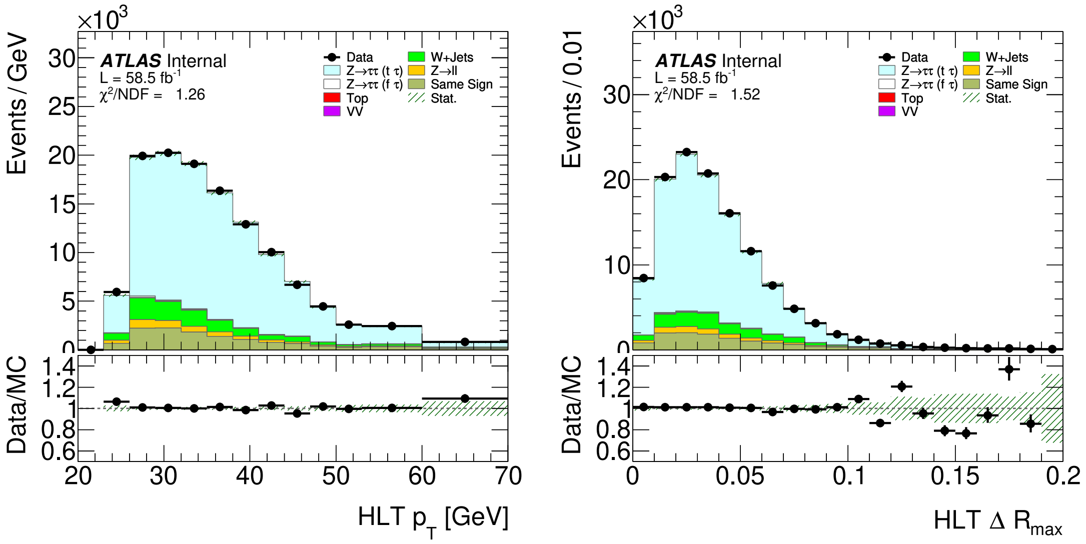
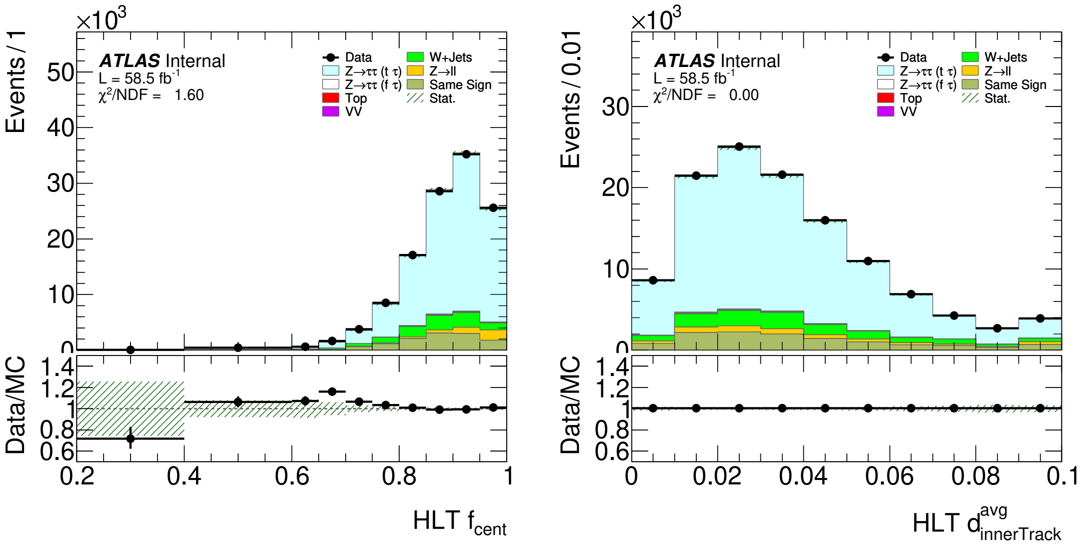


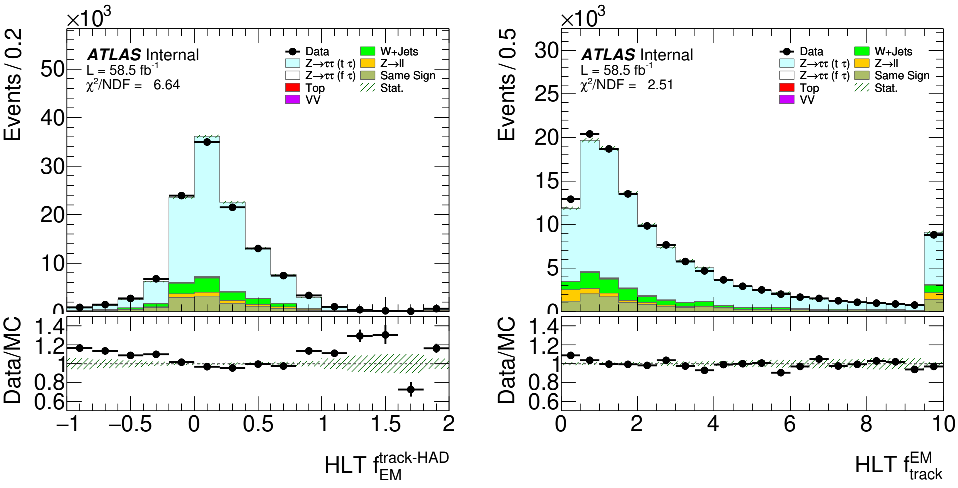
Figure 6.4.: The set of signal MC weights and their statistical error calculated from $\bar{d}_{innerTrack}$ in the 1-prong channel according to Formulas 6.1 and 6.2.



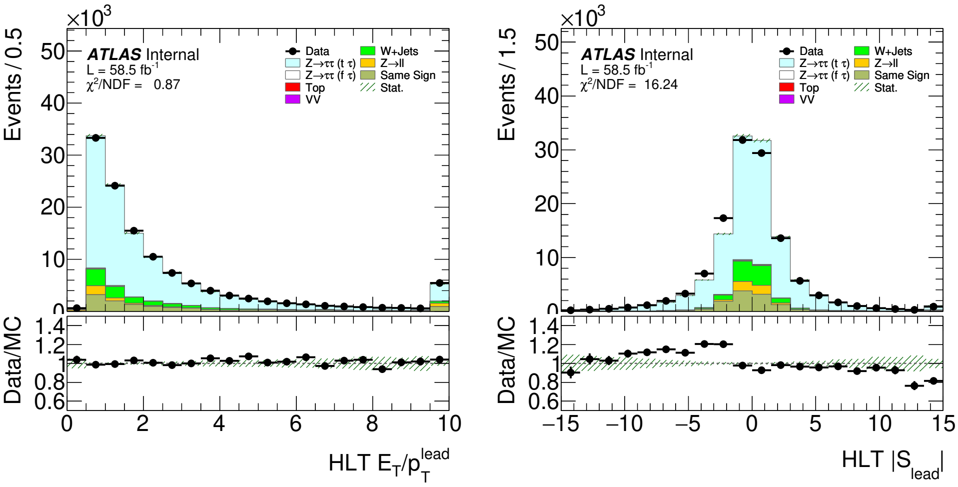
(a) p_T (b) ΔR_{max}



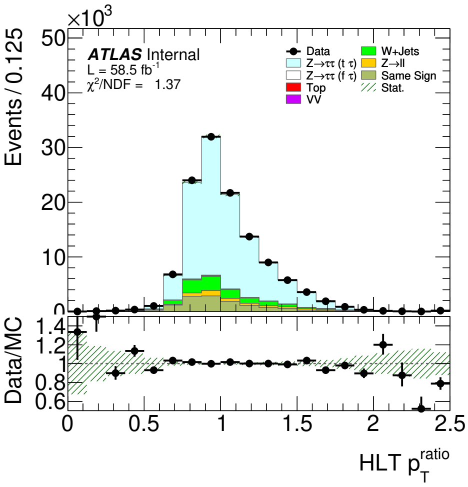
(c) f_{cent} (d) $\bar{d}_{innerTrack}$



(e) $f_{EM}^{track-HAD}$ (f) f_{track}^{EM}



(g) $\frac{E_T}{p_T^{lead}}$ (h) $|S_{lead}|$



(i) p_T^{ratio}

Figure 6.5.: Distributions of all the ID variables listed in e.g. Table 6.1 for the 1-prong channel with the $\bar{d}_{innerTrack}$ re-weighting applied.

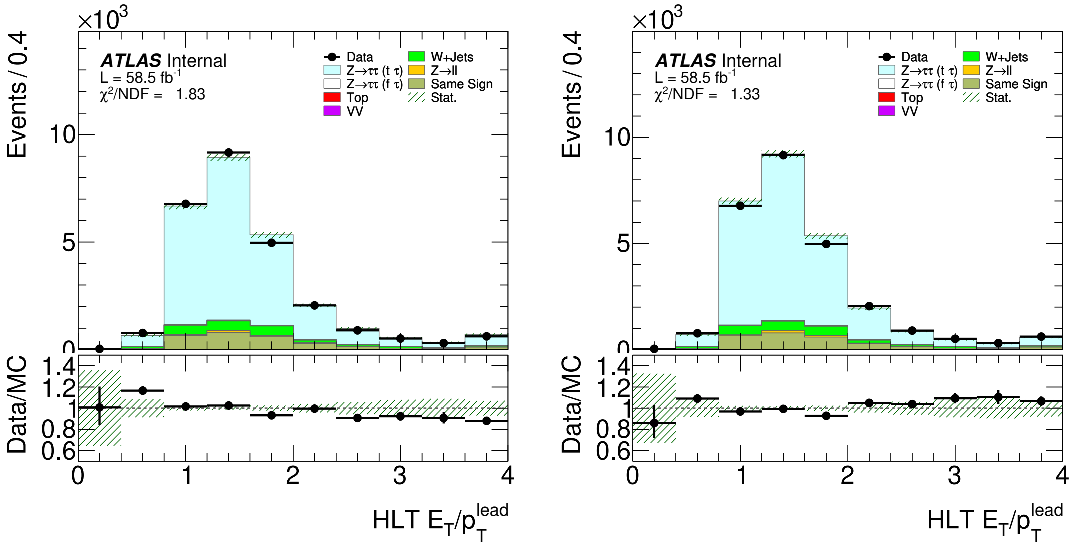
6.2.3. Results for the 3-prong Channel

From the results of Section 5.2 for the 3-prong HLT variables, it can be seen that there is a lot of potential to improve the modelling in this τ lepton decay channel. The 3-prong channel generally has less events available compared to the 1-prong channel and involves more QCD background, which makes the reconstruction more difficult.
The weights calculated from the f_{track}^{EM}, f_{cent} and $f_{EM}^{track-HAD}$ distributions (see e.g. Figures 5.4(c), 5.5(c) and 5.3(c).) have been applied to the signal region and the results of the data and MC agreement quantified by their respective χ^2/NDF values are summarised in Table 6.2.

	no weight	f_{track}^{EM}	f_{cent}	$f_{EM}^{track-HAD}$
p_T	1.2	1.3	1.1	1.7
ΔR_{max}	0.5	0.5	0.4	0.5
f_{cent}	7.6	5.9	0.0	8.7
$\bar{d}_{innerTrack}$	2.0	1.5	2.2	1.7
m_{track}	2.7	2.0	2.4	1.9
$f_{EM}^{track-HAD}$	4.9	2.1	6.0	0.0
f_{track}^{EM}	8.7	0.0	7.5	5.1
$\frac{E_T}{p_T^{lead}}$	1.8	1.3	1.2	0.8
S_T^{flight}	1.2	1.4	1.0	1.4
p_T^{ratio}	1.2	1.4	1.5	0.9

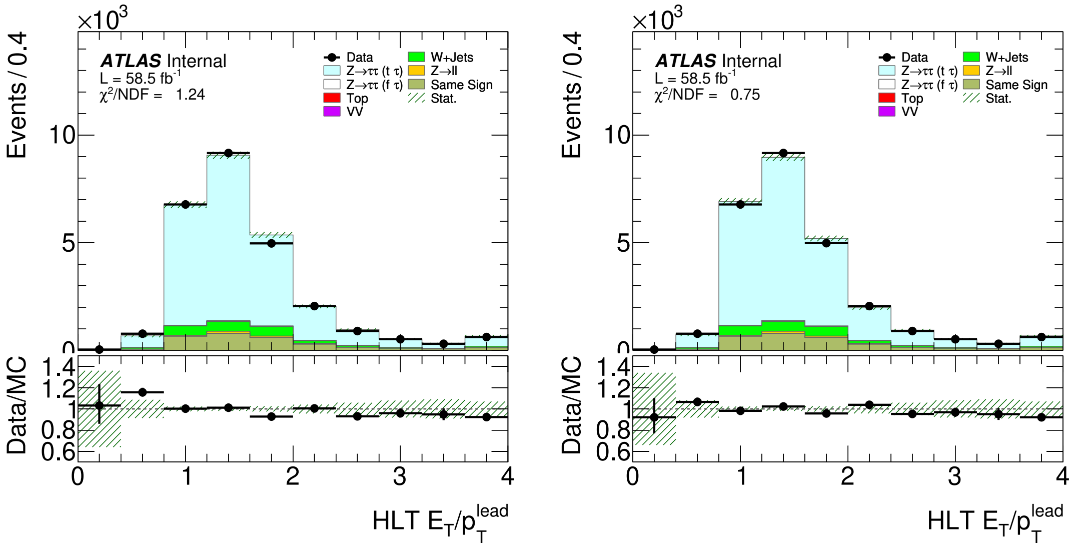
Table 6.2.: Results of the reduced χ^2 value for different re-weightings in the 3-prong channel. Based on these results, the f_{track}^{EM} re-weighting is recommended.

In Figure 6.6, the distributions for $\frac{E_T}{p_T^{lead}}$ with the different weight sets is shown. For this variable, all re-weightings improve the agreement between data and MC. The positive effects are more prominent in the higher bins that contain less events overall.
The next variable to consider in these investigations is $f_{EM}^{track-HAD}$, depicted in Figure 6.7. This variable is ranked with the third largest χ^2/NDF in the ID variable set. Applying the weights from f_{track}^{EM} improves the modelling in the region $[-0.5, 1.5]$, where most of the events can be found. Naturally, it is expected that the $f_{EM}^{track-HAD}$ weights create an exact agreement between data and MC. The f_{cent} weights in contrast affect this variable negatively. When the $f_{EM}^{track-HAD}$ weights are applied to f_{cent}, the quality of the modelling decreases as well. This is the reason why both the f_{cent} and the $f_{EM}^{track-HAD}$ re-weightings should be discarded.
To summarise the results of the re-weighting studies for 3-prong, there is more room to improve the modelling of the ID variable set. While the $f_{EM}^{track-HAD}$ re-weighting does not yield convincing results, both the f_{track}^{EM} and the f_{cent} re-weightings allow for improvement



(a) no weights

(b) f_{track}^{EM}

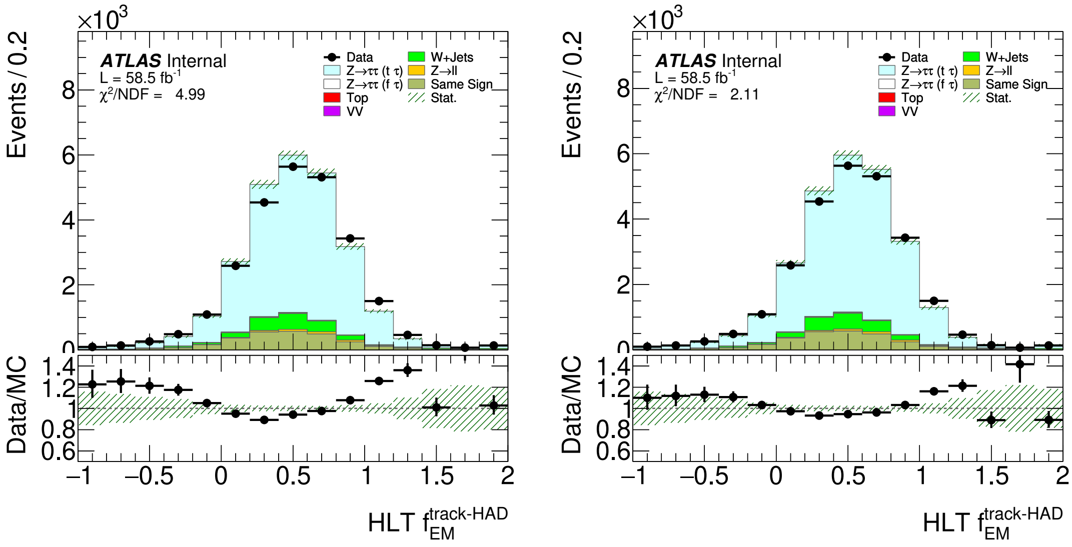


(c) f_{cent}

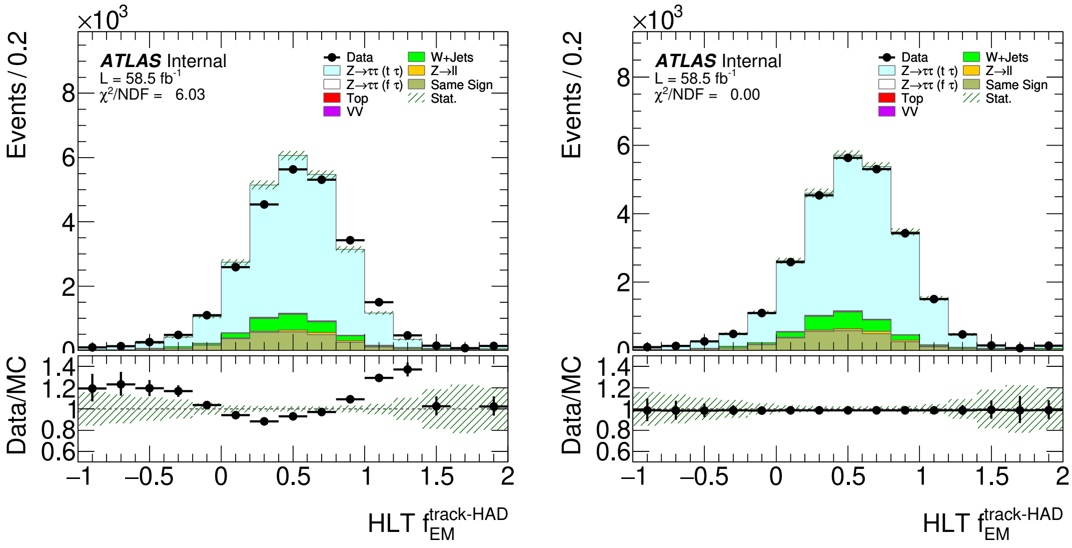
(d) $f_{EM}^{track-HAD}$

Figure 6.6.: Distributions of $\frac{E_T}{p_T^{lead}}$ for the 3-prong channel (a) without weights and with the three weight-sets calculated from (b) f_{track}^{EM}, (c) f_{cent} and (d) $f_{EM}^{track-HAD}$ applied.

for most of the variables. However, the f_{cent} weights negatively affect one of the worst modelled variables $f_{EM}^{track-HAD}$. Therefore, the f_{track}^{EM} re-weighting is recommended to be used for a 3-prong ID algorithm training. This set of weights is shown in Figure 6.8 together with the corresponding statistical uncertainty. It can be seen that the weight values deviate much further from one than for the calculated 1-prong weights, ranging in between the interval $[0.2, 1.4]$ compared to a range of $[0.85, 1.06]$ in Figure 6.4. The



(a) no weights (b) f_{track}^{EM}



(c) f_{cent} (d) $f_{EM}^{track-HAD}$

Figure 6.7.: Distributions of $f_{EM}^{track-HAD}$ for the 3-prong channel (a) without weights and with the three weight-sets calculated from (b) f_{track}^{EM}, (c) f_{cent} and (d) $f_{EM}^{track-HAD}$ applied.

statistical uncertainty once again increases for bins with low event counts for the last three bins. All ID variable distributions with the f_{track}^{EM} re-weighting applied can be found in Figures 6.9(a)-(j).

