FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

PhD Thesis

Milena Bajic

A search for lepton-flavor-violating decays of the Z boson into a τ-lepton and a light lepton with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 13$ TeV at the LHC

Supervisor: Stefania Xella

June, 2018

Acknowledgements

I would like to thank to my supervisor Stefania Xella, for her time and continuous support through the whole project. Then, for specific parts of the project, I would like to thank to Geert-Jan Besjes for the initial help and the introduction to the ATLAS software framework, to Wing Sheung Chan from Nikhef for technical discussions about HistFitter and to Ossama AbouZeid and Jason Koskinen for giving me feedback on results.

Abstract

In the original Standard Model with the massless neutrinos, the lepton flavour number is strictly conserved per each generation. But, in experimentally discovered neutrino oscillations neutrinos can change the flavour, leading to the violation of the lepton flavour number. As a consequence, lepton flavour number violation is possible in other processes via loop diagrams, corresponding to tiny branching ratios at the order of 10^{-54}.

Lepton flavour number violation is predicted in several Beyond the Standard Model theories such are certain Supersymmetric Models, multi Higgs doublet models and others. As those theories predict lepton flavour violation at much larger rates which are accessible by experiments, an observation of lepton-flavour violating process would be a sign of new Beyond the Standard Model physics.

The analysis presented in this thesis describes a search for lepton flavour violation in the $Z \rightarrow \tau e$ and $Z \rightarrow \tau\mu$ processes with hadronic τ-lepton decays. The search is performed using the 2015+2016 proton-proton collision corresponding to $36.1\,\mathrm{fb}^{-1}$, recorded by the ATLAS detector at the centre-of-mass energy $\sqrt{s} = 13\,\mathrm{TeV}$.

The backgrounds are estimated using a data-driven fake-factor method for the processes where the τ-lepton is faked by a jet and using the Monte Carlo simulation for the backgrounds contributing with a real τ-lepton or a τ-lepton faked by another lepton. Dedicated selections are used to suppress the backgrounds and to define a signal-enriched region where the final fit is performed.

The final discriminating variable is constructed using the neural networks approach, where a neural network is trained to differentiate between the signal and the major background processes by "learning" from their kinematical differences. Finally, the neural network score is fitted in the signal region and the upper limit on the branching ratio is set.

Contents

1 Theory 1
- 1.1 The Standard Model 1
  - 1.1.1 Quantum Chromodynamics 2
  - 1.1.2 Electroweak Interaction 2
  - 1.1.3 Higgs Mechanism 3
  - 1.1.4 Yukawa Interactions 4
- 1.2 Lepton Flavour Violation 5
  - 1.2.1 Neutrino Oscilliations 6
  - 1.2.2 Supersymmetric Models 7
  - 1.2.3 Two-Higgs Doublet Models 8
  - 1.2.4 Prior Searches 8

2 The LHC and the ATLAS experiment 9
- 2.1 The Large Hadron Collider 9
- 2.2 The ATLAS Detector 10
  - 2.2.1 The Detector Components 10
  - 2.2.2 The Trigger System 14
  - 2.2.3 The Luminosity Measurements 15
  - 2.2.4 The Luminosity and the Pile-Up Profiles 15

3 Monte Carlo Simulation 17
- 3.1 Monte Carlo Simulation within the ATLAS software framework 17
- 3.2 LFV $Z \to \tau e/\mu$ Signal 20
- 3.3 Standard Model Backgrounds 21
  - 3.3.1 Z+jets 21
  - 3.3.2 Drell-Yan 23
  - 3.3.3 W+jets 23
  - 3.3.4 Top 24
  - 3.3.5 Diboson 25
  - 3.3.6 QCD 25
  - 3.3.7 Higgs 26

4 Event Selection 28
- 4.1 Event Cleaning 28
- 4.2 Reconstruction and Definition of Physics Objects 29
  - 4.2.1 Electrons 29
  - 4.2.2 Muons 30
  - 4.2.3 Taus 31

4.2.4 Jets . . . 32
4.2.5 Missing transverse energy . . . 32
4.3 Overlap Removal . . . 33

5 Discriminating Variables **35**
5.1 Kinematic discriminant $\Delta\alpha$ . . . 35
5.2 Invariant Mass Reconstruction . . . 37
5.3 Other kinematical variables . . . 39

6 Signal Region Optimization **45**
6.1 Neural Network Approach . . . 45
6.2 Neural Network Training . . . 47
6.3 Signal Region . . . 60

7 Estimation of Fake Backgrounds **65**
7.1 Fake Factor Method . . . 65
7.2 Control Regions . . . 66
7.3 Fake Factor Method Results . . . 72
7.4 Fake-Factor Validation . . . 76

8 Fit Model and Results **85**
8.1 Statistical Treatment . . . 85
8.2 Fit Model . . . 88
8.3 Systematical Uncertainties . . . 89
8.4 Asimov Fit . . . 92
8.5 Sideband Fit . . . 97
8.6 Unblinded Fit . . . 102
8.6.1 Testing the discovery hypothesis . . . 110
8.6.2 Setting the Upper Limit . . . 110

9 Conclusion **112**

A List of Monte Carlo Samples **114**
A.1 Signal Samples . . . 114
A.2 Background Samples . . . 115
A.2.1 Z+jets production . . . 115
A.2.2 W+jets . . . 116
A.2.3 Top . . . 119
A.2.4 Diboson . . . 119
A.2.5 Higgs . . . 120
A.2.6 Drell-Yan . . . 120

B Additional Discriminating Variables **121**

C Asimov Fit Distributions **125**

Theory

1.1 The Standard Model

The Standard Mode [**Herrero:1998eql**] is a mathematical framework describing the strong, weak and the electromagnetic interaction. The Standard Model was developed during the 1960s and 1970s and incorporates quantum chromodynamics, a theory of strong interaction and the electroweak theory, in which the electromagnetic and the weak interactions were unified in the Glashow-Salam-Weinberg (GWS) model. The last missing piece of the GWS model was the explanation of how do particles acquire mass. The solution came in the form of the Higgs mechanism, which predicted a new particle, the Higgs boson which was discovered in 2012, at CERN.

The fundamental particles are classified into the spin-$\frac{1}{2}$ fermions, the spin-1 gauge bosons and the spin-0 Higgs boson. The fermions are further classified into quarks and leptons, depending on whether they interact via strong interaction. The interactions are mediated by the gauge bosons: strong interaction via eight massless gluons and the electroweak interaction via one massless photon and three massive bosons, W^+, W^- and Z.

An overview of the Standard Model particles is shown in Figure 1.1.

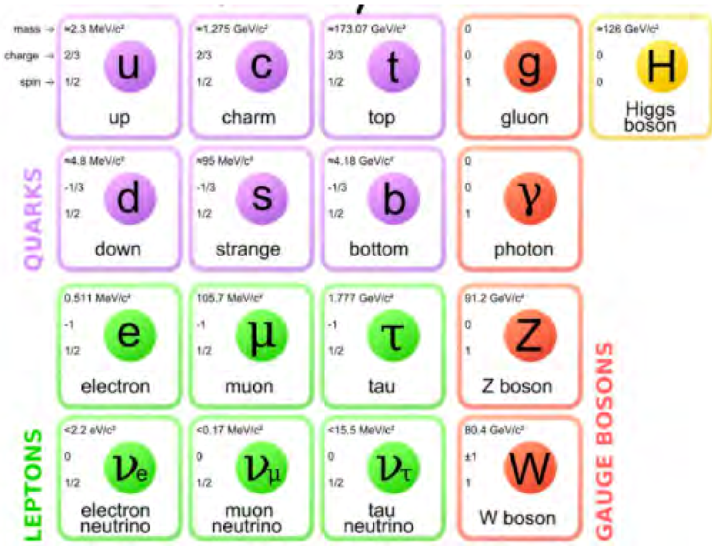


Figure 1.1: An overview of the Standard Model particles.

The interactions are mathematically represented by a Lagrangian whose terms describe the interactions between the particles. The possible ways for a process to occur can be formed from the combinations of the allowed vertices, which can schematically be represented by Feynman diagrams.

The amplitude of a process $\mathcal{M}$ can be computed by summing the Feynman diagrams, but since the amplitude for a process decreases with the number of vertices, the largest contribution will come from the diagrams containing the least number of vertices. The probability of a process can be computed as $|\mathcal{M}|^2$ and the cross section is proportional to the square of the amplitude $\sigma \backsim |\mathcal{M}|^2$.

The Lagrangian of the Standard Model can be split into multiple terms:

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{QCD}} + \mathcal{L}_{EW} + \mathcal{L}_{\mathrm{Higgs}} + \mathcal{L}_{\mathrm{Yukawa}} \tag{1.1}$$

where $\mathcal{L}_{\mathrm{QCD}}$ describes the strong interaction involving quarks and gluons, $\mathcal{L}_{EW}$ describes the electroweak interaction between fermions and the massless gauge bosons, the $\mathcal{L}^{\mathrm{Higgs}}$ describes the interactions between the Higgs boson and the gauge bosons, leading to the mass generation for the massive gauge bosons, while $\mathcal{L}_{\mathrm{Yukawa}}$ describes the Yukawa interactions between the Higgs boson and the fermions, leading to the mass generation for the fermions. The individual terms are described in more detail in the following sections.

1.1.1 Quantum Chromodynamics

Quantum chromodynamics is a mathematical framework describing the strong interaction, a fundamental interaction between the particles carrying the colour, namely the quarks and gluons, leading to the quark-gluon, but also the gluon-gluon self-interactions. The Lagrangian describing the strong interaction $\mathcal{L}_{\mathrm{QCD}}$ can be expressed as:

$$\mathcal{L}_{\mathrm{QCD}} = \sum_{fij} \bar{q}_{fi} i\gamma^{\mu} D_{\mu}^{ij} q_{fj} - \frac{1}{4} G^{a}_{\mu\nu} G_{a}^{\mu\nu} \tag{1.2}$$

where $\bar{q}_{fi} i\gamma^{\mu} D_{\mu}^{ij} q_{fj}$ describes the interactions while $G^{a}_{\mu\nu} G_{a}^{\mu\nu}$ describes the free field of gluons. The flavour is denoted with f, colour with i and j and the Lorentz indices with μ and ν.

The covariant derivative D_{μ} is defined as:

$$D_{\mu} = (\partial_{\mu} + i g_s G^{a}_{\mu} T_a) \tag{1.3}$$

where G_a^{μ} are the eight gluon fields, g_s is the strong coupling and T_a are the eight generators of the SU(3) group. Plugging in the covariant derivative defined in Equation 1.3 into the Equation 1.2 results in the terms describing the couplings of the quark-gluon and the gluon-gluon fields.

1.1.2 Electroweak Interaction

The electroweak sector of the Standard Model describes the weak interaction, mediated via the charged W^{+} and the W^{-} bosons and the neutral Z-boson and the electromagnetic interaction mediated by the photon. The weak interaction occurs between particles carrying weak isospin, namely between all fermions in the Standard Model and the Higgs boson, while the electromagnetic interactions between charged particles.

The weak interaction depends on the chirality of particles, namely the charged weak interactions occurs only between the left-handed particles and the right-handed antiparticles, but not between the right-handed particles and left-handed antiparticles, leading to the maximal parity violation. The neutral weak interaction involves both the left-handed and the right-handed particles, but the different coupling strengths leads to the parity violation as well.

The weak and the electromagnetic interaction were described together and unified in the Glashow-Salam-Weinberg (GWS) model. In order to develop the electroweak theory, the left-handed particles are placed into the weak-isospin SU(2) doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix} \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \tag{1.4}$$

$$\tag{1.5}$$

while the right-handed fermions are placed into the weak-isospin SU(2) singlets:

$$u_R, d_R, c_R, s_R, b_R, t_R, e_R, \mu_R, \tau_R \tag{1.6}$$

The Lagrangian, describing the electroweak interaction can be split into parts:

$$\mathcal{L}_{EW} = \mathcal{L}_{EW}^{\ell} + \mathcal{L}_{EW}^{q} + \mathcal{L}_{EW}^{\text{free}} \tag{1.7}$$

where $\mathcal{L}_{EW}^{\ell}$ and $\mathcal{L}_{EW}^{q}$ describes the electroweak interactions of the leptons and the quarks with the gauge bosons and the quarks, while the $\mathcal{L}_{EW}^{\text{free}}$ describes the free fields of the gauge bosons.

The electroweak interactions of the leptons are described by:

$$\mathcal{L}_{EW}^{\ell} = \sum_{j=1}^{3} \bar{\ell}_L^j i\gamma^\mu D_\mu \ell_L^j + \sum_{j=1}^{3} \bar{\ell}_R^j i\gamma^\mu D_\mu \ell_R^j \tag{1.8}$$

where ℓ_L^j are three left-handed lepton doublets, ℓ_R^j are the three right-handed lepton singlets and γ^μ are Dirac matrices. The right-handed neutrino singlet is not introduced as right-handed neutrinos are not predicted in the Standard Model. The covariant derivative D_μ introduces the interactions between the leptons and the gauge bosons.

The electroweak interactions of the quarks are described by:

$$\mathcal{L}_{EW}^{q} = \sum_{j=1}^{3} \bar{q}_L^j i\gamma^\mu D_\mu q_L^j + \sum_{j=1}^{6} \bar{q}_R^j i\gamma^\mu D_\mu q_R^j \tag{1.9}$$

where the q_L^j are the tree left-handed quark doublets, the q_R^j are the six right-handed quark singlets and the covariant derivative D_μ introduces the interactions between the quarks and the gauge bosons.

The term $\mathcal{L}_{EW}^{\text{free}}$ describing the free fields of the gauge bosons. In Equation 1.7, it is assumed that the gauge bosons do not have mass which is further introduced via the Higgs mechanism.

1.1.3 Higgs Mechanism

The Higgs mechanism describes the generation of the masses for the massive gauge bosons and for the fermions which are acquired via their interactions with the Higgs field.

In the Higgs model, introduced is a Higgs doublet $\phi(x)$, formed from two complex scalar fields ϕ^+ and ϕ^0:

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.10}$$

The Higgs Lagrangian, leading to the generation of masses for the gauge bosons, can be expressed as:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\phi)^\dagger(D_\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \tag{1.11}$$

where $(D_\mu\phi)^\dagger(D_\mu\phi)$ describes the interactions between the gauge bosons and the Higgs doublet, while $\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ is the Higgs potential.

The Lagrangian defined in Equation 1.11 is locally gauge invariant, but if a specific gauge is chosen, this leads to the spontaneous symmetry breaking. In the Higgs model, chosen is a gauge in which the Higgs field can be defined as:

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v+h \end{pmatrix}, \tag{1.12}$$

where v is the Higgs vacuum expectation value and h is the field of the Standard Model Higgs boson.

After plugging in the Higgs doublet defined as in Equation 1.12 into the Lagrangian defined in Equation 1.11, the Lagrangian contains the terms describing the interactions between the Higgs and the gauge bosons but also the terms quadratic in the gauge boson fields which describe the masses of the gauge bosons.

The masses of the gauge bosons can be expressed as:

$$m_W = \frac{1}{2}vg_W \tag{1.13}$$

$$m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2} \tag{1.14}$$

$$m_A = 0 \tag{1.15}$$

where m_W is the mass of $W^\pm$-boson, m_Z is the mass of the Z-boson and m_A is the mass of the photon (=0). The g_W and g' are the weak couplings and v is the Higgs vacuum expectation value.

1.1.4 Yukawa Interactions

The generation of the fermion masses is incorporated into the Standard Model via Yukawa interactions *i.e.* interactions between the fermions and the Higgs boson. Massive fermions in the Standard Model are quarks and charged leptons, while neutrinos do not have mass. Hence, the Yukawa interaction can be split into the term describing the interactions between the charged leptons with the Higgs field and the term describing the interactions between the quarks with the Higgs field:

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = \mathcal{L}^{\ell}_{\text{Yukawa}} + \mathcal{L}^{\text{q}}_{\text{Yukawa}} \tag{1.16}$$

where $\mathcal{L}^{\ell}_{\text{Yukawa}}$ is the Yukawa term for leptons, while $\mathcal{L}^{\text{q}}_{\text{Yukawa}}$ is the Yukawa term for quarks.

The lepton Yukawa interaction can be expressed as:

$$\mathcal{L}^{\ell}_{\text{Yukawa}} = \sum_{ij} y^{\ell}_{ij}(\bar{\ell}_{Li}\phi\ell_{Rj} + \bar{\ell}_{Ri}\phi^\dagger\ell_{Lj}) \tag{1.17}$$

where ℓ^j_L are left-handed lepton doublets, ℓ^j_R are the right-handed lepton singlets, ϕ is the Higgs doublet, the y^ℓ is the Yukawa matrix for the charged leptons and indices i and j denote the lepton generation. The Yukawa matrix for neutrinos is not defined, as neutrinos remain massless in the Standard Model. After the spontaneous symmetry breaking, arise the terms describing the interactions between the charged leptons and the Higgs boson and the terms describing the fermion masses.

The quark Yukawa interaction can be expressed as:

$$\mathcal{L}^{\mathrm{q}}_{\mathrm{Yukawa}} = \sum_{ij} y^d_{ij}(\bar{q}_{Li}\phi d_{Rj} + \bar{d}_{Ri}\phi^\dagger q_{Lj}) + y^u_{ij}(\bar{q}_{Li}\phi_C u_{Rj} + \bar{u}_{Ri}\phi^\dagger_C q_{Lj}) \tag{1.18}$$

where the $y^d(\bar{q}_L\phi d_R + \bar{d}_R\phi^\dagger q_L)$ term describes the generation of the masses for the down-type quarks, while $y^u(\bar{q}_L\phi_C u_R + \bar{u}_R\phi^\dagger_C q_L)$ describes the generation of the masses for the up-type quarks. The y^u and y^d are the Yukawa matrices for the up-type and the down-type quarks, while the field ϕ_C is the complex conjugate of the Higgs doublet ϕ.

The Lagrangians defined in Equations 1.16 and 1.17 are shown in the interaction basis, while if transformed to the mass basis, the Yukawa matrices are diagonalized and the fermion masses can be expressed as:

$$m_f = \frac{v\, y_f}{\sqrt{2}} \tag{1.19}$$

where y_f defines the Yukawa coupling between the fermion and the Higgs boson.

After the transformation to the mass basis, the $W^\pm$-bosons couple to quark mass eigenstates of different generations. The quark interaction states d^I, s^I and b^I can be related to mass eigenstates d, s and b via:

$$\begin{pmatrix} d^I \\ s^I \\ b^I \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.20}$$

where V_{ij} are the elements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. Due to the $W^\pm$ cross-generational couplings in the quark sector, the quark flavour is not conserved per generation, leading to flavour violation in the quark sector.

The CKM matrix is the only source of charge-parity (CP) violation in the Standard Model, where CP symmetry is defined by a product of charge-conjugation operation, performed when transforming the particles into antiparticles and parity operation, performed when inverting the coordinate system.

In the lepton sector, there are no $W^\pm$ cross-generational couplings and the lepton flavour is conserved per generation.

1.2 Lepton Flavour Violation

In the original Standard Model with the massless neutrinos, the lepton number is conserved per generation, leading to the lepton flavour number conservation. However, the experimentally discovered neutrino oscillations show that the neutrinos can change the flavour *i.e.* a neutrino produced as an electron neutrino can later interact as a muon neutrino. Hence, lepton flavour is not conserved in neutrino oscillations. Further, neutrino oscillations can lead to lepton flavour violation in other processes at the loop level, as is the $Z \to \tau e/\mu$ process, but a very small branching ratio.

Apart from neutrino oscillations, lepton flavour violation has been predicted in several Beyond the Standard Model theories at a larger rate as are certain supersymmetric models [1, 2], the models with more than one Higgs doublet [3, 4], composite Higgs models [5], Randall–Sundrum models [6] and others.

1.2.1 Neutrino Oscillations

Neutrino oscillations describing the flavour oscillations of the neutrinos, have not been predicted in the original Standard Model, but have been experimentally observed [7]. Neutrino oscillations lead to the violation of the lepton family numbers, namely to the violation of the electron lepton number Le, muon lepton number L_μ and the tau lepton number L_τ.

Neutrino oscillations are explained by the neutrino weak interaction states being different from the neutrino mass eigenstates. The weak eigenstates ν_e, ν_μ and ν_τ can be expressed in terms of the mass eigenstates ν_1, ν_2 and ν_3 as:

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu 1} & U_{\mu 2} & U_{\mu 3} \\ U_{\tau 1} & U_{\tau 2} & U_{\tau 3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} \tag{1.21}$$

where U_{ij} are the elements of the Pontecorvo–Maki– Nakagawa–Sakata (PMNS) matrix, an analog of the CKM matrix in the quark sector.

Neutrino oscillations are depicted in Figure 1.2, where an electron neutrino ν_e, produced together with a positron, can be later interacts as a muon neutrino ν_μ, due to the change of flavour, leading to the violation of both the electron and the muon lepton number.

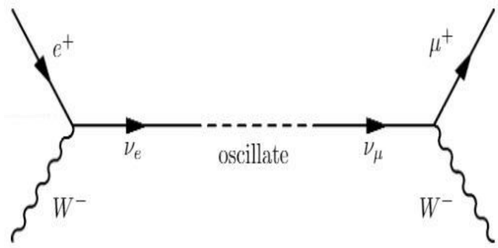


Figure 1.2: A neutrino produced as an electron neutrino further interacting as a muon neutrino, due to the change of flavour.

Neutrino oscillations occur only if the neutrinos are massive, but the origin of neutrino mass is not known. Neutrino mass can be introduced into the Standard Model via the Higgs mechanism, by including a term $-m(\bar{\nu}_R \nu_L + \bar{\nu}_L \nu_R)$ into the Standard Model Lagrangian. This implies the existence of the light right-handed neutrinos, which have not been detected so far and further results in very small Yukawa couplings. Other mechanisms are proposed such as the seesaw mechanism which in addition predicts new heavy neutrinos.

Due to neutrino oscillations, lepton flavour can be violated in other processes, such as in a Z-boson decay as shown in Figure 1.3. One of the neutrinos can change the flavour which results in producing a charged lepton of different flavour, compared to the initial. Hence, via this process, a Z-boson can be detected via two charged leptons of different flavour in the final state. However, the branching ratio for this process is very small, at the order of 10^{-54}.

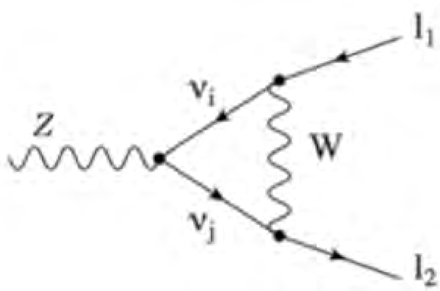


Figure 1.3: Lepton flavour-changing decay of the Z-boson, due to neutrino oscillations.

1.2.2 Supersymmetric Models

Lepton flavour violation is predicted in several supersymmetric models such as the supersymmetric seesaw model [2], which incorporates the seesaw mechanism for neutrino mass generation into the Minimal Supersymmetric Standard Model (MSSM), resulting in the prediction of new heavy right-handed neutrinos. Further, predicted are new couplings which enhance the lepton-flavour violating processes.

In Figure 1.4 shown are the Feynman diagrams, contributing to the flavour-changing decays of the Z-boson as predicted in the seesaw sypersymmetric model. Those Feynman diagrams result from the newly introduced neutral-current flavor-changing $Z\tilde{\nu}_X\tilde{\nu}_Y$, $Z\tilde{\ell}_X\tilde{\ell}_Y$, $\tilde{\chi}^0_\alpha \ell_I \tilde{\ell}_X$ and charged-current flavor-changing couplings $\tilde{\chi}^+_\alpha \ell_I \tilde{\nu}_X$.

Due to the new contributions, in this model, the branching ratio for the $Z \rightarrow \tau\mu$ process can be enhanced up to 10^{-8} [2].

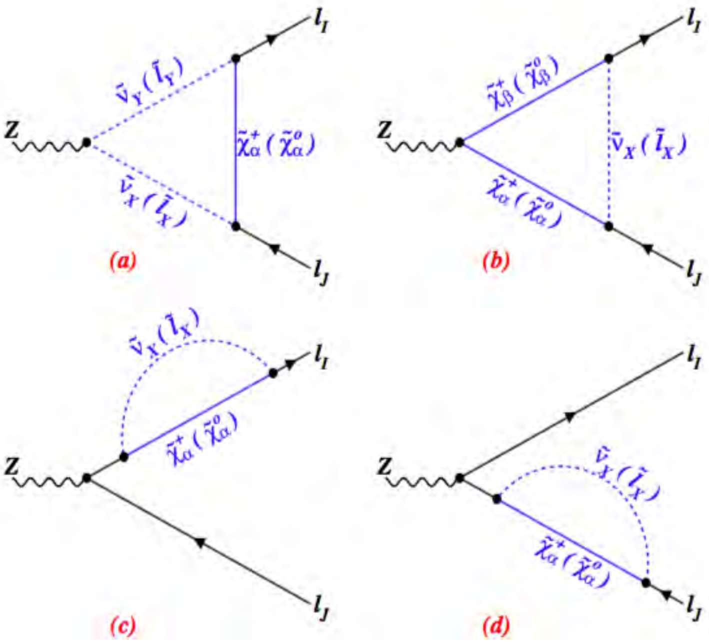


Figure 1.4: Feynman diagrams contributing to flavour-changing decays of the Z-boson, in seesaw sypersymmetric theories [2].

1.2.3 Two-Higgs Doublet Models

In the Two-Higgs Doublet Models (THDM) [4], introduced is a second Higgs doublet, resulting in five scalar particles in the Higgs sector: the three neutral h, H, A and a pair of charged $H^{\pm}$ particles.

The models predicting the coupling of the Higgs doublets to all fermions simultaneously introduce three-level flavour-changing couplings of the Higgs particles which can further contribute to other processes.

Due to the flavour-changing couplings of the Higgs particles, the flavour-violating decays of the Z-boson can occur at the loop level, as shown in Figure 1.5. the branching ratio for the $Z \rightarrow \tau\mu$ process can be enhanced up to $\backsim 10^{-6}$ [4].

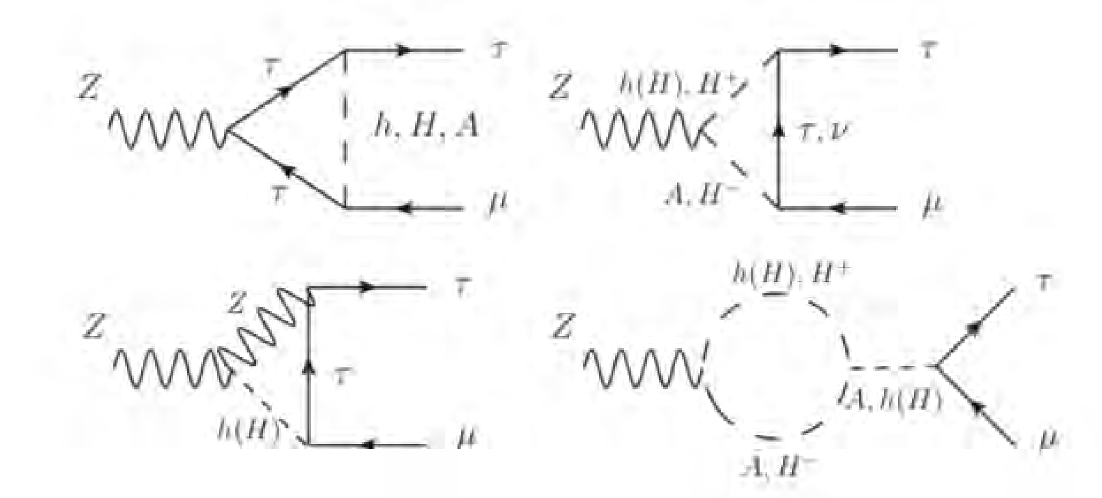


Figure 1.5: Feynman diagrams contributing to flavour-violating decays of the Z-boson, in two Higgs doublet models [4].

1.2.4 Prior Searches

The LEP experiments have set the upper limit on the branching ratio of the $Z \rightarrow \tau e$ process at $9.8 \cdot 10^{-6}$ [8] and on the $Z \rightarrow \tau\mu$ process at $1.2 \cdot 10^{-5}$ [9], at 95 % confidence level (CL).

The ATLAS experiment has set the upper limit on $Z \rightarrow \tau\mu$ at $1.7 \cdot 10^{-5}$ at 95 % CL [10] in the analysis done using the $20.3\,\mathrm{fb}^{-1}$ of proton-proton collision data at a center-of-mass energy of 8 TeV. No previously published limits on $Z \rightarrow \tau e$ process are set with ATLAS data. With on additional LHC data, new measurement of lepton flavour violation in Z decays is called for, which is the topic of this thesis.

The LHC and the ATLAS experiment

This chapter gives an overview of the Large Hadron Collider in Section 2.1, followed by the description of the ATLAS detector in Section 2.2. Described are the detector components, the trigger system and the luminosity measurements at ATLAS. Further, the luminosity and the pile-up profiles of proton-proton collision data at the center-of-mass energy $\sqrt{s} = 13\,\text{TeV}$ which corresponds to the data used in this analysis are given in Section 2.2.4.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [11] is the largest and the most powerful particle accelerator built so far, designed to explore current theories and to provide an insight into the theories which go beyond the Standard Model. The LHC was built in 2008, by the European Organization for Nuclear Research (CERN) in a tunnel of 27 km circumference, below the ground, near Geneva, Switzerland.

The LHC is designed for proton - proton, proton - lead and lead - lead ion collision[1]. The LHC operation started in 2009, at the proton-proton center-of-mass energies of $\sqrt{s} = 7\,\text{TeV}$, further increased to $\sqrt{s} = 8\,\text{TeV}$ in 2012 and early 2013 (Run-1 phase). During the Run-1, the Higgs boson, as the last missing piece of Standard Model was discovered which lead to the confirmation of the long foreseen Higgs Mechanism. In the following two years, the LHC was upgraded and restarted in 2015 at the proton-proton center-of-mass energies of $\sqrt{s} = 13\,\text{TeV}$ during the Run-2 phase (2015-2018).

Before being accelerated at the LHC, the particles are pre-accelerated at several accelerators. A scheme of the acceleration chain is shown in Figure 2.1. The proton-acceleration chain starts with the Linear Accelerator (LINAC 2) where the protons are accelerated up to 50 MeV and further injected into the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), providing the acceleration up to the energies of 1.4 GeV, 25 GeV and 450 GeV, respectively. The protons are further injected into the two beam pipes of the LHC, where they circulate in the clockwise and the anti-clockwise direction and collide at four intersection points where the two beams merge into one, and where the detectors are placed. Each beam consists of particles organized into trains of bunches, with each bunch containing approximately $1.15 \cdot 10^{11}$ protons, separated by a spacing of 25 ns in Run-2. The beams are bent into the circular path using dipole magnets positioned along the beam pipe. The focusing of the beam is done using the quadropole magnets and additional corrections are applied using sextupole and decapole magnets. At the end of each run, the beams are dumped on the dedicated dump targets, designed to absorb the beam energy.

There are four larger and several smaller detectors around the LHC ring. The two largest detectors are ATLAS [12] and CMS[13], which are general-purpose detectors, designed to independently conduct a wide range of searches and to enable cross-confirmation of results.

[1] In addition, in 2017, a short run of *xenon* - *xenon* collisions was conducted

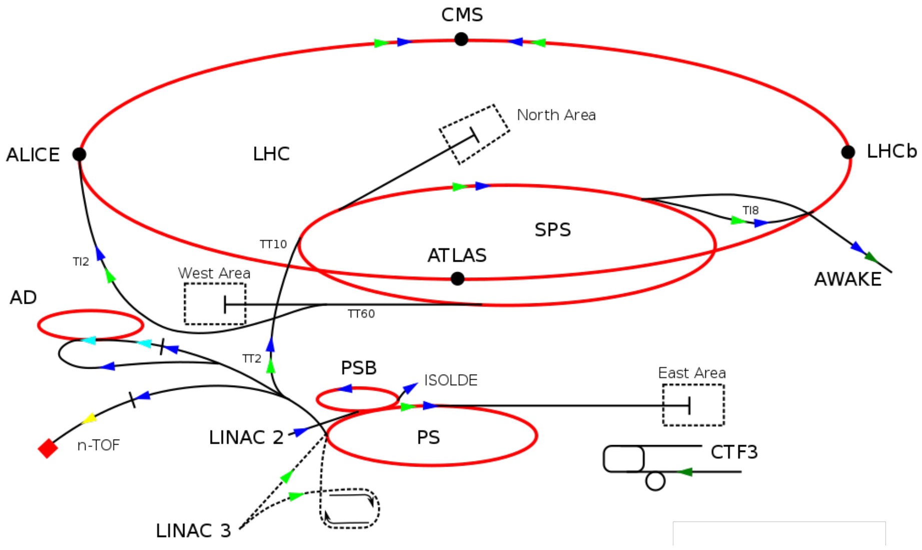


Figure 2.1: A scheme of the acceleration chain: the proton-acceleration chain starts with the Linear accelerator 2 (LINAC 2), proceeding with the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) before being injected into the LHC. Similarly, the acceleration of heavy ions starts with the Linear accelerator 3 (LINAC 3) and proceeds with Low Energy Ion Ring (LEIR), the PS and the SPS.