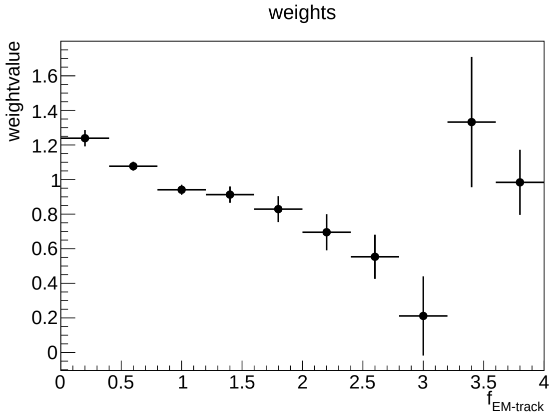
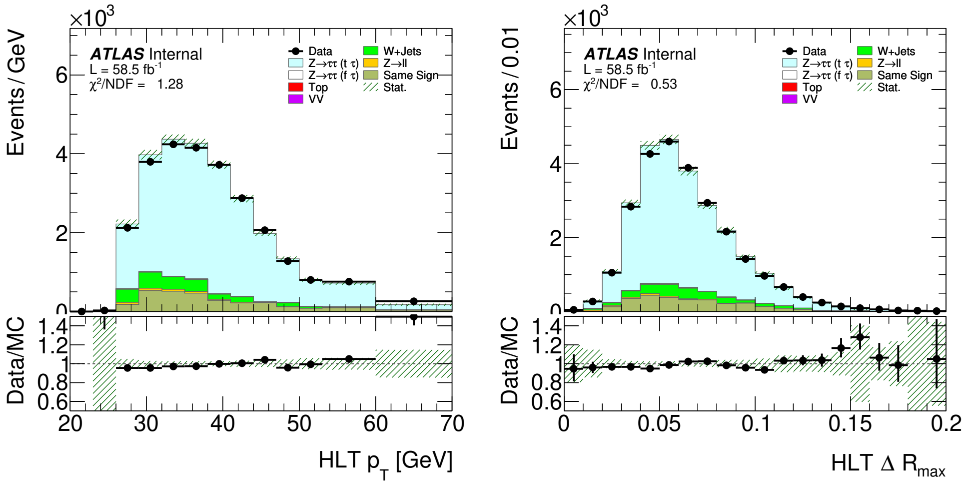
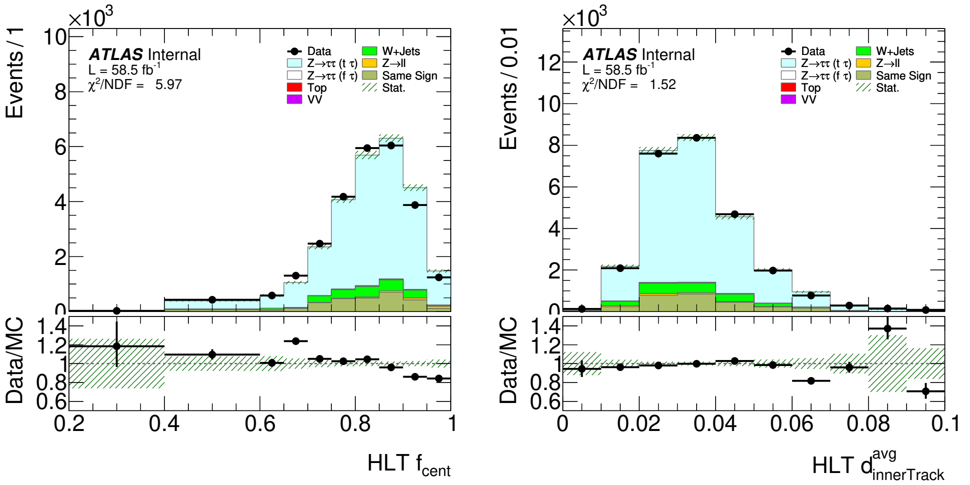


Figure 6.8.: The set of signal MC weights and their statistical error calculated from f_{track}^{EM} in the 3-prong channel according to Formulas 6.1 and 6.2.



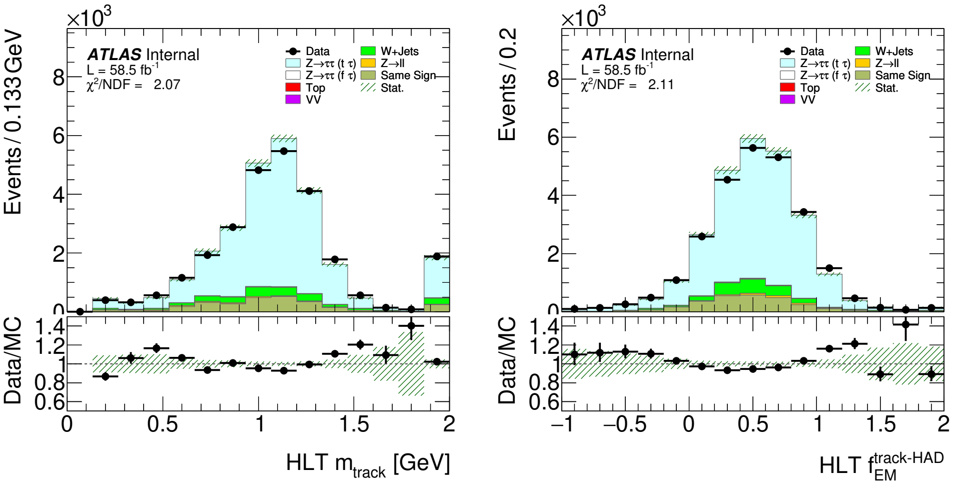
(a) p_T

(b) ΔR_{max}

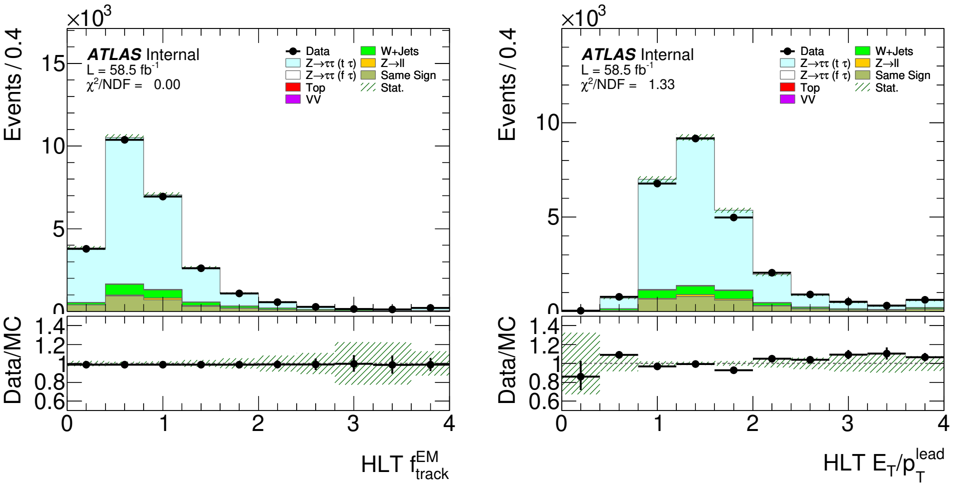


(c) f_{cent}

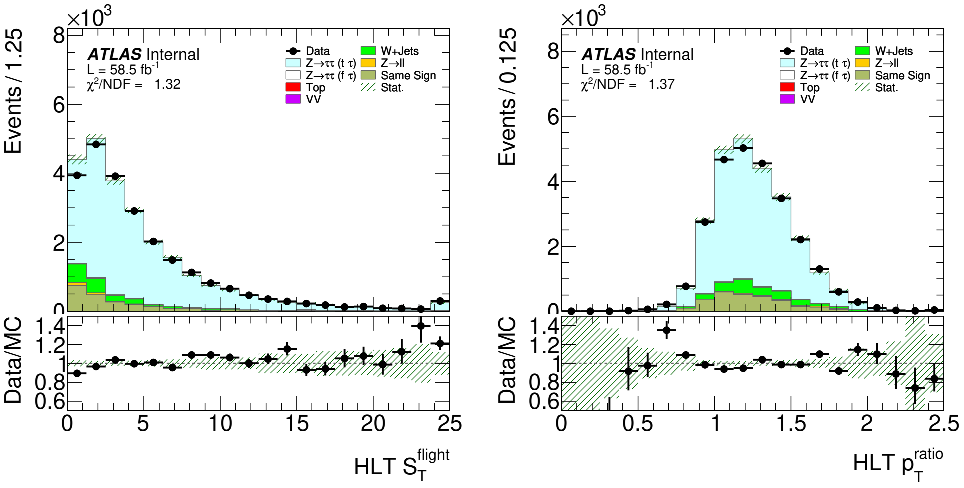
(d) $\bar{d}_{innerTrack}$



(e) m_{track} (f) $f_{EM}^{track-HAD}$



(g) f_{track}^{EM} (h) $\frac{E_T}{p_T^{lead}}$



(i) $|S_T^{flight}|$ (j) p_T^{ratio}

Figure 6.9.: Distributions of all the ID variables listed in e.g. Table 6.2 for the 3-prong channel with the f_{track}^{EM} re-weighting applied.

6.3. Expansion of the Re-Weighting Strategy into Two Variable Dimensions

In this Section, the signal re-weighting is expanded into two variable dimensions, where a joint weight set is calculated based on two ID variables. The adjusted strategy for two variables together with the selected sets of variables that have been used to perform the re-weighting are listed in Section 6.3.1. The results of this 2-dimensional signal re-weighting are presented for the 1-prong and 3-prong decay channels in Sections 6.3.2 and 6.3.3, respectively.

6.3.1. Variable Selection

Analogous to the previous section, a re-weighting of the signal sample for two dimensions has been performed. The strategy is to proceed in a similar way compared to the 1-dimensional re-weighting, by choosing the variables with the highest ranked χ^2/NDF as the first re-weighting variable. The second variable is determined by choosing distributions that are the least correlated to the first variable that has been selected in order to maximise the effect of the re-weighting to the SR. The two variables are then evaluated together in a 2-dimensional histogram, which is then used to calculate a joint set of weights akin to the calculations from Formula 6.1.

For 1-prong τ candidates, the weights that performed the best in the 1-dimensional case were calculated from $\bar{d}_{innerTrack}$ and f^{EM}_{track}. Thus, these two variables have been chosen as the first variables of the 2-dimensional distribution on which a set of weights was calculated.

In Table 6.3 the correlation factors for all the ID variable combinations for 1-prong τ candidates are listed. The variables that are the least correlated with $\bar{d}_{innerTrack}$ are f_{cent} and $|S_{lead}|$ with correlation factors of $r = 0.05$ and $r = 0.00$, respectively. Furthermore, these variables have been chosen to be investigated together with f^{EM}_{track} as well, which makes for four different combinations of two-dimensional weights for the 1-prong channel:

1. $\bar{d}_{innerTrack}$ and f_{cent},
2. $\bar{d}_{innerTrack}$ and $|S_{lead}|$,
3. f^{EM}_{track} and f_{cent},
4. f^{EM}_{track} and $|S_{lead}|$.

	$\bar{d}_{innerTrk}$	$f_{EM}^{trk-had}$	f_{cent}	f_{track}^{EM}	$\frac{E_T}{p_T^{lead}}$	ΔR_{max}	p_T^{ratio}	$\lvert S_{lead}\rvert$
$\bar{d}_{innerTrk}$	1	-0.21	0.05	0.48	0.48	0.84	0.11	0.00
$f_{EM}^{trk-had}$	-0.21	1	0.04	-0.23	-0.30	-0.14	0.81	-0.01
f_{cent}	0.05	0.04	1	0.21	0.24	0.06	0.11	0.00
f_{track}^{EM}	0.48	-0.23	0.21	1	0.93	0.39	-0.32	0.01
$\frac{E_T}{p_T^{lead}}$	0.48	-0.30	0.24	0.93	1	0.42	-0.39	0.01
ΔR_{max}	0.84	-0.14	0.06	0.39	0.42	1	-0.19	0.00
p_T^{ratio}	0.11	0.81	0.11	-0.32	-0.39	-0.19	1	-0.01
$\lvert S_{lead}\rvert$	0.00	-0.01	0.00	0.01	0.01	0.00	-0.01	1

Table 6.3.: Linear correlation factors of the ID variables for the 1-prong decay channel.

From Section 6.2.3, the variables with the highest χ^2/NDF values for 3-prong τ candidates are $f_{EM}^{track-HAD}$, f_{track}^{EM} and f_{cent}. These distributions will be used again to calculate the 2-dimensional weights. The correlation factors of the ID variables for the 3-prong channel are summarised in Table 6.4. It can be seen, that none of the variables mentioned before have high correlations with each other, resulting in three sets of weights to be investigated:

1. $f_{EM}^{track-HAD}$ and f_{cent},
2. f_{cent} and f_{track}^{EM},
3. $f_{EM}^{track-HAD}$ and f_{track}^{EM}.

	$\bar{d}_{inTrk}$	$f_{EM}^{trk-had}$	f_{cent}	f_{track}^{EM}	$\frac{E_T}{p_T^{lead}}$	ΔR_{max}	p_T^{ratio}	S_T^{flight}	m_{track}
$\bar{d}_{inTrk}$	1	-0.11	-0.13	0.19	0.14	0.5	-0.09	0.09	0.11
$f_{EM}^{trk-had}$	-0.11	1	0.09	-0.22	-0.38	-0.05	0.75	-0.01	0.29
f_{cent}	-0.13	0.09	1	0.18	0.11	-0.04	0.2	-0.06	-0.11
f_{track}^{EM}	0.19	-0.22	0.18	1	0.6	0.07	-0.05	0.00	-0.4
$\frac{E_T}{p_T^{lead}}$	0.14	-0.38	0.11	0.6	1	-0.1	-0.37	-0.00	-0.33
ΔR_{max}	0.5	-0.05	-0.04	0.07	-0.1	1	0.00	0.09	0.34
p_T^{ratio}	-0.09	0.75	0.2	-0.05	-0.37	0.00	1	-0.02	0.28
S_T^{flight}	0.09	-0.01	-0.06	0.00	-0.00	0.09	-0.02	1	0.15
m_{track}	0.11	0.29	-0.11	-0.4	-0.33	0.34	0.28	0.15	1

Table 6.4.: Linear correlation factors of the ID variables for the 3-prong decay channel.

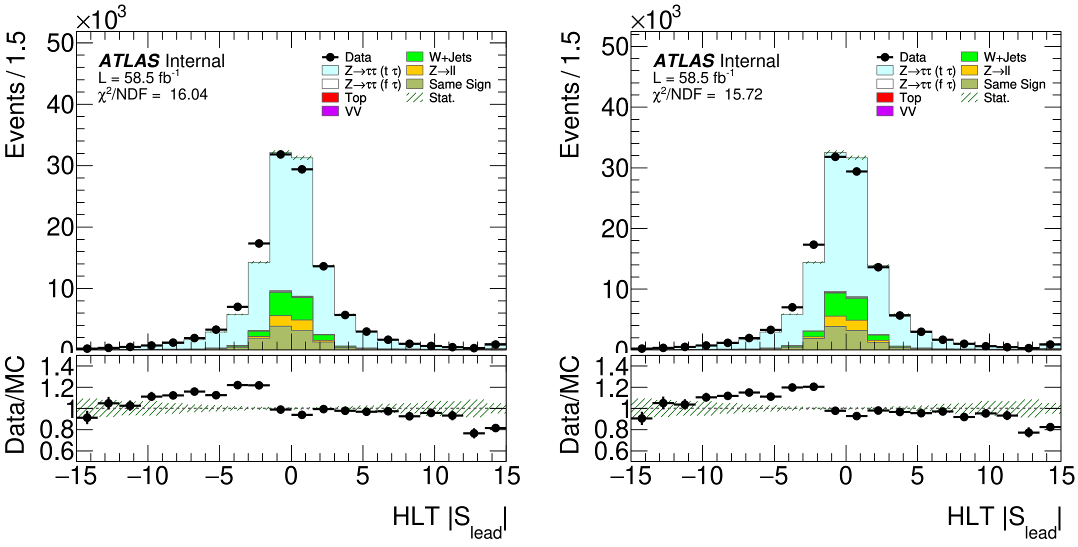
6.3.2. Results of the 2-Dimensional Signal Re-weighting for the 1-prong Channel

Based on the variable selection for the 1-prong decay from Section 6.3.1, 2-dimensional weights have been calculated with the variable sets $\bar{d}_{innerTrack}$+f_{cent}, $\bar{d}_{innerTrack}$+$|S_{lead}|$, f_{track}^{EM}+f_{cent} and f_{track}^{EM}+$|S_{lead}|$. Whether the re-weightings show an improvement on the modelling compared to before is again assessed by looking at the reduced χ^2 and the data and MC agreement of the variable distributions after the re-weighting. In Table 6.5, the reduced χ^2 for all the variables and weight configurations are listed. It can be noted, that all re-weightings seem to lower the χ^2/NDF values. There are two outliers for $\frac{E_T}{p_T^{lead}}$, where the f_{track}^{EM}+f_{cent} and f_{track}^{EM}+$|S_{lead}|$ re-weightings cause the χ^2/NDF to rise from 0.6 to 2.6 and 2.7, respectively. Therefore, these re-weightings will not be considered further.

Variable	no weight	f_{track}^{EM}+f_{cent}	f_{cent}+$\bar{d}_{innerTrack}$	$\bar{d}_{innerTrack}$ +$\lvert S_{lead}\rvert$	f_{track}^{EM}+$\lvert S_{lead}\rvert$
p_T	1.2	1.2	1.2	1.2	1.2
ΔR_{max}	3.6	2.2	1.5	1.5	2.1
f_{cent}	1.5	0.2	0.2	1.4	1.2
$\bar{d}_{innerTrack}$	3.7	1.3	0.0	0.0	1.3
$f_{EM}^{track-HAD}$	7.5	5.1	6.7	6.3	3.8
f_{track}^{EM}	3.9	0.0	2.4	2.5	0.0
$\frac{E_T}{p_T^{lead}}$	0.6	2.6	1.0	0.8	2.7
$\lvert S_{lead}\rvert$	16.0	15.9	15.7	0.0	0.0
p_T^{ratio}	1.4	1.8	1.3	1.2	1.2

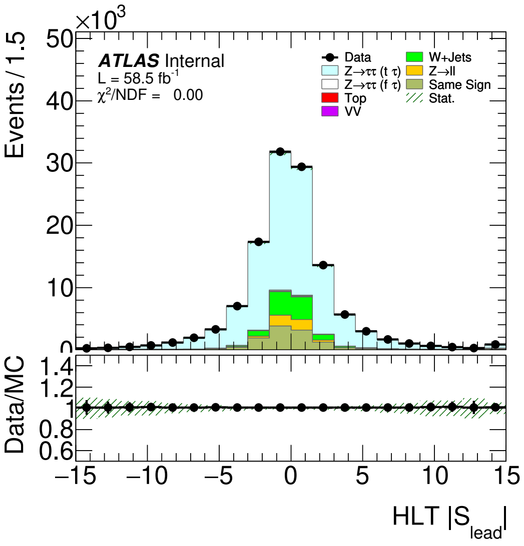
Table 6.5.: Summary of the reduced χ^2 for the 1-prong channel without weights and with the four weight sets calculated from f_{track}^{EM}+f_{cent}, f_{cent}+$\bar{d}_{innerTrack}$, $\bar{d}_{innerTrack}$ +$|S_{lead}|$ and f_{track}^{EM}+$|S_{lead}|$ applied. Based on these results, the $\bar{d}_{innerTrack}$ +$|S_{lead}|$ re-weighting is recommended for a 2-dimensional weighting.

The other two weight sets calculated from $\bar{d}_{innerTrack}$+f_{cent} and $\bar{d}_{innerTrack}$+$|S_{lead}|$ yield similar results from the updated reduced χ^2 values. However, in Figure 6.10, it is evident that $|S_{lead}|$ is the ID variable with the highest χ^2, which is highlighted by the data and MC agreement in Figure 6.10(a), especially in the range of $[-10, 0]$. Since $|S_{lead}|$ is not correlated to the other ID variables (see Table 6.3), the $\bar{d}_{innerTrack}$+f_{cent} re-weighting does not affect the modelling of this distribution, while the $\bar{d}_{innerTrack}$+$|S_{lead}|$ re-weighting yields perfect agreement between the data and MC samples, as can be seen in Figures 6.10(b) and (c). Hence, the weights calculated from $\bar{d}_{innerTrack}$ and $|S_{lead}|$ are seen to perform the best out of all the re-weighting for the 1-prong channel. It can be reasoned, that this result is consistent with Section 6.2.2 with an additional positive effect on the worst modelled variable.



(a) no weight

(b) f_{cent} $+\bar{d}_{innerTrack}$



(c) $\bar{d}_{innerTrack}$ $+|S_{lead}|$

Figure 6.10.: $|S_{lead}|$ for the 1-prong channel (a) without weights and with the four weight sets calculated from (b) $f_{cent}+\bar{d}_{innerTrack}$ and (c) $\bar{d}_{innerTrack}+|S_{lead}|$ applied.

The 2-dimensional distribution of the chosen re-weighting can be seen in Figure 6.11. It is evident, that for the 2-dimensional re-weighting, the number of weights that have to be calculated rises quickly with the number of bins per variable. In this case, the weight set has $10 \times 20 = 200$ weights compared to the 1-dimensional re-weighting from $\bar{d}_{innerTrack}$ with 10 weights. For bins with very little events, some weights can take on relatively high values, in Figure 6.11 there is a bin with a weight of 5.1.

All the ID variable distributions with the $\bar{d}_{innerTrack}$+$|S_{lead}|$ weights applied for the 1-prong channel can be found in Figures 6.12(a)-(i).

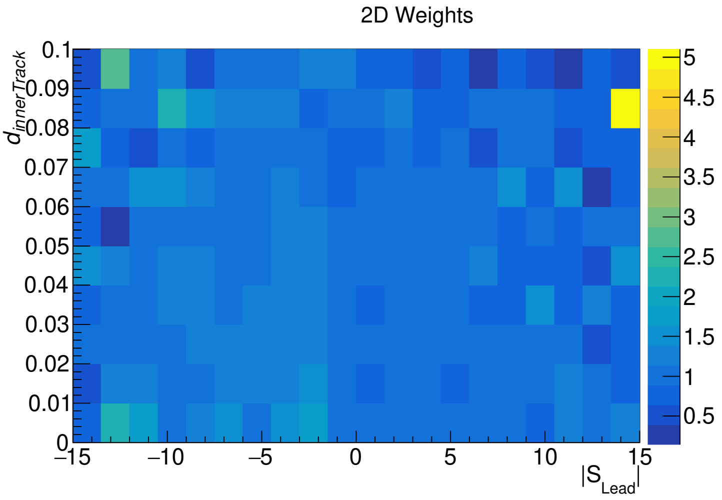
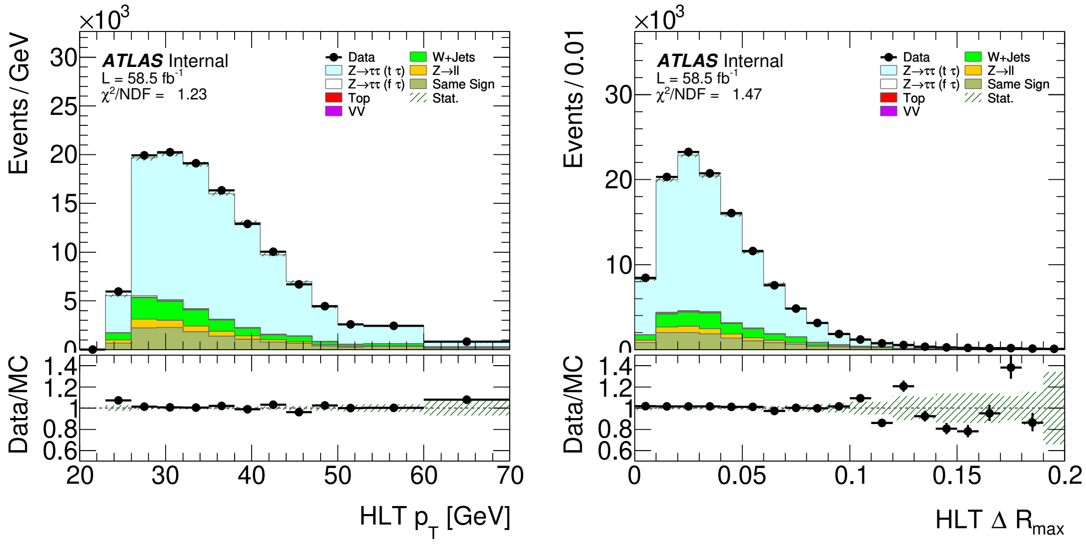
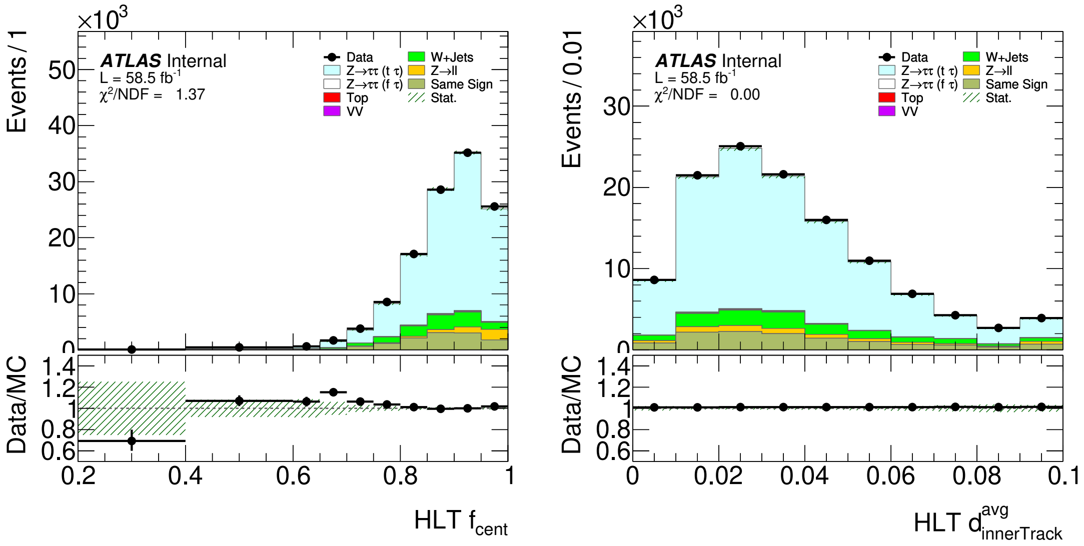


Figure 6.11.: The set of signal MC weights calculated from $\bar{d}_{innerTrack}$ +$|S_{lead}|$ in the 1-prong channel according to Formula 6.1.

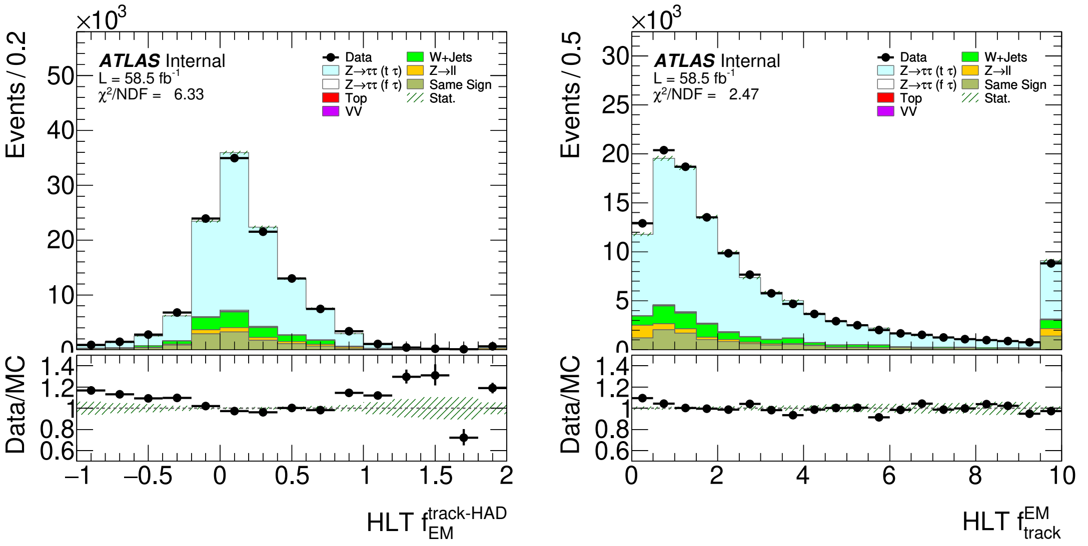


(a) p_T

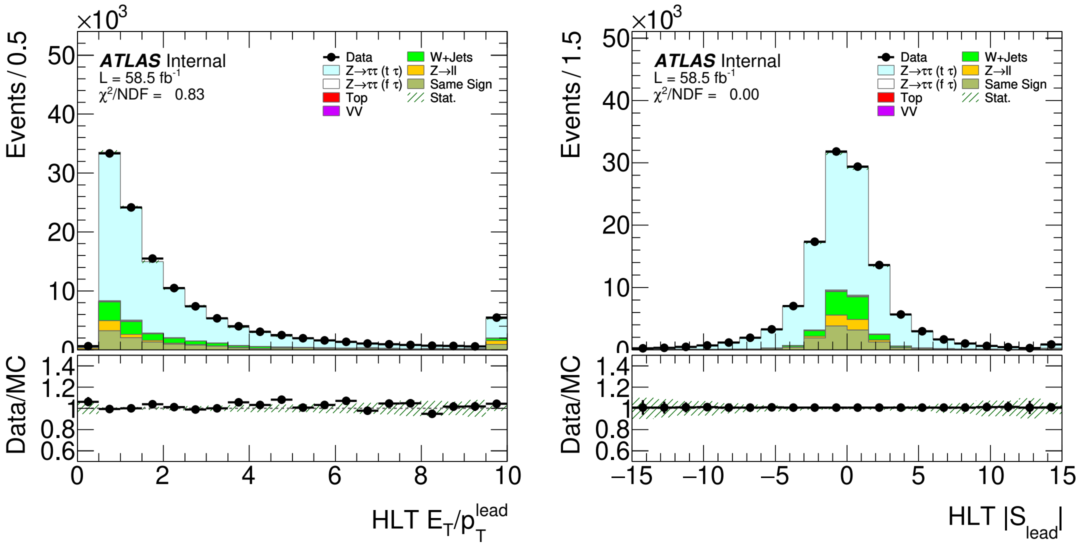
(b) ΔR_{max}



(c) f_{cent} (d) $\bar{d}_{innerTrack}$

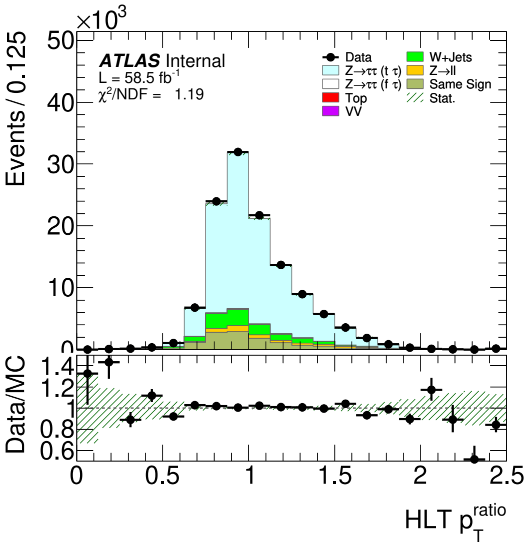


(e) $f_{EM}^{track-HAD}$ (f) f_{track}^{EM}



(g) $\frac{E_T}{p_T^{lead}}$

(h) $|S_{lead}|$



(i) p_T^{ratio}

Figure 6.12.: ID variable distributions for the 1-prong channel with the 2-dimensional $\bar{d}_{innerTrack}$+$|S_{lead}|$ re-weighting applied.

6.3.3. Results of the 2-Dimensional Signal Re-weighting for the 3-prong Channel

From the variable selection for 3-prong τ candidates from Section 6.3.1, 2-dimensional weights have been calculated with the variable sets $f_{cent}+f_{track}^{EM}$, $f_{track}^{EM} + f_{EM}^{track-HAD}$ and $f_{EM}^{track-HAD}+f_{cent}$. In Table 6.6, the reduced χ^2 for all the variables and weight configurations are listed. Based on the results from Table 6.6, it is apparent that the 2-dimensional re-weighting improves the χ^2/NDF values for most of the variables significantly.

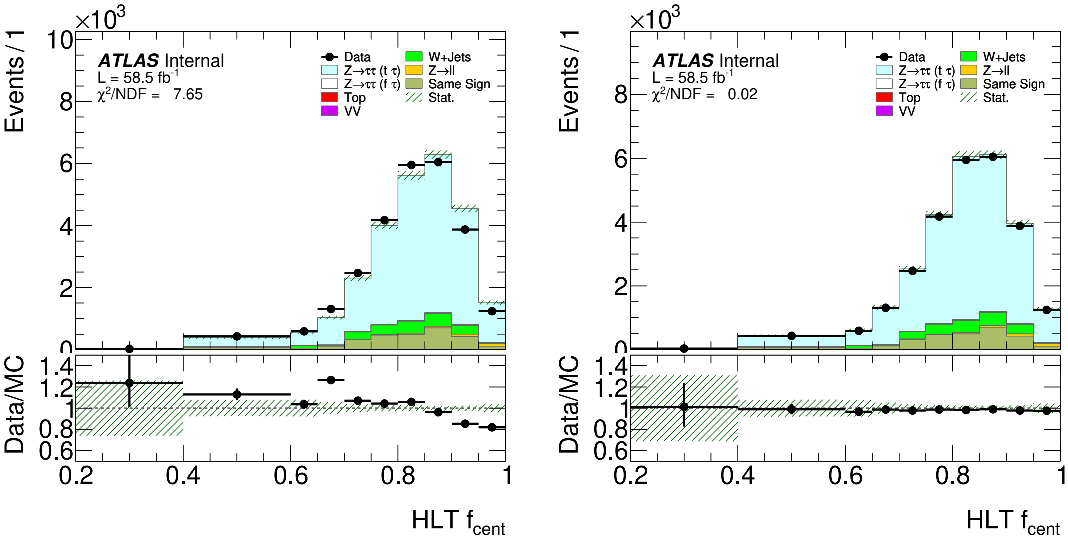
Variables	no weight	$f_{cent} + f_{track}^{EM}$	$f_{track}^{EM} + f_{EM}^{track-HAD}$	$f_{EM}^{track-HAD}+f_{cent}$
p_T	1.2	1.3	1.6	1.7
ΔR_{max}	0.5	0.4	0.5	0.5
f_{cent}	7.6	0.0	7.4	0.1
$\bar{d}_{innerTrack}$	2.0	1.7	1.6	1.8
m_{track}	2.7	1.9	1.7	1.9
$f_{EM}^{track-HAD}$	4.9	3.3	0.0	0.2
f_{track}^{EM}	8.7	0.6	1.2	4.4
$\frac{E_T}{p_T^{lead}}$	1.8	0.9	0.8	0.6
S_T^{flight}	1.2	1.2	1.4	1.2
p_T^{ratio}	1.2	1.7	1.4	0.7

Table 6.6.: Summary of the reduced χ^2 for 3-prong without weights and with the three weight sets calculated from $f_{cent}+f_{track}^{EM}$, $f_{track}^{EM} + f_{EM}^{track-HAD}$ and $f_{EM}^{track-HAD}+f_{cent}$ applied. Based on these results, the $f_{cent}+f_{track}^{EM}$ re-weighting is recommended for a 2-dimensional weighting.