Another two large detectors are ALICE [14], designed to explore the state of quark-gluon plasma, in lead ion collisions and the LHCb [15] detector, targeted at exploring the asymmetry between matter and antimatter. The three smaller detectors, sharing the cavern with ATLAS, CMS and LHCb are TOTEM [16], designed for precise cross section measurements, MoEDAL [17], designed for magnetic monopole searches and LHCf [18], designed for cosmic ray searches.

2.2 The ATLAS Detector

2.2.1 The Detector Components

The ATLAS detector [19, 12] is the largest LHC detector, designed as a general-purpose detector and involved in a wide range of physics searches. It consists of multiple subdetectors, specialized for tracking, energy measurements and the muon detection.

A scheme of the ATLAS detector is shown in Figure 2.2.

The **Inner Detector** is designed to provide precise measurements of the trajectories of the traversing charged particles. The measurement of the magnitude and the direction of the curvature in the magnetic field enable the determination of the momenta and the charge of the particles. Further, the interpolation of the trajectories towards the proton-proton interaction region enables the reconstruction of the primary and the secondary vertices and the determination of the transverse and the longitudinal track impact parameters.

The inner detector consists of three subsystems: the pixel detector, the semiconductor

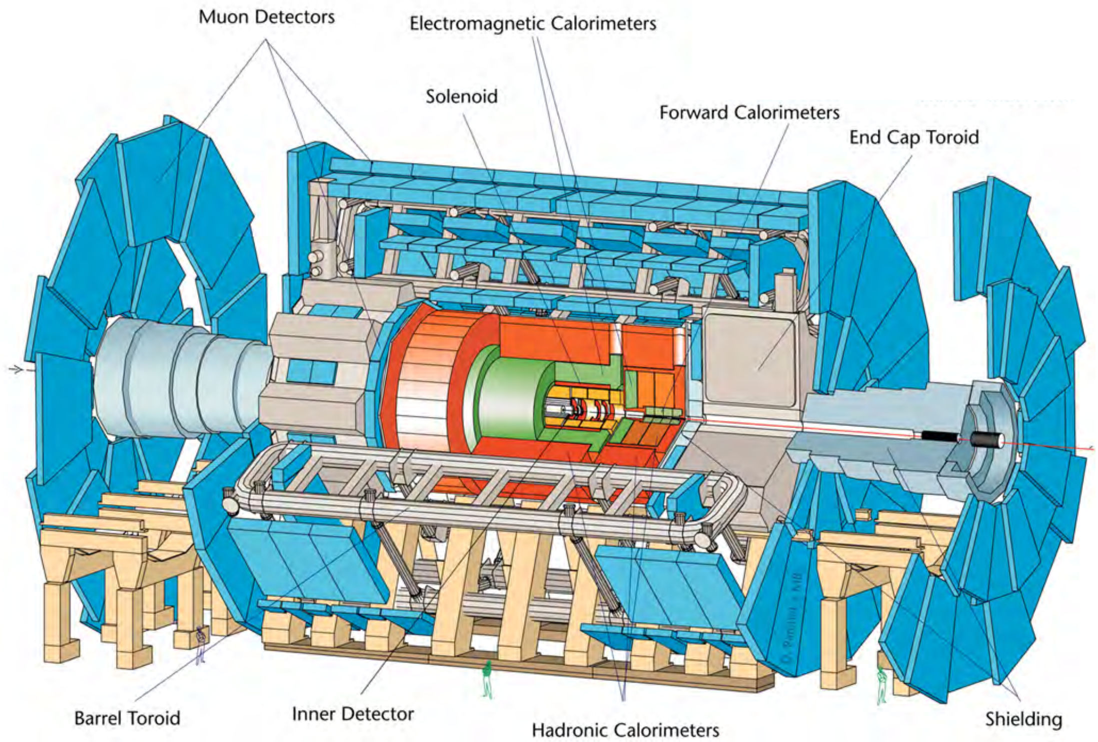


Figure 2.2: A scheme of the ATLAS detector.

tracker, and the transition radiation tracker each providing measurements of the *hits*, *i.e.* the signals produced due to the interaction of the charged particles with the detector material. The full inner detector is immersed into a 2T magnetic field, generated by the solenoid electromagnet which surrounds the inner detector and used to bend the particles.

Closest to the beam pipe are the pixel detector and the semiconductor tracker which provide the most precise space measurements. The pixel detector consists of silicon pixels organized into four barrel layers and three disks, placed in each end-cap region. The semiconductor tracker consists of silicon microstrip sensors organized into four barrel layers and nine disks, per end-cap. The semiconductor tracker is surrounded by the transition radiation tracker, consisting of a large number of gaseous drift tubes ("straws"), covering both the barrel and the end-cap region. The transition radiation tracker is used to detect transition radiation photons from charged particles which is especially important for correct electron identification. A scheme of the inner detector is shown in Figure 2.3.

The inner detector is encompassed by the **solenoid electromagnet**, used to generate the magnetic field to bend the particles inside the inner detector.

Above the solenoid are the **electromagnetic calorimeter** and the **hadronic calorimeter**, used to stop and measure the energy of charged and neutral particles. Particles are slow downed and stopped due to the interactions with the material of the calorimeter which results in showers of particles which are further collected to produce the output electrical signal. The material used to stop the particles is called the *passive* component, while the material used to collect the signal is called the *active* component.

The electromagnetic calorimeter is based on the Liquid Argon technology and it consists

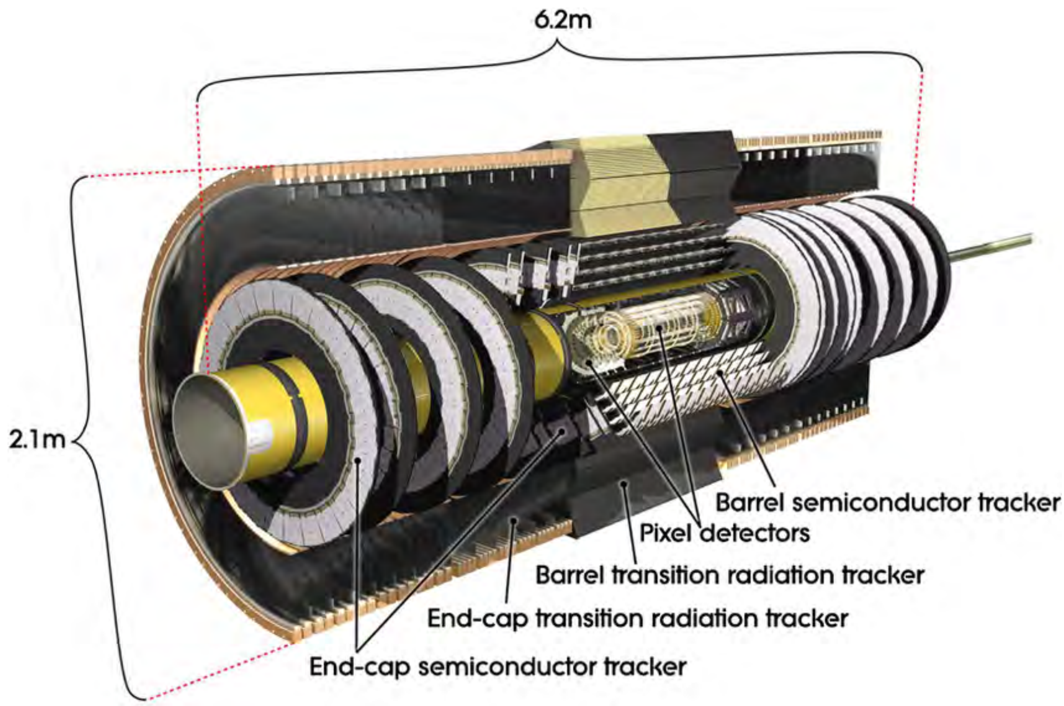


Figure 2.3: A scheme of the ATLAS inner detector, consisting of the pixel detector, the semiconductor tracker and the transition radiation tracker in the barrel and the two end-caps.

of the barrel and the two endcaps, both built using the lead plates as the passive material and the liquid argon as the active material. The hadronic calorimeter consists of the barrel and the two extended barrel parts, built using the steel as the passive material and the scintillating tiles as the active material. The hadronic calorimeter also contains two endcap calorimeters built using the copper and liquid argon.

In addition, the forward calorimeter (FCAL), specially designed for high fluxes, is positioned in the forward region, near the interaction point. It consists of tree modules: a module based on copper and liquid argon, optimized to measure the energy from the electromagnetic particles, and the two modules, based on tungsten and liquid argon, optimized to measure the energy from the hadrons.

A scheme of the ATLAS calorimeter system is shown in Figure 2.4.

Due to the low interaction rate, muons can traverse the large part of the detector without interaction. Hence, the **muon spectrometer**, specially designed for muon detection is added to the detector. The muon spectrometer is designed to determine the momenta of highly energetic muons, by measuring the curvature in the magnetic field, generated by the surrounding toroid magnet. It consist of muon chambers, organized into the tree cylindrical layers around the beam axis and the tree layers perpendicular to the beam, placed on each side.

The surrounding **toroid electromagnet** consists of the eight coils in the barrel and each end-cap regions, generating a magnetic field of 0.5T in the barrel and 1T in the end-cap region. The scheme of the toroid magnet is shown in Figure 2.5

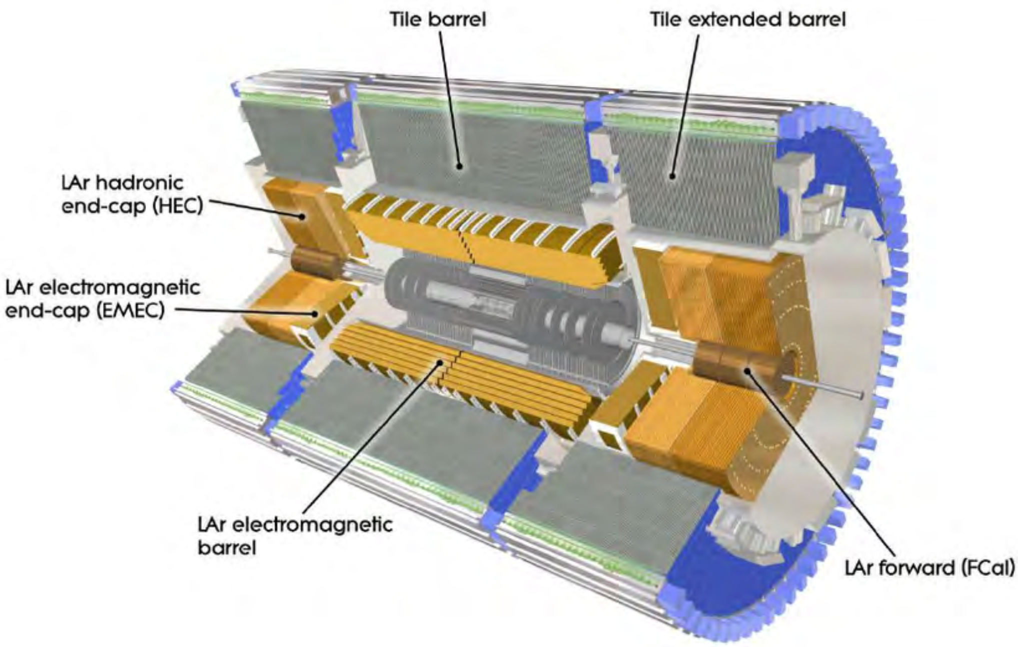


Figure 2.4: A scheme of the ATLAS calorimeter system: the electromagnetic calorimeter consists of the LAr based barrel and the two LAr based end-caps, while the hadronic calorimeter consists of the central and the two extended Tile barrel parts and the two LAr based end-caps. The LAr forward calorimeter consists of one module optimized for measurements due to the electromagnetic and two modules optimized for measurements due to the strong interaction.

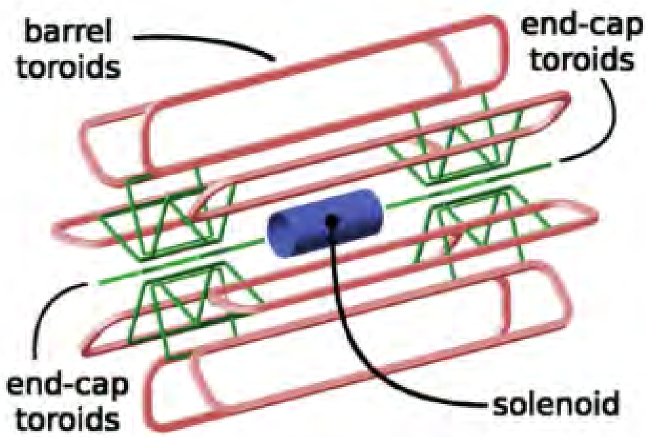


Figure 2.5: A scheme of the ATLAS toroid magnet.

2.2.2 The Trigger System

Due to the high collision rates at the LHC and the lower rates at which the events can be stored, only the collisions containing an events of interest are selected for storage. The system employed to make a fast decision on whether an event will be stored is called the trigger system. A parameter characterizing the promptness of trigger is called latency, defined as the time elapsed between a bunch-bunch collision and the trigger decision.

In order to make a decision, the bunch collision data are analysed and stored if a feature of interest is found. The features of interest are defined upon the requests of different physics analyses and form a trigger menu. The trigger menu consists of multiple triggers defined by the requirements imposed on the kinematics or identification quality of one or multiple objects. In this analysis, used are the single-electron and single-muon triggers defined by the requirements on the momenta and the identification quality.

A scheme of the ATLAS trigger system [20] is shown in Figure 2.6. It is implemented as a hardware-based low level trigger (L1) with a latency of ~2.5 µs and a software-based high level trigger (HLT) with a latency of ~550 ms. The L1 trigger uses the information from the calorimeters and from muon detectors to make a fast decision on the event which if accepted is passed to the HLT which further implements complex software algorithms to decide on the event. The L1 trigger reduces the rate from 40 MHz to roughly 100 kHz., while the HLT reduces it further to roughly 1 kHz.

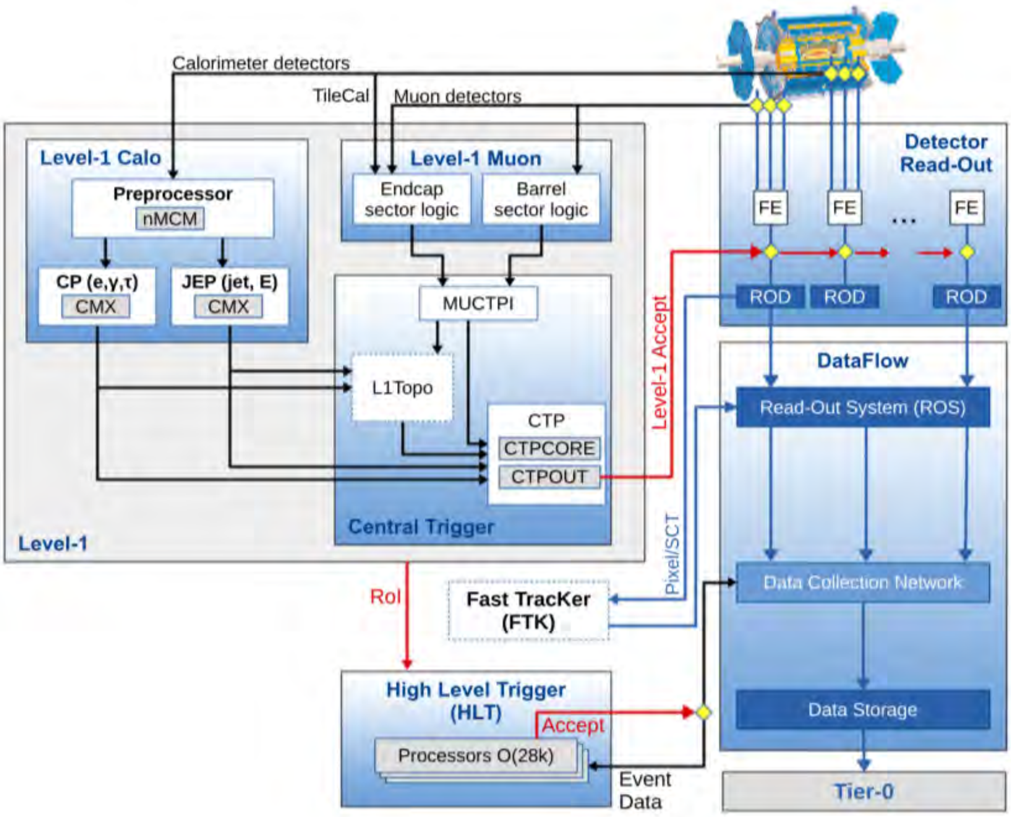


Figure 2.6: A scheme of the ATLAS trigger system [20]

2.2.3 The Luminosity Measurements

The luminosity is a measure of the number of (proton-proton) collisions which happen at the interaction point. The luminosity L can be expressed as:

$$L = \frac{f n_b n_p^2 g}{4\pi \epsilon_x \epsilon_x} \tag{2.1}$$

where n_b is the number of the bunches in the ring, n_p is the number of protons per bunch, f the revolution frequency , g is the geometrical factor related to the crossing angle of the two bunches and ϵ_x and ϵ_x are the beam widths in the horizontal and the vertical direction. The Equation 2.1 defines the instantaneous luminosity, while the luminosity over time is defined as $\int \mathcal{L} dt$.

The number of produced events N over time t for a specific process can be expressed as:

$$N = L\sigma \tag{2.2}$$

where σ is the cross section and L is the luminosity, integrated over time t.

In order to determine the cross sections, it is important to perform precise luminosity measurements. The luminosity can be determined directly from the measured parameters of the machine or indirectly from a known cross section and the measured number of events.

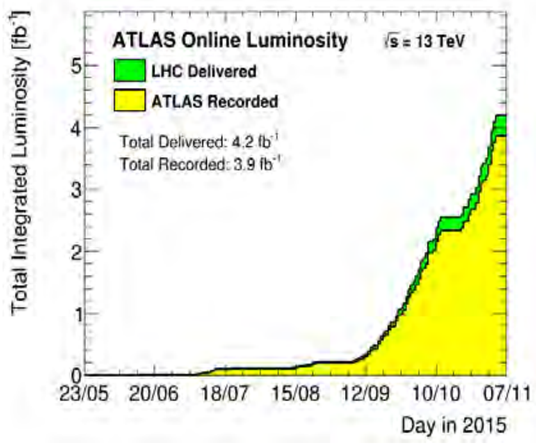
At ATLAS, the luminosity is measured using several different detectors. The major ones are the Luminosity Cherenkov Integrating Detector (LUCID) and the Beam Conditions Monitor (BCM). LUCID consists of two Cherenkov devices positioned at ±17 m from the ATLAS interaction point, while the Beam Conditions Monitor (BCM) is an ATLAS submodule which consists of diamond sensors placed at ± 1.8 m from the interaction point. The BCM is specially used to monitor the beam conditions and to quickly detect any beam irregularities. Since different detectors can have different efficiencies, they are calibrated by performing luminosity measurements for different separations between the beams in the so-called van der Meer scans.

The time periods during which the luminosity is considered constant are called luminosity blocks and they typically last for one minute. Multiple luminosity blocks during one stable beam define a run which typically lasts for 10-12 hours, while multiple runs during the same configuration define a period.

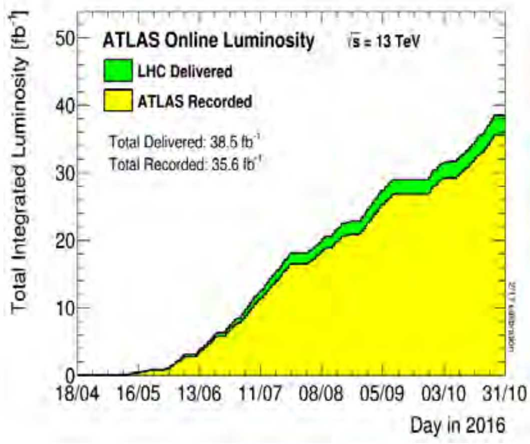
2.2.4 The Luminosity and the Pile-Up Profiles

The luminosity of the proton-proton collision data collected during 2015 and 2016 which corresponds to the data used in this analysis is shown in Figure 2.7. Shown is the luminosity delivered by the LHC and the luminosity recorded by the ATLAS detector which differ due to the inefficiencies in the data acquisition. The luminosity delivered by LHC corresponds to $4.2\,\mathrm{fb}^{-1}$ in 2015 and $38.5\,\mathrm{fb}^{-1}$ in 2016, while the luminosity recorded by the ATLAS corresponds to $3.9\,\mathrm{fb}^{-1}$ in 2015 and $35.6\,\mathrm{fb}^{-1}$ in 2016.

Due to the high instantaneous luminosity, multiple proton-proton interactions can occur in one bunch-bunch collision which can further form a background to the process of interest. The interactions can occur between the protons in the same collision (in-time pile-up) or between the protons from the previous or the following collision (out-of-time pile-up). The mean number of proton-proton interactions per bunch crossing in 2015 and 2016 data, weighted by the luminosity of the data, are shown in Figure 2.8.

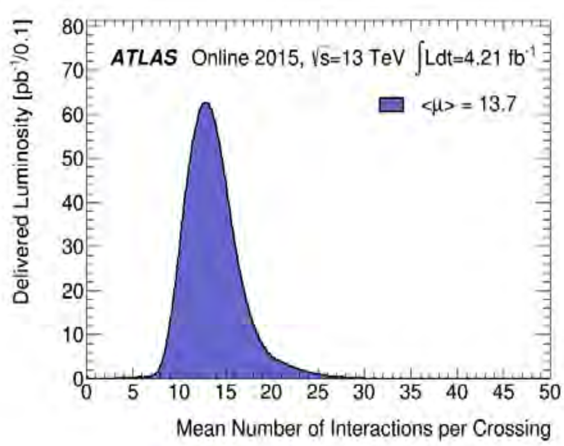


(a) 2015

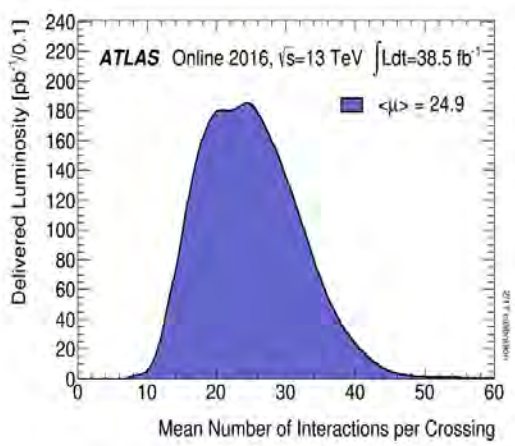


(b) 2016

Figure 2.7: The LHC delivered and the ATLAS recorded cumulative luminosity of the (a) 2015 and (b) 2016 proton-proton collision data collected at the centre-of-mass energy $\sqrt{s}$ = 13 TeV[21].



(a) 2015



(b) 2016

Figure 2.8: Luminosity-weighted distribution of the mean number of interactions per bunch crossing for proton-proton collisions at the center-of-mass energy $\sqrt{s}$ = 13 TeV, delivered by the LHC during a) 2015 and b) 2016 [21].

Monte Carlo Simulation

This chapter describes the characteristics of the signal and the Standard Model background processes together with their simulation within the ATLAS software framework. An overview of the Monte Carlo (MC) simulation within the ATLAS software framework is described in Section 3.1. The signature of the $Z \to \tau e/\mu$ signal process and its simulation is described in Section 3.2. The relevant background processes, the way each process forms the background and the details on their simulation are given in Section 3.3.

3.1 Monte Carlo Simulation within the ATLAS software framework

Theoretical predictions for physical processes are modelled using Monte Carlo generators (MC). MC simulation starts with the generation of a process of interest *i.e.* the hard process. To include the detector effects, the particles are propagated through a detailed detector simulation to produce hits in different parts of the detector which are then passed through the digitization algorithms. In the final stage of the simulation, the reconstruction algorithms are run and physics objects are formed in the same way as if the simulation was real data.

The full simulation chain is implemented within the ATLAS software framework, Athena [22]. An overview of the ATLAS simulation data flow [23] is shown in Figure 3.1.

The simulation flow starts with the simulation of the hard interaction *i.e. event generation*. This stage involves the calculation of the cross section for the hard process and computes the kinematics of the produced particles. The cross section for a hard process in proton-proton collision is given with the factorization formula [25, 26]:

$$\sigma = \int \int dx_1 dx_2 f_i^p(x_1) f_j^p(x_2) \hat{\sigma}(x_1 P_1, x_2 P_2) \tag{3.1}$$

where $\hat{\sigma}$ is the partonic cross section defined as the cross section for the interaction between the quarks and the gluons which consitute the protons. The partonic cross section depends on the amplitude for the process and on the phase space. The amplitude is calculated in perturbation theory as a power series in terms of the coupling constant and integrated over the available phase space to calculate the partonic cross section.

As partons can carry different fractions of the proton momentum, the probability for a parton of flavour i to carry fraction x of the proton momentum P_i is described by the parton distribution function (PDF) f_i^p. In the MC approach, the integral is computed by random sampling over the phase space which outputs the kinematics of the particles involved in the process.

In the next stage, QED and QCD radiation emitted by the particles in the hard process is simulated. If the radiation is emitted by the incoming particle before the the hard interaction,

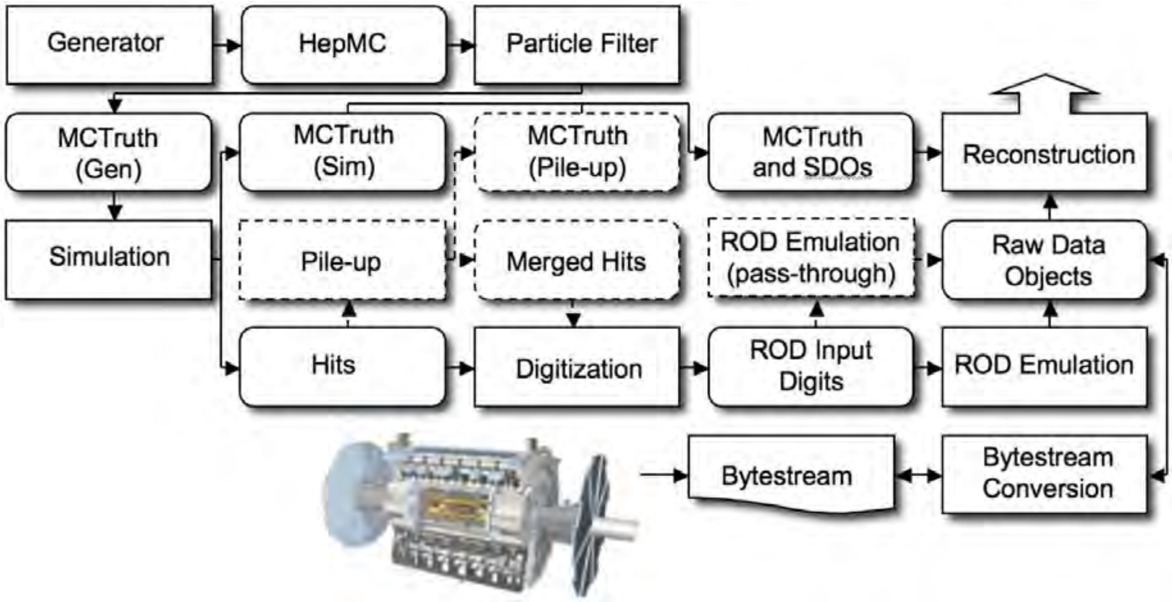


Figure 3.1: The ATLAS simulation data flow: the output of the MC generator is written to HepMC format [24]. At this stage, a particle filter can be applied to filter events of interest. A detailed output of the event generation is stored in the form of truth particles fwhich are then propagated trough the detector simulation to generate the hits in the detector. At this stage, a simulation of pile up hits can be overlayed with the hits from the hard process to produce the merged hits. The detector hits are further digitized to produce *Raw Data Object* (RAW) files. In the final stage, the reconstruction algorithms are run to form the physics objects.

it is called Initial-state radiation (ISR) while the radiation from the particles produced in the hard interaction is called Final-state radiation (FSR).

Additionally, since the quarks and gluons are not found as free particles, hadronization models are included to describe the formation of hadrons. Since hadronization happens at the low momentum scale, perturbation theory breaks down so phenomenological models are taken into account. In the final stage of the simulation, an event filter can be applied to select only events of interest for analysis.

After the event simulation, each particle is propagated trough the detector simulation. The ATLAS detector simulation is based on GEANT4 [27], a software package that provides a detailed description of the ATLAS detector including its geometry, the material and the models for physics interactions. The interactions of MC generated particles with the detector material produce hits in the detector which are stored as the output of the detector simulation stage. The detector simulation can be done as full, meaning that a very detailed and accurate detector description is used, or fast simulation [28], where a less detailed detector description is used but the simulation is faster.

Besides the particles from the hard process, additional traces in the detector are produced by the particles from inelastic soft interactions between the remnants of the colliding protons referred as the underlying event and by the particles originating from the pile-up.[1]. Particles from the underlying event and the pile-up produce additional hits in the detector which are

[1]Pile-up is described in Section ch:lumi$_{pu}$

overlayed with the hits left by the particles from the hard process. For a realistic description of a process, those additional signals need to be modelled and included in the simulation. In the ATLAS framework, pile-up interactions are simulated by minimum bias events.

Additionally, besides the underlying event and the pile-up, there are other sources of background like cavern background which are included in the simulation. [2]. The output of the detector simulation is written out as the hits file.

After the detector simulation, the readout electronics and the digitization of the signals are simulated. The simulation of the Low Level Trigger is included at this stage, as it is implemented in hardware. Events are not rejected at this stage, but the trigger decision is stored to be used for later analyses. The digitized signals are written to Raw Data Object (RAW) files which are taken as the input for the simulation of the High Level Trigger and for the reconstruction algorithms. After the reconstruction simulation, physics objects are formed and written to Event Summary Data (ESD) files which are mainly used for detector and reconstruction studies and Analysis Object Data (AOD) files which contain a less detailed output mainly used for physics analysis. If further data reduction is done, the output are Derived Analysis Object Data (dAOD) files. The data reduction can be done by removing events that do not pass certain selection criteria (skimming), by removing objects that fail some selection (thinning) or by removing variables not needed in the analysis (slimming).

In ATLAS, MC production is divided into campaigns which describe different run configurations.[3] The samples used for this analysis are simulated within the MC15 campaign which is a simulation of processes at $\sqrt{s}$ = 13 TeV runs, using the Run-2 conditions for 2015+2016 detector simulation.

During the simulation, several phenomenological models are used. The free parametes corresponding to those models are determined from data in the process called *tuning* which provides the set of parameters (*tunes*) used in the simulation.

The MC samples are normalized to the integrated luminosity of the used data samples using the following weight:

$$w = \mathcal{L} \cdot \frac{\sigma}{N_{gen}} \tag{3.2}$$

where $\mathcal{L}$ corresponds to the integrated luminosity of 36.1 fb^{-1}, σ is the cross section of the process while N_{gen} is the initial number of the generated AOD events, before any generator filter is applied.

All MC samples were reweighted with respect to the number of primary vertices by applying the pile-up weight [29] which removes the difference in the distribution of the number of primary vertices between the simulation and the data.

The generators used for the simulation of the samples in this analyses are PYTHIA 6 [30] and PYTHIA 8 [31] interfaced with POWHEG framework [32], for the parton shower, and to SHERPA, for the hard process. An overview of the used MC generators and the order at which the cross section is given in Table 3.1. The signal samples are simulated at the leading order (LO) and further normalized to $NNLO$, using the Z production cross section, derived from the $Z \to \tau\tau$ sample. Tables with the detailed description of used samples are given in Appendix A.

[2]Additionally, cosmic background and beam pipe gas(residual gas in the beam pipe) can leave traces in the detector, but as those effects are very small, they are not included in the simulation.

[3]Major campaigns correspond to the calendar year. Subversions are made for improvements in reconstruction software, trigger menu and/or pile-up simulation.

Table 3.1: Monte Carlo generators used in the simulation of the signal and the Standard Model background processes and the order at which the cross section was computed.