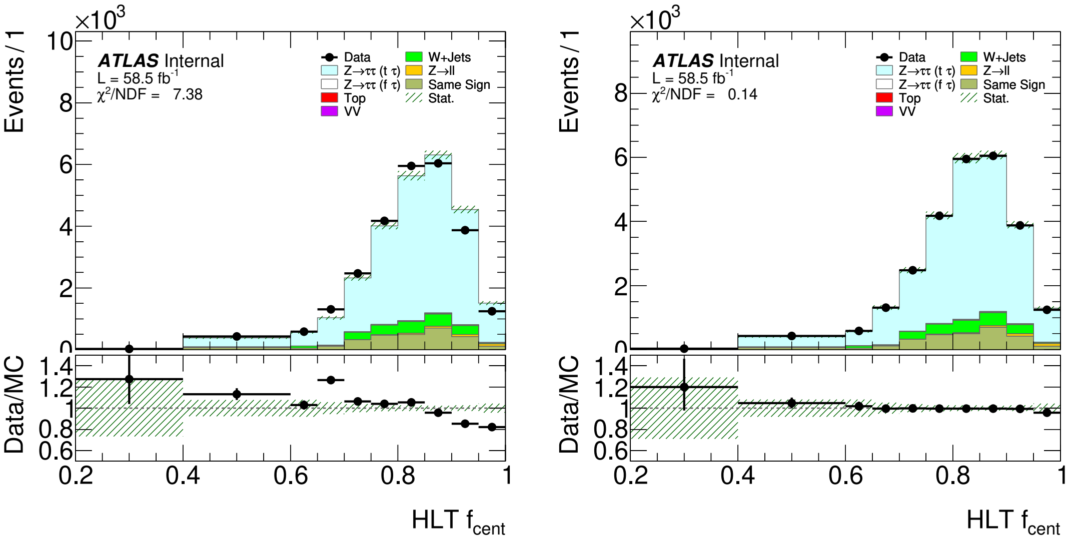
Figure 6.13 shows the different distributions for f_{cent} without and weights in 6.13(a) and the three re-weightings based on $f_{cent}+f_{track}^{EM}$, $f_{track}^{EM}+f_{EM}^{track-HAD}$ and $f_{EM}^{track-HAD}+f_{cent}$ in 6.13(b), (c) and (d), respectively. It can be seen, that the $f_{track}^{EM} + f_{EM}^{track-HAD}$ re-weighting in 6.13(c) does not affect the original distribution by much. With a χ^2/NDF value of 7.4, this variable has the highest value amongst all ID variables after any re-weighting. Given the high quality of the improvements of the 2-dimensional re-weightings, this is an argument to disregard the $f_{track}^{EM} + f_{EM}^{track-HAD}$ re-weighting.

Overall, the $f_{cent}+f_{track}^{EM}$ re-weighting has lower reduced χ^2 values with a peak value of $\chi^2/NDF = 3.3$ from $f_{EM}^{track-HAD}$ compared to the $f_{EM}^{track-HAD}+f_{cent}$ re-weighting with a peak χ^2/NDF value of 4.4 from f_{track}^{EM}.

Based on these results, the best 2-dimensional re-weighting in the 3-prong channel is chosen to be calculated from the f_{cent} and f_{track}^{EM} distributions. Re-weighting the signal samples with two variables with high χ^2/NDF values is quite profitable in terms of the overall modelling in the 3-prong channel.



(a) no weight (b) f_{cent} + f_{track}^{EM}



(c) f_{track}^{EM} + $f_{EM}^{track-HAD}$ (d) $f_{EM}^{track-HAD}$+f_{cent}

Figure 6.13.: f_{cent} for 3-prong (a) without weights and with the three weight sets calculated from (b) f_{cent}+f_{track}^{EM}, (c) f_{track}^{EM} + $f_{EM}^{track-HAD}$ and (d) $f_{EM}^{track-HAD}$+f_{cent} applied.

The best performing re-weightings are consistent with the behaviour of the 1-dimensional weight sets seen in Section 6.2. The re-weightings that are better overall between one and two dimensions for the 1-prong and 3-prong channels are discussed in the recommendation given in Section 6.5.

The 2-dimensional distribution of the chosen re-weighting can be seen in Figure 6.14. The weights calculated from f_{cent} and f_{track}^{EM} in the 3-prong channel have peak values of ~ 2.5, which are overall lower compared to the weights that can be seen in Figure 6.11. The majority of the weight values are of $\mathcal{O}(1.5)$.

All ID variable distributions with the f_{cent}+f_{track}^{EM} re-weighting applied can be found in Figures 6.15(a)-(j).

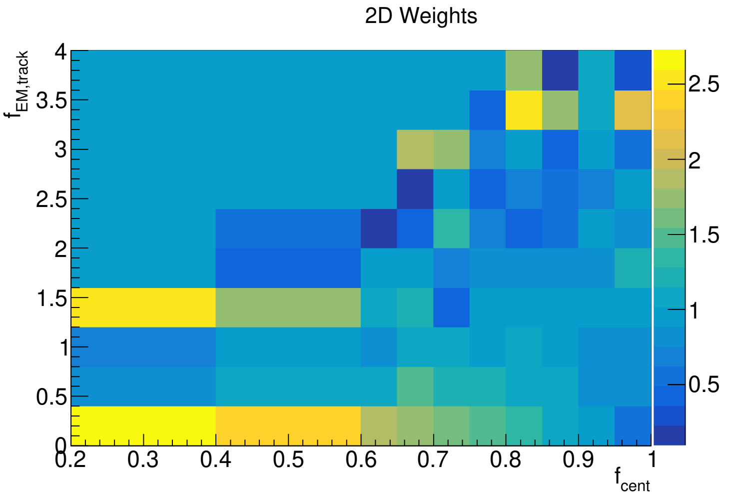
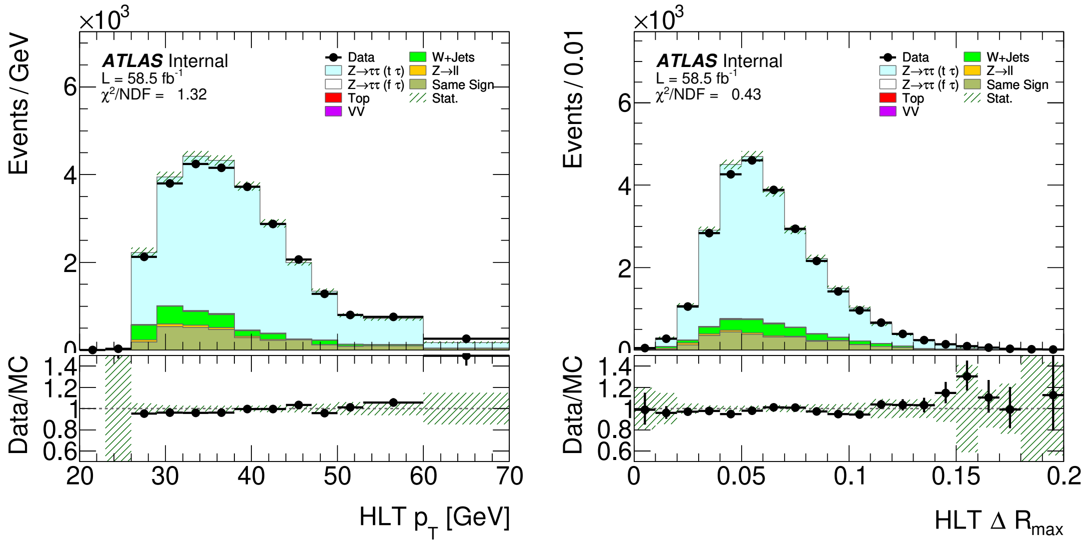
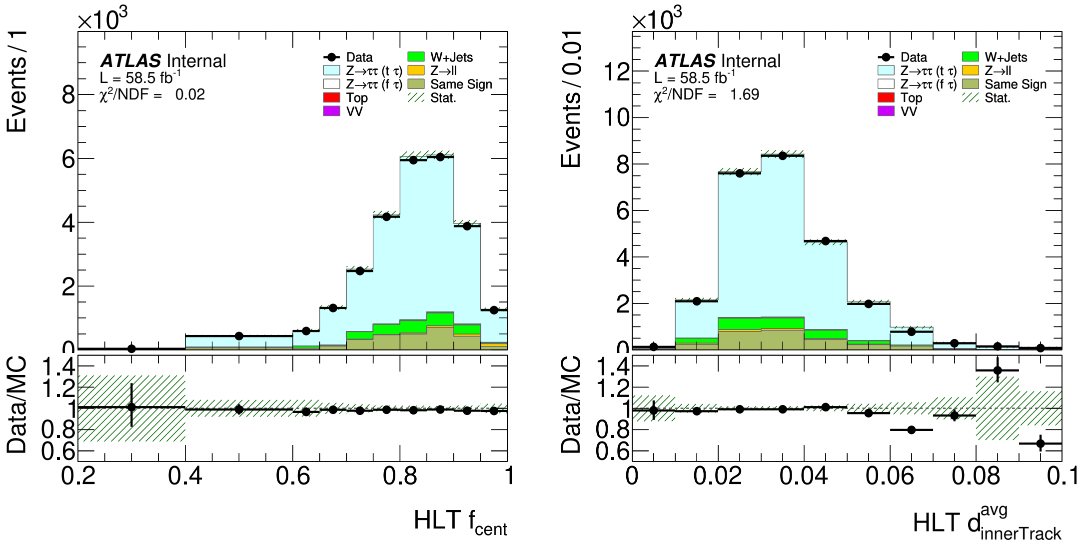


Figure 6.14.: The set of signal MC weights calculated from f_{cent}+f_{track}^{EM} in the 3-prong channel according to Formula 6.1.



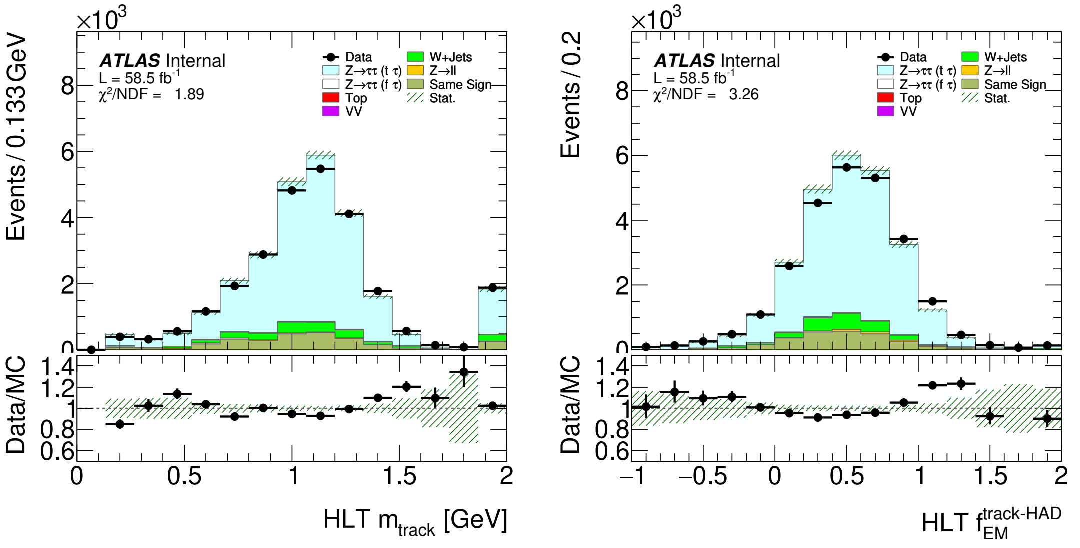
(a) p_T

(b) ΔR_{max}



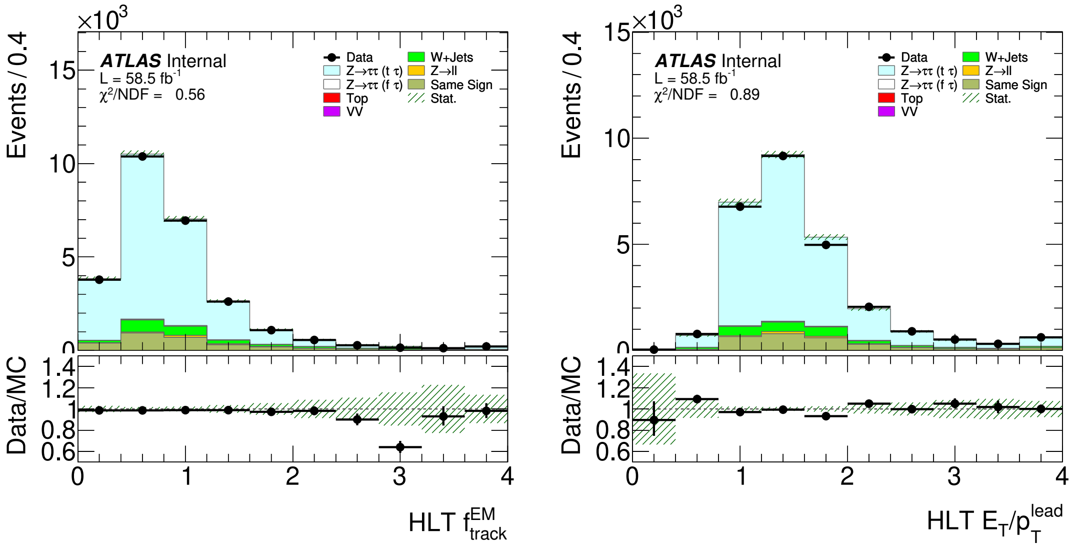
(c) f_{cent}

(d) $\bar{d}_{innerTrack}$

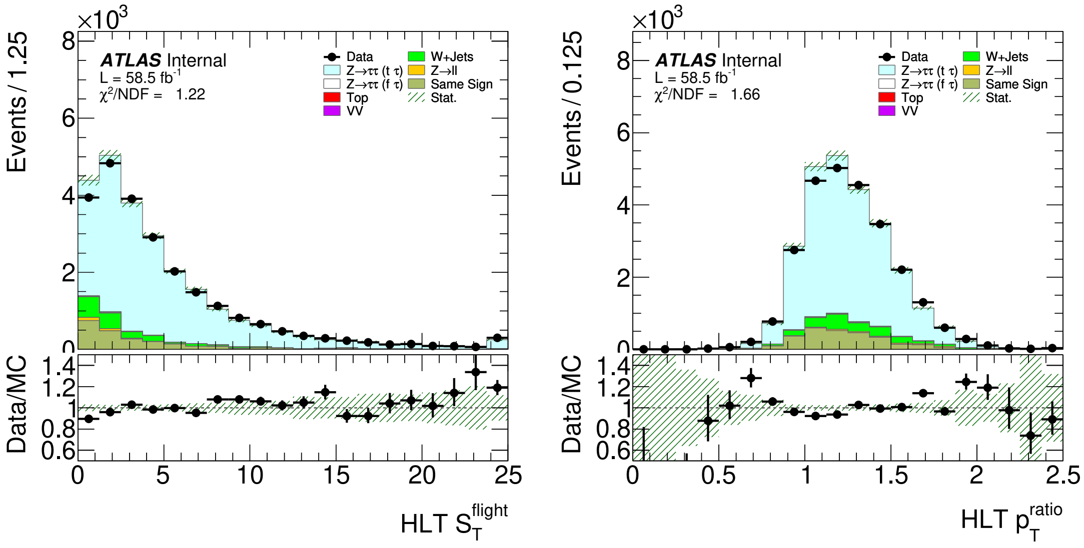


(e) m_{track}

(f) $f_{EM}^{track-HAD}$



(g) f_{track}^{EM} (h) $\frac{E_T}{p_T^{lead}}$



(i) S_T^{flight} (j) p_T^{ratio}

Figure 6.15.: ID variable distributions for the 3-prong channel with the 2-dimensional f_{cent}+f_{track}^{EM} re-weighting applied.

6.4. Cross-Checks and Modifications of the τ Lepton ID

In this Section, systematic variations are applied to the SR in order to test the robustness of the calculated 1-dimensional weights and systematic uncertainties will be discussed. To do that, the offline RNN cut will be adjusted to correspond to a medium or tight τ ID cut and some steps of the signal re-weighting analysis from Section 6.2 will be repeated in these SRs. Tighter RNN cuts result in a higher signal purity. It can be checked if the weights from the weights calculated from the loose RNN SR negatively affect the medium or tight regions, as the re-weighting is naturally sensitive to variations of the signal MC sample.

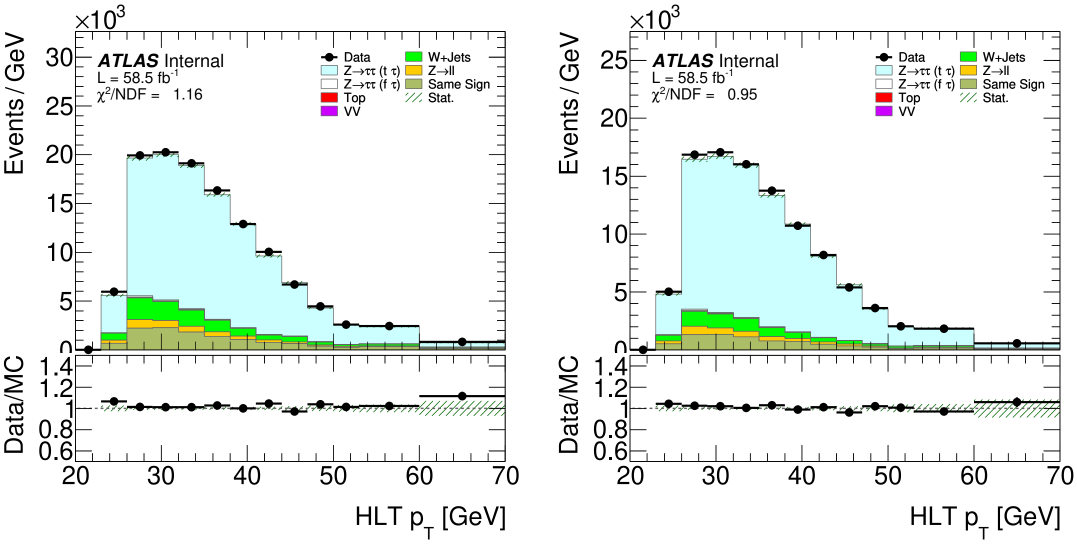
6.4.1. Robustness of the Weights

The first step in order to investigate the weights for different systematic variations is a change in the offline RNN τ ID cut, as explained in Section 4.3 and visualised in Figure 4.5. How a change in RNN ID cuts affects the events in the SR can also be seen in Figure 6.16, where the cuts for loose, medium and tight have been applied to the τ lepton p_T in Figures 6.16(a), (b) and (c), respectively. It can be seen that tightening the RNN cuts changes the total amount of events in the SR significantly. Additionally, for a tighter RNN selection, more background events are cut out, improving the signal purity. The modelling of the distributions from Figure 6.16 is consistent between RNN cuts.

	no weight	w_{loose}	w_{medium}
p_T	0.9	1.3	1.3
ΔR_{max}	2.9	0.9	0.7
f_{cent}	1.6	1.7	1.7
$\bar{d}_{innerTrack}$	3.6	0.3	0.0
$f_{EM}^{track-HAD}$	7.1	6.6	6.5
f_{track}^{EM}	3.3	2.2	2.6
$\frac{E_T}{p_T^{lead}}$	0.9	1.0	1.2
$\lvert S_{lead}\rvert$	12.3	12.9	12.5
p_T^{ratio}	1.4	1.4	1.4

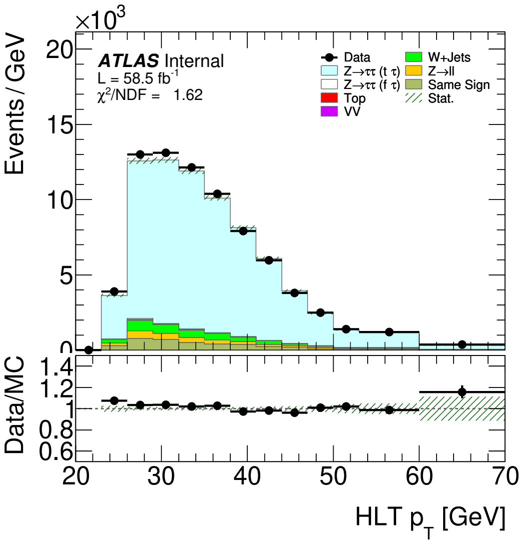
Table 6.7.: Summary of the updated χ^2/NDF for the 1-prong channel after applying the weights calculated from the loose or medium RNN SR in the medium RNN SR.

Table 6.7 shows a summary of the χ^2/NDF values for re-weightings in the medium RNN SR with w_{medium} weights that have been calculated from $\bar{d}_{innerTrack}$ in the medium RNN



(a) loose offline ID

(b) medium offline ID



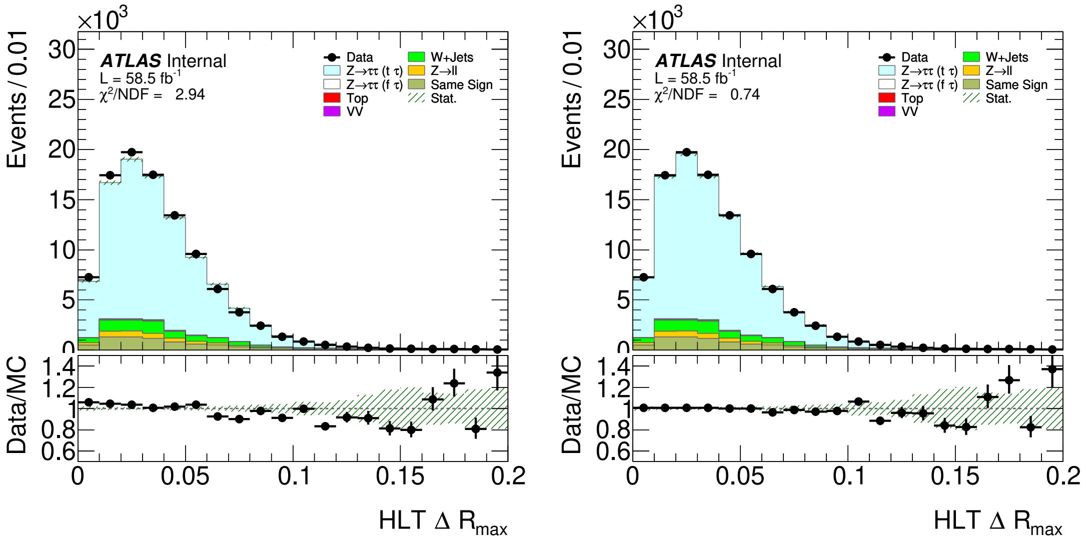
(c) tight offline ID

Figure 6.16.: τ lepton p_T for the different possible RNN ID cuts. The cuts correspond to (a) a loose, (b) a medium and (c) a tight selection.

SR and with the w_{loose} from $\bar{d}_{innerTrack}$ according to Section 6.2. The weights w_{medium} and also w_{tight} have been determined for the best performing variables for the 1-prong and 3-prong channels determined according to sections 6.2.2 and 6.2.3, respectively, and will be used for the considerations of systematic uncertainties in Section 6.4.2.

Figure 6.17(a) shows the ΔR_{max} distribution for the 1-prong SR without any weights and Figure 6.17(b) with w_{medium} applied to the signal sample. It can be seen, that the w_{medium} values influence the modelling of the variables in a similar manner to the loose

SR re-weighting in Figure 6.2. The data and MC agreement improves with the signal re-weighting applied.
Calculating the w_{medium} values from f^{EM}_{track} in the 3-prong channel is similar to the 1-prong channel consistent with the results presented in Section 6.2. Changing the RNN ID cut from loose to medium to tight slowly improves the modelling and signal purity. The weights w_{medium} and w_{tight} calculated from these SRs reduces the χ^2/NDF values even further, yielding results consistent to the results observed in Section 6.2.
The reduced χ^2 values calculated from the medium RNN cut SR without weights, together with the yields for applying w_{loose} and w_{medium}, can be found in Table 6.8.



(a) no weight

(b) $\bar{d}_{innerTrack}$

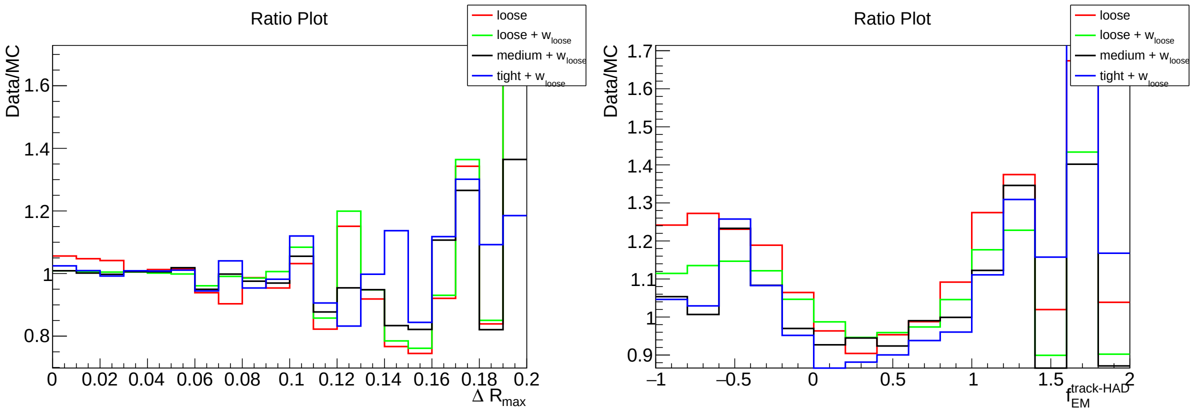
Figure 6.17.: Distributions of ΔR_{max} for the 1-prong channel, (a) without weights and (b) with weights applied, that have been calculated based on the $\bar{d}_{innerTrack}$ ID variable in the medium RNN SR.

The results from applying the w_{loose} values to the medium or tight offline RNN SRs can be used for an assessment of the robustness of the weights under systematic variations. Firstly, the reduced χ^2 values for the 1-prong and 3-prong channels in Tables 6.7 and 6.8, respectively, for w_{loose} indicate that the signal weights still yield an improvement of the ID variable modellings even after systematic variations. This can be confirmed by Figure 6.18, where the data and MC ratios are plotted in the loose, medium and tight RNN cut regions of Figure 6.18(a) ΔR_{max} in the 1-prong channel and Figure 6.18(b) $f^{track-HAD}_{EM}$ in the 3-prong decay channel. The data and MC ratio is consistently closer to one for all the signal regions with w_{loose} applied compared to the SR with a loose RNN cut without the e-weighting applied. This is especially true for bins with a high event count, for example

	no weights	w_{loose}	w_{medium}
p_T	0.7	0.7	0.9
ΔR_{max}	1.0	1.1	1.1
f_{cent}	7.5	6.0	5.8
$\bar{d}_{innerTrack}$	1.6	1.1	1.1
m_{track}	2.7	1.8	1.8
$f_{EM}^{track-HAD}$	5.1	2.3	2.3
f_{track}^{EM}	8.3	0.6	0.0
$\frac{E_T}{p_T^{lead}}$	1.7	1.0	1.0
S_T^{flight}	1.2	1.3	1.4
p_T^{ratio}	1.1	1.3	1.3

Table 6.8.: Summary of the updated χ^2/NDF for the 3-prong channel after applying the weights calculated from the loose or medium RNN SR in the medium RNN SR.

for $0 < \Delta R_{max} < 0.1$ in Figure 6.18(a).



(a) ΔR_{max}, 1-prong

(b) $f_{EM}^{track-HAD}$, 3-prong

Figure 6.18.: Distributions of the data and MC ratios without and with w_{loose} applied to the different considered RNN cuts. The (a) ΔR_{max} is in the 1-prong channel and (b) $f_{EM}^{track-HAD}$ in the 3-prong decay channel.

6.4.2. Systematic Uncertainties

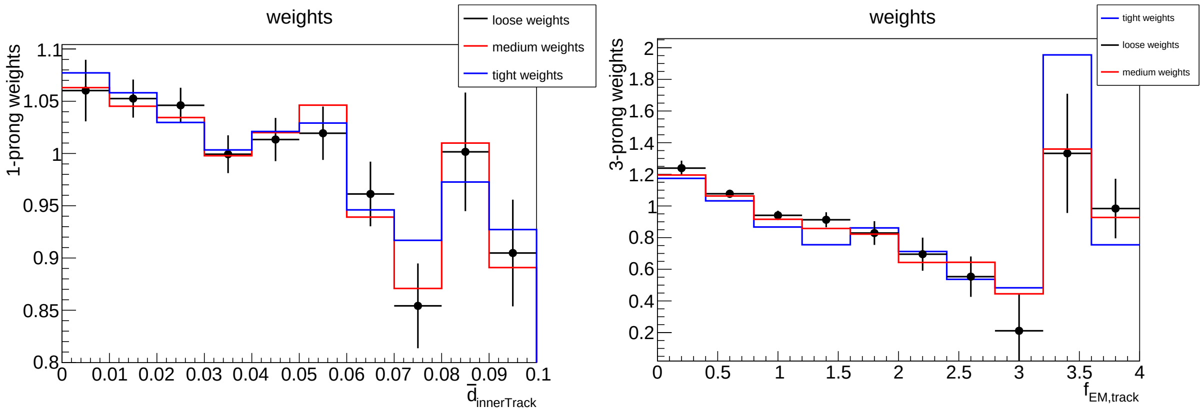
The final step of the signal re-weighting analysis is to discuss systematic uncertainties for the signal MC weights. As mentioned earlier in Section 6.4.1, two different 1-dimensional weight sets w_{medium} and w_{tight} have been calculated in both the 1-prong and the 3-prong channel based on the choice of ID variables for the signal MC re-weightings determined in Section 6.2.

Figure 6.19 displays the weight sets w_{loose}, w_{medium} and w_{tight} calculated from $\bar{d}_{innerTrack}$ for the 1-prong decay in 6.19(a) and f^{EM}_{track} for the 3-prong decay in 6.19(b). These distributions confirm what could already be observed in e.g. Figure 6.18, that the weights are stable under systematic variations, which implies that the weight sets w_{loose}, w_{medium} and w_{tight} are going to be relatively similar. The deviations between the weight sets rise for weights that have been calculated from bins with low events. The differences between w_{medium} or w_{tight} and w_{loose} are suitable to use as systematic uncertainties on w_{loose}.

As seen in Figure 6.19(a), the systematic variations in weights for the 1-prong channel are low throughout the distribution. Given both the w_{medium} and w_{tight} weight sets, systematic uncertainties of $2\% - 8\%$ can be extracted.

For the 3-prong channel on the other hand, higher deviations between the weight sets in some bins e.g. the bin of $[3.2, 3.5]$ in Figure 6.19(b) can be observed. Ignoring these low event count bins, the rest of the distribution has average systematic uncertainties between 5% and 10%.

For both the 1-prong and 3-prong channels, a majority of the higher uncertainty values stem from the w_{tight} weights. Therefore, this set of values should be used as a systematic uncertainty. More systematic uncertainties can be reasoned from slight variations of the SR cuts for future RNN tunings or the amount of data available from the upcoming data-taking period at the LHC.



(a) $\bar{d}_{innerTrack}$, 1-prong

(b) f^{EM}_{track}, 3-prong

Figure 6.19.: Distributions of signal weights calculated from the loose, medium and tight RNN SRs with the distributions of (a) $\bar{d}_{innerTrack}$ for 1-prong and (b) f^{EM}_{track} for the 3-prong decay.

6.5. Recommendation for a $Z \to \tau\tau$ Re-Weighting

To conclude Chapter 6, a recommendation about the signal MC re-weighting based on the results presented throughout this thesis is given.

It can be summarised for all the investigated strategies, that both signal re-weightings for one and two dimensions yield convincing results in terms of an improvement of the data and MC agreement. For each τ decay channel, one 1-dimensional and one 2-dimensional weight set has been ranked the highest of all variable combinations that were determined:

- $\bar{d}_{innerTrack}$ and $\bar{d}_{innerTrack}$+$|S_{lead}|$ for 1-prong τ candidates and
- f_{track}^{EM} and f_{track}^{EM}+f_{cent} for 3-prong τ candidates.

For the 1-prong channel, a summary of the χ^2/NDF values with and without weights applied is given in Table 6.9. It is evident that a normalisation of the $\bar{d}_{innerTrack}$ distribution improves the overall data and MC agreement for the ID variables. $\bar{d}_{innerTrack}$ has a relatively high correlation to other relatively badly modelled variables such as ΔR_{max}, which is reflected in the differences of the respective reduced χ^2 before and after the weightings. When considering $|S_{lead}|$ in addition for the 2-dimensional re-weighting, all of the χ^2/NDF values reduce slightly or stay the same compared to the 1-dimensional re-weighting. The two variable weight set negates the fact that the $\bar{d}_{innerTrack}$ weights increase the χ^2/NDF value of $|S_{lead}|$, the worst modelled variable in the 1-prong channel from the start. The ID variable $f_{EM}^{track-HAD}$, which is also a highly ranked variable in χ^2/NDF, also improves through application of 2-dimensional weights.

Therefore, a recommendation of a signal MC re-weighting in the 1-prong channel is to use a 2-dimensional weight set calculated from the ID variables $\bar{d}_{innerTrack}$ and $|S_{lead}|$.

	no weight	$\bar{d}_{innerTrack}$	$\bar{d}_{innerTrack}$+$\lvert S_{lead}\rvert$
p_T	1.2	1.3	1.2
ΔR_{max}	3.6	1.5	1.5
f_{cent}	1.5	1.7	1.4
$\bar{d}_{innerTrack}$	3.7	0.0	0.0
$f_{EM}^{track-HAD}$	7.5	6.8	6.3
f_{track}^{EM}	3.9	2.5	2.5
$\frac{E_T}{p_T^{lead}}$	0.6	0.9	0.8
$\lvert S_{lead}\rvert$	16.0	16.4	0.0
p_T^{ratio}	1.4	1.4	1.2

Table 6.9.: Summary of the χ^2/NDF values of the ID variable distributions for the 1-prong channel without weights and with the best 1-dimensional and 2-dimensional signal re-weightings applied.

The results of the χ^2/NDF calculations for the 3-prong decay channel without and with either the f^{EM}_{track} or the f^{EM}_{track}+f_{cent} weight sets applied are presented in Table 6.10. Comparable to the 1-prong channel, the re-weightings for both dimensions yield significant improvements of the ID variable modelling compared to the distributions without any signal weights. This is evident in the χ^2/NDF values of e.g. f_{cent} or $f^{track-HAD}_{EM}$ before and after applying weights.

A downside to the 2-dimensional re-weighting calculated from f^{EM}_{track}+f_{cent} is that the 1-dimensional re-weighting has a higher improvement of the reduced χ^2 from the $f^{track-HAD}_{EM}$ variable. The f^{track}_{EM}+f_{cent} re-weighting does naturally improve the data and MC agreement of the variable with the second highest χ^2/NDF value.

The reduction of the χ^2/NDF of 7.6 for the distribution without weights to 5.9 with f^{EM}_{track} weights could presumably be improved even further by readjusting the binning of f_{cent}. Thus, for the 3-prong channel a 1-dimensional re-weighting based on f^{EM}_{track} is recommended.

	no weights	f^{EM}_{track}	f^{EM}_{track}+f_{cent}
p_T	1.2	1.3	1.3
ΔR_{max}	0.5	0.5	0.4
f_{cent}	7.6	5.9	0.0
$\bar{d}_{innerTrack}$	2.0	1.5	1.7
m_{track}	2.7	2.0	1.9
$f^{track-HAD}_{EM}$	4.9	2.1	3.3
f^{EM}_{track}	8.7	0.0	0.6
$\frac{E_T}{p^{lead}_T}$	1.8	1.3	0.9
S^{flight}_T	1.2	1.4	1.2
p^{ratio}_T	1.2	1.4	1.7

Table 6.10.: Summary of the χ^2/NDF values of the ID variable distributions for the 3-prong channel without weights and with the best 1-dimensional and 2-dimensional signal re-weightings applied.

7. Conclusion and Outlook

Throughout this thesis, the modelling of the τ lepton HLT has been investigated using a $Z \rightarrow \tau_{lep}\tau_{had}$ tag and probe analysis. Datasets from the LHC Run-2 collisions recorded in the years 2017 and 2018 have been utilised, corresponding to 44.3 fb^{-1} and 58.5 fb^{-1}, respectively.

The properties of the $Z \rightarrow \tau_{lep}\tau_{had}$ SR and the modelling of the HLT ID variables have been presented in Chapter 4 and Chapter 5. The main goal of this analysis was to use the knowledge about the HLT modelling to search for a strategy to improve the modelling and performance for future RNN identification trainings of τ lepton triggers in Run-3.
The suggested strategy is to apply a data-driven re-weighting on the signal MC sample as presented in Chapter 6. For that, the HLT ID variables have been ranked according to the modelling of their distributions in both the 1-prong and 3-prong channels. The worst modelled variables have been chosen to base the weight calculations on. Additionally, both 1-dimensional and 2-dimensional re-weighting strategies have been investigated.
In general, 1-dimensional and 2-dimensional re-weightings are both viable strategies to improve the data and MC agreement with little disadvantages or computing cost if the correct variables are chosen to calculate the signal weight from. As summarised in Section 6.5, the re-weightings that have proven to perform the best are:

- the 2-dimensional re-weighting strategy based on the ID variables $\bar{d}_{innerTrack}$ and $|S_{lead}|$ for the 1-prong channel and
- the 1-dimensional re-weighting strategy based on the ID variable f^{EM}_{track} for the 3-prong channel.

Therefore, these signal MC re-weightings are recommended to use when training an τ lepton ID algorithm for future Run-3 tunings.

Beyond the scope of this analysis, there is still room to improve the modelling and therefore the efficiency of future HLT RNNs for particle identification. It is still possible to fine tune the selections of the variables for the re-weightings, especially the 2-variable re-weightings by investigating the efficiency improvements of the RNN training. There are also possible weight sets from other ID variables or different combinations for 2-dimensional re-weightings available that have not been considered within this thesis.