Sample	**Generator**	Order(σ)
$Z \to \tau e/\mu$	POWHEG +PYTHIA 8	LO, normalized to NNLO
$Z/\gamma* \to ee$+jets	POWHEG + PYTHIA 8	NLO
$Z/\gamma* \to \mu\mu$+jets	POWHEG + PYTHIA 8	NLO
$Z/\gamma* \to \tau\tau$+jets	SHERPA 2.2.1	NNLO
$W \to e\nu_e$+jets	SHERPA 2.2.1	NNLO
$W \to \mu\nu_\mu$+jets	SHERPA 2.2.1	NNLO
$W \to \tau\nu_\tau$+jets	SHERPA 2.2.1	NNLO
$t\bar{t}$	POWHEG + PYTHIA 6	NLO
single-top	POWHEG + PYTHIA 6	NLO
diboson	SHERPA 2.2.1	NNLO
Higgs	POWHEG + PYTHIA 8	NLO
low-mass Drell-Yan	POWHEG + PYTHIA 8	NLO

3.2 LFV $Z \to \tau e/\mu$ Signal

The signal signature of a lepton flavour violating $Z \to \tau e/\mu$ process consists of an oppositely charged $(\tau, e/\mu)$ pair, with a hadronically decaying τ. In the Standard Model, τ-lepton decays via an emission of a virtual W-boson and an ν_τ. Depending on the final products, the τ decays are classified as leptonic or hadronic. Feynman diagrams for leptonic and hadronic τ decays are shown in Figure 3.2.

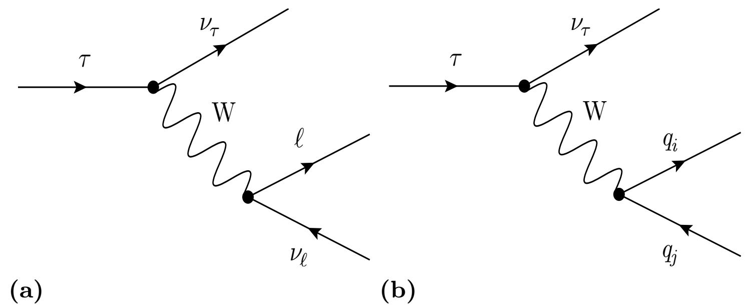


Figure 3.2: Feynman diagram for the leptonic (a) and hadronic (b) decay of the τ-lepton in the Standard Model. Labels q_i and q_j refer to different quark flavours.

Final state of a leptonically decaying τ-leptons is characterized by the lepton and by $E_{\mathrm{T}}^{\mathrm{miss}}$ due to the two neutrinos (ν_τ and ν_ℓ from the W-boson decay) which escape the detection.

Hadronic decays are characterized by $E_{\mathrm{T}}^{\mathrm{miss}}$ due to ν_τ and by one or more hadronic jets, formed during the hadronization of the quarks originating from the W-boson. As the τ mass is heavy enough [4], τ is the only lepton which can decay to lighter hadrons *i.e.* $\pi^{\pm}$ or $K^{\pm}$ mesons. Since the τ-lepton is a charged particle, the summed charge of the final particles after the decay has to be ± 1 meaning that the final state can have only an odd number of charged mesons like π^{+}, K^{+} (so called 1-prong decay) $\pi^{+}\pi^{-}\pi^{+}$ (so called 3-prong decay) etc. Neutral mesons as π^{0} can be produced together with the charged mesons.

Since τ-leptons have mass that is low compared to the LHC energies, they travel at high speeds and are highly boosted particles. Due to the high boost, in the laboratory frame, their decay products are produced inside very narrow angular cones. As jets have larger mass, they are not as highly boosted, meaning that their decays products are produced inside wider cones. This property is later on used to reject the backgrounds where the jets are faking τ-leptons.

As a conclusion, the $Z \to \tau e/\mu$ signal signature with a hadronically decaying τ-lepton is characterized by an opposite sign $(\tau, e/\mu)$ pair where the τ-lepton is reconstructed as a narrow cone of jets and with $E_{\mathrm{T}}^{\mathrm{miss}}$ due to ν_{τ}. As the Z-boson is not highly boosted due its large mass, the second lepton is at a large angular distance with respect to the τ.

Signal processes are simulated with PYTHIA 8. Both the leptonic and the hadronic decays are included in the simulation, but the hadronically decaying τ-leptons are selected for the analysis. Detailed tables describing the signal samples are given in Appendix A.1.

3.3 Standard Model Backgrounds

This section describes the relevant background processes and their simulation. The Standard Model processes which leave a similar signature in the detector as the signal process are referred as background processes. A process can be classified as background if it has the same signature as the signal (real, irreducible background) or if the particles in the process can be misidentified as signal particles (ake, reducible backgrounds). The $Z/\gamma*$+jets, W+jets, diboson, top and Higgs processes are included as the background. As the $Z/\gamma*$+jets samples include a low mass cut on the invariant mass of the Z-boson, additional Drell-Yan samples for low mass are included.

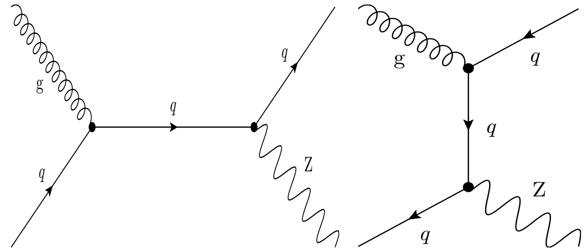
3.3.1 Z+jets

$Z/\gamma*$+jets processes include the production of a Z-boson or a virtual γ photon alone or in association with jets. This process can contribute as fake background in $Z \to ee$ and $Z \to \mu\mu$ processes and as real background in $Z \to \tau\tau$ processes. Fake background in $Z \to ee$ and $Z \to \mu\mu$ is formed if one of the e/μ-leptons or a jet is misidentified as a τ-lepton, while the second e/μ is correctly reconstructed. Real background in $Z \to \tau\tau$ is formed when one of the τ-leptons decays leptonically to the $(e, \nu_e)/(\mu, \nu_\mu)$ pair and the e/μ is correctly identified, while the second τ-lepton decays hadronically and is reconstructed as a hadronic τ-jet. The major background is originating from the Z+0 jets processes.

Examples eading order diagrams for the production of a Z-boson with an associated jet is shown in Figure 3.3.

$Z/\gamma*$+jets were separately simulated for the $Z \to ee$ and $Z \to \mu\mu$ and $Z \to \tau\tau$ processes. The $Z \to ee$ and $Z \to \mu\mu$ samples were produced with PYTHIA 8 and POWHEG while the $Z \to \tau\tau$ samples were produced with SHERPA 2.2.1, interfaced with POWHEG and sliced in $\max(H_{\mathrm{T}}, p_{\mathrm{T}}(Z))$ where H_T is the scalar sum of jet transverse momentum, while $p_{\mathrm{T}}(Z)$ is the

[4] 1.776 GeV



(a) **(b)**

Figure 3.3: Examples of leading order Feynman diagrams for $Z/\gamma * +1$ jet production.

transverse momentum of the Z-boson. The samples include a cut on minimum invariant mass of the Z-boson. A list of used MC samples is given in Appendix A.2.1.

3.3.2 Drell-Yan

Drell-Yan is an electroweak process in which $q/\bar{q}$ pair annihilates into a $Z/\gamma*$ which further decays into a lepton pair. The process is shown via Feynman diagram in Figure 3.4.

This process forms a minor real or fake background as described in Section 3.3.1.

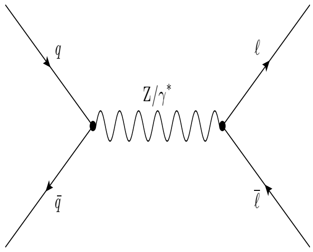


Figure 3.4: Drell-Yan process: $q/\bar{q}$ annihilates into a $Z/\gamma*$ which further decays into a lepton pair.

Drell-Yan process is simulated for the low Z-boson mass using PYTHIA 8 interfaced with POWHEG, with the AZNLOCTEQ6L1 tune and CT10 PDF set. A list of used MC samples is given in Appendix A.2.6.

3.3.3 W+jets

W+jets processes include the production of a W-boson alone or in association with jets. Processes with leptonically decaying W-bosons are considered, as they can leave the same signature as the signal. W+jets form the fake background with the main contribution from a jet misidentified as a τ-lepton and a real e/μ in $W \rightarrow e\nu_e$+jets and $W \rightarrow \mu\nu_\mu$+jets processes and with a small contribution from a fake τ-lepton and a real e/μ in the $W \rightarrow \tau\nu_\tau$+jets process. Hence, the main contribution is from the processes which contain at least one jet.

Examples of leading order Feynman diagrams for W+jets production are shown in Figure 3.5.

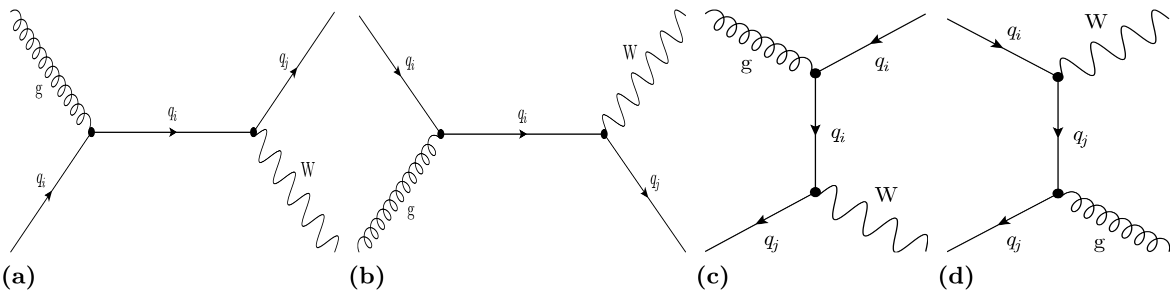


Figure 3.5: Examples of leading order Feynman diagrams for W+1 jet production. Labels q_i and q_j refer to different quark flavours.

W+jets samples were produced using SHERPA 2.2.1, and POWHEG with the NNPDF3.0 PDF set and the NNLO tune. The samples are split into $W \rightarrow e\nu_e$+jets, $W \rightarrow \mu\nu_\mu$+jets and $W \rightarrow \tau\nu_\tau$+jets processes and sliced in $\max(H_{\mathrm{T}}, p_{\mathrm{T}}(W))$ where H_T is the scalar sum of jet transverse momentum, while $p_{\mathrm{T}}(W)$ is the transverse momentum of the W-boson. Each slice and process are further separated into the samples with a light hadron filter, c-filter and a b-filter. Light hadron filter selects processes with a light hadron, while c-filter and b-filter select processes with B and C-hadrons. A list of used MC samples is given in Appendix A.2.2.

3.3.4 Top

Top background includes $t\bar{t}$ pair and single $t/\bar{t}$ production (single-top). It contributes to the fake background with a jet faking the τ-lepton and to the real background in the $t\bar{t}$ processes when the two W-bosons decay to the (τ,ν_τ),$(e/\mu,\nu_e e/\nu_\mu)$ pairs.

t-quark decays to a d,s or b-quark, but since the V_{td} and V_{ts} parameters in the CKM matrix are very small, t-quark mainly decays to a b-quark. b- quark then hadronizes into a B-meson which decays into ligther mesons, like π-mesons. The lifetime of a B-meson is short enough to decay inside the detector, so the lighter mesons from a b-quark can be detected. As the hadronically decaying τ-lepton decays to charged mesons, the charged mesons from the B-meson decay can be misidentified as coming from a τ decay. Another possibility of forming the background is if one of the leptons fakes the τ. Leading order processes for the $t\bar{t}$ pair production are shown in Figure 3.6.

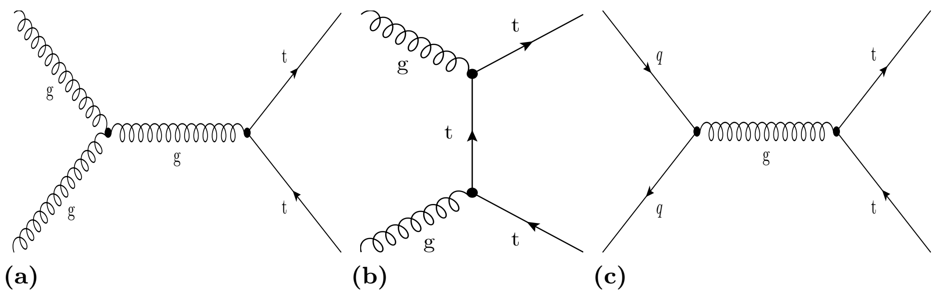


Figure 3.6: Examples of leading order Feynman diagrams for $t\bar{t}$ pair production via gluon scattering in the s-channel (a) and the t-channel (b) and quark scattering in the s-channel (c).

Single-top background is mainly formed in Wt process (a production of a $t/\bar{t}$ with an associated W-boson) when a jet formed from the decay of the $t/\bar{t}$ is misidentified as a τ and an e/μ from a W-boson is correctly reconstructed. Leading order Feynman diagrams for the single-top production are shown in Figure 3.7.

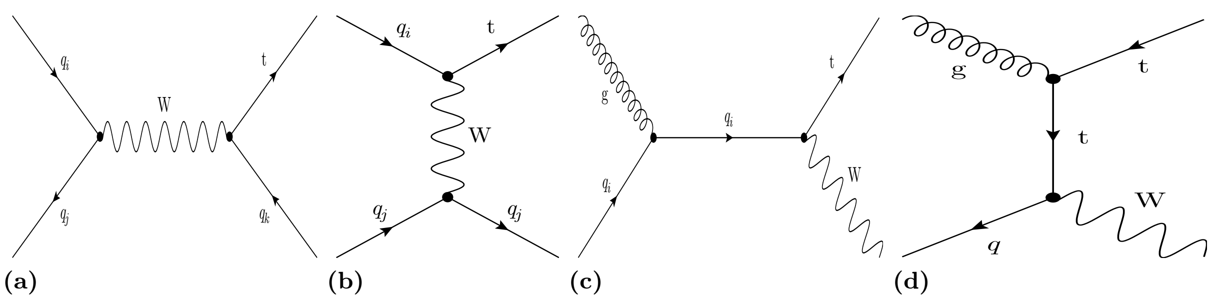


Figure 3.7: Examples of leading order Feynman diagrams for single-top production in the s-channel (a) and the t-channel (b) and the associated production with a W-boson in the s-channel (c) and the t-channel (d). Labels q_i and q_j refer to different quark flavours.

Top processes were simulated using POWHEG interfaced with PYTHIA 6, with the CT10 PDF set, $h_{damp} = m_{top}$[5] and PERUGGIA 2012 tune[6].

$t\bar{t}$ samples are available for several filters: nonallhad which selects processes with at least one lepton in the final state *i.e.* processes in which both t and $t\bar{t}$ decay leptonically and/or semileptonically, allhadronic where both t and $\bar{t}$ decay hadronically and dileptonwhere both decay leptonically. As this analyses is based on the signature with a hadronically decaying τ and a lepton, chosen is the sample with the nonallhad filter.

In single-top processes, a lepton filter is used.

A detailed list of used MC samples is given in Appendix A.2.3.

3.3.5 Diboson

The production of a WW, ZZ, $\gamma\gamma$, WZ, $W\gamma$ and $Z\gamma$ pair forms the diboson background. This is a smaller background contributing both to the fake and the real background. An example of a contribution to the real background is a WW pair decaying leptonically to a $(\tau,\nu_\tau)(e/\mu,\nu_e/\nu_\mu)$ pair with the τ-lepton and the e/μ mimicking the signal signature. A fake background can be formed if one of the W-bosons decays to a $(e/\mu,\nu_e)$ pair while another one decays hadronically. As a jet can fake the τ-lepton, this is another source of the background.

Feynman diagrams for the leading order production are shown in Figure 3.8. The s-channel diagram includes a triple gauge coupling (TGC) which is allowed between the WWZ and $WW\gamma$, while other couplings like ZZZ, $ZZ\gamma$ are not allowed in the Standard Model. Therefore WW, WZ and $W\gamma$ pairs can be produced in both the s and the t-channel, while the $Z\gamma$ and $\gamma\gamma$ pairs can be produced only in the t-channel.

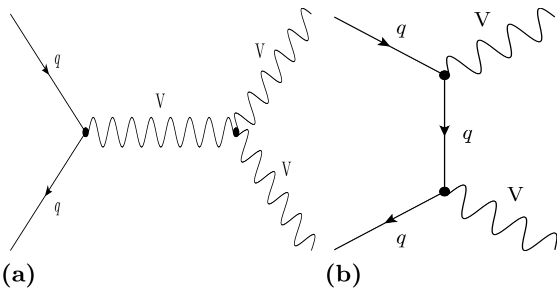


Figure 3.8: Examples of leading order Feynman diagrams for diboson production in the s-channel with the TGC vertex (a) and the t-channel(b). Allowed TGC vertices in SM are WWZ and $WW\gamma$.

Diboson processes were simulated using SHERPA with the CT10 PDF set. Each process and each decay type were simulated separately. A list of used MC samples is given in Appendix A.2.4.

3.3.6 QCD

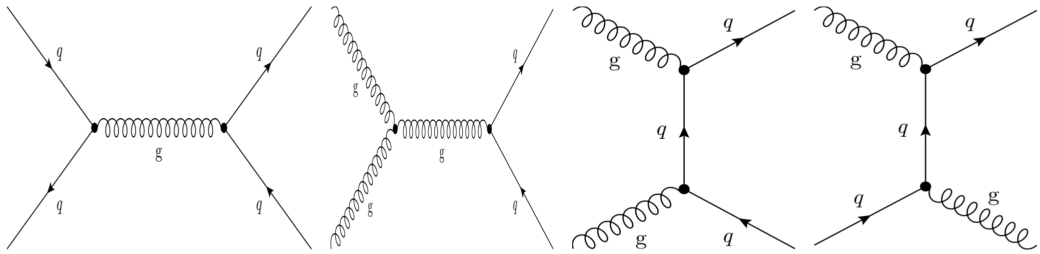
QCD background includes the processes where two (dijets) or more jets (multijets) are produced and it forms one of the largest backgrounds at the LHC. In this analysis, QCD events forms the background if both the e/μ and the τ are faked by a jet or if the τ is faked by a

[5] h_{damp} is a POWHEG parameter which controls the parton shower scale.

[6] Perrugia 2012 sis a set of paramets used for MC generation[33]

jet and a e/μ produced in the QCD event is defined as the signal e/μ . In both cases, the background is fake and can be reduced by good identification and isolation of particles.

Examples of leading order diagrams for a production of two jets are shown in Figure 3.9. Multijet production proceeds in a similar way, with additional gluon emissions.



(a) **(b)** **(c)** **(d)**

Figure 3.9: Examples of leading order Feynman diagrams for dijet production in the s-channel via q-$\bar{q}$ annihilation (a) triple gluon interaction (b) and in the t-channel via gluon (c) and q-$\bar{q}$ scattering (d).

As QCD processes have very large cross sections and many ways of being formed, it is hard to produce enough Monte Carlo statistics to match the data. Dijet Monte Carlo samples were tested, but as the QCD statistics was very small after the initial selection, QCD Monte Carlo samples were not used. QCD was modeled using data driven methods instead.

3.3.7 Higgs

Although Higgs production is a very rare process, it is included in the background processes. The main channels for the Higgs production at the LHC are gluon and vector boson fusion. Those processes are shown in Figure 3.10. Gluon fusion $gg \to H$ is the dominant process for the Higgs production. It proceeds as a gluon interaction, but as the Higgs bosons do not couple to gluons, but do couple to quarks, the lowest order process involves a quark loop. As the t-quark is the heaviest quark, the main contribution is from a t-quark.The second most important process is vector boson fusion in which a ZZ or WW pair couples to the Higgs boson.

Higgs processes were simulated using PYTHIA 8 and POWHEG, with the AZNLOCTEQ6L1 tune and CT10 PDF set. The samples are separately simulated for different production modes and different decay channels. A list of used MC samples is given in Appendix A.2.5.

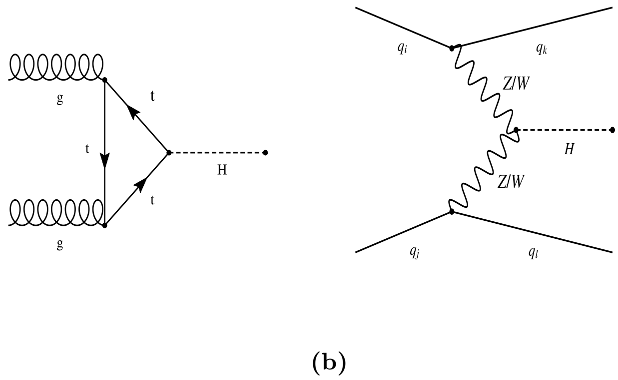


Figure 3.10: The dominant channels for the Higgs production at the LHC: gluon fusion (a) and vector boson fusion (b).

Event Selection

This chapter describes the the event cleaning procedure in Section 4.1. Further, the reconstruction and the definition of physics objects are described in Section 4.2 while the overlap removal procedure is described in Section 4.3.

4.1 Event Cleaning

The event cleaning starts with the data quality checks which include the Good Run List (GRL) requirement and the detector cleaning which includes the additional checks of the state of the subdetectors and a rejection of the non-collision background.

The luminosity blocks during which all subdetectors worked properly are listed in a Good run list (GRL). As the data used in physics analysis is required to pass the GRL, its luminosity is calculated by adding the luminosity blocks during which the full detector operated properly and by taking into account the prescales of the triggers required in the analysis. Furthermore, as the GRL does not include certain subdetector defects like noise bursts in the calorimeter, those events are additionaly removed and the luminosity is corrected to take into account only the events without the additional subdetector defects. The final combined 2015+2016 dataset used in the analysis correspond to an integrated luminosity of $36.1\,\mathrm{fb}^{-1}$ where $32.9\,\mathrm{fb}^{-1}$ corresponds to the 2016 dataset while $3.2\,\mathrm{fb}^{-1}$ is from the 2015 dataset.

GRL requires that the event was taken during a good luminosity block. Since GRL does not include certain subdetector defects like calorimeter noise bursts or data corruption, an event is additionally checked for LAr, Tile, SCT or Core defects and rejected if any of those is found. Additionaly, the events originating from a non-collision background as are the collisions of the residual beam gas or the cosmic background are also rejected.

Both the data and the MC events are required to contain a primary vertex with at least two associated tracks. Furthermore, it is required that at least one of the lowest unprescaled single-electron/muon triggers as defined in Table 4.1 fired and that the object which fired the trigger is matched to a reconstructed object of the same type and used further in the analysis.

The event cleaning efficiency is shown in Table 4.2.

Table 4.1: The list of lowest unprescaled single-electron/muon triggers used in the analysis. The HLT trigger name is typically formed by the name of the object $(e/\mu), followed by the p_{\mathrm{T}}$ $cut(e.g.e24means p_{\mathrm{T}}$ >24 GeV$and the identification requirement(e.g. lhmedium =$ $likelihood medium). In addition,$ HLT_e24_lhmedium_L1EM20VH$indicates an unusual requirement on the e$

Year	Electron triggers	Muon triggers
2015	HLT_e24_lhmedium_L1EM20VH HLT_e60_lhmedium HLT_e120_lhloose	HLT_mu20_iloose_L1MU15 HLT_mu40
2016	HLT_e26_lhtight_nod0_ivarloose HLT_e60_lhmedium_nod0 HLT_e140_lhloose_nod0	HLT_mu26_ivarmedium HLT_mu50

Table 4.2: The event cleaning efficiency for MC and data events. The table shows efficiencies with respect to the number of processed events. The GRL and Detector cleaning are applied to data only, while other selections are applied to both the data and the MC.

Selection	$Z\to\tau e$[%]	$Z\to\tau\mu$[%]	$Z\to\ell\ell$[%]	$Z\to\tau\tau$[%]	W+jets[%]	Top[%]	Other[%]	Data[%]
Processed	100	100	100	100	100	100	100	100
GRL	100	100	100	100	100	100	100	98
Detector cleaning	100	100	100	100	100	100	100	97
Primary vertex check	100	100	100	100	100	100	100	97
Trigger	82	77	87	66	77	72	69	54

4.2 Reconstruction and Definition of Physics Objects

The objects used in the analysis are defined by following the official recommendations of the relevant ATLAS Combined Performance (CP) groups. A summary of the applied object definitions is given in Table 4.3.

4.2.1 Electrons

Electron reconstruction [34] starts with a search for a cluster in the electromagnetic calorimeter (ECAL), using the sliding-window algorithm. The $\eta-\phi$ plane is scanned to find the candidate electron above a certain energy threshold in a predefined window size. After the clusters (candidate electrons) have been formed, the reconstructed tracks from the inner detector are extrapolated to the clusters and the matching of the tracks and the clusters is attempted. If a tracks and a cluster are matched, the object is tagged as a reconstructed electron.

The purity of prompt electrons is low at this stage since there is a large background, mainly from the misreconstructed pions and kaons.

To select the prompt electrons while suppressing the backgrounds, the electron identification criteria are used. The standard discrimination is based on the likelihood-ratio test where the likelihoods for the signal and the background hypothesis are computed using several cluster and track variables. The cut on the discriminant is optimized to provide the desired signal efficiency which further defines the working point.

Several identification working points are provided, each corresponding to a certain signal efficiency: Loose, Medium and Tight. The loose working point provides the highest signal efficiency but the lowest purity and the lowest background rejection, while the tight working point provides the lowest signal efficiency but the highest purity and the highest background rejection.

For further background rejection, to single out electrons without other surrounding activity, the isolation variable is used to measure the activity around the electron candidate. The track-based isolation is defined as the sum of the transverse momenta of all tracks in a cone around the electron candidate excluding its own transverse momenta while the calorimeter-based isolation is defined as the sum of the calorimeter clusters in a cone around the electron candidate, having subtracted the electron energy. Provided are looseTrackOnly, loose and tight working points and the gradient and gradientLoose with the E_T dependent selections.

Each working point comes with an efficiency scale factor SF which accounts for the difference between the MC and the data:

$$\mathrm{SF} = \frac{\epsilon^{\mathrm{data}}}{\epsilon^{\mathrm{MC}}} \tag{4.1}$$

where ϵ^{data} is the efficiency mesured in data, while ϵ^{MC} is the efficiency measured in MC for specific working point.

The total electron scale factor is defined as the combined scale factor for the reconstruction, identification, isolation and the trigger:

$$\mathrm{SF}(e)^{\mathrm{tot}} = \mathrm{SF}(e)^{\mathrm{rec}}\mathrm{SF}(e)^{\mathrm{id}}\mathrm{SF}(e)^{\mathrm{iso}}\mathrm{SF}(e)^{\mathrm{trigger}} \tag{4.2}$$

Electrons selected for the analysis are required to pass the medium identification and the gradient isolation working point. They are further required to have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.47$ with the excluded $1.37 < |\eta| < 1.52$ crack region. The recommended selections on the transverse and the longitudinal impact parameter are applied in order to reject the cosmic background. The electron distributions are further corrected by applying the total scale factor defined in Equation 4.2.

4.2.2 Muons

Muons are independently reconstructed[35] in the inner detector (ID) and the muon spectrometer (MS). The track reconstruction in the muon spectrometer starts with a search for the hits inside the muon chambers (MDT) to form the segments. The hits from the segments are then used to form the track and they are combined with the inner detector hits. Several muon types are defined depending on this combination: combined muons for which the reconstruction is done independently in MS and ID to form two separate tracks which if matched are refit to form the global track, segment-tagged muons for which the inner detector track is extrapolated to the muon spectrometer and it is checked if an associated track is found, calorimeter-tagged muons for which the inner detector track is extrapolated to the calorimeter and it is checked if it can be matched to a small energy deposit from the muon and extrapolated muons for which the tracks are reconstructed only in the muon spectrometer, but requiring that the track is originating from the primary vertex. Extrapolated muons are mainly used to extend the muon acceptance into the $2.5 < |\eta| < 2.7$ region.

Muon identification is done by applying a set of quality requirements aimed to reject the fake muons, mainly from pions and kaons. Four working points are provided: loose, medium, tight and high-pt. The default selection is the medium selection which is using combined tracks with the requirement of at least 3 hits in at least 2 MDT layers and to extend the acceptance outside of the ID coverage, the extrapolated tracks in the $2.5 < |\eta| < 2.7$ are used with the requirement to have hits in all 3 MDT layers and to originate from the primary vertex. The loose selection is applying loose requirements while considering all track types. It is providing the highest efficiency, but the lowest purity while the tight selection is applying the stringest selections which results in the highest purity but the lowest efficiency. The high-pt selection is specially optimized for muons with $p_T > 100$GeV and is mainly used in searches for very massive particles.

Isolation is a good discriminant between the prompt and the non-prompt muons, mainly the muons originating from semileptonic decays of heavy mesons. The isolation working points are defined based on the combined or selective cut on the calorimeter and track-based isolation variables. Several working points are provided, including fixed-cut and the gradient working point with the p_T dependent cuts.

The total muon scale factor is defined as the combined scale factor for reconstruction, isolation and the trigger:

$$\mathrm{SF}(\mu)^{\mathrm{tot}} = \mathrm{SF}(\mu)^{\mathrm{reco}}\mathrm{SF}(\mu)^{\mathrm{iso}}\mathrm{SF}(\mu)^{\mathrm{trigger}} \tag{4.3}$$

The muons selected for the analysis are required to pass the mediumidentification and the gradient isolation working point .They are further required to have $p_T > 30\,\mathrm{GeV}$, $|\eta| < 2.5$ and to pass the recommended selection on the impact parameters aimed to reject the muons from the cosmic background. The muon distributions are further corrected by applying the total scale factor defined in Equation 4.3.

4.2.3 Taus

Leptonic tau decay products can not be distinguished from prompt electrons or muons and will not be considered.

Reconstruction of hadronic taus [36] starts from jets, formed using the Anti-k(t) [37] algorithm from the locally calibrated topo-clusters[1]. Further, the inner detector tracks formed by the charged pions are associated to the jets and the secondary vertex is reconstructed. The tau candidate is then calibrated to the tau energy scale (TES) by applying the corrections which bring the reconstructed to the real visible tau four-momentum. Depending on whether one or three tracks are associated to the tau, the taus are classified as 1-prong and 3-prong.

The signature of a hadronic tau can be faked by QCD jets, electrons and muons. To discriminate between the taus and the QCD jets, the Boosted Decision Tree (BDT) algorithm is used. Three working points are provided: loose, medium and tight, from the lowest to the highest signal purity. To suppress the electrons and the muons, the taus which are overlapping with an electron or a muon which pass the basic quality requirements are vetoed.

Taus selected for the analysis are required to be charged, have one or three associated tracks and to pass the $p_T > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$ selection, excluding the crack region. Additionally, for QCD jet suppression, the taus are required to pass the tight BDT working point. For the electron and muon suppression, vetoes are applied, but as the $e \rightarrow \tau$ fakes formed a significant background, even after the electron veto, applied is a dedicated electron-tau BDT at a working point corresponding to 85% signal efficiency for 1-prong and 95% signal efficiency is for 3-prong events.

[1]**Topocluster is a 3D topological cluster of neighboring calorimeter cells**.

4.2.4 Jets

Different types of jets are reconstructed from various inputs, using the Anti-k(t) [37] algorithm. The inputs can be the energy deposits in the hadronic and the electromagnetic calorimeter (calorimeter jets), the tracks reconstructed in the inner detector (track jets). Since the track jets are formed for charged objects only, the standardly used jets in ATLAS are calorimeter jets.

The reconstruction of calorimeter jets starts with the formation of topological clusters (topo-clusters) [38], clusters of neighboring calorimeter cells with a high signal-to-noise ratio. The topo-clusters are then used as the inputs to the Anti-k(t) algorithm, which further forms the jets based on the distances between the topo-cluster objects. The jet algorithm can form EM or LC jets, depending on whether the inputs are uncalibrated or locally calibrated (LC) topo-clusters.

The jet calibration is done to correct the energy of the reconstructed jet to the Jet Energy Scale (JES) *i.e.* to the energy deposited by the real jet. It is important to apply the corrections to the reconstructed energy so that the objects with correct energies are used for further analysis. The differences arise due to the different effects as are overlays of pile-up energy deposits and the various detector effects as are energy loss due to the interactions with the material, the dead calorimeter cells and the detector noise. The corrections are derived from the MC simulation by comparing the reconstructed jets with the truth jets and are applied to both the data and the MC samples. In the last step, applied are the corrections due to the different detector response between the data and the MC simulation. The jet calibration procedure is commonly applied at the topo-cluster level and the corrected topo-clusters are then used to form the calibrated jets.

As the pile-up jets [39] form the background to the jets from the hard process, it is important to discriminate between the two. Commonly used discriminators are the jet vertex fraction (JVF) and Jet Vertex Tagger (JVT) [40]. To construct the JVF, a narrow cone around the jet axis is formed and the JVF is defined as the ratio of the transverse momenta of all the tracks inside a cone which are originating from the primary vertex and all the tracks inside the cone:

$$\mathrm{JVF} = \frac{\sum_{\text{tracks inside the cone from PV}} \bar{p}_T}{\sum_{\text{all tracks inside the cone}} \bar{p}_T} \tag{4.4}$$

JVT is a newer multivariate tagger based on JVF and additional calorimeter and track variables.