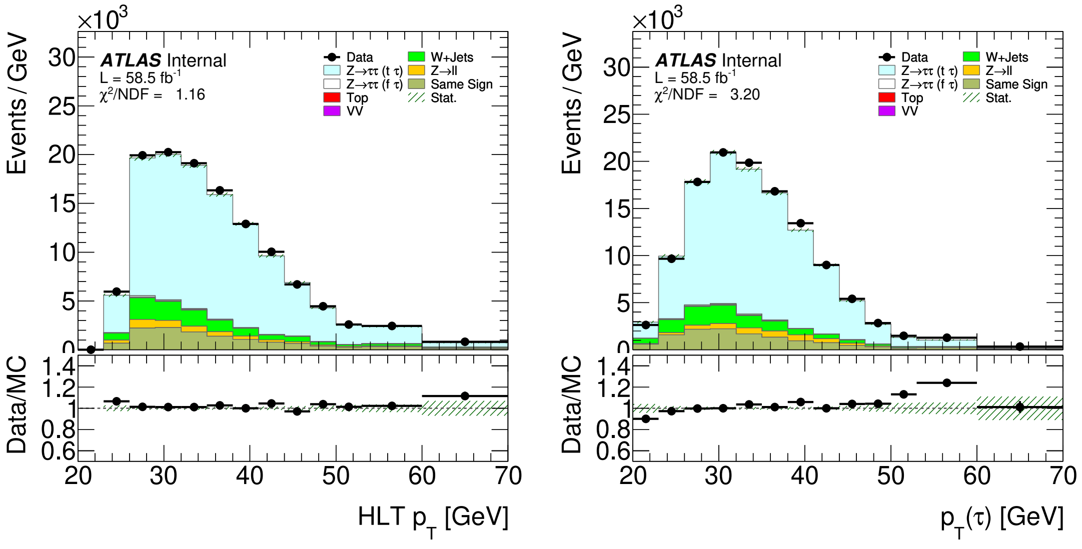
There are also methods that might have a slight positive effect, e.g. readjustments of the bins of the variable distributions.

The results data-driven signal MC re-weighting indicate high potential to improve the variable modelling for regions with high signal purity to maximise the information output gained from these signatures for HLT ID algorithm trainings.

Another way to investigate a region with a high signal purity besides the $Z \rightarrow \tau_{lep}\tau_{had}$ channel is to employ a $t\bar{t}$ tag and probe analysis (see Appendix A). This analysis supplements the $Z \rightarrow \tau_{lep}\tau_{had}$ tag and probe analysis with events in the higher $\tau_{had,vis}$ p_T region. The $t\bar{t}$ tag and probe analysis could also profit from applications of signal sample weights, similarly to the re-weightings presented throughout this thesis.

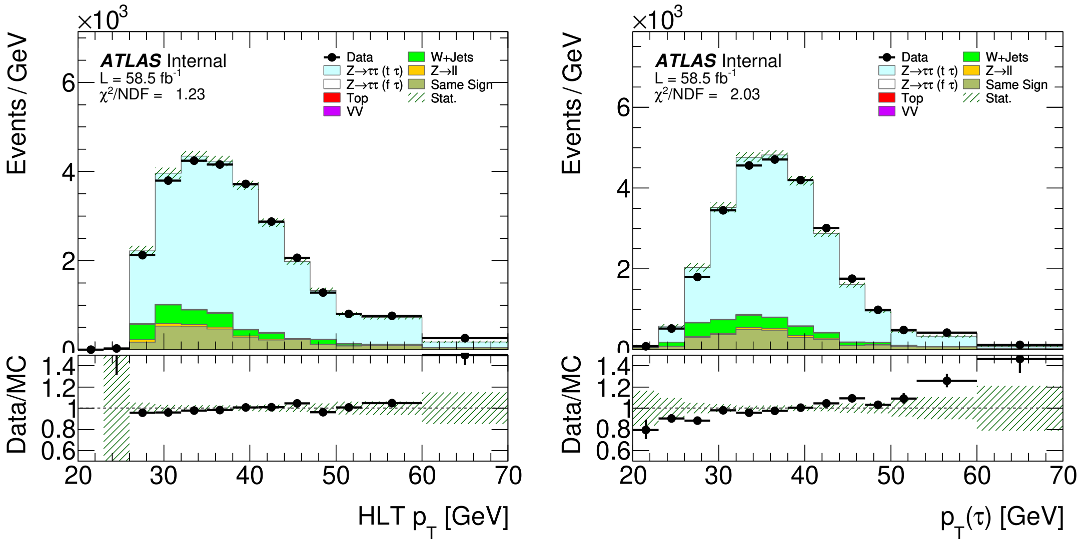
The results presented in this thesis provide a good basis for tunings and improvements regarding ID methods of trigger system with the $Z \rightarrow \tau\tau$ SR. Efficient particle identification and a high signal purity of RNN trainings allow the triggers to handle the requirements of future data-taking periods in a more precise manner. Especially for Higgs boson physics, a high precision and resolution of the detector is needed. In addition, a precise machinery is key to investigate other physics processes, such as processes including physics beyond the SM.

A. Additional Plots



(a) HLT, 1p

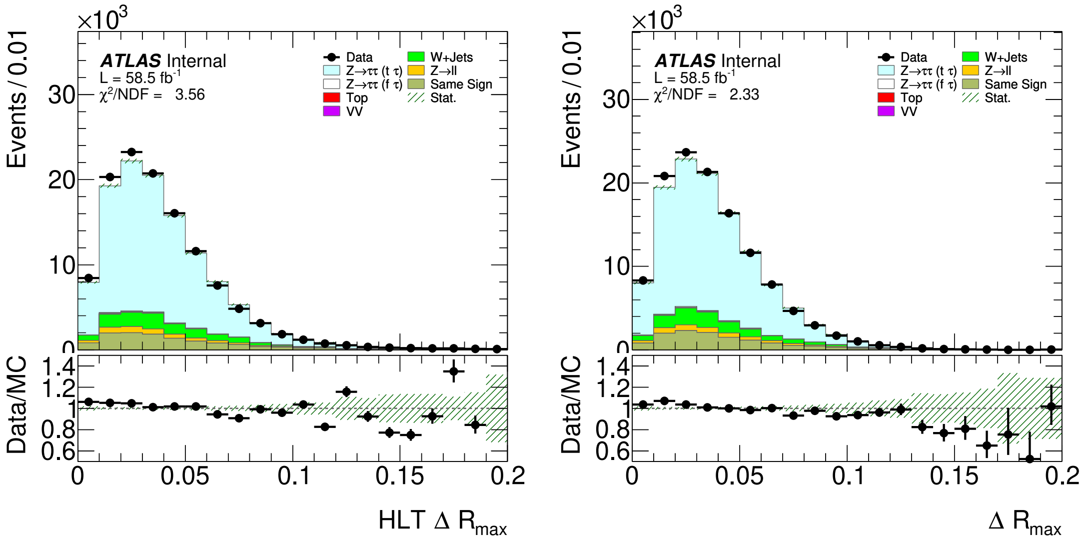
(b) Offline, 1p



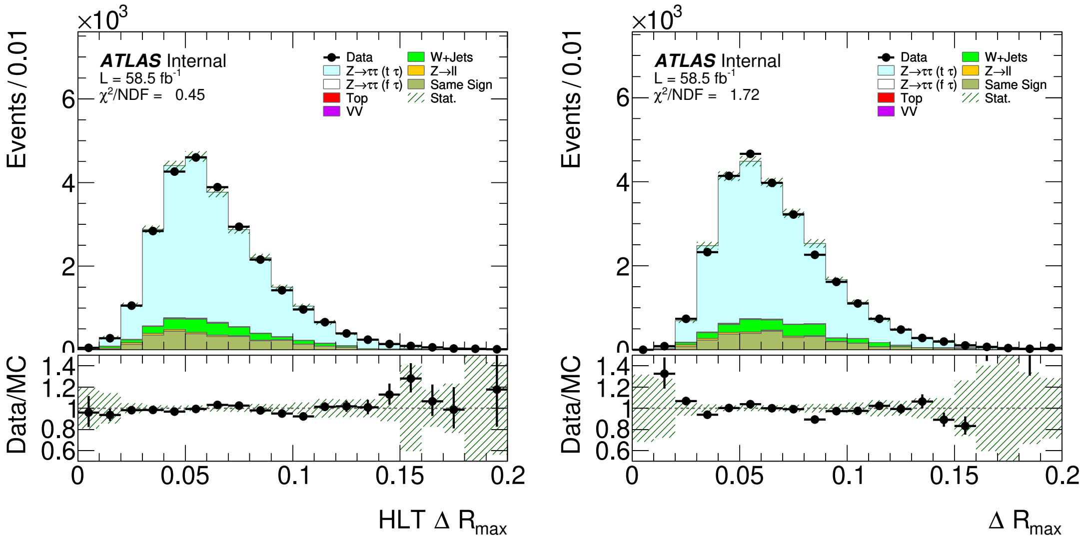
(c) HLT, 3p

(d) Offline, 3p

Figure A.1.: Different distributions of p_T for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

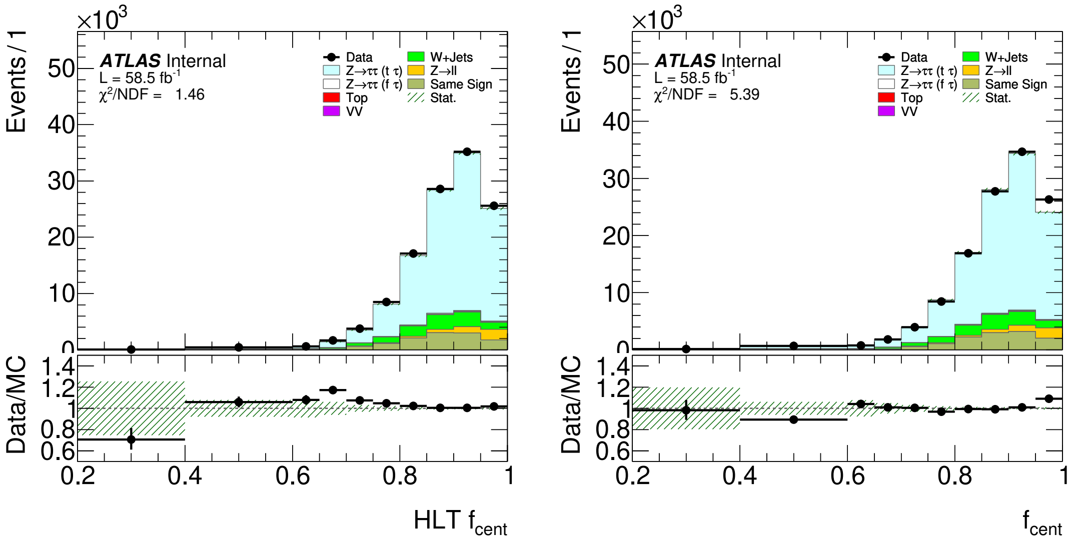


(a) HLT, 1p (b) Offline, 1p

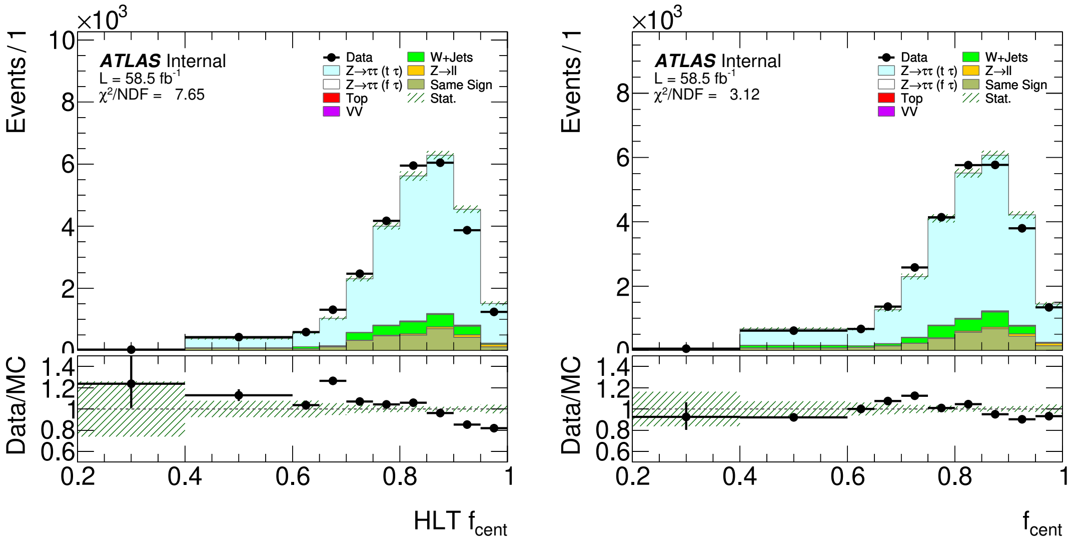


(c) HLT, 3p (d) Offline, 3p

Figure A.2.: Different distributions of ΔR_{max} for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

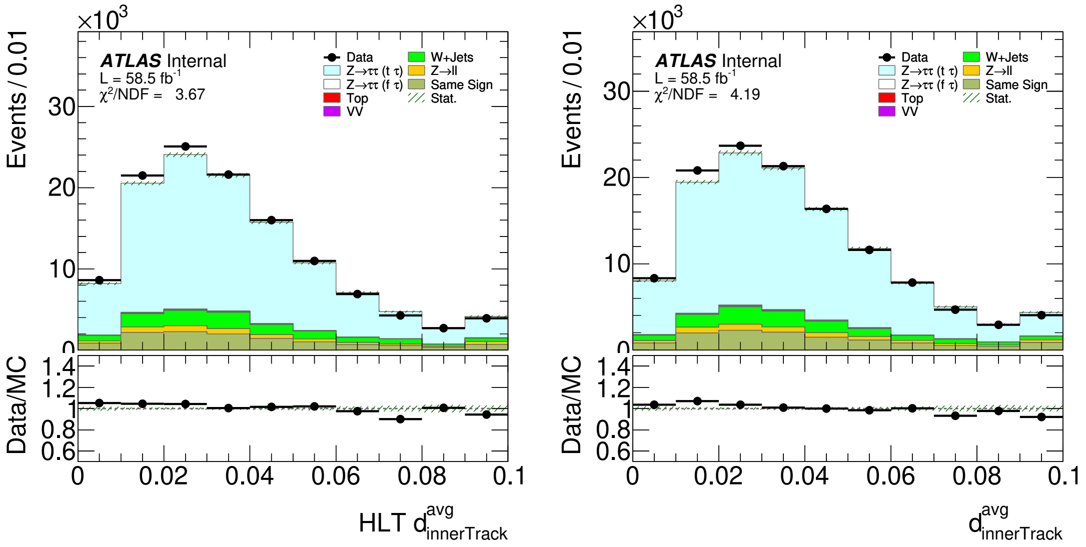


(a) HLT, 1p (b) Offline, 1p

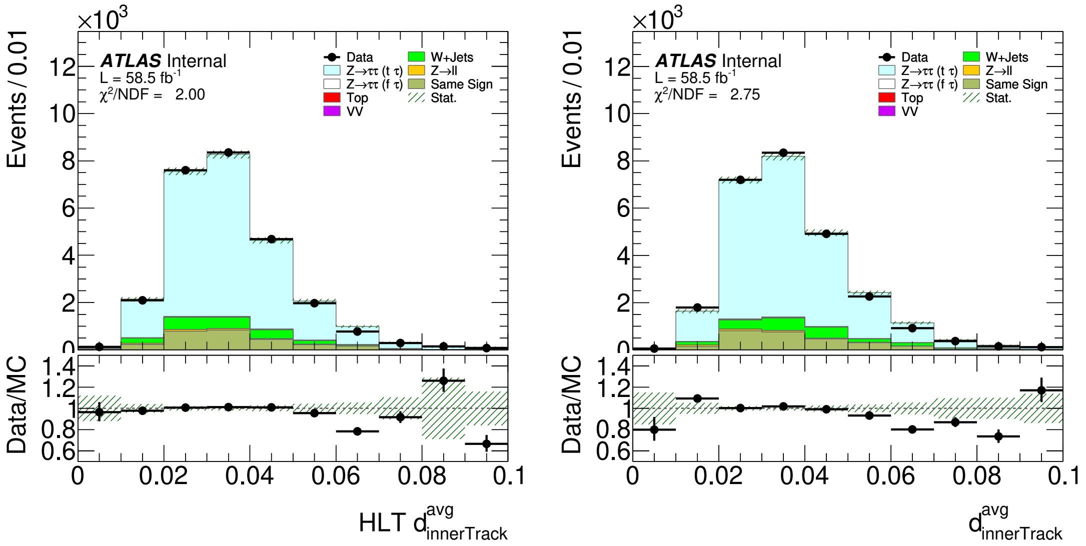


(c) HLT, 3p (d) Offline, 3p

Figure A.3.: Different distributions of f_{cent} for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