The identification of jets containing b-hadrons is exploiting the measurements of the secondary vertex *i.e.* the vertex of the b-meson decay and the track impact parameters. This analysis is using a BDT-based MV2c10 tagger[41]

Jets selected for the analysis are formed from the LC calibrated topo-clusters which are required to pass the $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.5$ and the medium(JVT) working point, to suppress the pile-up jets. The b-jets are identified using the MV2c10 b-tagger at the 77% b-jet efficiency working point.

4.2.5 Missing transverse energy

Missing transverse energy [42] is reconstructed as the negative sum of transverse momenta of all selected and calibrated objects

$$\vec{E}_{\mathrm{T}}^{\mathrm{miss}} = -(\vec{p}_T(e) + \vec{p}_T(\mu) + \vec{p}_T(\tau) + \vec{p}_T(\mathrm{jet}) + \vec{p}_T(\gamma) + \vec{p}_T(\mathrm{soft})) \tag{4.5}$$

The energy deposits which are not associated with any of the defined objects are taken into account by including the soft term which can be computed from the sum of the tracks (track soft term) or the sum of the calorimeter energy deposits (calo soft term) which are not assigned to any of the objects. The calo soft term is more sensitive to pile-up, while the track soft term does not include the neutral particles. The standard approach is to use to track soft term, combined with the calo soft term to account for the neutral particles.

Table 4.3: An overview of the object definitions used in the analysis.

Object	**Requirements**
Electrons	Medium LLH $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.47$, excl. $1.37 < \lvert\eta\rvert < 1.52$ Gradient isolation $\lvert d_0/\sigma(d_0)\rvert < 5.0$ $\lvert z_0 \sin\theta\rvert < 0.5$
Muons	Medium quality $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.5$ Gradient isolation $\lvert d_0/\sigma(d_0)\rvert < 3.0$ $\lvert z_0 \sin\theta\rvert < 0.5$
Taus	Tight BDT $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.5$, excl. $1.37 < \lvert\eta\rvert < 1.52$ $\lvert q\rvert = 1$ $N_{\mathrm{tracks}} = 1, 3$ `EleOLR` and `MuonOLR` veto 85% WP electron veto BDT for 1p taus 95% WP electron veto BDT for 3p taus
Jets	$p_{\mathrm{T}} > 20\,\mathrm{GeV}$ $\lvert\eta\rvert < 2.5$ Medium JVT WP
b-jet tagging	MV2c10 FixedCutBEff_77

4.3 Overlap Removal

Physics objects which are reconstructed at the same spatial position in the detector are referred as the overlapping objects. The criteria for the overlap removal procedure is based on the removal of multiple objects reconstructed within a cone of radious $\Delta R \leq 0.2$ where ΔR is defined as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. Based on the recommendations of the ATLAS harmonization group [43], the overlap removal is performed in the following order:

- If a baseline tau and a baseline electron are found within $\Delta R < 0.2$, the 'tau' is ignored.
- If a baseline tau and a baseline muon are found within $\Delta R < 0.2$, the 'tau' is ignored.

- If a baseline electron and a baseline muon share the same ID track, the electron is ignored.
- If a baseline electron and a jet are found within $\Delta R < 0.2$ and a jet is not b-tagged with MV2c10 85% efficiency working point, the object is interpreted as an electron and the overlapping jet is ignored.
- If a baseline electron and a jet are found within $\Delta R < 0.4$ and the jet is not flagged as pileup jet ($p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$ and JVT < 0.59), the object is interpreted as a jet and the nearby electron is ignored.
- If a baseline muon and a jet are ghost-associated or found within $\Delta R < 0.2$, the object is treated as a muon and the overlapping jet is ignored if the jet and the muon satisfy either of the following criteria:
  - the number of tracks with $p_{\mathrm{T}} > 500\,\mathrm{MeV}$ that are associated to the jet is less than three
  - the jet is not b-tagged with MV2c10 85% efficiency working point
- If a baseline muon and a jet are found within $\Delta R < \min(0.4, 0.04 + 10\,\mathrm{GeV}/p_{\mathrm{T}}^{\mathrm{muon}}$, the object is treated as a jet and the overlapping muon is removed if the jet is not flagged as pileup jet.
- If a baseline tau and a jet are found within $\Delta R < 0.2$, the 'tau' is ignored in case the jet is b-tagged with MV2c10 85% efficiency working point; otherwise, the jet is ignored.

Discriminating Variables

This chapter describes the variables used in the discrimination between the signal and the backgrounds processes. Section 5.1 describes the construction of the $\Delta\alpha$ variable built to suppress the $Z \to \tau\tau$ background, by exploiting the kinematical differences between the signal process and this background. The following Section 5.2 describes the reconstruction of the invariant mass in the presence of neutrinos. As neutrinos escape detection, the calculation cannot be done exactly and several different approaches including the collinear approximation and the likelihood-based MMC method are commonly used and briefly explained, with the focus on the collinear approach used in the analysis. Section 5.3 describes other kinematical variables used either in the definitions of the signal or control regions or as the input neural networks. Such regions and the neural network approach are described in the following chapters.

5.1 Kinematic discriminant $\Delta\alpha$

It is possible to construct a kinematical variable to suppress the large irreducible $Z \to \tau\tau$ background, by exploiting the difference in the number of neutrinos in the signal and the background process which further affectis the energies of the visible particles. The method is derived in [44] and briefly explained here.

The method is assuming two approximations, namely that the τ-lepton decay products are collinear due to the high τ-lepton boost and that the transverse momentum of the Z-boson is negligible which holds for majority of events.

For the **signal** process, in the collinear approximation, the four-momentum of the τ_{had}-e/μ system can be expressed in terms of the visible component of the decay as:

$$p_{\tau_{\text{had}}} = p_{\tau_{\text{had-vis}}} + p_{\nu_\tau} \equiv \alpha p_{\tau_{\text{had-vis}}}. \tag{5.1}$$

where the $p_{\tau_{\text{had-vis}}}$ is the four-momentum of the τ-jet (the visible part of the hadronic τ-lepton decay), the p_{ν_τ} is the four-momentum of the ν_τ while α is showing the ratio between the four-momentum of the τ-lepton and the visible part of its decay.

From the energy-momentum conservation, It holds that:

$$p_Z = p_{\tau_{\text{had}}} + p_{e/\mu}. \tag{5.2}$$

By taking into account the collinear approximation $\tau_{\text{had}} \equiv \alpha p_{\tau_{\text{had-vis}}}$ and by neglecting the mass of the e/μ-lepton, it can be obtained:

$$m_Z^2 - m_{\tau_{\text{had}}}^2 = 2p(\tau_{\text{had}})p(e/\mu) = 2\alpha p(\tau_{\text{had-vis}})p(e/\mu).,\tag{5.3}$$

and α can be expressed as:

$$\alpha = \frac{m_Z^2 - m(\tau_{\text{had}})^2}{2p(\tau_{\text{had-vis}})p(e/\mu)} \tag{5.4}$$

The second assumption neglects the transverse momentum of the Z-boson and gives the following in the collinear case $\tau_{\text{had}} \equiv \alpha p_{\tau_{\text{had-vis}}}$:

$$|p_{\text{T}}(e/\mu)| = |p_{\text{T}}(\tau_{\text{had}})| \tag{5.5}$$
$$|p_{\text{T}}(e/\mu)| = \alpha\,|p_{\text{T}}(\tau_{\text{had-vis}})|\,, \tag{5.6}$$

and α can be expressed as:

$$\alpha = \left|\frac{p_{\text{T}}(e/\mu)}{p_{\text{T}}(\tau_{\text{had-vis}})}\right| \tag{5.7}$$

By subracting the relations defined in Equations 5.7 and 5.4, the $\Delta\alpha$ can be expressed as:

$$\Delta\alpha = \left|\frac{p_{\text{T}}(e/\mu)}{p_{\text{T}}(\tau_{\text{had-vis}})}\right| - \frac{m_Z^2 - m(\tau_{\text{had}})^2}{2p(\tau_{\text{had-vis}})p(e/\mu)} = \alpha - \alpha = 0 \tag{5.8}$$

By construction, the $\Delta\alpha$ peaks at zero for the signal process.

For the $\boldsymbol{Z \to \tau\tau}$, due to another neutrino from the second τ-lepton decay, the distribution of the electron or a muon will be softer than in the signal process and this difference is exploited in the $\Delta\alpha$ construction. In the collinear approximation, each τ-lepton can be expressed in terms of the visible component as:

$$p_{\tau_{\text{had}}} = p_{\tau_{\text{had-vis}}} + p_{\nu_\tau} \equiv \alpha p_{\tau_{\text{had-vis}}}. \tag{5.9}$$
$$p_{\tau_{\text{lep}}} = p_{e/\mu} + p_{\nu_\tau} + p_{\nu_{e/\mu}} \equiv \beta p_{e/\mu}. \tag{5.10}$$

where β is showing the ratio of the four-momentum of the τ-lepton and the electron or a muon, depending on the channel.

From the energy-momentum conservation, it holds that:

$$p_Z = p_{\tau_{\text{had}}} + p_{\tau_{\text{lep}}} \tag{5.11}$$

which in the collinear approximation gives the following:

$$m_Z^2 - 2m_{\tau_{\text{had}}}^2 = 2p_{\tau_{\text{had}}}p_{\tau_{\text{lep}}} = 2\alpha\beta p_{\tau_{\text{had-vis}}}p_{e/\mu}. \tag{5.12}$$

By using the assumption that the transverse momentum of the Z-boson is zero, it can be written:

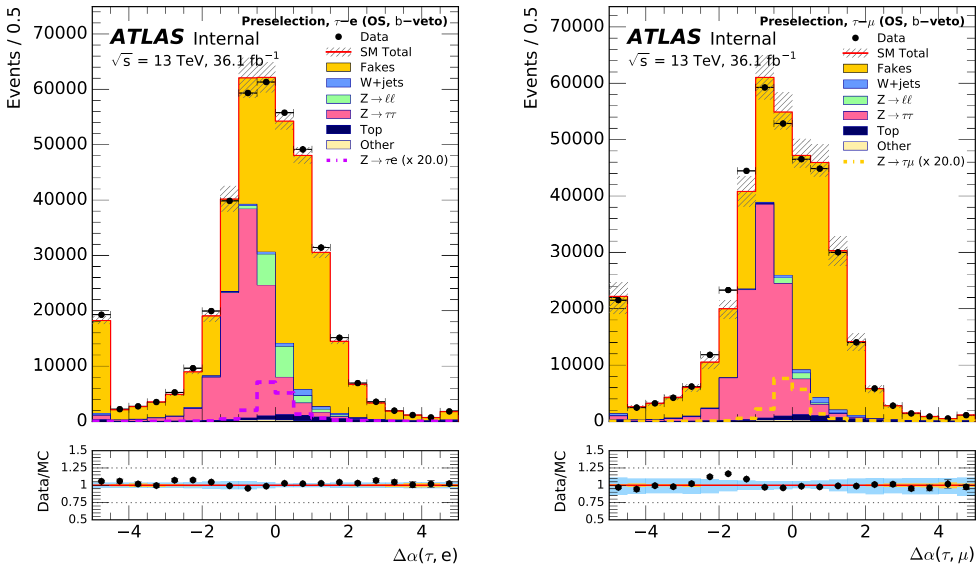
$$|p_{\text{T}}(\tau_{\text{had}})| = \left|p_{\text{T}}(\tau_{\text{lep}})\right| \tag{5.13}$$
$$\alpha\left|p_{\tau_{\text{had-vis}}}\right| = \beta\,|p_{\text{T}}(e/\mu)| \tag{5.14}$$

Combining the Equations 5.12 and 5.14 and using the same definition of $\Delta\alpha$ as in Equation 5.8:

$$\Delta\alpha = \left|\frac{p_{\text{T}}(e/\mu)}{p_{\text{T}}(\tau_{\text{had-vis}})}\right| - \frac{m_Z^2 - m(\tau_{\text{had}})^2}{2p(\tau_{\text{had-vis}})p(e/\mu)} \approx \alpha/\,\beta - \alpha\beta \tag{5.15}$$

Since both α and β are positive and larger than one, the $\Delta\alpha$ distribution for the $Z \to \tau\tau$ process peaks at a negative value and provides a good way of discrimination. The $\Delta\alpha$ distribution for the signal and the $Z \to \tau\tau$+jets, after the trigger, the requirement for one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets is shown in Figure 5.1[1]

The $\Delta\alpha$ variable is used as the input to the neural networks[2] trained to classify between the signal and the $Z \to \tau\tau$ processes. Apart from the $Z \to \tau\tau$, the variable has shown to give an improvement in the discrimination against other backgrounds and is used also as the input to neural networks trained to discriminate the signal from the $Z \to ee/\mu\mu$ and W+jets backgrounds.



(a) **(b)**

Figure 5.1: The $\Delta\alpha(\tau, \ell)$ distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20 to be visible. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. The underflow and the overflow are merged into the first and the last bin.

5.2 Invariant Mass Reconstruction

The invariant mass of the Z-boson in the signal sample with a hadronically decaying τ-lepton is defined as the invariant mass of the $\tau_{\text{had-vis}} + e/\mu + \nu_\tau$ system, where $\tau_{\text{had-vis}}$, e/μ and ν_τ are reconstructed objects. As the longitudinal component of the neutrino momentum is not reconstructed, the Z-boson invariant mass can not be directly calculated. In the presence of

[1] In this and the following Figures, the background model is already shown as determined for the analysis, while being described in Chapter 7. What is called Fakes is a data driven background for the processes where the tau hadronic decay signature is faked by a jet. The second main background is $Z \to \tau\tau$, derived from MC samples. The additional backgrounds (very minor) are background processes which contribute with a real lepton and a real hadronic tau decay, they are estimated via MC samples where the reconstructed leptons can be matched in truth to real leptons.

[2] Neural networks are described in Chapter 6.3

several neutrinos, the situation is even more complicated since the neutrinos momenta along the beam axis can cancel out each other.

Several approaches as the Missing mass calculator (MMC) [45] and collinear approximation [46] are commonly used to solve this problem. The MMC method is a likelihood based technique used to calculate the invariant mass in processes where a heavy resonance is decaying to $\tau\tau$ or $\tau\ell$ pairs [3]. The collinear approximation is a simpler method based upon the assumption that due to the high boost of the τ-lepton, the τ-neutrino and the visible τ decay products travel in the same direction.

In the collinear approximation, the τ-neutrino direction is assumed to be collinear with the $\tau_{\text{had-vis}}$ direction $\vec{p}(\tau_{\text{had-vis}})/\left|\vec{p}(\tau_{\text{had-vis}})\right|$. For processes with one τ-lepton, the neutrino energy $E(\nu_\tau)$ can be computed from either of the two equations:

$$E_{\mathrm{T_x}}^{\text{miss}} = E(\nu_\tau)\vec{e}_{\nu_\tau} \cdot \vec{e}_x = E(\nu_\tau)\frac{\vec{p}(\tau_{\text{had-vis}})}{\left|\vec{p}(\tau_{\text{had-vis}})\right|} \cdot \vec{e}_x = E(\nu_\tau)\frac{p_x(\tau_{\text{had-vis}})}{\left|\vec{p}(\tau_{\text{had-vis}})\right|} \tag{5.16}$$

$$E_{\mathrm{T_y}}^{\text{miss}} = E(\nu_\tau)\vec{e}_{\nu_\tau} \cdot \vec{e}_y = E(\nu_\tau)\frac{\vec{p}(\tau_{\text{had-vis}})}{\left|\vec{p}(\tau_{\text{had-vis}})\right|} \cdot \vec{e}_y = E(\nu_\tau)\frac{p_y(\tau_{\text{had-vis}})}{\left|\vec{p}(\tau_{\text{had-vis}})\right|} \tag{5.17}$$

where $p_x(\tau_{\text{had-vis}})$ and $p_y(\tau_{\text{had-vis}})$ are the momentum components and $\vec{p}(\tau_{\text{had-vis}})$ is the spatial momentum of the τ-jet. Once the neutrino energy is computed, the collinear mass is computed as:

$$m_{\text{coll}}^2 = (p(\tau_{\text{had-vis}}) + p(\nu_\tau) + p(e/\mu))^2 \tag{5.18}$$

In the processes where the two τ-leptons are produced, as are the $Z \to \tau\tau$ and $H \to \tau\tau$, the transverse components of the missing energy are computed as the sums of the energies of the two neutrinos:

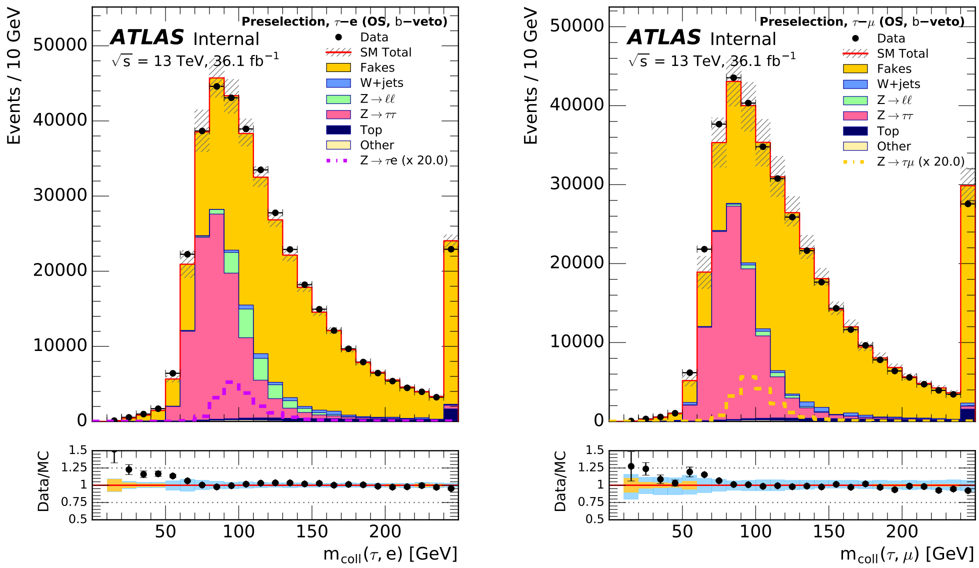
$$E_{\mathrm{T_x}}^{\text{miss}} = E(\nu_{\tau,1})\frac{p_x(\tau_{\text{had-vis},1})}{\left|\vec{p}(\tau_{\text{had-vis},1})\right|} + E(\nu_{\tau,2})\frac{p_x(\tau_{\text{had-vis},2})}{\left|\vec{p}(\tau_{\text{had-vis},2})\right|} \tag{5.19}$$

$$E_{\mathrm{T_y}}^{\text{miss}} = E(\nu_{\tau,1})\frac{p_y(\tau_{\text{had-vis},1})}{\left|\vec{p}(\tau_{\text{had-vis},1})\right|} + E(\nu_{\tau,2})\frac{p_y(\tau_{\text{had-vis},2})}{\left|\vec{p}(\tau_{\text{had-vis},2})\right|} \tag{5.20}$$

This system of equations does not have a solution if the two $\tau_{\text{had-vis}}$ obejcts are travelling back-to-back which causes a certain inefficiency of the method.

The collinear mass distribution is computed using the Equation 5.18 and is shown in Figure 5.2. The collinear mass is used as the input to the neural networks built to discriminate between the signal and the $Z \to ee/\mu\mu$, $Z \to \tau\tau$ and W+jets backgrounds.

[3] *MMC* was originally developed for the $H \to \tau\tau$ analysis. Subsuquentally, the LFV mode was added.



(a) **(b)**

Figure 5.2: The collinear mass distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20 to be visible. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. The underflow and the overflow are merged into the first and the last bin.

5.3 Other kinematical variables

The **visible mass** is defined as the invariant mass of the visible part of the τ-lepton decay and the e/μ-lepton:

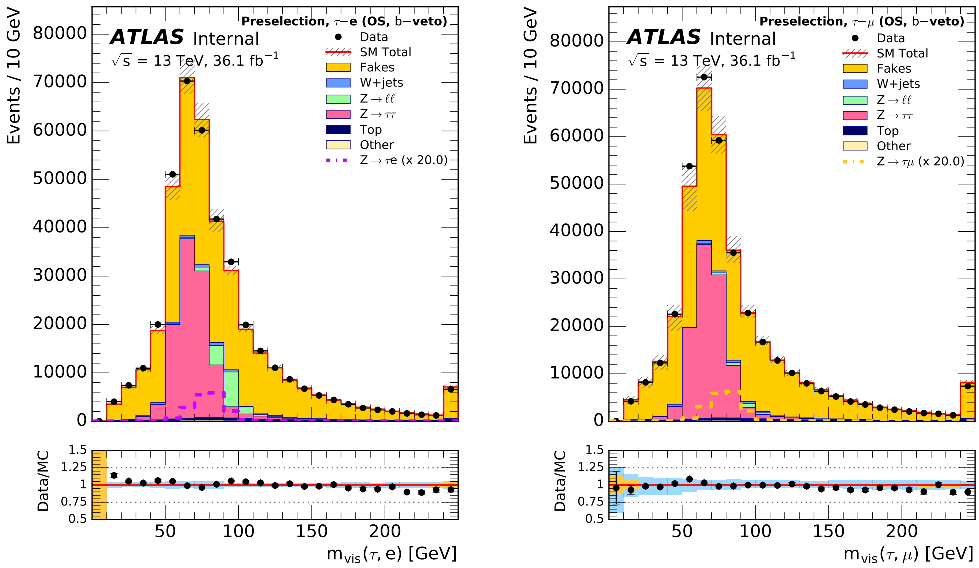
$$m_{\text{vis}}(\tau, e/\mu) = \sqrt{m^2_{\tau_{\text{had}}-vis} + 2p_{e/\mu}\sqrt{m^2_{\tau_{\text{had}}-vis} + p^2_{\tau_{\text{had}}-vis}} - 2p_{\tau_{\text{had}}-vis}p_{e/\mu}cos\Delta\Omega_{\tau_{\text{had}}-vis,e/\mu}} \tag{5.21}$$

The visible mass distribution is shown in Figure 5.3. It is used to define the signal region and as input to the the neural networks.

The **transverse mass** of the two particles p_1 and p_2 is defined as:

$$m_T(p_1, p_2) = \sqrt{2p_{\text{T}}(p_1)p_{\text{T}}(p_2)(1 - \cos(\Delta\phi(p_1, p_2))} \tag{5.22}$$

In the $Z \to \tau e/\mu$ process, due to the boost of the τ-lepton, the $\Delta\phi(\tau_{\text{had-vis}}, E_{\text{T}}^{\text{miss}})$ angle between the visible part and the neutrino from the τ-lepton decay process is very small and the transverse mass of the $\tau_{\text{had-vis}} - E_{\text{T}}^{\text{miss}}$ system is small. As the Z-boson is mainly produced with low p_{T}, the τ-lepton and the e/μ are mainly produced in the back-to-back configuration. Hence the angle between the e/μ and the neutrino from the τ-lepton decay is large and the transverse mass of the $e/\mu - E_{\text{T}}^{\text{miss}}$ system is large in the $Z \to \tau e/\mu$ process.



(a) **(b)**

Figure 5.3: The visible mass distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20 to be visible. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. The underflow and the overflow are merged into the first and the last bin.

In the $Z \to \tau\tau$ process, due to the presence of several neutrinos, the angle between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the τ-jet is larger, meaning that the transverse mass of the $\tau_{\text{had-vis}} - E_{\mathrm{T}}^{\mathrm{miss}}$ is larger while the transverse mass of the $e/\mu - E_{\mathrm{T}}^{\mathrm{miss}}$ system is smaller than in the signal sample. In the W+jets process, the angle between the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the e/μ is large and the $e/\mu - E_{\mathrm{T}}^{\mathrm{miss}}$ transverse mass is large, while the angle between the jet which fakes the τ-lepton and the $E_{\mathrm{T}}^{\mathrm{miss}}$ is not correlated. Therefore the use of $m_{\mathrm{T}}(e/\mu, E_{\mathrm{T}}^{\mathrm{miss}})$ and $m_{\mathrm{T}}(\tau, E_{\mathrm{T}}^{\mathrm{miss}})$ can be used for the definition of regions rich in signal or Z/W + jets backgrounds.

The distribution of the transverse mass of the $\tau_{\text{had-vis}} - E_{\mathrm{T}}^{\mathrm{miss}}$ versus the transverse mass of the $e/\mu - E_{\mathrm{T}}^{\mathrm{miss}}$ system for the signal and the MC backgrounds is shown in Figure 5.4 for the τ-e channel and in Figure 5.5 for the τ-μ channel.

The **invariant mass of the (τ track, e/μ) pair** is defined as the invariant mass of the leading track from the τ-jet and the e/μ-lepton:

$$m(\tau_{\mathrm{trk}}, e/\mu) = \sqrt{2p_{e/\mu}\sqrt{p_{\mathrm{trk}}(\tau)^2} - 2p_{\mathrm{trk}}(\tau)p_{e/\mu}cos\Delta\Omega_{\mathrm{trk}}(\tau), e/\mu} \tag{5.23}$$

The distribution is showed in Figure 5.6. It is used to further suppress the $Z \to ee/\mu\mu$ backgrounds which occur in 1-prong events. It is mainly targeted at $Z \to \mu\mu$ which is a non negligible reducible background in the analysis.

In most of the $Z \to ee/\mu\mu$ where an electron/muon fakes the 1-prong tau, the momentum of the track associated to the 1-prong tau candidate corresponds to the original momentum of the electron/muon while the energy deposited in the calorimeter measures the energy of electron/muon originating mainly from photon radiation. Therefore, events in which the $m(\tau_{\mathrm{trk}}, e/\mu)$ is compatible with the Z boson mass are rejected. In particular, events with a 1-prong tau candidate are accepted when $m(\tau_{\mathrm{trk}}, e/\mu)$ <84 GeV or $m(\tau_{\mathrm{trk}}, e/\mu)$ >105 GeV if $|\eta(\tau)| < 2$ and when $m(\tau_{\mathrm{trk}}, e/\mu)$ <80 GeV or $m(\tau_{\mathrm{trk}}, e/\mu)$ >105 GeV if $|\eta(\tau)| > 2$

A wider range in $m(\tau\mathrm{trk}(\tau), e/\mu)$ is rejected at high $|\eta(\tau)|$ because of the smaller signal contribution and the higher $Z \to ee/\mu\mu$ background rate. Additionally, events in which the invariant mass of the 1-prong τ candidate and thee/μ is in 80 GeV < $m(\tau\mathrm{trk}(\tau), e/\mu)$ <100 GeV are required to have $m(\tau\mathrm{trk}(\tau), e/\mu)$ >40 GeV.

Additional discriminating variables are given in Appendix B.

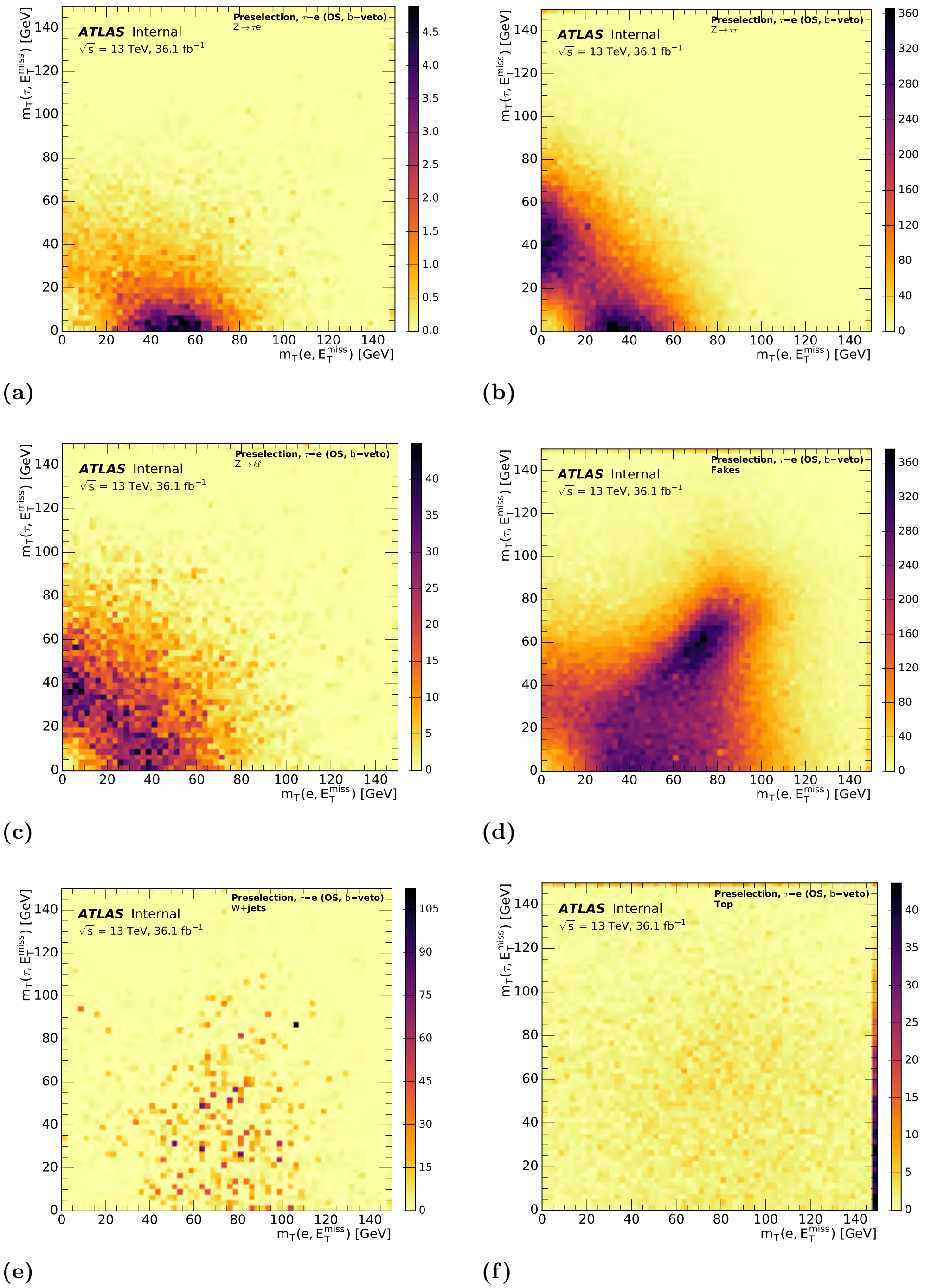


Figure 5.4: The transverse mass distribution of the ($\tau_{\text{had-vis}}$,$E_{\text{T}}^{\text{miss}}$) versus the (e,$E_{\text{T}}^{\text{miss}}$) system in the τ-e channel for the: (a) $Z \to \tau e$, (b) $Z \to \tau\tau$+jets, (c) $Z \to ee/\mu\mu$, (d) Fakes, (e) W+jets and (f) top after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20 to be visible. The underflow and the overflow are merged into the first and the last bin. The color axis shows the number of events per bin.