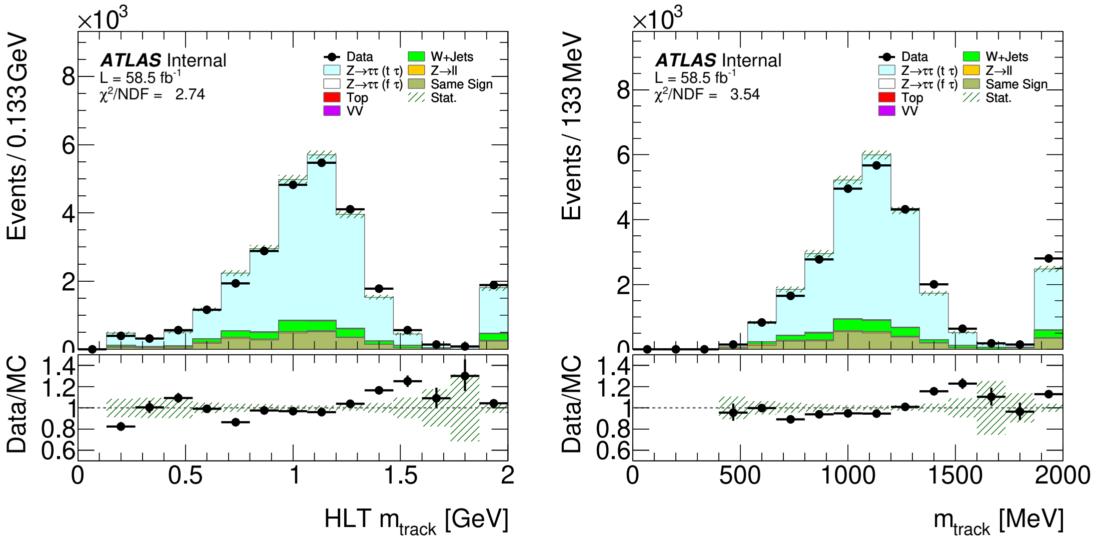


(a) HLT, 1p (b) Offline, 1p



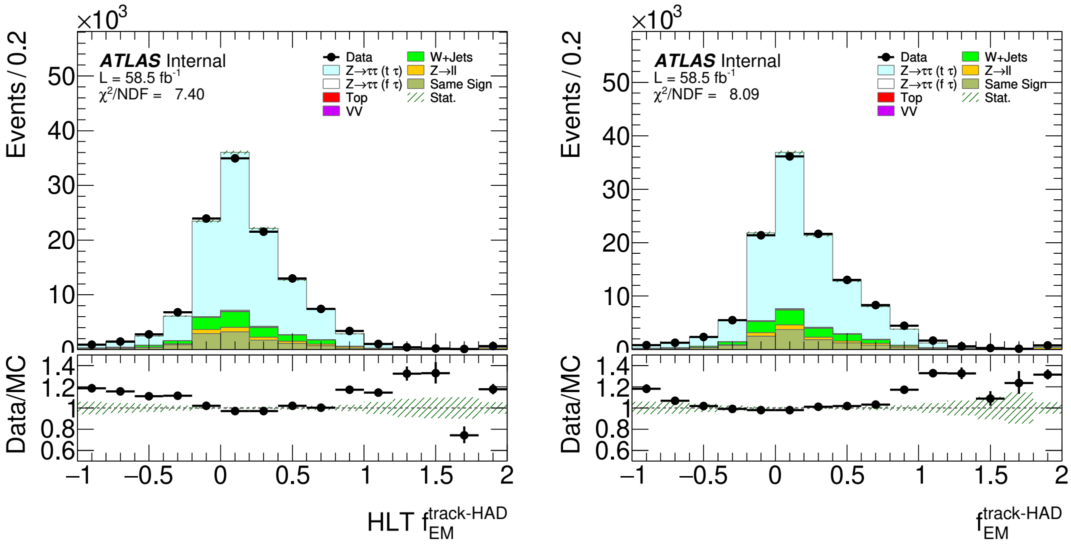
(c) HLT, 3p (d) Offline, 3p

Figure A.4.: Different distributions of $\bar{d}_{innerTrack}T$ for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

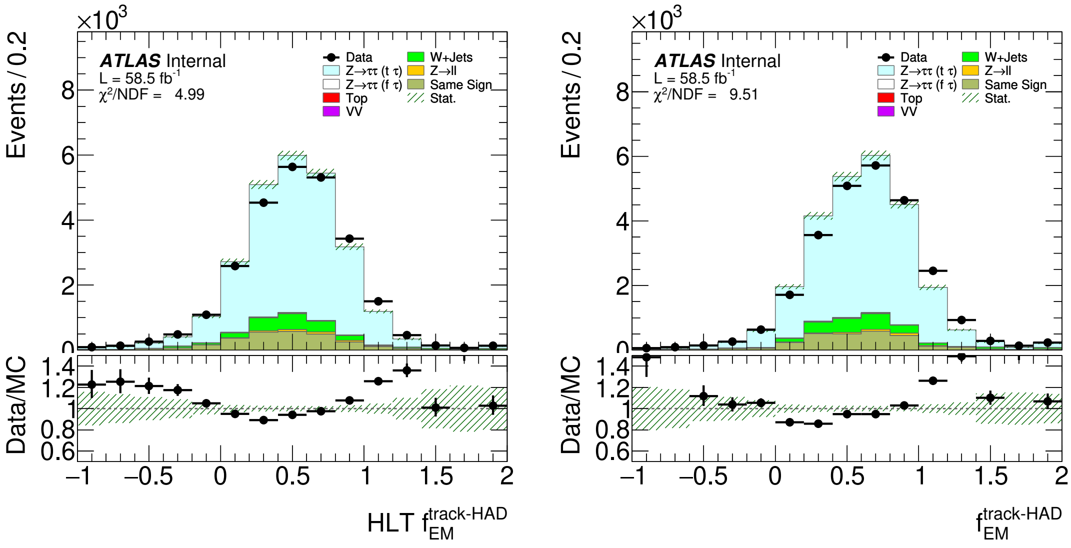


(a) HLT, 3p (b) Offline, 3p

Figure A.5.: Different distributions of m_{track} for HLT and offline reconstruction in the 3-prong channel. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

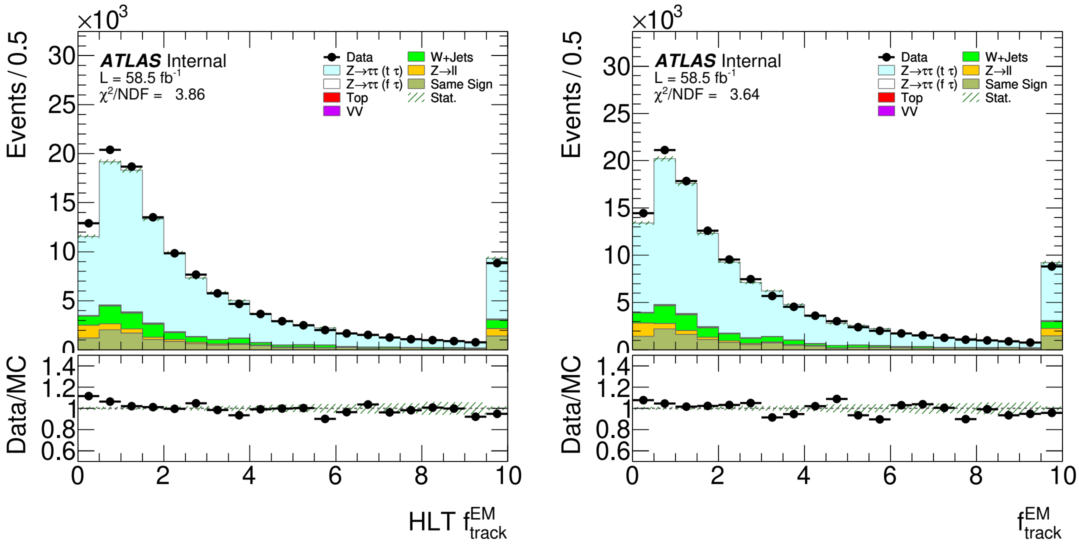


(a) HLT, 1p (b) Offline, 1p

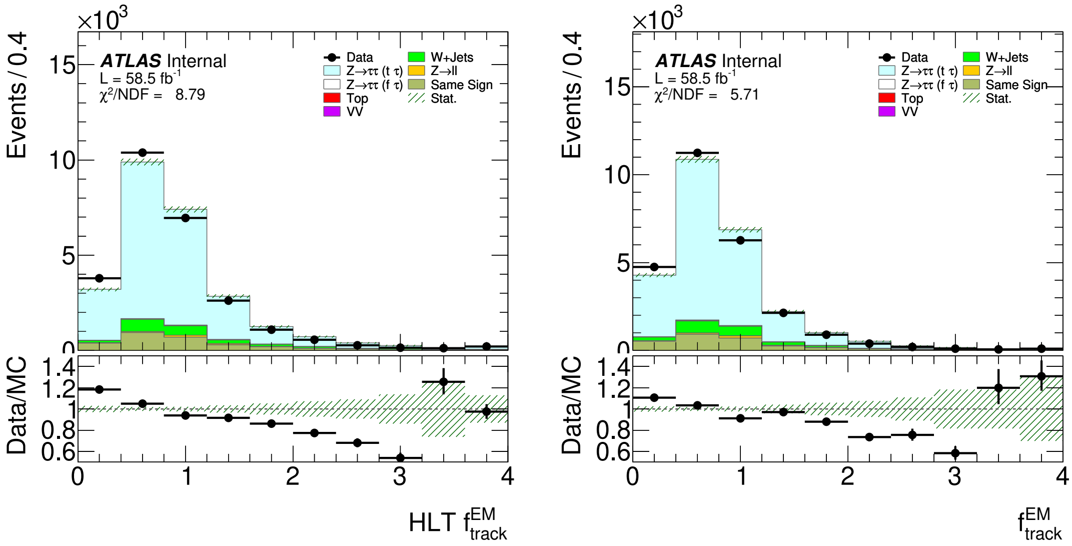


(c) HLT, 3p (d) Offline, 3p

Figure A.6.: Different distributions of $f_{EM}^{track-HAD}$ for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

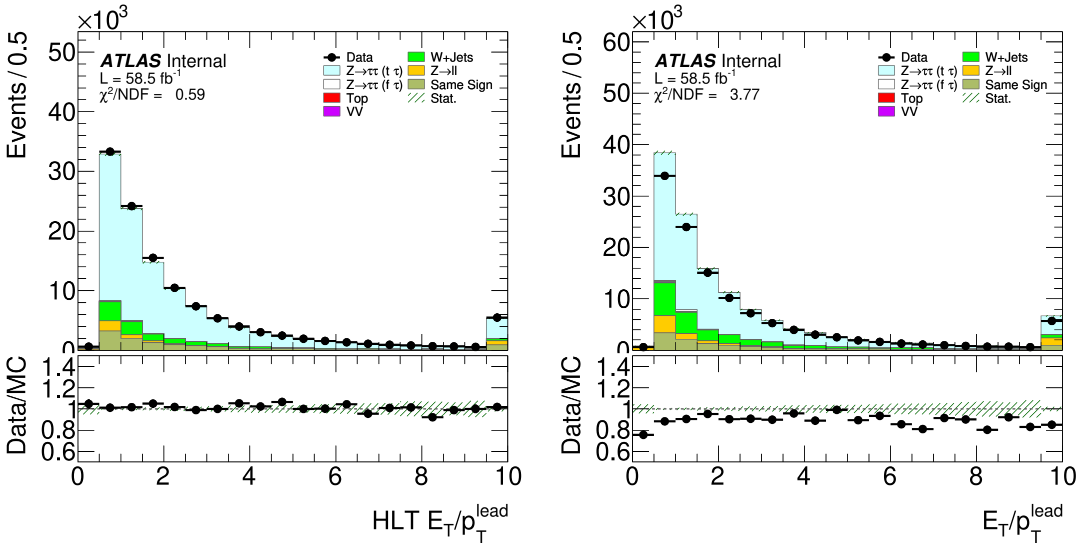


(a) HLT, 1p (b) Offline, 1p



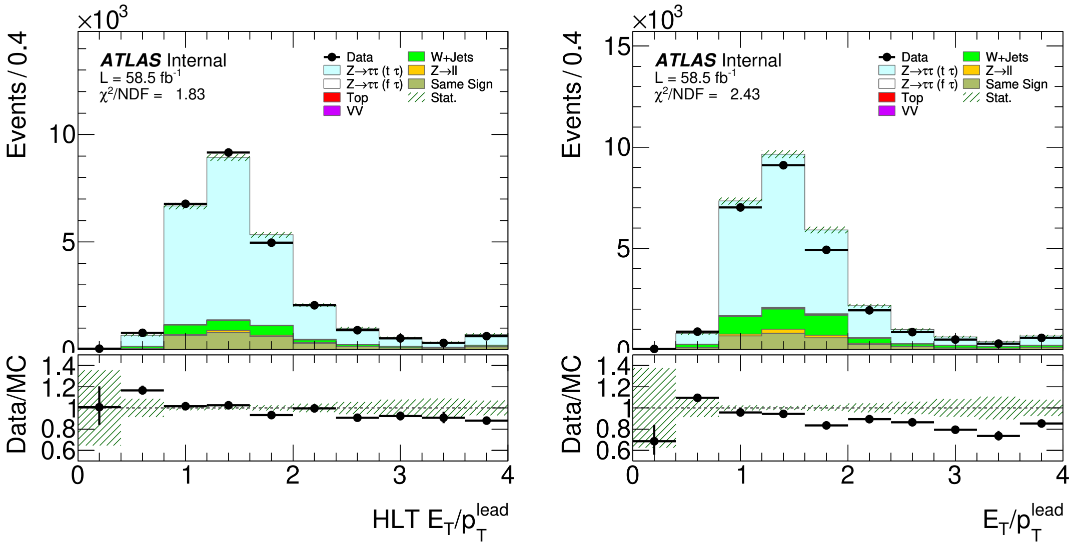
(c) HLT, 3p (d) Offline, 3p

Figure A.7.: Different distributions of f_{track}^{EM} for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.



(a) HLT, 1p

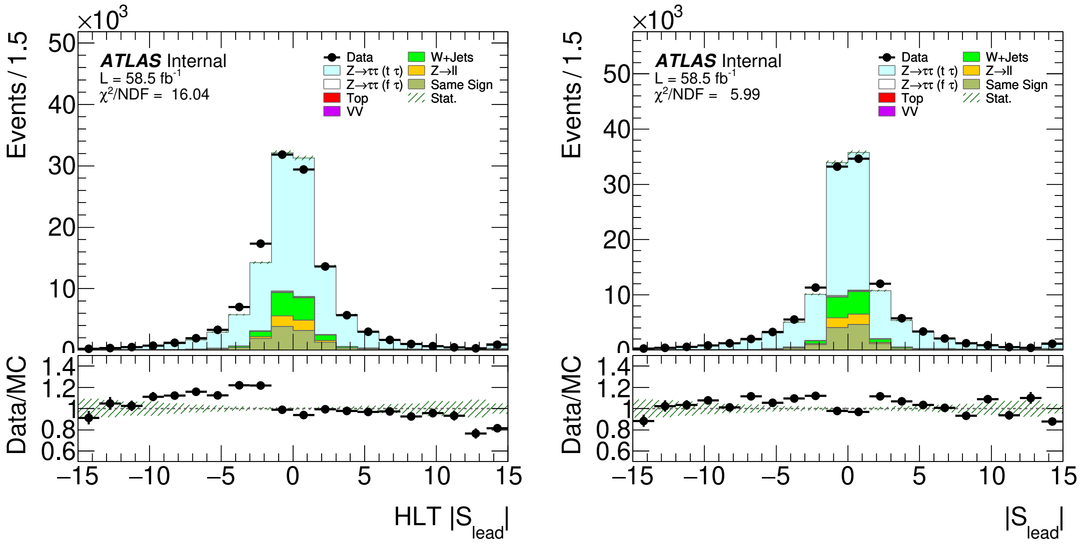
(b) Offline, 1p



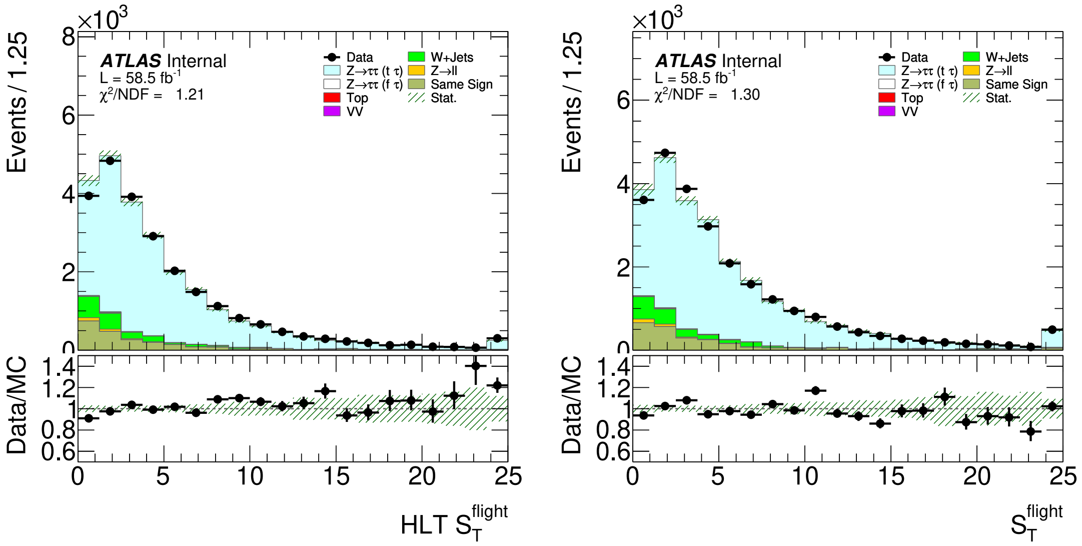
(c) HLT, 3p

(d) Offline, 3p

Figure A.8.: Different distributions of $\frac{E_T}{p_T^{lead}}$ for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

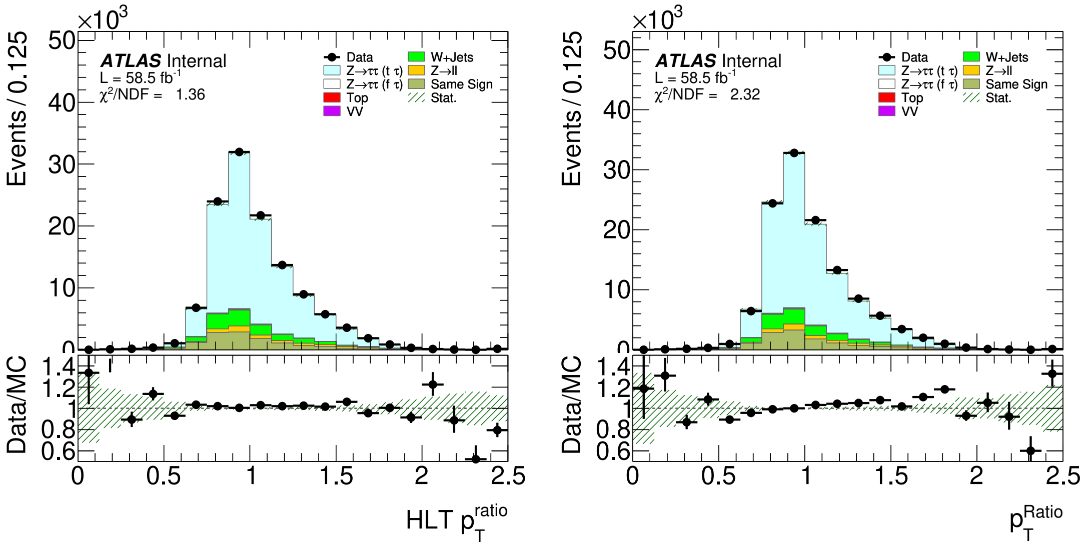


(a) HLT, 1p (b) Offline, 1p

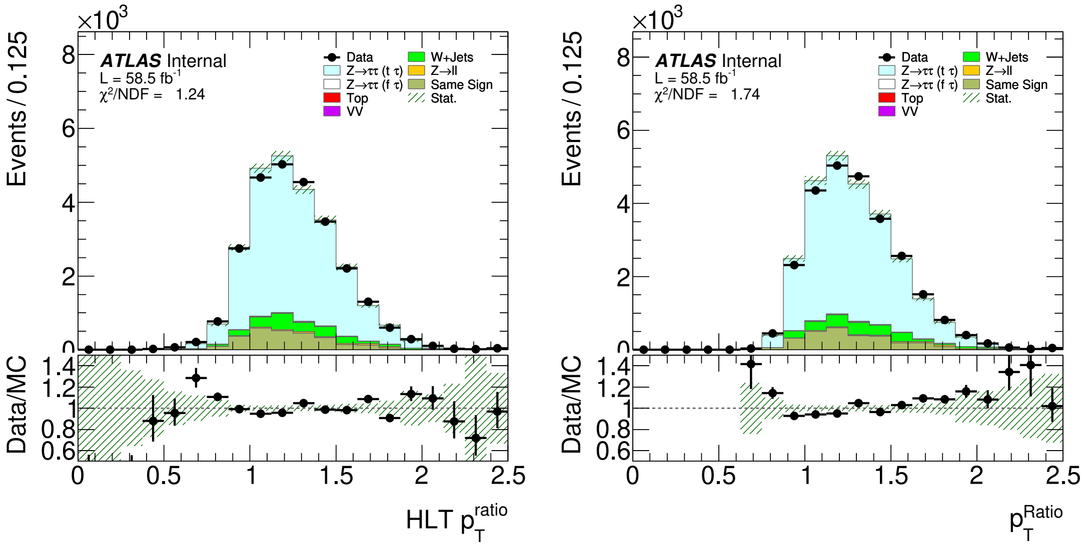


(c) HLT, 3p (d) Offline, 3p

Figure A.9.: Different distributions of $|S_{lead}|$ in the 1-prong channel for (a) HLT and (b) offline and S_T^{flight} in the 3-prong channel for (c) HLT and (d) offline reconstruction. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.