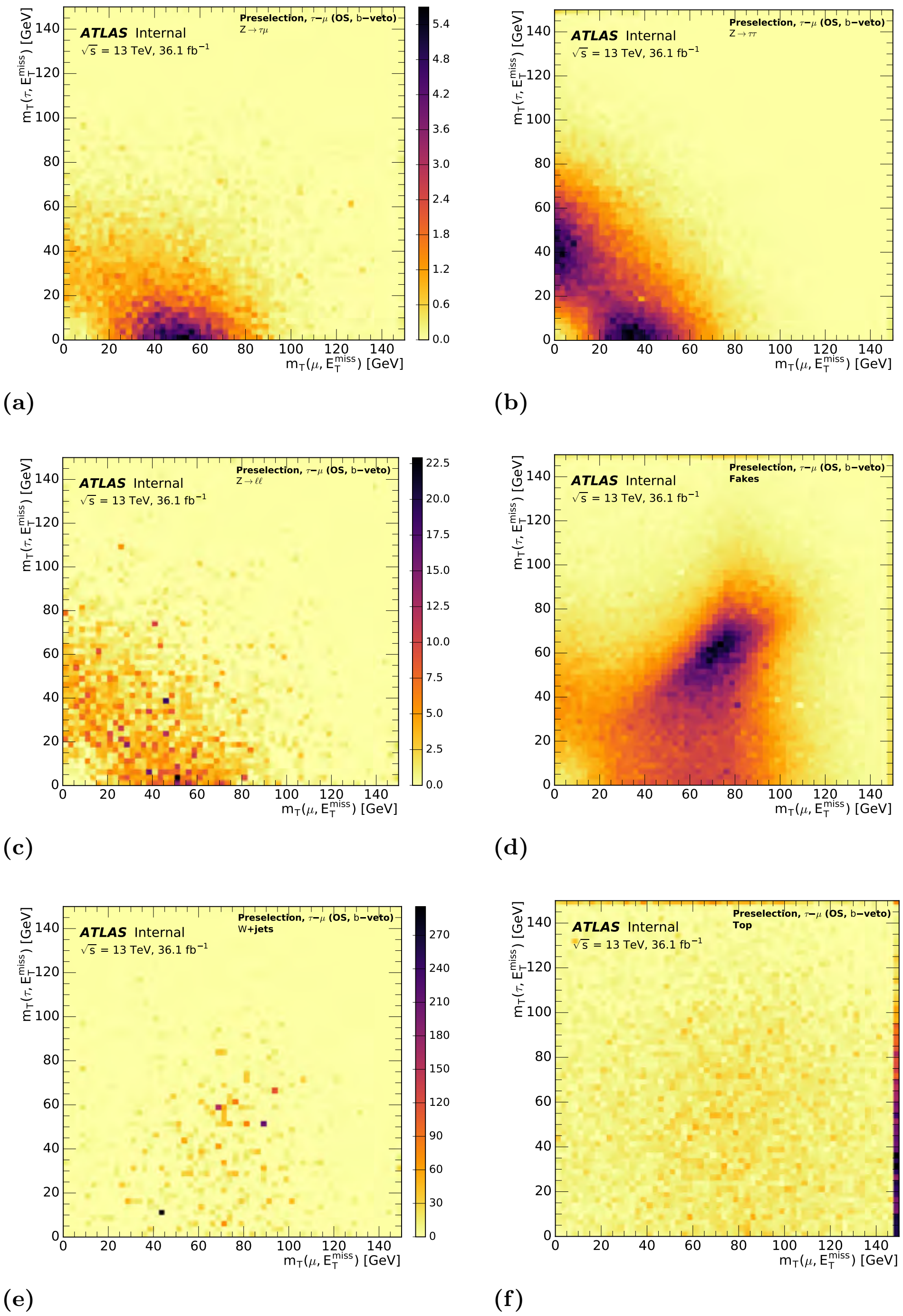


Figure 5.5: The transverse mass distribution of the ($\tau_{\text{had-vis}}$,$E_{\text{T}}^{\text{miss}}$) versus the (μ,$E_{\text{T}}^{\text{miss}}$) system in the τ-μ channel for the: (a) $Z \rightarrow \tau\mu$, (b) $Z \rightarrow \tau\tau$+jets, (c) $Z \rightarrow ee/\mu\mu$, (d) Fakes, (e) W+jets and (f) top after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20 to be visible. The underflow and the overflow are merged into the first and the last bin. The color axis shows the number of events per bin.

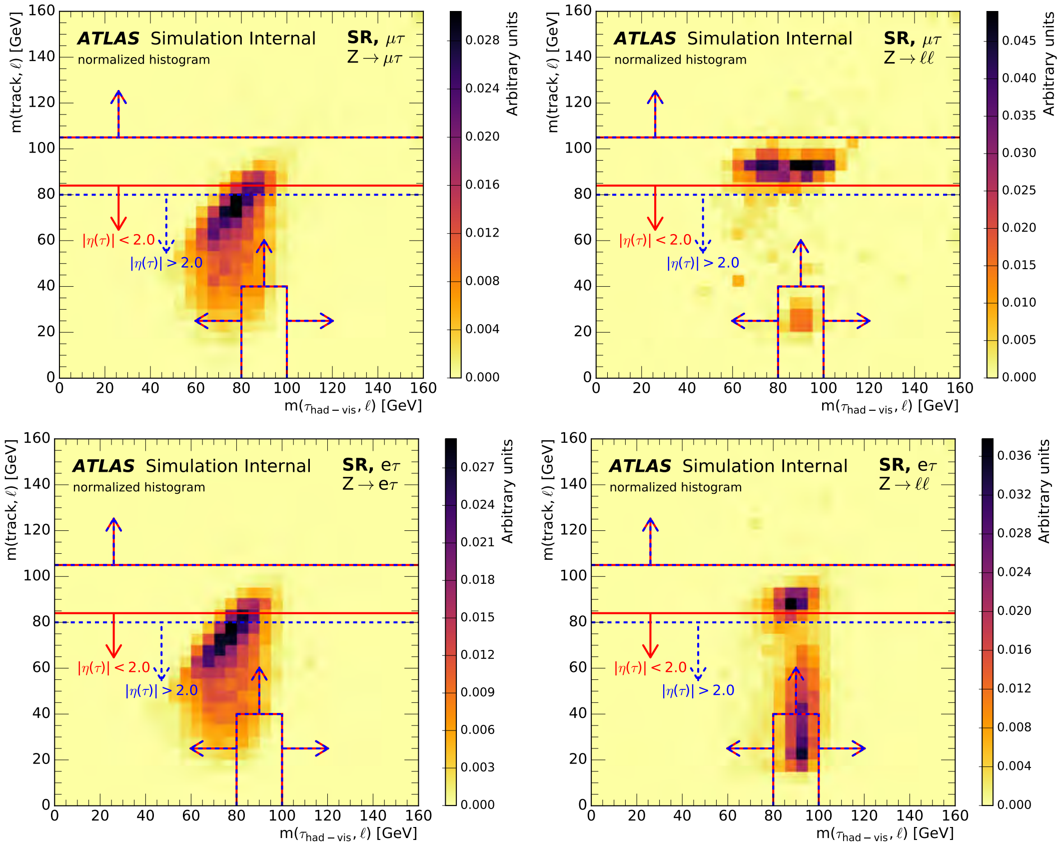


Figure 5.6: MC distributions of $m(\tau_{\mathrm{trk}}, e/\mu)$ versus $m_{\mathrm{vis}}(\tau, e/\mu)$ in signal (left) and $Z \to \ell\ell$ (right) events with 1-prong τ candidate in the $\tau - \mu$ (top) and $\tau - e$ (bottom) channel after the signal region selection except for the cuts on those two variables (shown in Tables 6.4 and 6.5) and is used to motivate the selection on these two variables. Shown $Z \to \ell\ell$ is combined $Z \to ee$ and $Z \to \mu\mu$ background. The histogram is normalized to 1.

Signal Region Optimization

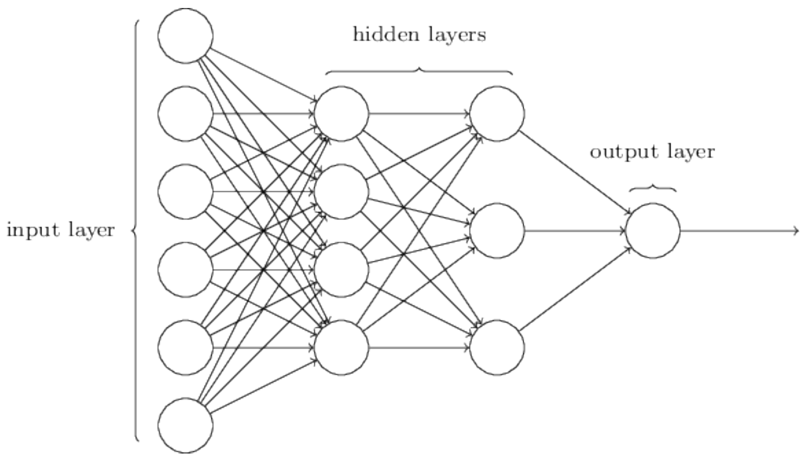
This analysis uses neural network approach to constructs a neural network score variable, which by exploiting the differences in the kinematics between the signal and the background processes helps in distinguishing between them. The neural network approach is described in Section 6.1 while the training, the model and the results are given in Section 6.2. Further, the neural network score is used as the final discriminating variable in the fit, which is done in the signal region *i.e.* a signal-enriched region of phase space defined to enhance the signal, while suppressing the backgrounds. The definition of the signal region and the event yields for the signal and the background processes are given in Section 6.3.

6.1 Neural Network Approach

Neural networks [47] (NN) are a set of computational models used in machine learning for classification and discrimination between different patterns. Several different architectures are used, but the basic architecture contains an input layer used to store the values of the inputs and an output layer which computes the output score as a combination of the scores from the previous layer.

In addition to the input and the output layer, many neural networks have one or many hidden layers which compute the score in the same way as the output layer. A simple neural network architecture with two hidden layers is shown in Figure 6.1.

Figure 6.1: A scheme of neural network architecture. The first layer is called the input layer. It is followed by two hidden layers and the output layer. Neurons from different layes are connected to each of the neurons from the previous and the next layer.



The smallest processing unit in a neural network is called a neuron [1]. Neurons are organized into multiple layers and each neuron in a layer is connected to all neurons in the previous and the next layer and the strength of the connection being parametrized by a weight. The score s of a neuron is calculated as the combination of the scores of all neurons in the previous layer:

$$s_j = A(\sum_i w_{ji} \cdot s_i) \tag{6.1}$$

where s_i is the score of a neuron in the previous layer, w_{ji} is the weight of the connection between the j and i neurons and sA is the activation function which is used to map the score to the range between 0 and 1.

The weights are adjusted in the process called training which uses a specific dataset in which correct labels are known. In the first iteration of the training process, the weights are initialized to small random values and the output score is calculated for all classes. As the correct label is known in training, the output scores for all incorrect classes are then compared to the score for the correct class. If the incorrect classes have higher scores than the correct class, the loss function is large, while if the correct class has a higher score than the incorrect classes, the loss function is small. Different loss functions can be used, but a common definition of the loss function per event is:

$$L_{event} = \sum_{i \neq k} max(0, s_i - s_k + 1) \tag{6.2}$$

where i runs over all classes, except for the correct class k. If a higher score was assigned to an incorrect class than to the correct class, the difference $s_i - s_k$ will be large and the loss function will be large. If a higher score was assigned to the correct class than to an incorrect class, the $s_i - s_k$ will be negative and 0 will be selected.

The average loss per training dataset is quantied as:

$$L = \frac{1}{N_{events}} \sum_{events} L_{event} \tag{6.3}$$

where the sum runs over all events in the training dataset, while the L_{event} is the loss function in an event.

The weights w_i are then optimized to minimize the average loss function in each iteration. The next iteration then uses the weights set in the previous iteration. If the classification was correct, the loss function per event L_{event} is set to zero, meaning that the total loss function is not increased. If the classification was incorrect, the loss function is increased. The weights are optimized in each iteration by finding the values which minimize the loss function. A standard way to minimize the loss function is called gradient descent, a method which finds the direction in which the loss function is getting smaller by computing the gradient $\nabla_w L_{event}$. As the negative direction of the gradient $\nabla_w L_{event}$ shows the negative slope of L_{event} , by moving towards this direction, the L_{event} gets smaller. In each iteration, the weights are shifted by $\frac{\partial L_{event}}{\partial w_i} d_{step}$ and used as initial weights in the next iteration. At the end of the training process,after enough iterations, the weights are optimized and loss function minimized.

As the loss function and its gradient have to be computed for all events in each iteration, in datasets with many events, the gradient descent method is very slow and other methods

[1] Neural networks were inspired by the way human brain works.

like stochastic gradient descent or adam are commonly used. In stochastic gradient descent method, instead of computing the loss function per whole dataset in each iteration, the dataset is randomly samples to produce a minibatch of events and the loss function of the dataset is computed as the average loss per this minibatch. The gradient is also calculated using this batch and the weights are updated based on subset of events. The sampling is done per each iteration, so that different events are used.

After the training process, the predictions are tested on a different set of data called validation data. During the training process, it can happen that the weights were optimized to describe the specific characteristics of the training dataset but do not give good predictions on a new dataset. This is called overtraining and is tested by finding the differences in the performance with the training and a validation dataset. If the performance is good for training data and worse for validation data this would imply that the fit was optimized too well for the specific dataset used in training.

The overtraining is solved using regularization, by adding regularization terms to the loss function which gives penalty if very compex models are used in the fit meaning that If a simpler model fits the data properly, it will be preferred over a more complex model.

6.2 Neural Network Training

In the process of training, neural networks were optimized to differentiate between different processes, based on their kinematical differences. Three neural networks were trained for the three largest backgrounds, leading to the following neural network scores:

- NN_W: optimized to differentiate between the signal and W+jets process.
- $\mathrm{NN}_{Z\tau\tau}$: optimized to differentiate between the signal and $Z \to \tau\tau$ process.
- $\mathrm{NN}_{Z\ell\ell}$: optimized to differentiate between the signal and $Z\ell\ell$ process, namely between the $Z \to \tau e$ and $Z \to ee$ in the $\tau - e$ channel and $Z \to \tau\mu$ and $Z \to \mu\mu$ in the τ-μ channel. The training was done using 1-prong events only, since the $Z \to \mu\mu$ and $Z \to ee/\mu\mu$ mainly contribute to backgrounds in 1-prong events.

The training was done separately for different backgrounds since it is expected that due to the differences in kinematics, different selections on the same variable but also different variables would lead to a good separation. Further, since the three backgrounds are formed in a different way, this could additionally lead to different optimal discriminating variables and selections. Namely, the W+jets are mainly formed from jet $\to \tau$ fakes, $Z \to \tau\tau$ due to a real τ and a real e/μ in the final state and $Z \to \ell\ell$ due to $e/\mu \to \tau$ fakes.

The training was done independently for the τ-e and the τ-μ channel. The $\mathrm{NN}_W$and $\mathrm{NN}_{Z\tau\tau}$were trained using combined 1-prong and 3-prong events since the W+jets and $Z \to \tau\tau$ contribute to the background for both event types, while the $\mathrm{NN}_{Z\ell\ell}$was trained using 1-prong events, with an additional requirement that the selected signal tau is matched to a true electron in the $Z \to ee$ or a true muon in the $Z \to \mu\mu$ process, since the $Z \to ee$ and $Z \to \mu\mu$ mainly contribute to 1-prong events with $e/\mu \to \tau$ fakes.

Kinematical variables which show a good discrimination between the signal and the backgrounds are used as the input to the neural networks. An overview of the inputs per each neural network is given in Table 6.1, while the variable definitions are given and motivated in Chapter 5.

The Monte Carlo events used for training were required to pass the initial selection chosen to enhance the number of signal events, as summarized in Table 6.2. For 1-prong events,

Table 6.1: Lists of kinematical variables given as inputs to neural networks. The scores NN_W, $\mathrm{NN}_{Z\tau\tau}$and $\mathrm{NN}_{Z\ell\ell}$refer to different networks trained to discriminate between the signal and the W+jets, $Z \to \tau\tau$ and $Z\ell\ell$ processes.

Variable	$\mathrm{NN}_{Z\tau\tau}$	$\mathrm{NN}_{Z\ell\ell}$	NN_W
$\hat{E}^{\mathrm{lep}}$	•	•	•
$\hat{p}_x^{\tau_h}$	•	•	•
$\hat{p}_z^{\tau_h}$	•	•	•
$\hat{E}^{\tau_h}$	•	•	•
$\hat{p}_z^{\mathrm{miss}}$	•	•	•
$\hat{E}^{\mathrm{miss}}$	•	•	•
$p_{\mathrm{T}}^{\mathrm{tot}} = (p^{\mathrm{lep}} + p^{\tau_h} + E_{\mathrm{T}}^{\mathrm{miss}})_{\mathrm{T}}$	•	•	•
m_{coll}	•	•	•
$\Delta\alpha$	•	•	•
$m_{\mathrm{vis(lep},\tau_h)}$		•	

applied is a BDT selection on the tau, optimized to discriminate between the true taus and the $e \to \tau$ fakes. Further, applied is a requirement for exactly one signal tau and one signal electron/muon with opposite charges and the p_{T} selection on the signal electron/muon. The training was done in this signal-like region in order for the neural network to learn to discriminate between the signal and the signal-like background events, instead of all background events. The number of events per process, after the applied selections is shown in Table 6.3.

Table 6.2: Selection applied to the events used for training: the common selection is applied in both channels, while the additional selections are applied per a specific channel.

Common Selection	
1-prong τ -lepton with eveto BDT 85% or 3-prong τ-lepton b-jet veto	
τ-e channel	τ-μ channel
1 electron, no muons OS (τ, e) pair $p_{\mathrm{T}}(e) > 30\,\mathrm{GeV}$	1 muon, no electrons OS (τ, μ) pair $p_{\mathrm{T}}(\mu) > 30\,\mathrm{GeV}$

Table 6.3: Number of Monte Carlo events used for the training of the neural-network classifiers in the $e\tau$ and $\mu\tau$ channels.

Process	$e\tau$	$\mu\tau$
Signal	38236	37890
$Z \to \tau\tau$	235663	194466
$Z \to ee/\mu\mu$	27446	4526
W+jets	76800	91044

For an optimal training, the NN inputs are pre-processed to normalize their magnitudes and to remove symmetries. The pre-processing consists of the boost *i.e.* the momenta of the particles are boosted to the frame to their center-of-mass frame, then of their rotation and the scaling, in which each input variable is scaled by subtracting its mean and by dividing by its standard deviation. Further, the correlations between the input variables were tested and shown in Figures 6.2 - 6.7[2].

In order to obtain one discriminating variable which is going to be used in the final fit, a combined neural network score $\mathrm{NN}_{\mathrm{comb}}$ is computed from the scores for individual background.

For 3-prong events, the combined score is computed from NN_W and $\mathrm{NN}_{Z\tau\tau}$:

$$\mathrm{NN}_{\mathrm{comb}} = 1 - \sqrt{(1 - \mathrm{NN}_W)^2 + (1 - \mathrm{NN}_{Z\tau\tau})^2}/\sqrt{2} \tag{6.4}$$

while for 1-prong events, the combined score is computed from NN_W, $\mathrm{NN}_{Z\tau\tau}$ and $\mathrm{NN}_{Z\ell\ell}$:

$$\mathrm{NN}_{\mathrm{comb}} = 1 - \sqrt{(1 - \mathrm{NN}_W)^2 + (1 - \mathrm{NN}_{Z\tau\tau})^2 + (1 - \mathrm{NN}_{Z\ell\ell}}/\sqrt{3}. \tag{6.5}$$

In the training, each sample was split into a subsample with even and a subsample with odd events. The training was done on the subsample with even events and tested on the subsample with odd events and vice versa. The events not used in training were split into validation and test samples. The validation samples were used to optimise the neural-network model, while the test sample was used to evaluate the performance. The optimized model uses two hidden layers and sixteen nodes in each layers[3].

Futher, the modelling of neural-network inputs at the training stage is shown in Figures 8.12 and 8.13 for $\tau - e$ and 8.14 and 8.15 for $\tau - \mu$ channel. The shown figures are used to validate the estimation of fake backgrounds using the fake-factor method which is introduced and described in Chapter 7. As the inputs are used to build the final discriminating variable *i.e.* the neural-network output $\mathrm{NN}_{\mathrm{comb}}$, it was important to check their modelling using the final background model *i.e.* the model of Monte Carlo predictions for $e/\mu/\tau \to \tau$ backgrounds and the data-driven approach for jet $\to \tau$ backgrounds.

The modelling of NN inputs at the beginning of the training process is shown in Figures 8.12 and 8.13 for $\tau - e$ and 8.14 and 8.15 for $\tau - \mu$ channel. The Figures 8.13-8.14 show good overall modelling of the input distributions with small differences between the background prediction and the data which are further going to be adjusted in the fit.

[2]Figures 6.2 - 6.7 are produced by Daniele Zanzi, a member of the analysis team, and taken from the published paper on the analysis [48]

[3]The optimization of the model was done by Daniele Zanzi, a member of the analysis team.

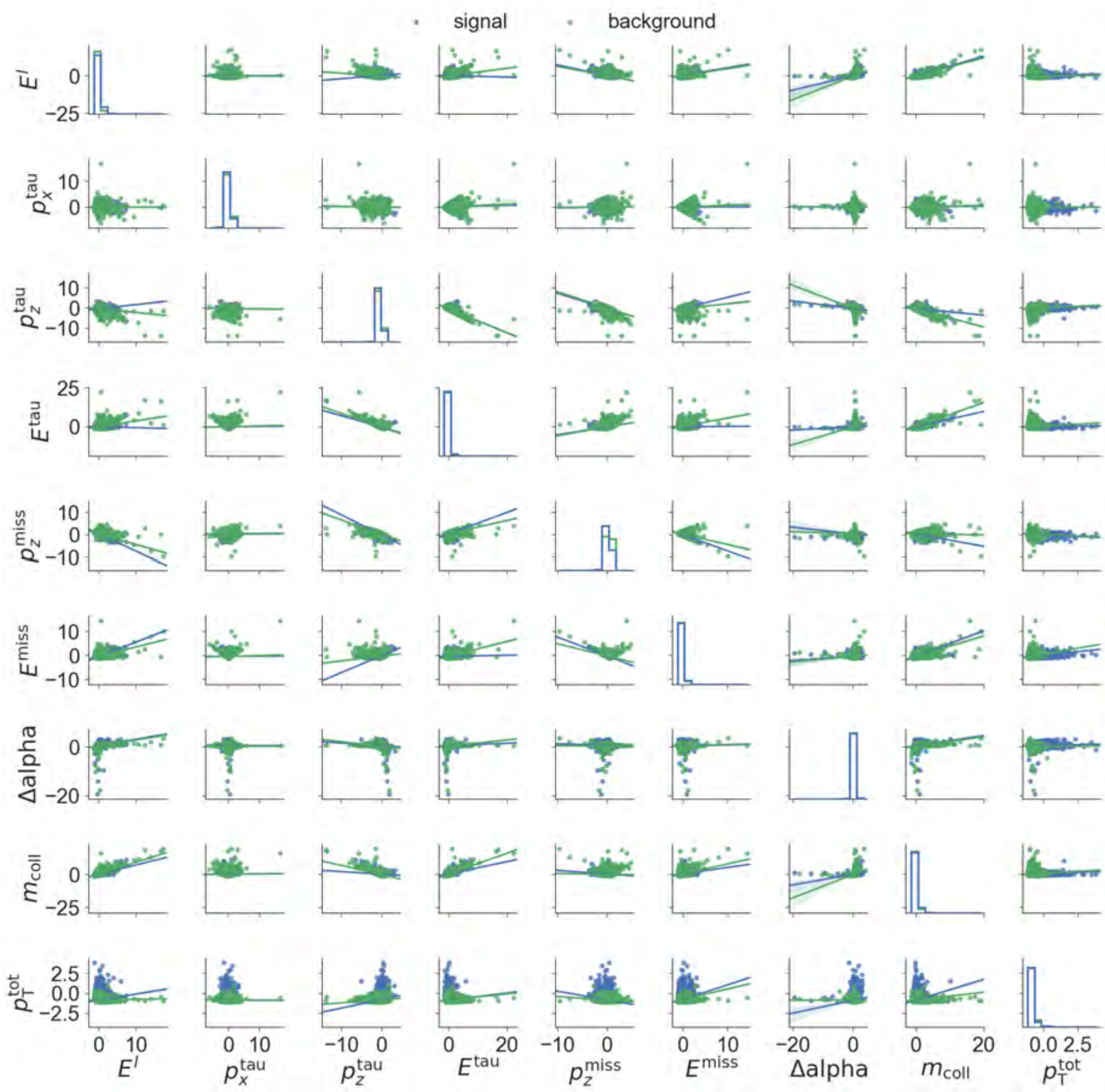


Figure 6.2: Distributions of the input features for $NN_{Z\tau\tau}$ in the signal (signal) and $Z \rightarrow \tau\tau$(background) events used in the training for the $e\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48]

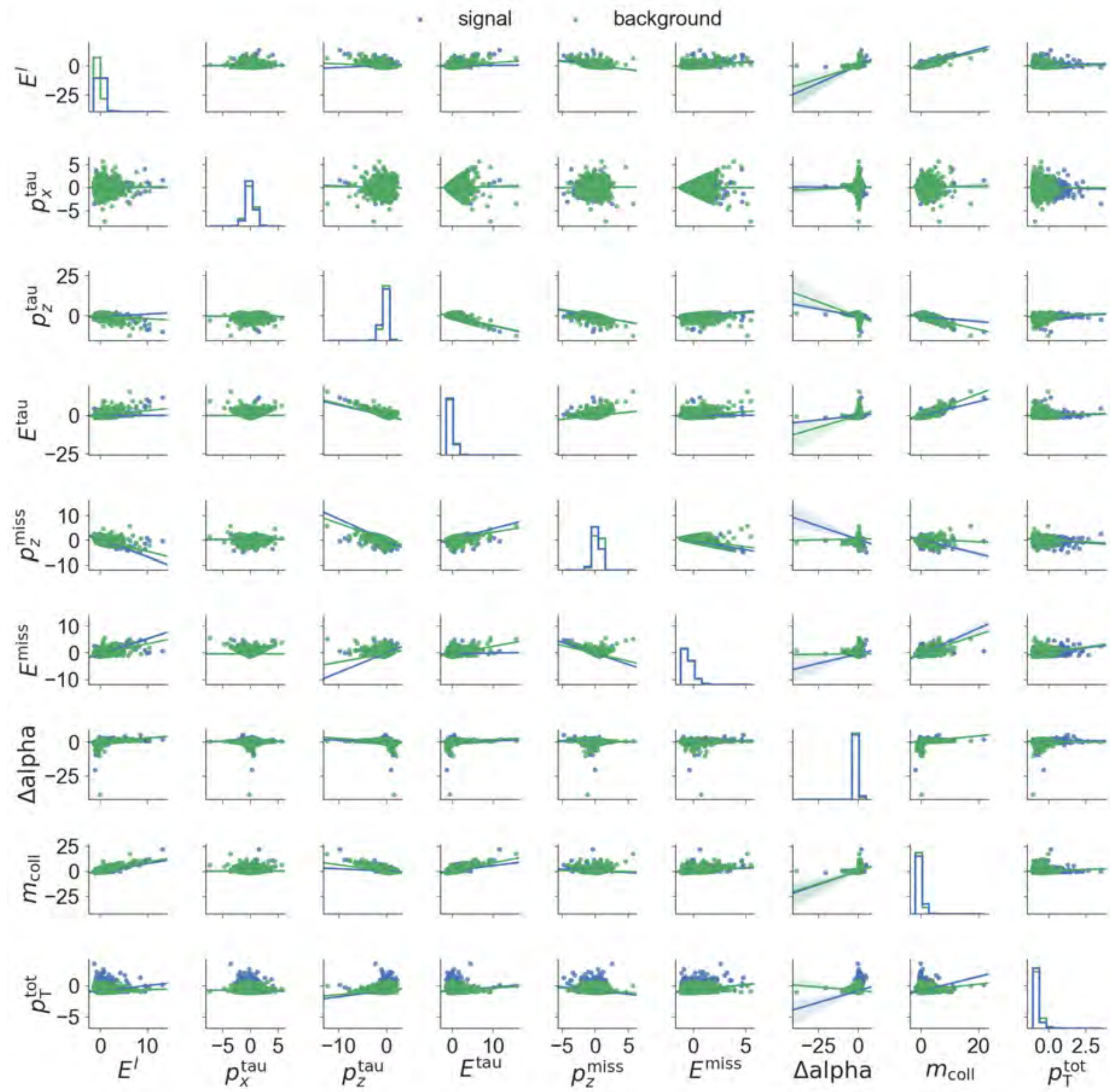


Figure 6.3: Distributions of the input features for $\text{NN}_{Z\tau\tau}$ in the signal (signal) and $Z \to \tau\tau$(background) events used in the training for the $\mu\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48].

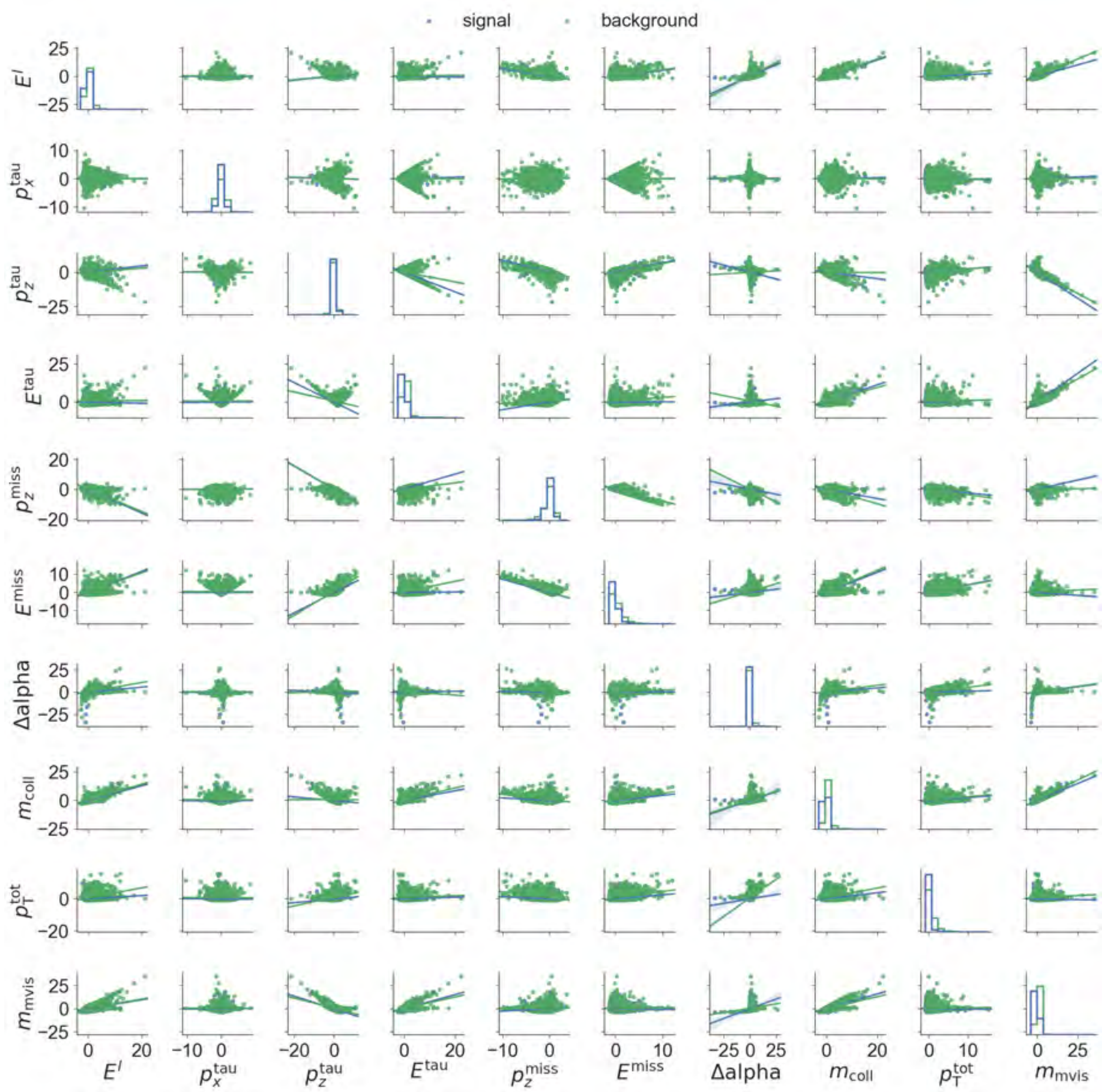


Figure 6.4: Distributions of the input features for $\mathrm{NN}_{Z\ell\ell}$ in the signal (signal) and $Z \to ee/\mu\mu$ (background) events used in the training for the $e\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48]

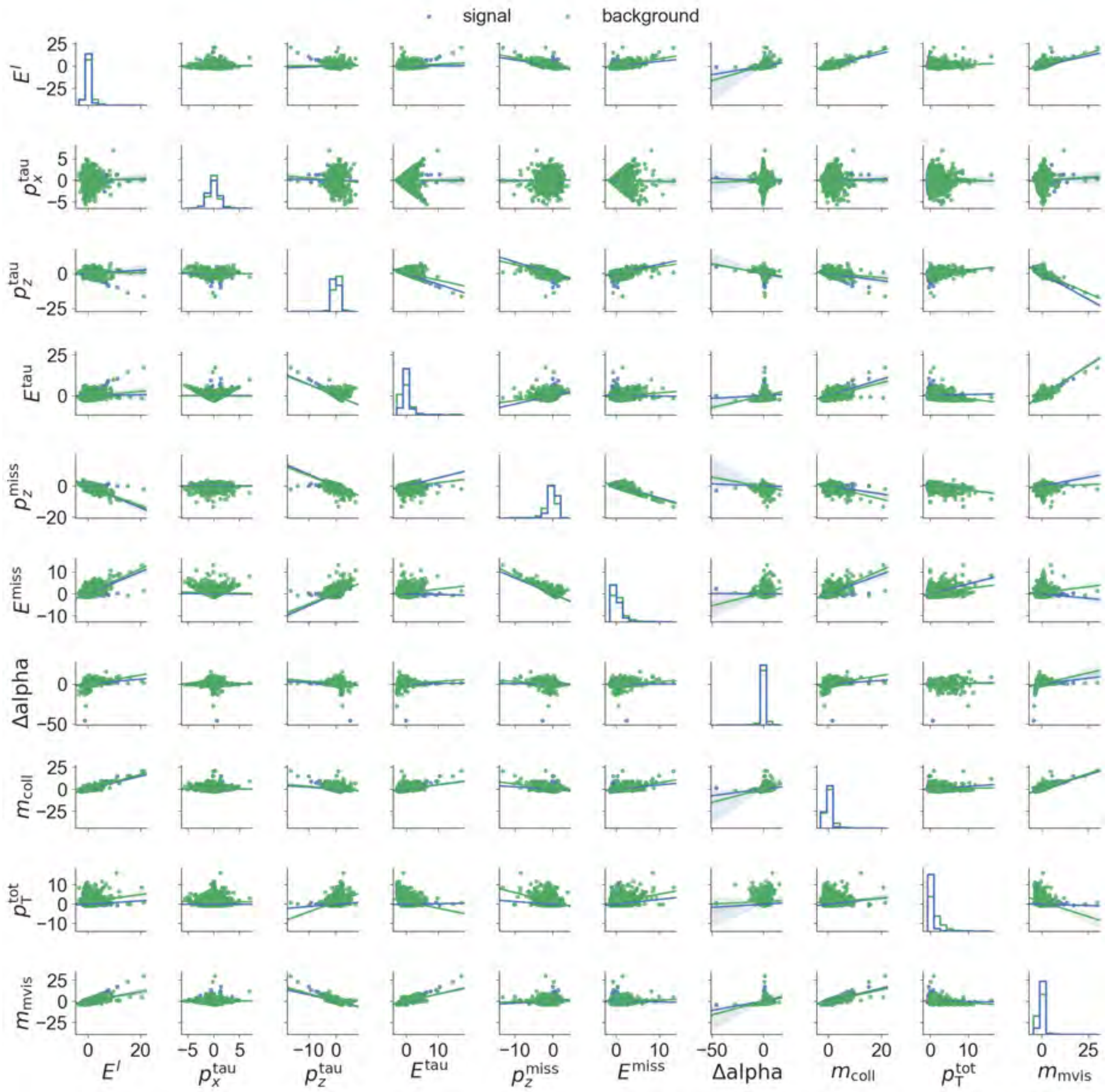


Figure 6.5: Distributions of the input features for $\mathrm{NN}_{Z\ell\ell}$ the signal (signal) and $Z \to ee/\mu\mu$ (background) events used in the training for the $\mu\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48]

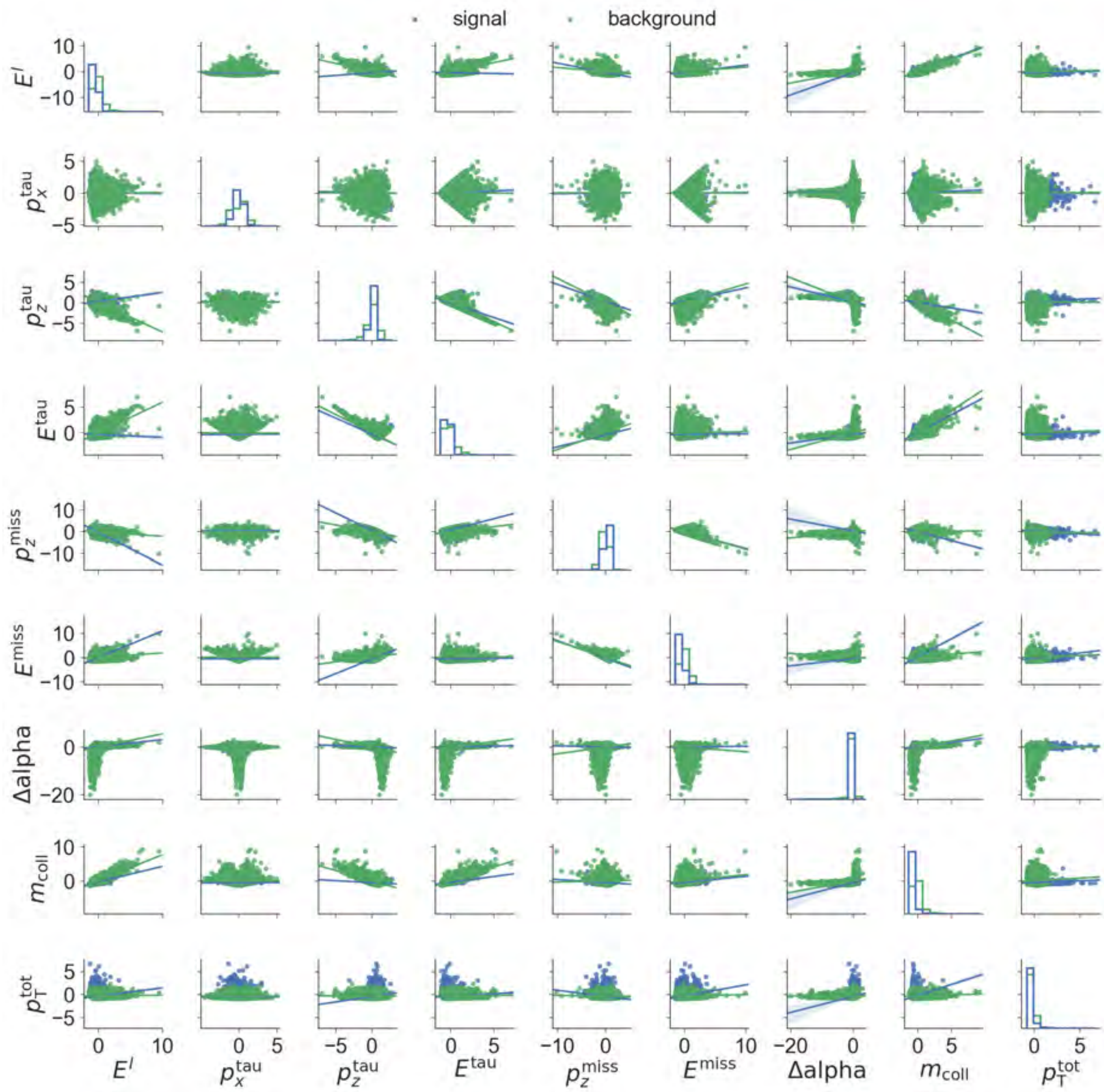


Figure 6.6: Distributions of the input features for NN_W in the signal (signal) and W+jets (background) events used in the training for the $e\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48]

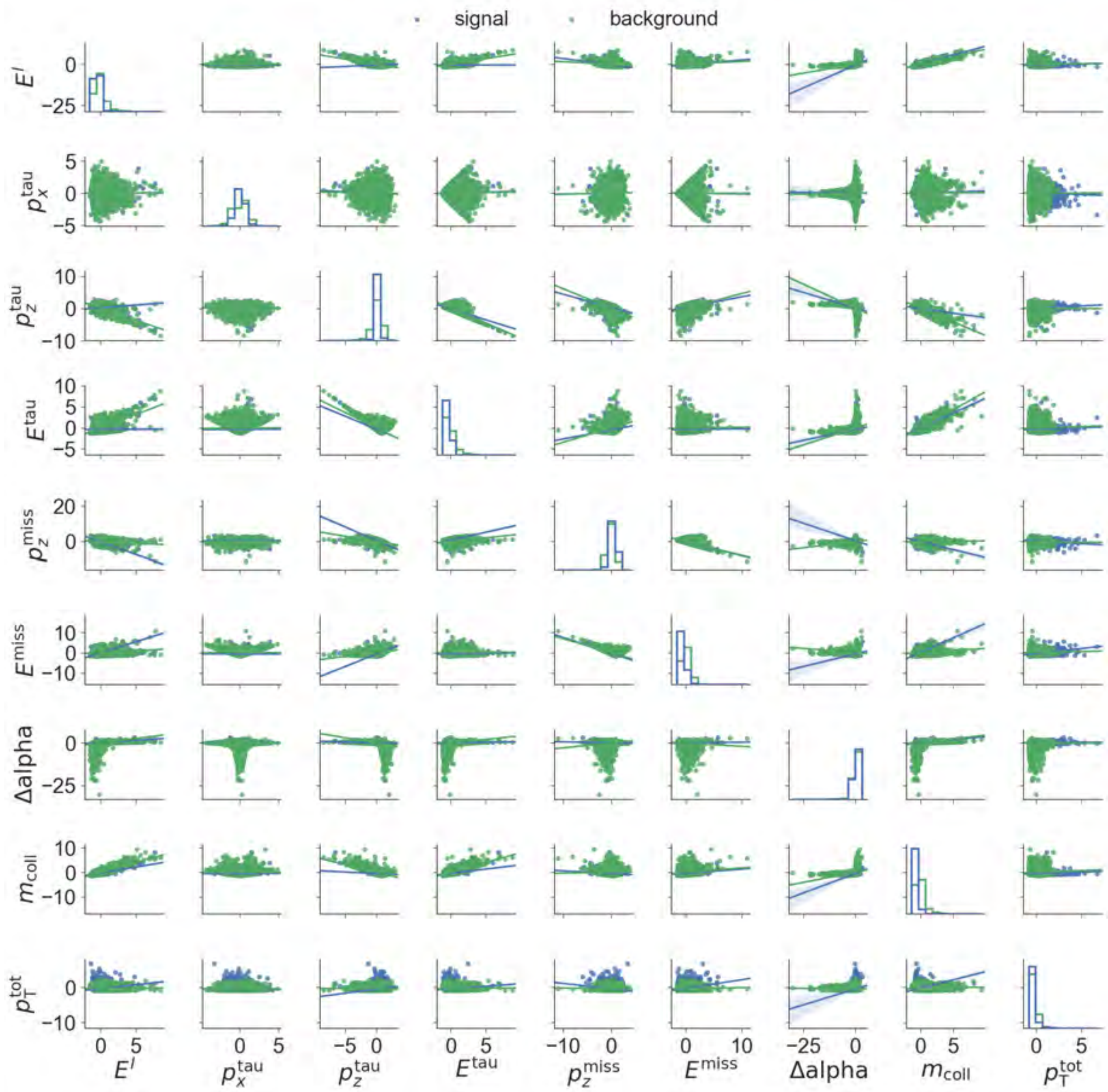
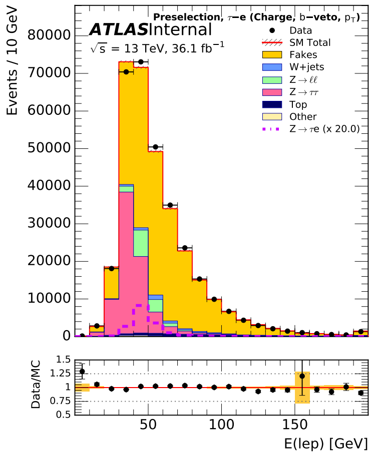
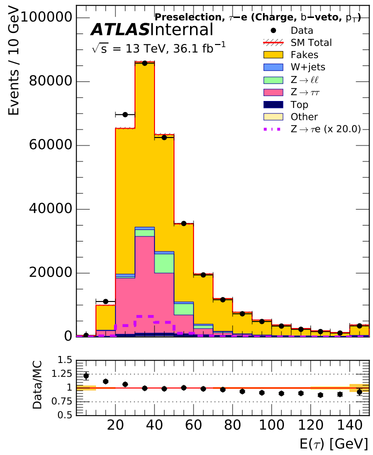


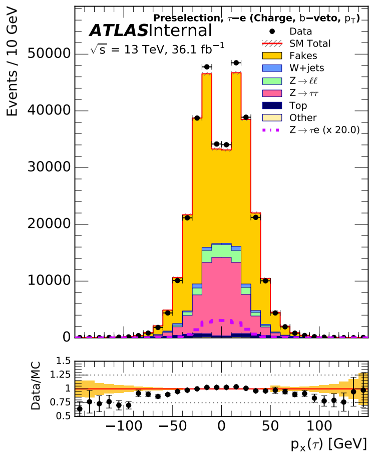
Figure 6.7: Distributions of the input features for NN_W in the signal (signal) and W+jets (background) events used in the training for the $\mu\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48]



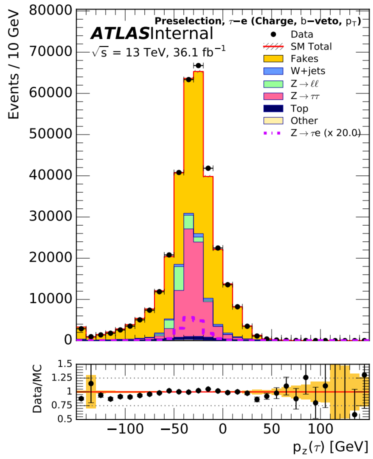
(a) $\hat{E}^{\text{lep}}$



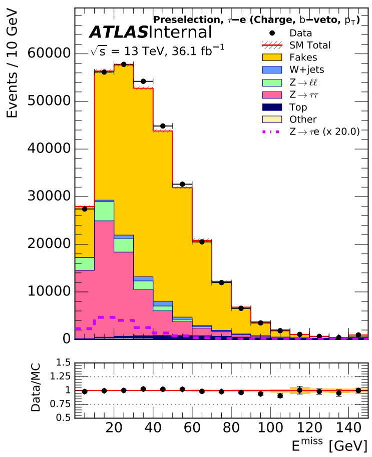
(b) $\hat{E}^{\tau_h}$



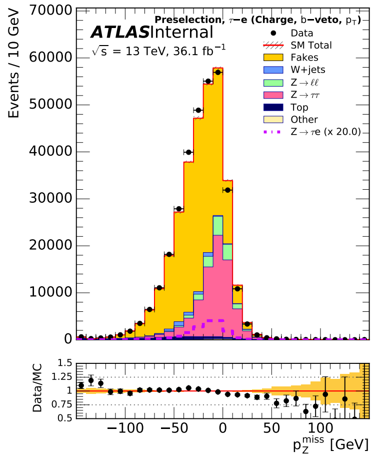
(c) $\hat{p}_x^{\tau_h}$



(d) $\hat{p}_z^{\tau_h}$

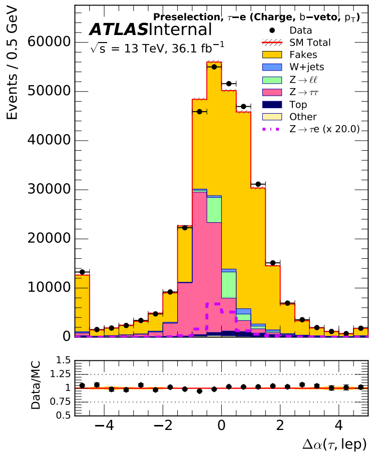


(e) $\hat{E}^{\text{miss}}$

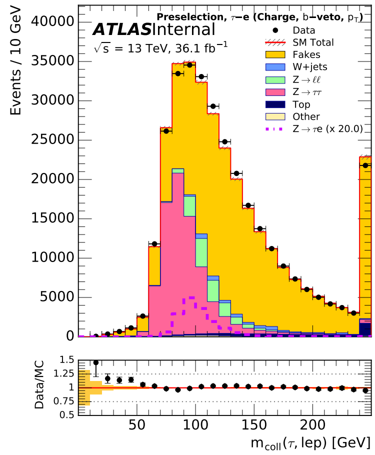


(f) $\hat{p}_z^{\text{miss}}$

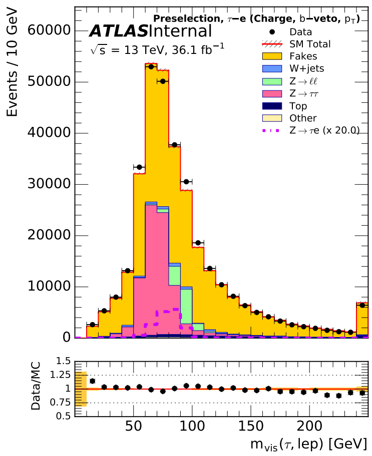
Figure 6.8: Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - e$ channel.



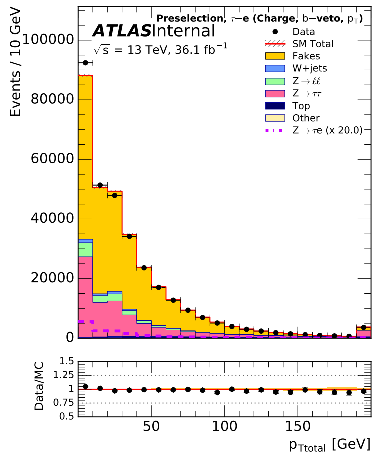
(a) $\Delta\alpha$



(b) m_{coll}

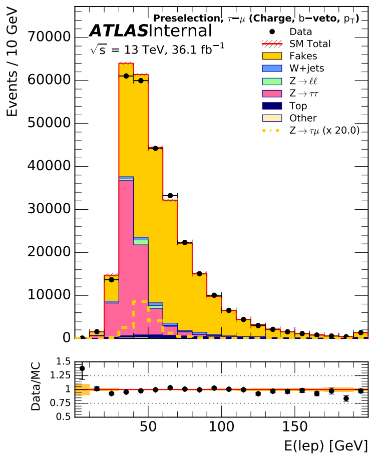


(c) $m_{\text{vis}(\text{lep},\tau_h)}$

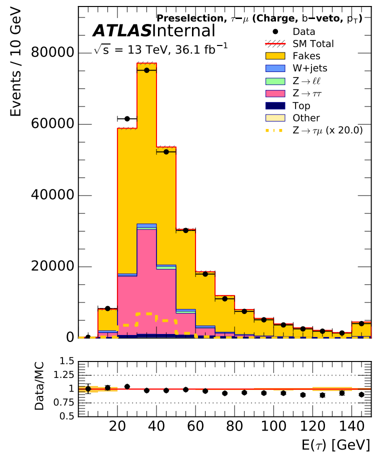


(d) $p_{\text{T}}^{\text{tot}}$

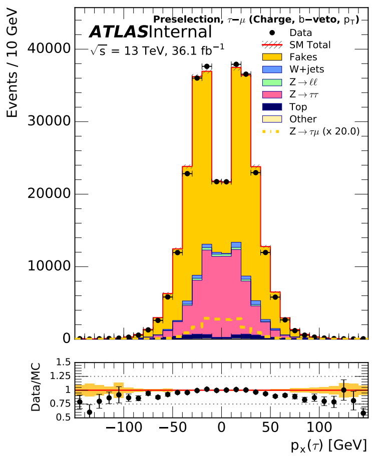
Figure 6.9: Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - e$ channel.



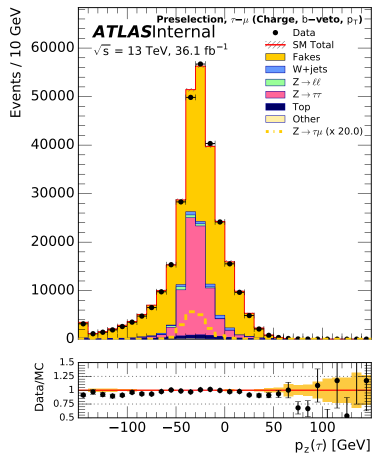
(a) $\hat{E}^{\text{lep}}$



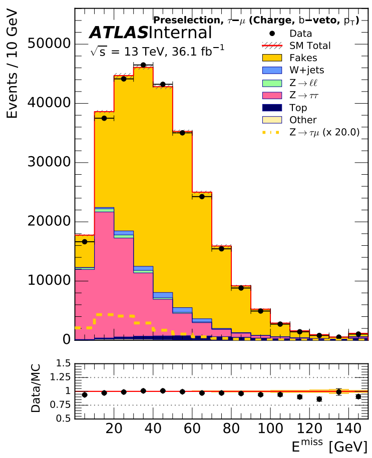
(b) $\hat{E}^{\tau_h}$



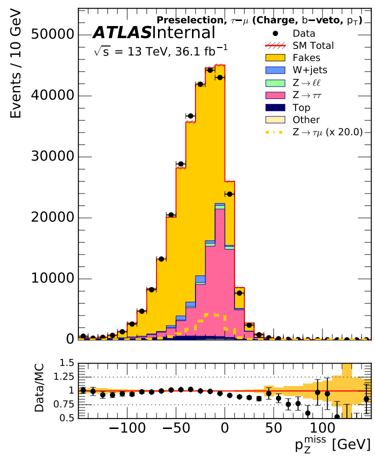
(c) $\hat{p}_x^{\tau_h}$



(d) $\hat{p}_z^{\tau_h}$

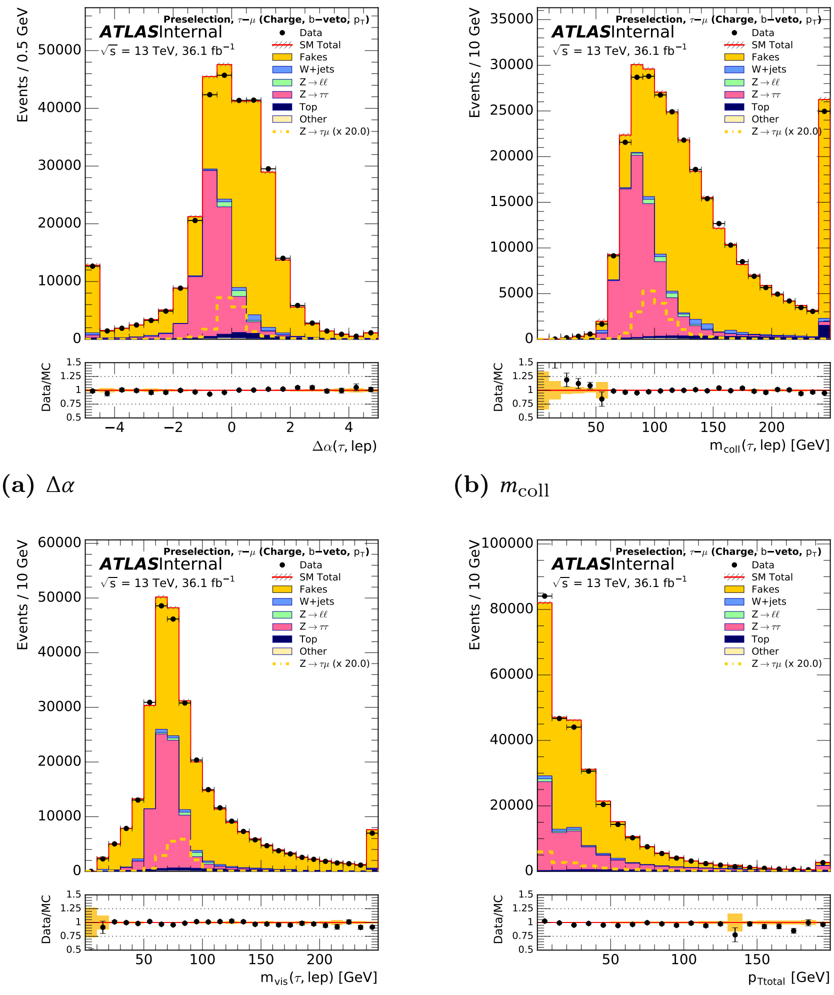


(e) $\hat{E}^{\text{miss}}$



(f) $\hat{p}_z^{\text{miss}}$

Figure 6.10: Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - \mu$ channel.



(a) $\Delta\alpha$

(b) m_{coll}

(c) $m_{\mathrm{vis(lep},\tau_h)}$

(d) $p_{\mathrm{T}}^{\mathrm{tot}}$

Figure 6.11: Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - \mu$ channel.

6.3 Signal Region

The signal region is defined to enhance the signal, while suppressing the backgrounds processes. In order to do this, multiple selections are applied on variables which show good discrimination between the signal and the background processes.

In each event, it is required that an electron or a muon trigger fired, depending on the channel and that the object which fired the trigger is matched to a reconstructed particle. Further, the events are required to have one signal tau which passes the tight BDT working point, developed to discriminate between the real taus and the taus faked by jets. In order to discriminate between the real taus and the taus faked by electrons, the tau is required to pass the eveto BDT point, corresponding to 85% signal efficiency. In the τ-e channel, one signal electron of the opposite charge with respect to the tau and no signal muons are required, while in the τ-μ an oppositely charged muon and no electrons are required.

Events with one or more b-tagged jets are rejected in order to suppress the background due to the top-quark production. Further, a p_{T} selection on the signal electron/ is applied due to the increase in significance. The $m_{\mathrm{T}}(\tau)$ selection is targeted at the rejection of the W+jets and $Z \to \tau\tau$ backgrounds while the selections involving $m(\tau_{\mathrm{track}}, e)$ were targeted at the $Z \to \mu\mu$ background and applied to 1-prong events only.

A summary of the selections is shown in Tables 6.4 and 6.5. Shown is the number of initially simulated Monte Carlo events (Initial), the number of events after the event cleaning and a requirement of having at least one muon or an electron passing the object definition requirements and the overlap removal (After obj. def.), the event number after the Monte Carlo weight and the pile-up weight are applied and after the events are normalized to data luminosity of 36.1fb^{-1} (After obj. def. (weighted)). Further the events with a bug in the storage of the pile-up weight are removed (Pileup weight) and the described selections are applied.

The distributions of the neural network scores NN_W, $\mathrm{NN}_{Z\tau\tau}$, $\mathrm{NN}_{Z\ell\ell}$ and $\mathrm{NN}_{\mathrm{comb}}$ in the signal region are shown in Figures 6.12 and 6.13. The shown figures are used to validate the fake background estimation, described in Chapter 7. The data and the background predictions are compared in the sideband of the signal region where a good modelling is observed. Small differences are due to the background normalization factors which are further adjusted in the fit.

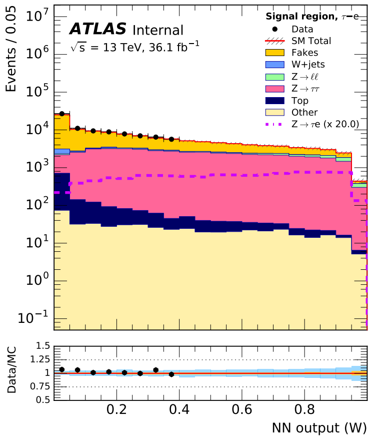
The data is not shown for the high values of neural network score distributions in order to not bias the analysis by looking at the data in the region where the signal is expected, before the ft model is defined and tested.

Table 6.4: Cutflow for τ-e channel, normalised to 36.1fb^{-1}. The signal branching ratio is assumed to be 1.2×10^{-5}.

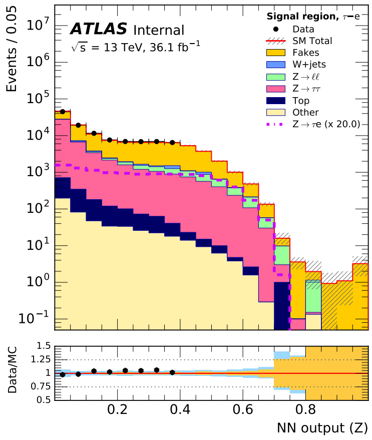
Selection	$Z \to \tau e$	ε_{tot}	S	Other	ε_{tot}	Top	ε_{tot}	$Z \to \tau\tau$	ε_{tot}	$Z \to ee$	ε_{tot}	W+jets	ε_{tot}
Initial	498 400.00	1.00	0.00	52 177 600.00	1.00	71 327 200.00	1.00	36 024 350.00	1.00	156 543 397.00	1.00	404 759 600.00	1.00
After obj. def.	72 643.00	1.00	0.00	3 343 319.00	1.00	4 928 338.00	1.00	4 162 222.00	1.00	23 156 925.00	1.00	10 372 252.00	1.00
After obj. def. (weighted)	2701.81	1.00	0.42	172 310.98	1.00	1 546 076.97	1.00	763 505.76	1.00	20 767 148.54	1.00	18 669 709.75	1.00
Pileup weight	2701.81	1.00	0.42	172 310.98	1.00	1 546 076.97	1.00	763 505.76	1.00	20 767 148.54	1.00	18 669 709.75	1.00
Lowest single-electron trigger	2268.31	0.84	0.47	85 150.04	0.49	738 764.27	0.48	270 987.86	0.35	14 322 197.17	0.69	7 599 248.63	0.41
Leading tau tight	1122.20	0.42	1.29	5688.27	0.03	72 502.40	0.05	147 386.30	0.19	238 991.49	0.01	297 216.85	0.02
eveto BDT 85%	1030.38	0.38	1.41	4797.88	0.03	64 635.09	0.04	134 512.02	0.18	68 073.60	0.00	263 721.38	0.01
1 electron, no muons	938.84	0.35	1.40	3896.15	0.02	58 595.72	0.04	116 878.32	0.15	45 530.59	0.00	227 063.09	0.01
$\lvert\eta(\tau_{1\text{P}})\rvert < 2.2$	891.61	0.33	1.38	3604.50	0.02	56 246.25	0.04	112 137.85	0.15	36 763.77	0.00	208 189.62	0.01
OS (τ, e) pair	880.76	0.33	1.52	2682.79	0.02	50 908.12	0.03	111 090.78	0.15	24 947.66	0.00	147 798.64	0.01
b-jet veto	858.91	0.32	1.60	2510.60	0.01	8165.20	0.01	107 869.62	0.14	24 418.93	0.00	144 257.55	0.01
$p_{\text{T}}(e) > 30\,\text{GeV}$	813.00	0.30	1.65	2252.38	0.01	7751.32	0.01	78 452.35	0.10	22 611.44	0.00	131 470.33	0.01
$m_{\text{T}}(\tau) < 35\,\text{GeV}$	648.74	0.24	1.88	932.89	0.01	2396.97	0.00	48 646.37	0.06	12 960.08	0.00	54 459.40	0.00
1-prong: $\text{NN}_{\text{comb}} > 0.15$	642.61	0.24	2.03	765.02	0.00	1713.56	0.00	46 119.11	0.06	10 702.36	0.00	40 309.20	0.00
1-prong and $\lvert\eta(\tau)\rvert < 2 (> 2)$: $m(\tau_{\text{trk}}, e) < 84(80)$ or $m(\tau_{\text{trk}}, e) < 100$	597.33	0.22	1.95	697.93	0.00	1579.55	0.00	45 560.95	0.06	8732.60	0.00	37 197.89	0.00
1-prong and $80 < m_{\text{vis}}(\tau, e) < 100$: $m(\tau_{\text{trk}}, e) > 40$	576.91	0.21	1.91	681.52	0.00	1561.10	0.00	45 183.10	0.06	6533.95	0.00	36 895.28	0.00

Table 6.5: Cutflow for τ-μ channel(s), normalised to 36.1 fb^{-1}. The signal branching ratio is assumed to be 1.2×10^{-5}.