(a) HLT, 1p (b) Offline, 1p



(c) HLT, 3p (d) Offline, 3p

Figure A.10.: Different distributions of p_T^{ratio} for HLT and offline reconstruction in the 1-prong and 3-prong channels. These distributions correspond to the 58.5 fb^{-1} dataset from 2018.

B. $t\bar{t}$ Tag and Probe Analysis

An important orthogonal analysis to the $Z \rightarrow \tau\tau$ tag and probe analysis is a tag and probe method in the $t\bar{t}$ signal region. The $t\bar{t}$ SR is used to expand the $Z \rightarrow \tau\tau$ tag and probe analysis in the higher p_T regions of the $\tau_{had,vis}$ candidate between 70 GeV and 350 GeV. A short summary of the event selection and the data and MC agreement will be presented throughout this Chapter.

B.1. Event Selection

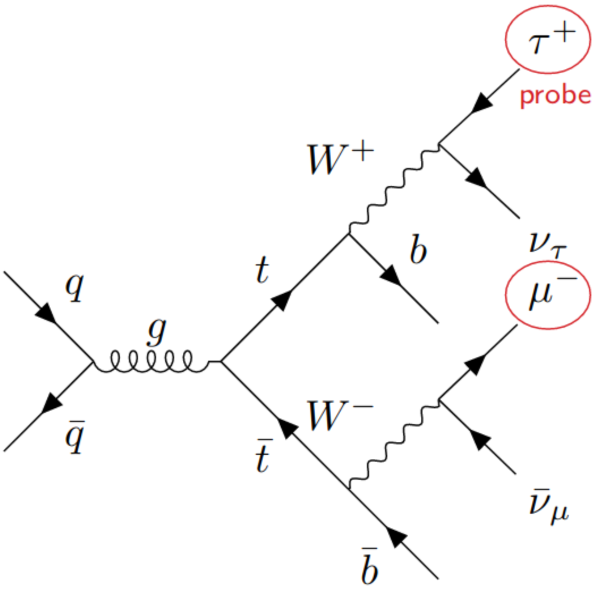


Figure B.1.: Feynman diagram of the $t\bar{t} \rightarrow [b\mu\nu][b\tau_{had-vis}2\nu]$ signature that is exploited in the $t\bar{t}$ tag and probe analysis.

Figure B.1 shows the Feynman diagram of the $t\bar{t} \rightarrow [b\mu\nu][b\tau_{had-vis}2\nu]$ signature. The muon, that results as a decay product from the $\bar{t}$-quark is tagged, whereas a the rest of the signature is searching for two jets with at least one jet stemming from a b-quark to identify the b-quark and tag the $\tau_{had,vis}$ candidate with a charge opposite to the muon charge that are both decay products from the other top quark. This method works very similar to the tag and probe analysis discussed in Chapter 4.

The main backgrounds for the $t\bar{t}$ SR are QCD processes that get misidentified as $\tau_{had,vis}$ candidates. Therefore, an OS-SS estimation is applied to this signal region. This estimation is analogous to the explanations in Section 4.3. For $t\bar{t}$, only a calculation of the r_{QCD} normalisation factor is necessary, as there is just a QCD QR.

The requirements for the $t\bar{t}$ SR selection compared to the $Z \to \tau\tau$ SR are summarised in Table B.1.

Variable	$Z \to \tau\tau$ SR	$t\bar{t}$ SR	QCD CR
Muon isolation	yes	yes	inverted
Opposite sign	-1	-1	-1
$m_T(\mu, E_T^{miss})$ [GeV]	< 50	-	-
$\sum \cos(\Delta\phi)$	> -0.15	-	-
$m_{vis}(\mu, \tau_{had-vis})$ [GeV]	[45,90]	-	-
b-tagged jets	veto	> 0	> 0

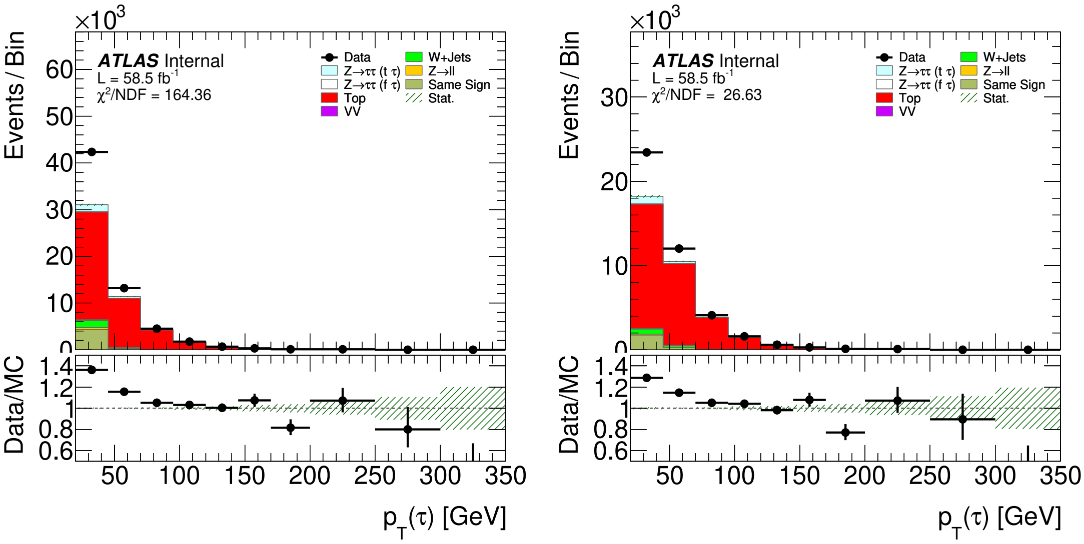
Table B.1.: Summary of cut requirements for the $Z \to \tau\tau$ and $t\bar{t}$ signal regions and the QCD CR corresponding the the $t\bar{t}$ SR. The charge of the tagged muon and the probed τ lepton as well as the isolation of the muon are considered.

B.2. Data/Monte Carlo Agreement

For the $t\bar{t}$ SR, several different trigger cuts are available. These cuts mimic a selection chain from the HLT online selection for offline analyses. They require the $\tau_{had,vis}$ candidate to have at least e.g. $p_T > 25$ GeV and decide the ID process, in this case a medium BDT cut and the track algorithm. A few examples of three possible trigger cuts are displayed in Figure B.2 with "tau25_medium1_tracktwo" in B.2(b), "tau60_medium1_tracktwo" in B.2(c) and "tau160_medium1_tracktwo" in B.2(d). These distributions highlight that there are a lot of events available in lower p_T regions. For example Figure B.2(a), the $p_T < 100$ GeV have a total number of events of $\mathcal{O}(10^4)$, while the higher p_T bins only contain a few hundred events, which can be seen especially in B.2(d).

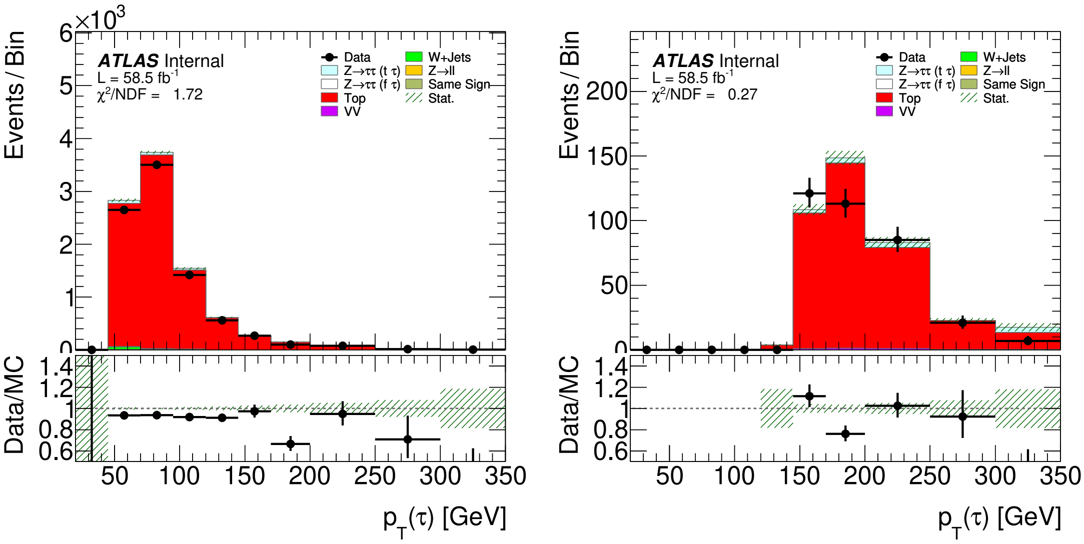
Regarding the data and MC agreement of this SR, it can be seen that the prediction underestimates the data measurement in the low p_T bins. Although the number of events available in the bins in the interval $[100, 350]$ GeV is relatively low, the modelling of the p_T distributions Figure B.2(c) and (c) is quite accurate.

Especially for the less accurately modelled regions of the $t\bar{t}$ tag and probe analysis, there is potential to improve the data and MC agreement and a signal MC re-weighting similar to the one presented in Chapter 6 would be a good method to do so.



(a) no trigger cut

(b) tau25_medium1_tracktwo



(c) tau60_medium1_tracktwo

(d) tau160_medium1_tracktwo

Figure B.2.: Distributions of the τ lepton p_T for a high p_T range up to 350 GeV (a) without a trigger cut, (b) a *tau*25_*medium*1_*tracktwo* trigger cut, (c) a *tau*60_*medium*1_*tracktwo* trigger cut and (d) a *tau*160_*medium*1_*tracktwo* trigger cut applied.

Bibliography

[1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22(4)**, 579 (1961)

[2] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C **680519**, 367 (1968)

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

[4] G. Aad, et al. (The ATLAS Collaboration), *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, S08003 (2008)

[5] L. Evans, P. Bryant, *LHC Machine*, JINST (2008), 10.1088/1748-0221/3/08/S08001

[6] A. Ruiz-Martinez, et al. (The ATLAS Collaboration), *The Run-2 ATLAS Trigger System,* ATL-DAQ-PROC-2016-003, `https://cds.cern.ch/record/2133909`

[7] R. E. Owen (The ATLAS Collaboration), *The ATLAS Trigger System* (2018), `https://cds.cern.ch/record/2302730`

[8] M. Thomson, *Modern Particle Physics*, Cambridge University Press (2013)

[9] C. Patrignani, et al. (Particle Data Group), *Review of Particle Physics*, Chin. Phys. C **40(10)**, 100001 (2016)

[10] G. Altarelli, *A QCD primer*, AIP Conf. Proc. **631(1)**, 70 (2002), `hep-ph/0204179`

[11] G. Arnison, et al. (UA1 Collaboration), *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV*, Phys. Lett. **B122**, 103 (1983)

[12] G. Arnison, et al. (UA1 Collaboration), *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. **B126**, 398 (1983)

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

[14] W. de Boer (The CMS Collaboration), *The Discovery of the Higgs Boson with the CMS Detector and its Implications for Supersymmetry and Cosmology*, Technical report, CERN, Geneva (2013), `https://cds.cern.ch/record/1597055`

[15] C. Arroyo, et al., *Precise measurement of the weak mixing angle in neutrino-nucleon scattering.*, Phys. Rev. Lett. **72**, 3452 (1994)

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

[17] S. Chatrchyan, et al. (The CMS Collaboration), *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716(1)**, 30 (2012)

[18] D. de Florian, et al. (LHC Higgs Cross Section Working Group), *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs (2016), `https://doi.org/10.23731/CYRM-2017-002`

[19] M. Tanabashi, et al. (Particle Data Group), *Status of Higgs Boson Physics,* `Phys. Rev.D98,030001(2018)`

[20] G. L. Fogli, et al., *Observables sensitive to absolute neutrino masses: Constraints and correlations from world neutrino data*, Phys. Rev. D **70**, 113003 (2004)

[21] M. Aaboud, et al. (The ATLAS Collaboration), *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. Lett. **121(19)**, 191801 (2018)

[22] A. Sirunyan, et al., *Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV*, JHEP **12(2019)**, 61 (2019)

[23] S. Chatrchyan, et al. (The CMS Collaboration), *The CMS experiment at the CERN LHC*, JINST **3(08)**, S08004 (2008)

[24] A. Augusto Alves Jr, et al. (The LHCb Collaboration), *The LHCb Detector at the LHC*, JINST **3(08)**, S08005 (2008)

[25] K. Aamodt, et al. (The ALICE Collaboration), *The ALICE experiment at the CERN LHC*, JINST **3(08)**, S08002 (2008)

[26] G. Aad, et al. (The ATLAS Collaboration), *Studies of the performance of the ATLAS detector using cosmic-ray muons*, Eur. Phys. J. C **C71**, 1593 (2011), `1011.6665`

[27] A. Tumasyan, et al. (The ATLAS Collaboration), *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment,* ATL-PHYS-PUB-2019-033, `https://cds.cern.ch/record/2688062`

[28] *The ATLAS Tau Trigger in Run 2,* ATLAS-CONF-2017-061, `https://cds.cern.ch/record/2274201`

[29] G. Aad, et al. (The ATLAS Collaboration), *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s} = 8$ *TeV*, Eur. Phys. J. C **75(7)** (2015)

[30] M. Aaboud, et al. (The ATLAS Collaboration), *Performance of the ATLAS trigger system in 2015*, Eur. Phys. J. C **77(5)** (2017)

[31] *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC,* ATL-PHYS-PUB-2015-045, `https://cds.cern.ch/record/2064383`

[32] M. Huebner (The ATLAS Collaboration), *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ TeV*,* ATLAS-CONF-2017-029, `https://cds.cern.ch/record/2261772`,

[33] G. Aad, et al. (The ATLAS collaboration), *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, Eur Phys. J. C **74(11)** (2014)

[34] G. Aad, et al. (The ATLAS Collaboration), *Search for new resonances in mass distributions of jet pairs using 139 fb*$^{-1}$ *of pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **2003** (2020), `1910.08447`, URL `https://cds.cern.ch/record/2694189`

[35] P. Nason, *A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms*, JHEP **2004(11)** (2004)

[36] S. Frixione, P. Nason, C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **2007(11)** (2007)

[37] S. Alioli, P. Nason, C. Oleari, E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **2010(6)** (2010)

[38] S. Alioli, P. Nason, C. Oleari, E. Re, *NLO vector-boson production matched with shower in POWHEG*, JHEP **2008(07)**, 060 (2008)

[39] S. Alioli, P. Nason, C. Oleari, E. Re, *Vector boson plus one jet production in POWHEG*, JHEP **2011(1)**, 095 (2011)

[40] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, C.-P. Yuan, *New parton distributions for collider physics*, Phys. Rev. D **82(7)** (2010)

[41] T. Sjoestrand, S. Mrenna, P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178(11)**, 852 (2008)

[42] G. Aad, et al. (The ATLAS Collaboration), *Measurement of the Z/γ boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **2014(9)**, 145 (2014)

[43] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, C.-P. Yuan, *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78(1)** (2008)

[44] P. Golonka, Z. Was, *PHOTOS Monte Carlo: a precision tool for QED corrections in Z and W decays*, The Eur. Phys. J. C **45(1)**, 97 (2006)

[45] N. Davidson, T. Przedzinski, Z. Was, *PHOTOS Interface in C++; Technical and Physics Documentation* (2010), arXiv:1011.0937

[46] J. M. Campbell, R. K. Ellis, P. Nason, E. Re, *Top-pair production and decay at NLO matched with parton showers*, JHEP **2015(4)**, 114 (2015)

[47] T. Sjoestrand, S. Mrenna, P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **2006(05)**, 026 (2006)

[48] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D **82(7)** (2010)

[49] P. Artoisenet, et al., *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, JHEP **2013(3)**, 015 (2013)

Acknowledgements

Firstly, I would like to thank Prof. Dr. Stan Lai. He has been a key person throughout my whole academic career and shaped me into the particle physicist that I am today by supporting me through several lectures, a bachelor thesis, an internship and finally the past year and this thesis. While always helping me to improve myself as a scientist, he still accepted that in my world, "today" means "some time in between 2am and 6am the next day". I am sad that I will have to move on from working in his group and will cherish all the knowledge I have gained from him in the past few years.

Furthermore, my thanks goes out to the whole AG Lai and the particle physics institute in Göttingen. Above all, Dr. Kira Abeling and Andrés Melo have supported me throughout this year with help, interesting discussions and important impulses to help me with my research.

Special thanks to Dr. Antonio di Maria, who provided me with the technical infrastructure that I needed to execute my research.

I would also like to thank all the people I had the pleasure to have discussions about my work with, for example the tau trigger group.

Believe in yourself; it doesn't hurt to try.

Liebe Mama, lieber Rainer,

Ich bin so froh über die ganzen Zeiten in meinem Studium, die ihr an mich geglaubt habt, wenn ich es nicht getan habe. Ohne euch, wäre es wahrscheinlich nur bei einer "Schnapsidee" geblieben. Danke, dass ihr immer für mich da seid.

Für Max. Danke, dass du dieses Jahr mein Rücken warst.

Erklärung nach §17(9) der Prüfungsordnung für den Bachelor-Studiengang Physik und den Master-Studiengang Physik an der Universität Göttingen: Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht habe.

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

Göttingen, den 29. März 2023

(Vanessa Annabelle Grauer)