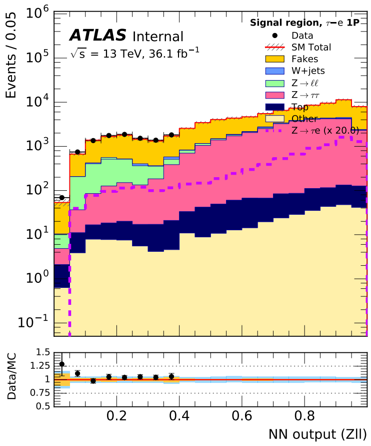
Selection	$Z \to \tau\mu$	ε_{tot}	S	Other	ε_{tot}	Top	ε_{tot}	$Z \to \tau\tau$	ε_{tot}	$Z \to \mu\mu$	ε_{tot}	W+jets	ε_{tot}
Initial	499 800.00	1.00	0.00	52 177 600.00	1.00	71 327 200.00	1.00	36 024 350.00	1.00	156 543 397.00	1.00	404 759 600.00	1.00
After obj. def.	68 288.00	1.00	0.00	3 343 319.00	1.00	4 928 338.00	1.00	4 162 222.00	1.00	23 156 925.00	1.00	10 372 252.00	1.00
After obj. def. (weighted)	2541.76	1.00	0.39	172 310.98	1.00	1 546 076.97	1.00	763 505.76	1.00	20 767 148.54	1.00	18 669 709.75	1.00
Pileup weight	2541.76	1.00	0.39	172 310.98	1.00	1 546 076.97	1.00	763 505.76	1.00	20 767 148.54	1.00	18 669 709.75	1.00
Lowest single-muon trigger	2271.95	0.89	0.65	65 620.22	0.38	638 120.74	0.41	301 935.39	0.40	3 128 976.04	0.15	8 258 385.59	0.44
Leading tau tight	1196.59	0.47	1.52	5394.66	0.03	68 623.19	0.04	170 760.70	0.22	56 235.42	0.00	318 208.31	0.02
eveto BDT 85%	1094.94	0.43	1.47	4820.54	0.03	61 179.75	0.04	155 413.14	0.20	50 463.37	0.00	284 161.02	0.02
1 muon, no electron	994.44	0.39	1.48	3576.18	0.02	54 362.54	0.04	114 236.26	0.15	28 073.89	0.00	250 502.29	0.01
$\lvert\eta(\tau_{1\text{P}})\rvert > 0.1$	957.28	0.38	1.46	3427.24	0.02	51 773.78	0.03	110 080.51	0.14	25 086.33	0.00	240 278.49	0.01
OS (τ, μ) pair	947.63	0.37	1.63	2513.18	0.01	46 866.80	0.03	109 188.34	0.14	13 825.14	0.00	166 165.06	0.01
b-jet veto	925.10	0.36	1.71	2348.03	0.01	7713.30	0.00	106 313.00	0.14	13 547.77	0.00	162 475.36	0.01
$p_{\text{T}}(\mu) > 30\,\text{GeV}$	870.98	0.34	1.76	2081.39	0.01	7263.93	0.00	76 070.98	0.10	12 253.04	0.00	146 770.96	0.01
$m_{\text{T}}(\tau) < 30\,\text{GeV}$	647.77	0.25	2.05	708.18	0.00	1897.93	0.00	42 150.32	0.06	5864.32	0.00	48 622.34	0.00
1-prong: $\text{NN}_{\text{comb}} > 0.15$	646.80	0.25	2.06	695.89	0.00	1822.01	0.00	41 933.77	0.05	5812.91	0.00	47 745.02	0.00
1-prong and $\lvert\eta(\tau)\rvert < 2(> 2)$: $m(\tau_{\text{trk}}, \mu) < 84(80)$ or $m(\tau_{\text{trk}}, \mu) < 100$	597.42	0.24	1.99	611.25	0.00	1608.15	0.00	41 289.29	0.05	4249.76	0.00	41 982.27	0.00
1-prong and $80 < m_{\text{vis}}(\tau, \mu) < 100$: $m(\tau_{\text{trk}}, e) > 40$	576.31	0.23	1.93	599.54	0.00	1594.50	0.00	40 975.68	0.05	4051.87	0.00	41 561.73	0.00



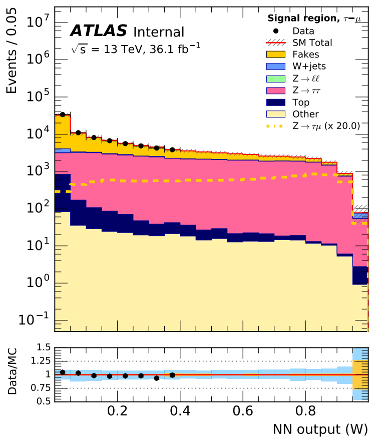
(a) $\mathrm{NN}_W(e, \tau_{\mathrm{had}})$



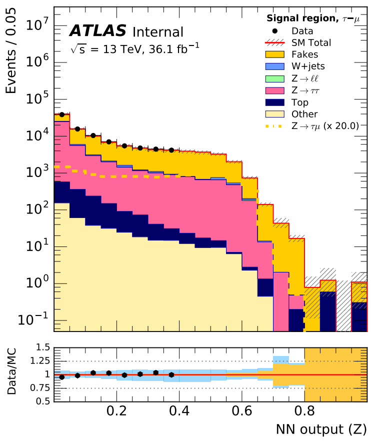
(b) $\mathrm{NN}_{Z\tau\tau}(e, \tau_{\mathrm{had}})$



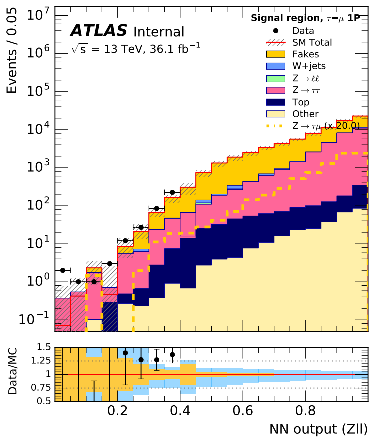
(c) $\mathrm{NN}_{Z\ell\ell}(e, \tau_{\mathrm{had}}\ 1\mathrm{p})$



(d) $\mathrm{NN}_W(\mu, \tau_{\mathrm{had}})$

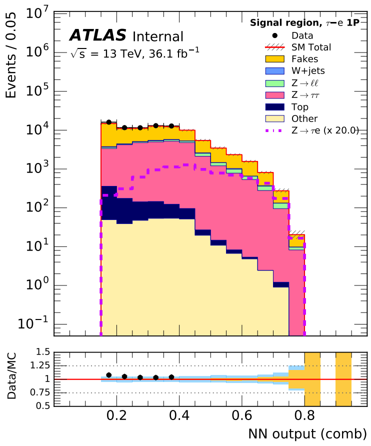


(e) $\mathrm{NN}_{Z\tau\tau}(\mu, \tau_{\mathrm{had}})$

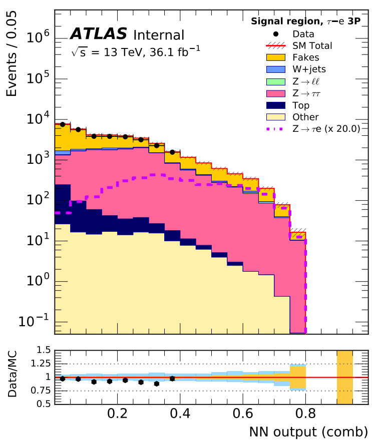


(f) $\mathrm{NN}_{Z\ell\ell}(\mu, \tau_{\mathrm{had}}\ 1\mathrm{p})$

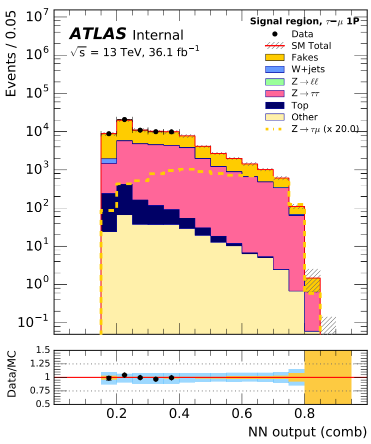
Figure 6.12: Monte Carlo and data distributions of the neural network scores $\mathrm{NN}_W, \mathrm{NN}_{Z\tau\tau}$ and $\mathrm{NN}_{Z\ell\ell}$ in the signal region. To test if the modelling is good, the data and the Monte Carlo predictions are compared in the low sideband of the distributions ($\mathrm{NN}_i > 0.4$), done in order to not unblind the data where the signal is expected, before the full statistical model is built.



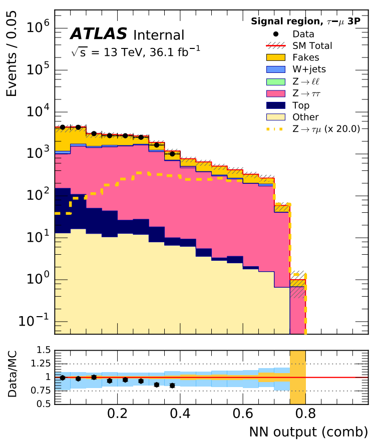
(a) NN Comb (e, τ_{had} 1p)



(b) NN Comb (e, τ_{had} 3p)



(c) NN Comb (μ, τ_{had} 1p)



(d) NN Comb (μ, τ_{had} 3p)

Figure 6.13: Monte Carlo and data distributions of the combined neural network scores NN_{comb} in the signal region. To test if the modelling is good, the data and the Monte Carlo predictions are compared in the low sideband of the distributions ($\text{NN}_{\text{comb}} > 0.4$), done in order to not unblind the data where the signal is expected, before the full statistical model is built.

Estimation of Fake Backgrounds

Background processes are categorized according to the type of the τ-lepton candidate into the backgrounds with a real, correctly identified τ-lepton (*real* backgrounds) and the backgrounds where the τ-lepton candidate is a misidentified jet, electron or a muon (fake backgrounds). The fake backgrounds are not accurately modelled with the MC simulation and therefore the data-driven techniques are implemented. The largest contribution to the fake backgrounds is originating from the jet $\to \tau$ fakes which are estimated using the fake-factor method which is described in Section 7.1. To derive fake-factors, background enriched regions of phase space *i.e. control regions* are defined and described in Section 7.2. Fake-factor results are given in Section 7.3 and the validation is described in Section 7.4. Other backgrounds are estimated using Monte Carlo simulations and further corrected in the fit, together with the normalization of the fake backgrounds.

7.1 Fake Factor Method

The fake-factor method is a data-driven technique used to estimate the backgrounds originating from the particle misidentification. In this analysis, the fake-factor method is used to model the jet $\to \tau$ backgrounds originating from W+jets, $Z \to ee/\mu\mu$+jets, top and multijet processes.

The number of jet $\to \tau$ fakes in a region (the signal region) is defined as:

$$N^{\text{SR, pass}}_{\text{jet}\to\tau} = F \cdot N^{\text{SR, fail}}_{\text{jet}\to\tau} \tag{7.1}$$

where F is the fake factor and $N^{\text{SR, fail}}_{\text{jet}\to\tau}$ is the number of jet $\to \tau$ fakes in a *fail* region, a fakes enriched region where the τ-lepton is required to fail the tight, but to pass the *loose* identification working point.

The number of jet $\to \tau$ fakes in the *fail* region is estimated as:

$$N^{\text{SR, fail}}_{\text{jet}\to\tau} = N^{\text{SR, fail}}_{\text{data}} - N^{\text{SR, fail}}_{\text{MC, not jet}\to\tau} \tag{7.2}$$

where $N^{\text{SR, fail}}_{\text{data}}$ is the number of data events and the $N^{\text{SR, fail}}_{\text{MC, not jet}\to\tau}$ is the number of MC events where the selected tau candidate is matched to a truth electron, muon or a tau in the *fail* region. The jet $\to \tau$ matching in MC is not used since the MC modelling of jet $\to \tau$ misidentification is less accurate.

The fake factor F is computed as a sum of the individual background fake factors, weighted with the relative contribution of the background in the region (the signal region):

$$F = \sum_i R_i F_i + R_{\text{QCD}} F_{\text{QCD}} = R_W F_W + R_Z F_Z + R_{\text{top}} F_{\text{top}} + R_{\text{QCD}} F_{\text{QCD}}, \tag{7.3}$$

where F_W, F_Z, F_{top} and F_{QCD} are W+jets, Z+jets, top and multijet fake factors, while R_W, R_Z, R_{top} and R_{QCD} are relative contributions of each background process in the region of interest.

The individidual fake factors F_i are derived per background in background enriched regions (*control regions*), as ratios of events in the control region in which the τ-lepton passed the *tight* identification working point (*pass control region*) and the fakes enriched control region where the τ-lepton is required to fail the tight, but to pass the *loose* identification working point. Since the control regions are providing high purity of the targeted background process, the number of events originating from the targeted process is computed from data, while subtracting the MC contributions from other processes and the MC contribution from the targeted process, which is not originating from jet $\rightarrow \tau$ fakes. Hence, individual fake factors are computed as:

$$F_i = \frac{N^{\text{pass}}_{\text{CR}_i,\text{data}} - N^{\text{pass}}_{\text{CR}_i,\text{MC}_{j\neq i}} - N^{\text{pass}}_{\text{CR}_i,\text{MC}_i,\text{not jet}\rightarrow\tau}}{N^{\text{fail}}_{\text{CR}_i,\text{data}} - N^{\text{fail}}_{\text{CR}_i,\text{MC}_{j\neq i}} - N^{\text{fail}}_{\text{CR}_i,\text{MC}_i,\text{not jet}\rightarrow\tau}}. \tag{7.4}$$

where $N^{\text{pass/fail}}_{\text{CR}_i,\text{MC}_{j\neq i}}$ are MC contributions of other background processes and the $N^{\text{pass/fail}}_{\text{CR}_i,\text{MC}_i,\text{not jet}\rightarrow\tau}$ is the contribution of the targeted process, which is not due to the jet $\rightarrow \tau$ misidentification.

The relative contributions R_i are computed in region where the fake contribution is being determined (the signal region) as:

$$R_i = \frac{N^{\text{fail}}_{\text{SR},\text{MC}_\text{i},\text{jet}\rightarrow\tau}}{\left(N^{\text{fail}}_{\text{SR},\text{data}} - N^{\text{fail}}_{\text{SR},\text{MC},\text{not jet}\rightarrow\tau}\right)}. \tag{7.5}$$

where $\text{SR}, \text{MC}_\text{i}, \text{jet} \rightarrow \tau$ is the number of the jet $\rightarrow \tau$ events originating from the targeted process and the $\left(N^{\text{fail}}_{\text{SR},\text{data}} - N^{\text{fail}}_{\text{SR},\text{MC},\text{not jet}\rightarrow\tau}\right)$ is the overall number of the jet $\rightarrow \tau$ events in the fakes enriched *fail* region. The fraction of QCD events is defined as $R_{\text{QCD}} \equiv 1 - \sum_i R_i$.

The R-factor can be corrected by applying the correction factor k which accounts for the difference between the MC modelling and the data and is derived per background in its control regions as:

$$k_i = \frac{N^{\text{fail}}_{\text{CR}_i,\text{data}} - N^{\text{fail}}_{\text{CR}_i,\text{MC}_{j\neq i}} - N^{\text{fail}}_{\text{CR}_i,\text{MC}_i,\text{not jet}\rightarrow\tau}}{N^{\text{fail}}_{\text{CR}_i,\text{MC}_i,\text{jet}\rightarrow\tau}} \tag{7.6}$$

The individual fake factors can be corrected by using the $k_i R_i$ instead of R_i and by rederiving R_{QCD} accordingly. In this analysis, the fake-factor is derived separately for 1-prong and 3-prong events. Since the fake-factor is dependent on kinematical variables, the binning in several variables was attempted and the predictions were tested in dedicated validation regions[1]. As the 2-dimensional binning in $p_{\text{T}}(\tau)$ and $p_{\text{T}}(\tau\text{-track})$ has shown to be necessary to provide good modelling of the $m(\tau_{\text{track}}, e/\mu)$ distribution, it is used in 1-prong fake-factor since the signal region applies a selection on $m(\tau_{\text{track}}, e/\mu)$ for 1-prong events. As $m(\tau_{\text{track}}, e/\mu)$ selection is not applied on 3-prong events, the 1-dimensional $p_{\text{T}}(\tau)$ binning is used.

7.2 Control Regions

The control regions are background enriched regions of phase space, defined to provide the high purity of the targeted background process. Control regions for W+jets, $Z \rightarrow ee/\mu\mu$+jets,

[1] W+jets control region and the same-sign signal region are used to validate the fake-factor estimates.

top and multijet processe are defined by adding or inverting the loose signal region selections, where the *loose signal region* is defined as the signal region (Tables 6.4 and 6.5), but without the additional requirements for 1-prong events *i.e.* the removal of events with the $\mathrm{NN_{comb}} > 0.15$ and the $(\tau_{\mathrm{track}}, \mu)$.

The W+jets control region (CRW) is defined by exploiting the difference in the orientation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ with respect to the τ-lepton and the electron/muon. A window providing the highest significance in the $m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}})$-$m_{\mathrm{T}}(e/\mu, E_{\mathrm{T}}^{\mathrm{miss}})$ plane is selected. The $Z \to ee/\mu\mu$+jets control region (CRZll) is defined by requiring an additional electron/muon and by requiring the invariant mass of the two-electron/muon system to be near the Z-mass peak. The requirement for one τ-lepton still holds in CRZll and since the two objects cannot overlap, by requiring one τ and two electrons/muons, it is ensured to select the jet $\to \tau$ and not the $e, \mu \to \tau$ fakes in the $Z \to ee/\mu\mu$+jets process. The top control region (CRT) is defined by requiring at least two b-jets in the event. Finally, the multijet control region (CRQ) is defined by inverting the e/μ isolation. In the multijet control region, the multijet background is defined as the difference between the data and the MC prediction for other backgrounds.

The control region definitions are summarized in Table 7.1 together with the expected purities in each control region for the target process which are estimated from MC simulation. The purity is computed as the number of events from the targeted process with respect to all events in control region:

$$P_{\mathrm{CR}_i,\mathrm{MC}_i} = \frac{N_{\mathrm{CR}_i,\mathrm{MC}_i}}{N_{\mathrm{CR}_i,\mathrm{data}}} \tag{7.7}$$

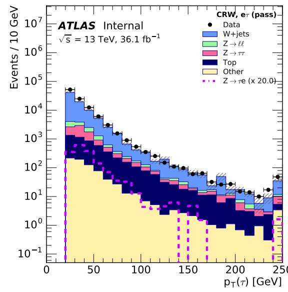
The purity of multijet background is computed as:

$$P_{\mathrm{QCD}} = \frac{N_{\mathrm{CRQ,data}} - N_{\mathrm{CRQ,MC}}}{N_{\mathrm{CRQ,data}}} \tag{7.8}$$

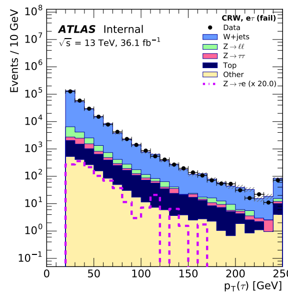
Table 7.1: Contol regions used to derive fake-factors. Differences from the loose SR selection are listed together with the purities for the target processes

Region	**Change relative to the loose SR selection**	**Purity [%]**	
		$e\tau$	$\mu\tau$
CRZll	Two same-flavor opposite-sign light leptons with $81 < m_{\ell\ell} < 101$ GeV	98	98
CRW	$m_{\mathrm{T}}(\ell, E_{\mathrm{T}}^{\mathrm{miss}}) > 40$ GeV and $m_{\mathrm{T}}(\tau_{\text{had-vis}}, E_{\mathrm{T}}^{\mathrm{miss}}) > 30(35)$ GeV in $\mu\tau$ $(e\tau)$ events	84	85
CRT	$N_{b\text{-jets}} \geq 2$	98	98
CRQ	Inverted light lepton isolation	75	37

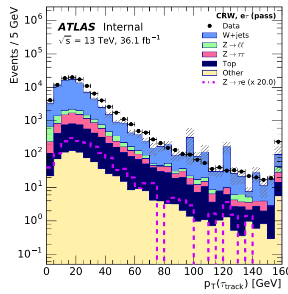
The purity in the control regions is shown in Figures 7.1 and 7.2 for the τ-e and 7.3 and 7.4 for the τ-μ channel. The distributions of the $p_{\mathrm{T}}(\tau)$ and the $p_{\mathrm{T}}(\tau\text{-track})$ are shown per *pass* and the *fail* control regions.



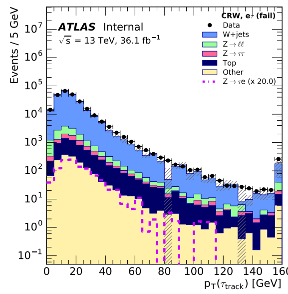
(a) CRW pass: $p_{\mathrm{T}}(\tau)$



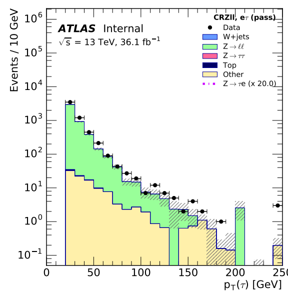
(b) CRW fail: $p_{\mathrm{T}}(\tau)$



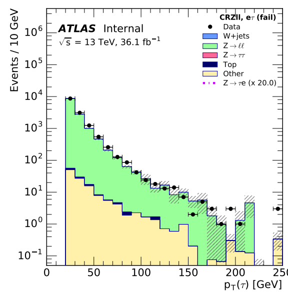
(c) CRW pass: $p_{\mathrm{T}}(\tau\text{-track})$



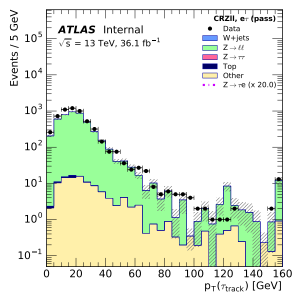
(d) CRW fail: $p_{\mathrm{T}}(\tau\text{-track})$



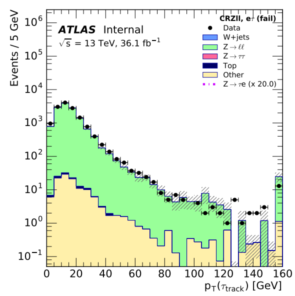
(e) CRZll pass: $p_{\mathrm{T}}(\tau)$



(f) CRZll fail: $p_{\mathrm{T}}(\tau)$

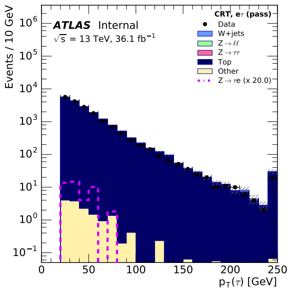


(g) CRZll pass: $p_{\mathrm{T}}(\tau\text{-track})$

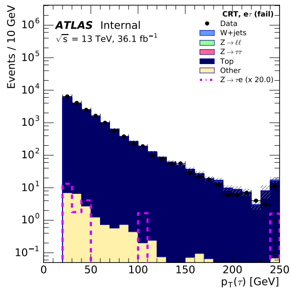


(h) CRZll fail: $p_{\mathrm{T}}(\tau\text{-track})$

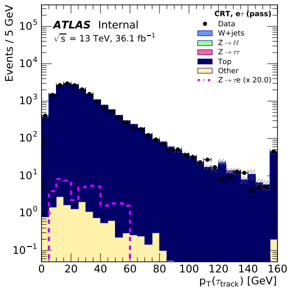
Figure 7.1: Data and Standard Model background distributions of the $p_{\mathrm{T}}(\tau)$ and $p_{\mathrm{T}}(\tau\text{-track})$ in the CRW and CRZll regions of the τ-e channel. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The errors and statistical.



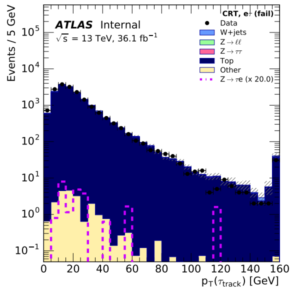
(a) CRT pass: $p_{\mathrm{T}}(\tau)$



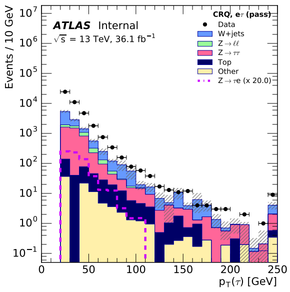
(b) CRT fail: $p_{\mathrm{T}}(\tau)$



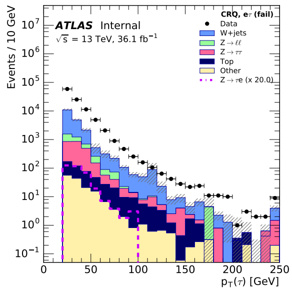
(c) CRT pass: $p_{\mathrm{T}}(\tau\text{-track})$



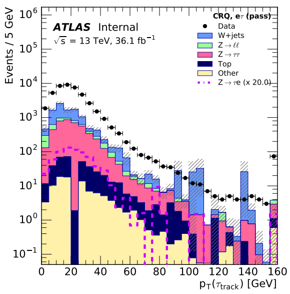
(d) CRT fail: $p_{\mathrm{T}}(\tau\text{-track})$



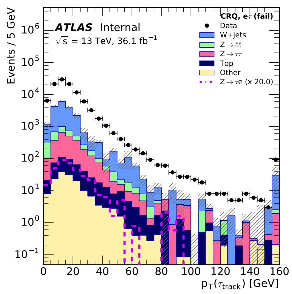
(e) CRQ pass: $p_{\mathrm{T}}(\tau)$



(f) CRQ fail: $p_{\mathrm{T}}(\tau)$

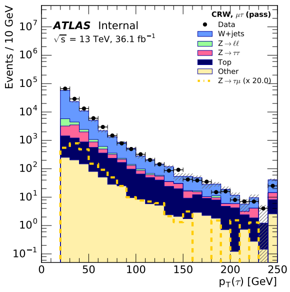


(g) CRQ pass: $p_{\mathrm{T}}(\tau\text{-track})$

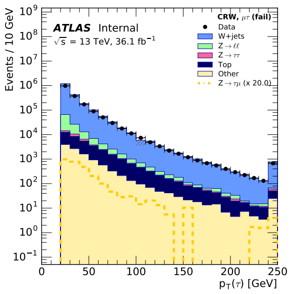


(h) CRQ fail: $p_{\mathrm{T}}(\tau\text{-track})$

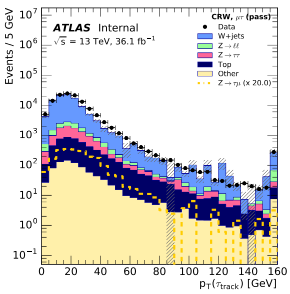
Figure 7.2: Data and Standard Model background distributions of the $p_{\mathrm{T}}(\tau)$ and $p_{\mathrm{T}}(\tau\text{-track})$ in the CRT and CRQ regions of the τ-e channel. The multijet events in CRQ are shown as the difference between the data and the MC prediction for other backgrounds. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The errors and statistical.



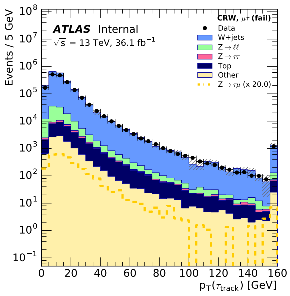
(a) CRW pass: $p_\mathrm{T}(\tau)$



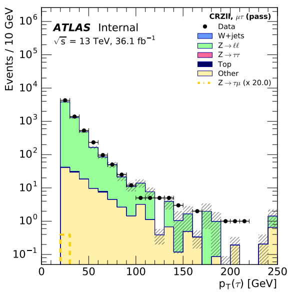
(b) CRW fail: $p_\mathrm{T}(\tau)$



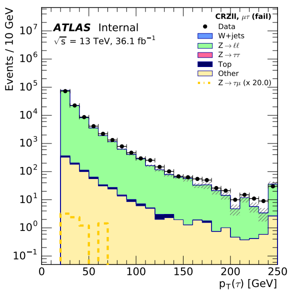
(c) CRW pass: $p_\mathrm{T}(\tau\text{-track})$



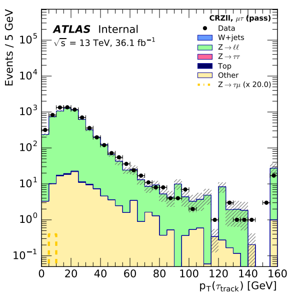
(d) CRW fail: $p_\mathrm{T}(\tau\text{-track})$



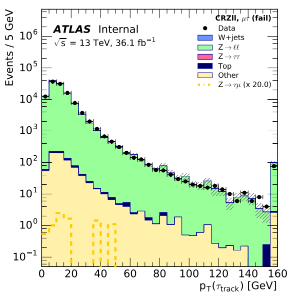
(e) CRZll pass: $p_\mathrm{T}(\tau)$



(f) CRZll fail: $p_\mathrm{T}(\tau)$

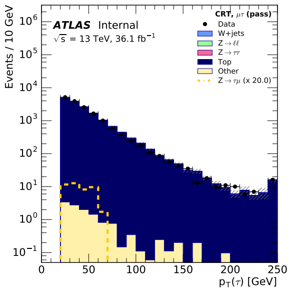


(g) CRZll pass: $p_\mathrm{T}(\tau\text{-track})$

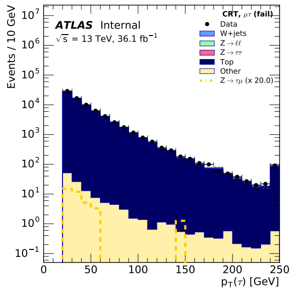


(h) CRZll fail: $p_\mathrm{T}(\tau\text{-track})$

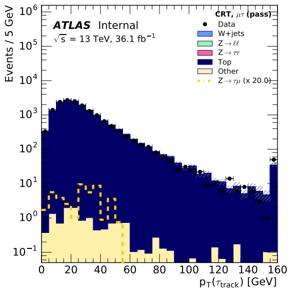
Figure 7.3: Data and Standard Model background distributions of the $p_\mathrm{T}(\tau)$ and $p_\mathrm{T}(\tau\text{-track})$ in the CRW and CRZll regions of the τ-μ channel. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20. The errors and statistical.



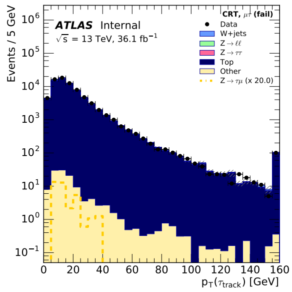
(a) CRT pass: $p_{\mathrm{T}}(\tau)$



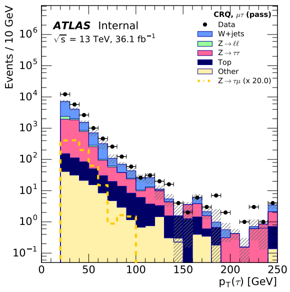
(b) CRT fail: $p_{\mathrm{T}}(\tau)$



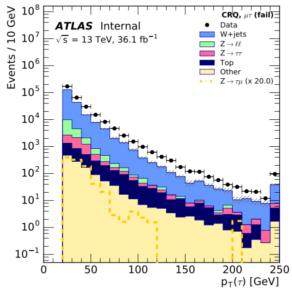
(c) CRT pass: $p_{\mathrm{T}}(\tau\text{-track})$



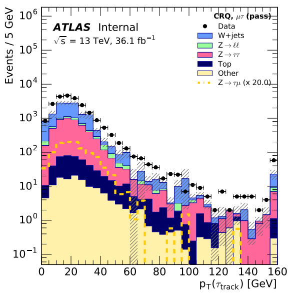
(d) CRT fail: $p_{\mathrm{T}}(\tau\text{-track})$



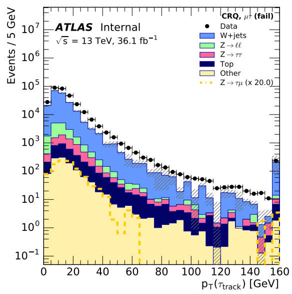
(e) CRQ pass: $p_{\mathrm{T}}(\tau)$



(f) CRQ fail: $p_{\mathrm{T}}(\tau)$



(g) CRQ pass: $p_{\mathrm{T}}(\tau\text{-track})$



(h) CRQ fail: $p_{\mathrm{T}}(\tau\text{-track})$

Figure 7.4: Data and Standard Model background distributions of the $p_{\mathrm{T}}(\tau)$ and $p_{\mathrm{T}}(\tau\text{-track})$ in the CRT and CRQ regions of the τ-μ channel.The multijet events in CRQ are shown as the difference between the data and the MC prediction for other backgrounds. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The errors and statistical.

7.3 Fake Factor Method Results

The individual fake-factors per each background are derived in dedicated control regions and shown in Tables 7.2 and 7.3. Due to the low statistics, in few bins ,the fake-factor is > 1. The MC correction k-Factors are derived in control regions and shown in Tables 7.4 and 7.5. The relative background contribution in the signal regions of the two channels are shown in Tables 7.6 and 7.7. Finally, the fake factor in the signal regions is shown in Table 7.8. Computed is the fake-factor without the k-correction as $F = \sum_i R_i F_i$ where $R_{\mathrm{QCD}} = 1 - \sum_{\text{other bkg}} R$ and the k-corrected fake-factor as $F_{\mathrm{corr}} = \sum_i R_i k_i F_i$ where $R_{\mathrm{QCD}} k_{\mathrm{QCD}} = 1 - \sum_{\text{other bkg}} Rk$. The k-corrected fake-factor is used in the analysis.

Table 7.2: Fake-factor per background computed in the dedicated control regions of the τe channel.

p_{T} (τ_{had})	track p_{T} (τ_{had})	FF(Wjets)	FF(top)	FF(Zll)	FF(QCD)
1-prong					
20-30	<15	0.300 ± 0.004	0.297 ± 0.034	0.299 ± 0.010	0.263 ± 0.022
20-30	15-20	0.532 ± 0.009	0.489 ± 0.064	0.601 ± 0.028	0.552 ± 0.038
20-30	20-60	1.402 ± 0.022	1.349 ± 0.136	1.386 ± 0.068	1.250 ± 0.064
20-30	>60	1.008 ± 0.133	0.453 ± 0.485	0.628 ± 0.286	0.141 ± 0.476
30-40	<15	0.330 ± 0.012	0.195 ± 0.064	0.330 ± 0.024	0.277 ± 0.031
30-40	15-20	0.408 ± 0.013	0.487 ± 0.139	0.408 ± 0.043	0.506 ± 0.056
30-40	20-60	0.782 ± 0.025	0.603 ± 0.092	0.759 ± 0.046	0.828 ± 0.048
30-40	>60	1.921 ± 0.377	0.000 ± 8.027	2.172 ± 2.082	2.632 ± 3.134
>40	<15	0.303 ± 0.012	0.144 ± 0.077	0.387 ± 0.044	0.276 ± 0.033
>40	15-20	0.331 ± 0.014	0.043 ± 0.146	0.354 ± 0.054	0.314 ± 0.044
>40	20-60	0.530 ± 0.010	0.203 ± 0.072	0.503 ± 0.034	0.514 ± 0.035
>40	>60	0.704 ± 0.030	0.474 ± 0.332	0.833 ± 0.138	0.606 ± 0.102
3-prong					
20-30	-	0.202 ± 0.004	0.203 ± 0.029	0.187 ± 0.010	0.238 ± 0.017
30-40	-	0.210 ± 0.006	0.148 ± 0.038	0.204 ± 0.015	0.233 ± 0.022
>40	-	0.204 ± 0.005	0.045 ± 0.044	0.195 ± 0.017	0.192 ± 0.016

Table 7.3: Fake-factor per background computed in the dedicated control regions of the $\tau\mu$ channel.

p_T (τ_{had})	track p_T (τ_{had})	FF(Wjets)	FF(top)	FF(Zll)	FF(QCD)
1-prong					
20-30	<15	0.287 ± 0.005	0.333 ± 0.035	0.277 ± 0.009	0.701 ± 0.396
20-30	15-20	0.524 ± 0.008	0.429 ± 0.064	0.504 ± 0.021	0.613 ± 0.186
20-30	20-60	1.392 ± 0.023	1.118 ± 0.178	1.399 ± 0.062	0.935 ± 0.285
20-30	>60	0.889 ± 0.092	1.090 ± 0.739	0.616 ± 0.176	0.000 ± 0.200
30-40	<15	0.334 ± 0.009	0.243 ± 0.066	0.311 ± 0.021	0.172 ± 0.135
30-40	15-20	0.395 ± 0.011	0.399 ± 0.100	0.311 ± 0.033	0.479 ± 0.313
30-40	20-60	0.784 ± 0.014	0.630 ± 0.087	0.782 ± 0.042	0.501 ± 0.217
30-40	>60	1.568 ± 0.298	0.346 ± 0.383	0.000 ± 1.804	0.119 ± 0.263
>40	<15	0.295 ± 0.010	0.003 ± 0.078	0.318 ± 0.035	0.206 ± 0.137
>40	15-20	0.320 ± 0.016	0.000 ± 0.114	0.189 ± 0.033	0.324 ± 0.227
>40	20-60	0.542 ± 0.011	0.386 ± 0.075	0.535 ± 0.031	0.363 ± 0.109
>40	>60	0.664 ± 0.031	0.178 ± 0.229	0.687 ± 0.114	1.254 ± 0.388
3-prong					
20-30	-	0.197 ± 0.005	0.200 ± 0.029	0.185 ± 0.009	0.298 ± 0.064
30-40	-	0.211 ± 0.006	0.294 ± 0.041	0.171 ± 0.012	0.447 ± 0.345
>40	-	0.196 ± 0.005	0.122 ± 0.044	0.163 ± 0.016	0.206 ± 0.057

Table 7.4: MC correction k-factor per background computed in the dedicated control regions of the τe channel.

p_T (τ_{had})	track p_T (τ_{had})	k(Wjets)	k(top)	k(Zll)
1-prong				
20-30	<15	0.867 ± 0.029	0.966 ± 0.043	0.949 ± 0.023
20-30	15-20	1.109 ± 0.077	0.989 ± 0.062	0.846 ± 0.035
20-30	20-60	1.342 ± 0.122	1.080 ± 0.077	1.089 ± 0.060
20-30	>60	1.046 ± 0.443	1.128 ± 0.514	0.690 ± 0.340
30-40	<15	1.039 ± 0.081	1.068 ± 0.075	1.150 ± 0.065
30-40	15-20	0.912 ± 0.101	0.758 ± 0.091	0.987 ± 0.083
30-40	20-60	1.615 ± 0.154	0.990 ± 0.071	0.987 ± 0.056
30-40	>60	-0.913 ± 1.221	0.266 ± 0.679	0.145 ± 0.126
>40	<15	1.007 ± 0.091	1.106 ± 0.088	1.313 ± 0.127
>40	15-20	1.150 ± 0.178	0.840 ± 0.114	1.330 ± 0.156
>40	20-60	1.158 ± 0.060	0.802 ± 0.050	1.086 ± 0.062
>40	>60	1.697 ± 0.181	0.646 ± 0.130	1.051 ± 0.172
3-prong				
20-30	-	0.892 ± 0.035	1.249 ± 0.054	1.146 ± 0.034
30-40	-	1.015 ± 0.063	1.273 ± 0.061	1.234 ± 0.055
>40	-	0.888 ± 0.038	1.033 ± 0.052	1.275 ± 0.066

Table 7.5: MC correction k-factor per background computed in the dedicated control regions of the $\tau\mu$ channel.

p_T (τ_{had})	track p_T (τ_{had})	k(Wjets)	k(top)	k(Zll)
1-prong				
20-30	<15	0.861 ± 0.025	1.011 ± 0.042	0.996 ± 0.021
20-30	15-20	0.968 ± 0.059	0.932 ± 0.058	0.929 ± 0.033
20-30	20-60	1.475 ± 0.133	1.247 ± 0.131	1.018 ± 0.051
20-30	>60	0.781 ± 0.317	1.051 ± 0.580	1.428 ± 0.442
30-40	<15	0.972 ± 0.064	1.001 ± 0.075	1.105 ± 0.053
30-40	15-20	0.867 ± 0.089	0.995 ± 0.099	1.112 ± 0.087
30-40	20-60	1.400 ± 0.113	1.049 ± 0.064	1.039 ± 0.056
30-40	>60	0.469 ± 0.221	1.343 ± 0.625	0.589 ± 0.354
>40	<15	0.960 ± 0.085	1.068 ± 0.092	1.275 ± 0.103
>40	15-20	1.039 ± 0.110	1.083 ± 0.126	1.390 ± 0.143
>40	20-60	1.166 ± 0.089	0.852 ± 0.050	1.085 ± 0.054
>40	>60	1.855 ± 0.209	0.929 ± 0.158	1.025 ± 0.161
3-prong				
20-30	-	0.907 ± 0.033	1.388 ± 0.065	1.104 ± 0.030
30-40	-	0.923 ± 0.049	1.191 ± 0.060	1.306 ± 0.055
>40	-	0.849 ± 0.033	1.113 ± 0.056	1.235 ± 0.061

Table 7.6: R-factor in the signal region of the τe channel.

p_T (τ_{had})	track p_T (τ_{had})	R(Wjets)	R(top)	R(Zll)	R(QCD)
1-prong					
20-30	<15	0.750 ± 0.037	0.004 ± 0.000	0.054 ± 0.002	0.193 ± 0.037
20-30	15-20	0.617 ± 0.053	0.003 ± 0.000	0.054 ± 0.003	0.326 ± 0.053
20-30	20-60	0.395 ± 0.058	0.003 ± 0.000	0.054 ± 0.004	0.548 ± 0.058
20-30	>60	0.871 ± 0.733	0.003 ± 0.003	0.124 ± 0.065	0.002 ± 0.736
30-40	<15	0.643 ± 0.072	0.005 ± 0.001	0.052 ± 0.003	0.300 ± 0.072
30-40	15-20	0.492 ± 0.102	0.005 ± 0.001	0.060 ± 0.006	0.443 ± 0.102
30-40	20-60	0.331 ± 0.072	0.005 ± 0.001	0.061 ± 0.005	0.604 ± 0.072
30-40	>60	0.929 ± 0.709	0.000 ± 0.000	0.000 ± 0.000	0.071 ± 0.709
>40	<15	0.451 ± 0.085	0.005 ± 0.001	0.046 ± 0.007	0.497 ± 0.085
>40	15-20	0.792 ± 0.174	0.008 ± 0.001	0.042 ± 0.006	0.158 ± 0.174
>40	20-60	0.565 ± 0.093	0.008 ± 0.001	0.058 ± 0.005	0.370 ± 0.093
>40	>60	0.304 ± 0.109	0.006 ± 0.003	0.067 ± 0.013	0.623 ± 0.110
3-prong					
20-30	-	0.657 ± 0.028	0.005 ± 0.000	0.046 ± 0.001	0.292 ± 0.028
30-40	-	0.563 ± 0.041	0.006 ± 0.000	0.044 ± 0.002	0.388 ± 0.041
>40	-	0.605 ± 0.044	0.008 ± 0.000	0.052 ± 0.002	0.335 ± 0.044

Table 7.7: R-factor in the signal region of the $\tau\mu$ channel.

p_{T} (τ_{had})	track p_{T} (τ_{had})	R(Wjets)	R(top)	R(Zll)	R(QCD)
1-prong					
20-30	<15	0.910 ± 0.041	0.004 ± 0.000	0.071 ± 0.002	0.015 ± 0.041
20-30	15-20	0.815 ± 0.064	0.005 ± 0.000	0.071 ± 0.003	0.109 ± 0.064
20-30	20-60	0.542 ± 0.074	0.004 ± 0.000	0.062 ± 0.004	0.391 ± 0.075
20-30	>60	0.549 ± 0.497	0.004 ± 0.003	0.068 ± 0.020	0.379 ± 0.497
30-40	<15	0.870 ± 0.098	0.007 ± 0.001	0.072 ± 0.004	0.051 ± 0.098
30-40	15-20	0.908 ± 0.116	0.007 ± 0.001	0.065 ± 0.006	0.021 ± 0.116
30-40	20-60	0.685 ± 0.089	0.008 ± 0.001	0.082 ± 0.006	0.225 ± 0.090
30-40	>60	0.905 ± 0.786	0.002 ± 0.002	0.081 ± 0.046	0.013 ± 0.787
>40	<15	0.789 ± 0.112	0.011 ± 0.001	0.071 ± 0.006	0.129 ± 0.112
>40	15-20	0.832 ± 0.217	0.009 ± 0.001	0.108 ± 0.024	0.050 ± 0.219
>40	20-60	0.689 ± 0.067	0.011 ± 0.001	0.081 ± 0.006	0.219 ± 0.068
>40	>60	0.414 ± 0.158	0.008 ± 0.002	0.130 ± 0.028	0.448 ± 0.160
3-prong					
20-30	-	0.931 ± 0.093	0.006 ± 0.000	0.058 ± 0.002	0.005 ± 0.094
30-40	-	0.831 ± 0.058	0.007 ± 0.000	0.062 ± 0.003	0.099 ± 0.058
>40	-	0.921 ± 0.063	0.011 ± 0.001	0.059 ± 0.003	0.009 0.063

Table 7.8: Fake-factor in the signal region for the τe and $\tau\mu$ channels.

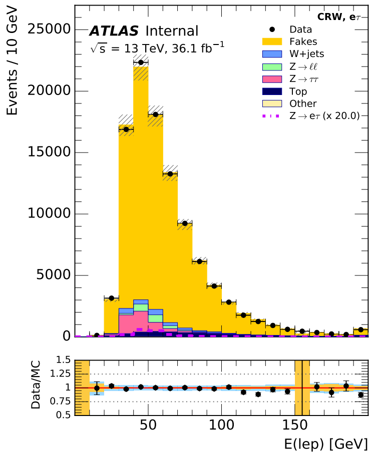
p_{T} (τ_{had})	track p_{T} (τ_{had})	FF(τe)	k-corr. FF(τe)	FF($\tau\mu$)	k-corr. FF($\tau\mu$)
1-prong					
20-30	<15	0.292 ± 0.015	0.289 ± 0.017	0.293 ± 0.032	0.345 ± 0.064
20-30	15-20	0.542 ± 0.043	0.540 ± 0.059	0.532 ± 0.056	0.535 ± 0.069
20-30	20-60	1.318 ± 0.116	1.339 ± 0.175	1.213 ± 0.168	1.331 ± 0.225
20-30	>60	0.957 ± 0.757	0.972 ± 0.986	0.534 ± 0.451	0.446 ± 0.393
30-40	<15	0.314 ± 0.033	0.315 ± 0.041	0.323 ± 0.038	0.320 ± 0.043
30-40	15-20	0.452 ± 0.071	0.456 ± 0.074	0.391 ± 0.073	0.400 ± 0.091
30-40	20-60	0.807 ± 0.088	0.798 ± 0.147	0.719 ± 0.097	0.782 ± 0.138
30-40	>60	1.972 ± 2.349	2.632 ± 5.284	1.421 ± 1.274	0.813 ± 0.713
40-80	<15	0.293 ± 0.039	0.294 ± 0.043	0.282 ± 0.045	0.281 ± 0.050
40-80	15-20	0.327 ± 0.081	0.330 ± 0.113	0.303 ± 0.101	0.298 ± 0.113
40-80	20-60	0.520 ± 0.070	0.522 ± 0.084	0.501 ± 0.051	0.522 ± 0.067
40-80	>60	0.650 ± 0.121	0.672 ± 0.186	0.927 ± 0.287	0.717 ± 0.439
3-prong					
20-30	-	0.212 ± 0.010	0.214 ± 0.012	0.197 ± 0.034	0.202 ± 0.033
30-40	-	0.218 ± 0.016	0.217 ± 0.019	0.232 ± 0.045	0.242 ± 0.060
40-80	-	0.198 ± 0.014	0.197 ± 0.015	0.193 ± 0.019	0.194 ± 0.020

7.4 Fake-Factor Validation

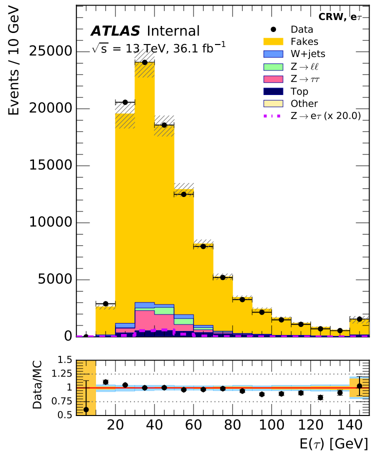
The estimates obtained using the fake-factor method are validated in two regions: W+jets control region (CRW) since the W+jets are the largest jet $\rightarrow \tau$ background and the same-sign signal region *i.e.* a region with the same selection as the signal region, but with the inverted charge selection.

The distributions of NN-inputs in CRW and same-sign signal region are shown in Figures 7.5, 7.6, 7.7 and 7.8 for the $Z \rightarrow \tau e$ and 7.9, 7.10, 7.11 and 7.12 for the $Z \rightarrow \tau\mu$ channel.

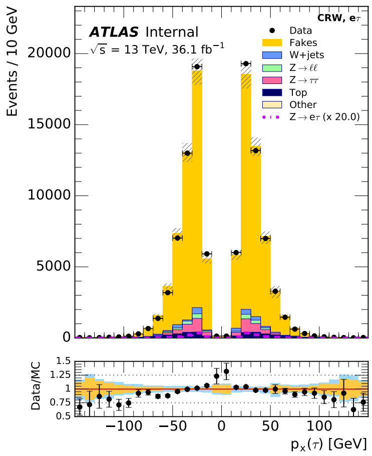
Overall, the CRW distributions serve as a closure test and show a good predictions in the control region. The distributions in the same-sign signal region show certain discrepancy and the necessity to correct the fake-factor normalization *i.e.* to use the *Fakes* normalization factor as a free parameter in the final template fit.



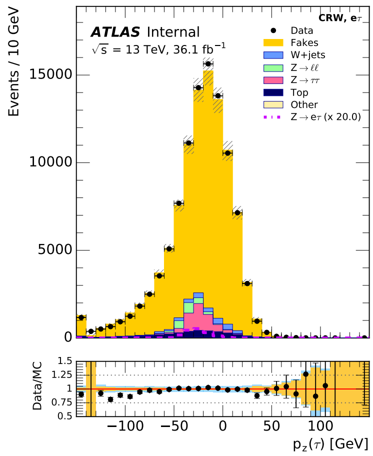
(a) $\hat{E}^{\mathrm{lep}}$



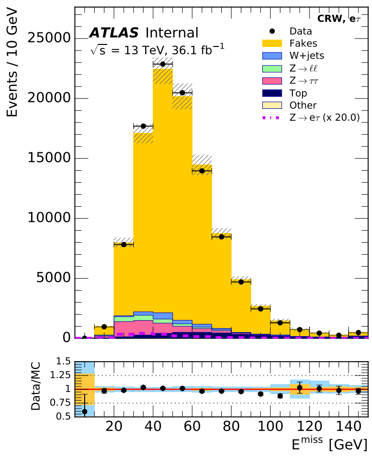
(b) $\hat{E}^{\tau_h}$



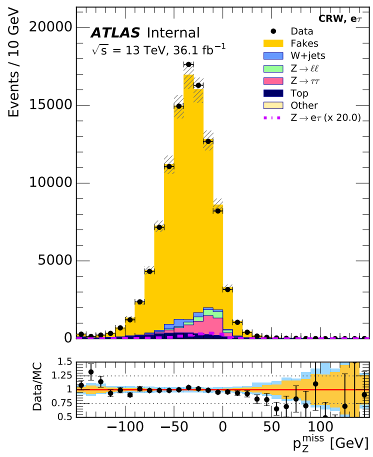
(c) $\hat{p}_x^{\tau_h}$



(d) $\hat{p}_z^{\tau_h}$

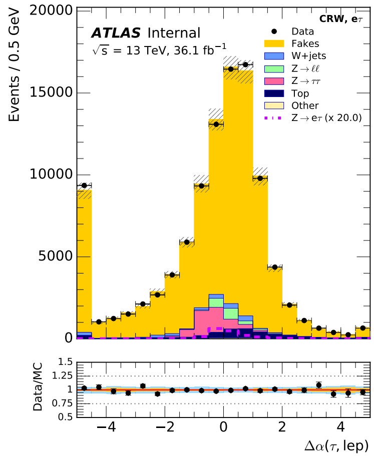


(e) $\hat{E}^{\mathrm{miss}}$

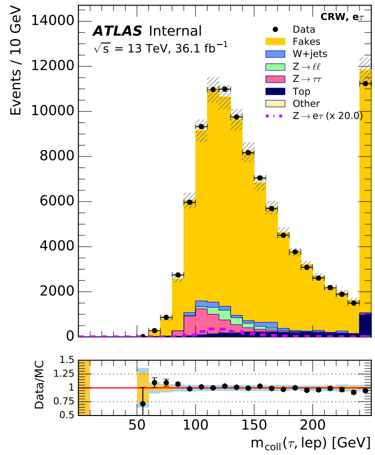


(f) $\hat{p}_z^{\mathrm{miss}}$

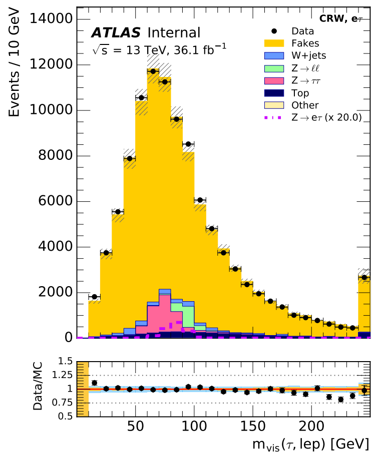
Figure 7.5: Data and Standard Model background distributions of the NN inputs in the W+jets control region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.



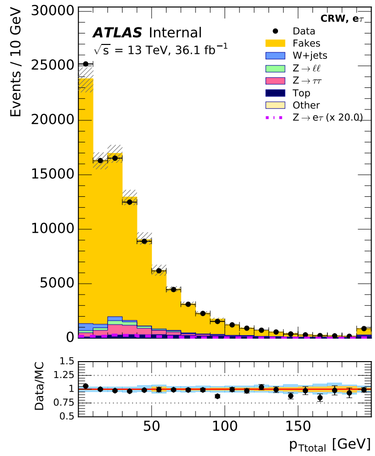
(a) $\Delta\alpha$



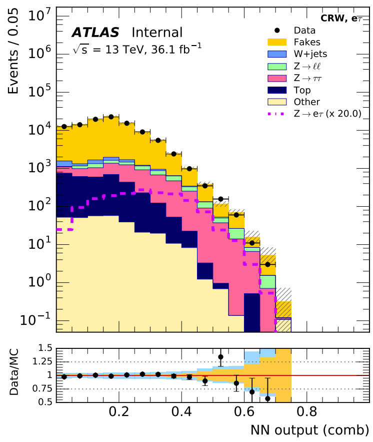
(b) m_{coll}



(c) $m_{\text{vis(lep},\tau_h)}$

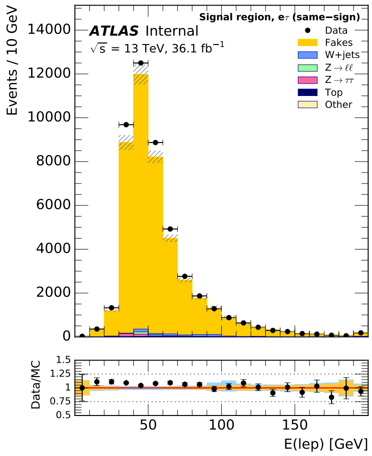


(d) $p_{\text{T}}^{\text{tot}}$

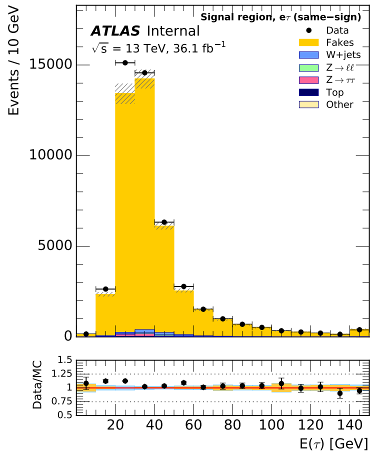


(e) $p_{\text{T}}^{\text{tot}}$

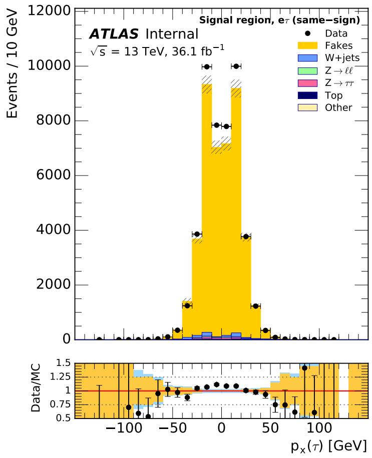
Figure 7.6: Data and Standard Model background distributions of the NN inputs and the combined output in the W+jets control region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.



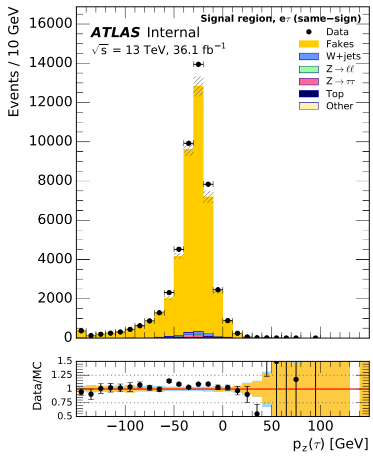
(a) $\hat{E}^{\text{lep}}$



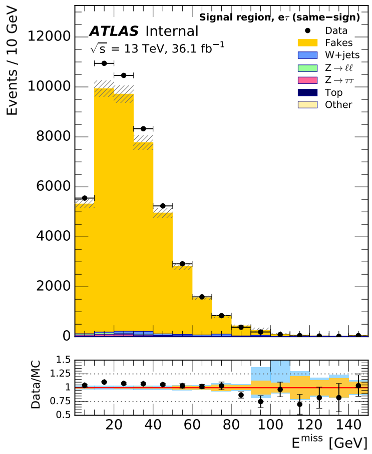
(b) $\hat{E}^{\tau_h}$



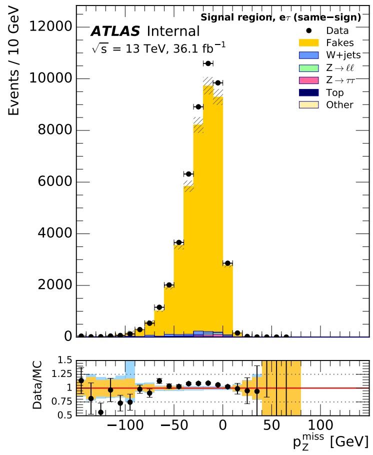
(c) $\hat{p}_x^{\tau_h}$



(d) $\hat{p}_z^{\tau_h}$

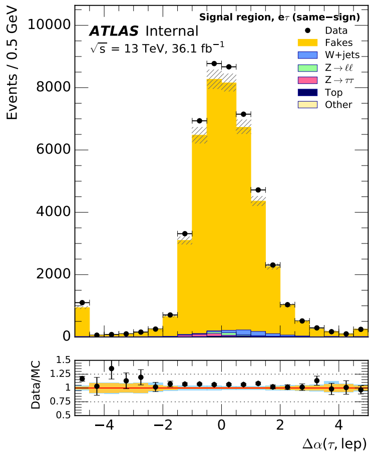


(e) $\hat{E}^{\text{miss}}$

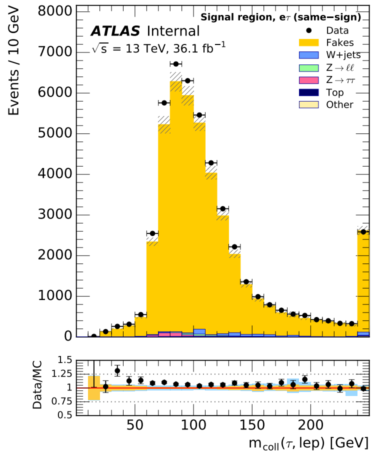


(f) $\hat{p}_z^{\text{miss}}$

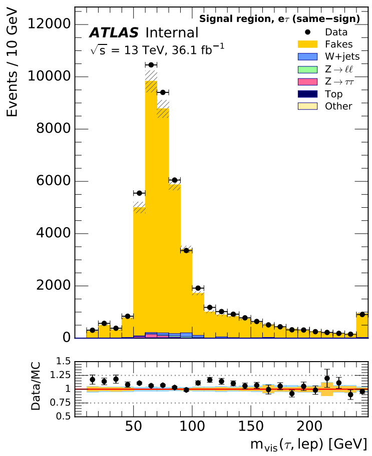
Figure 7.7: Data and Standard Model background distributions of the NN inputs in the same-sign signal region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.



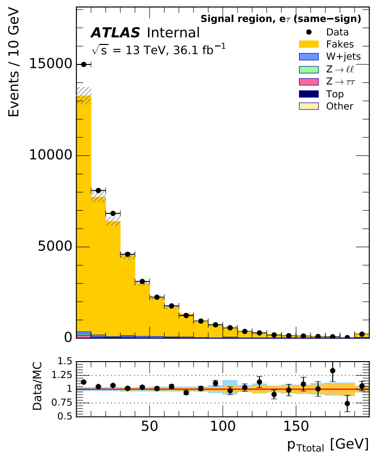
(a) $\Delta\alpha$



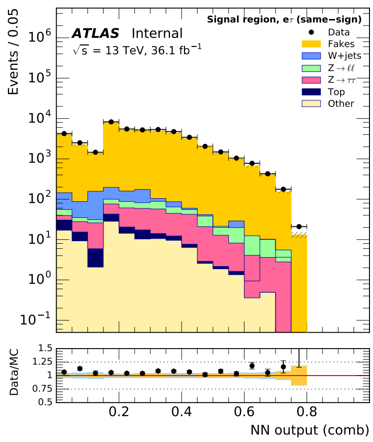
(b) m_{coll}



(c) $m_{\mathrm{vis(lep},\tau_h)}$

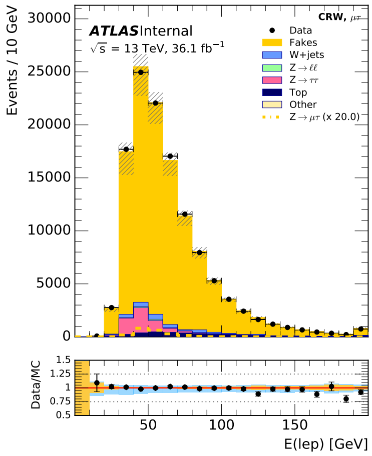


(d) $p_{\mathrm{T}}^{\mathrm{tot}}$

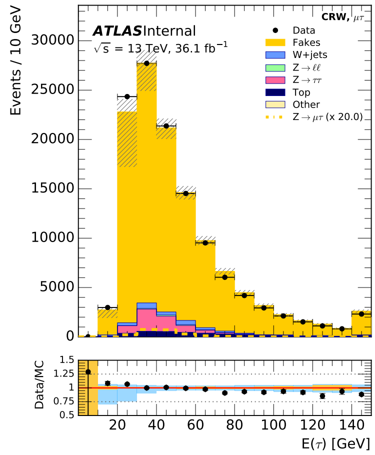


(e) $p_{\mathrm{T}}^{\mathrm{tot}}$

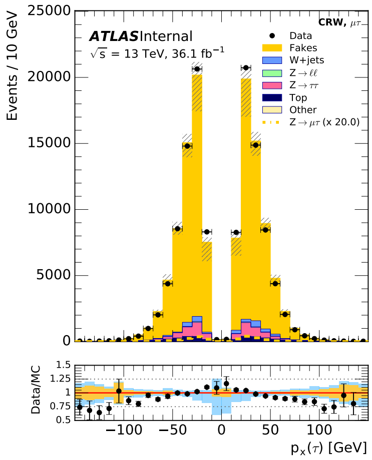
Figure 7.8: Data and Standard Model background distributions of the NN inputs and the combined output in the same-sign signal region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.



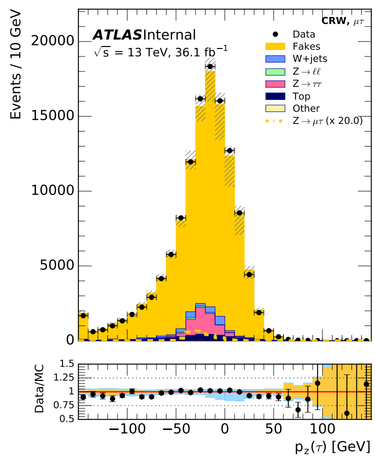
(a) $\hat{E}^{\mathrm{lep}}$



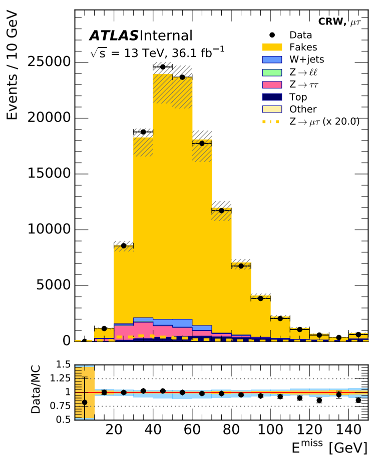
(b) $\hat{E}^{\tau_h}$



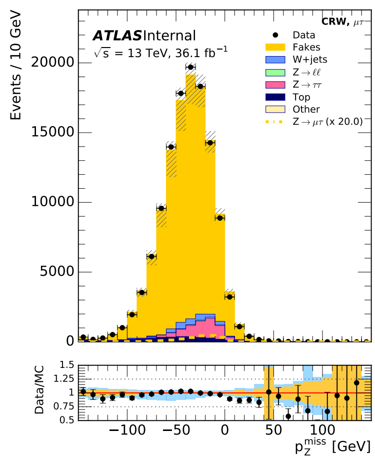
(c) $\hat{p}_x^{\tau_h}$



(d) $\hat{p}_z^{\tau_h}$

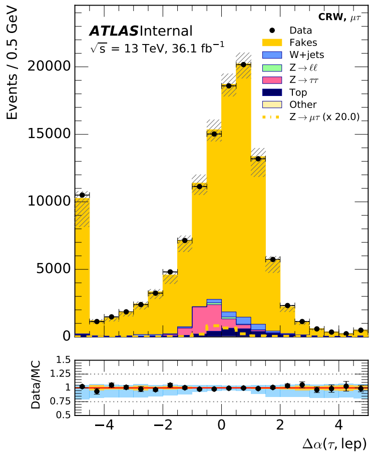


(e) $\hat{E}^{\mathrm{miss}}$

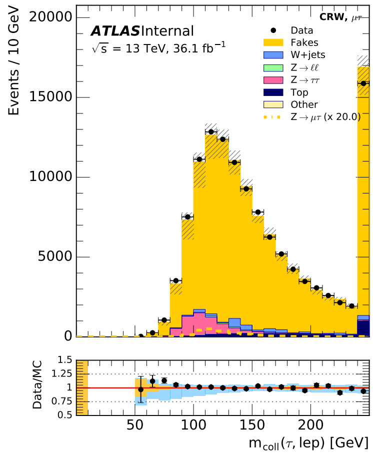


(f) $\hat{p}_z^{\mathrm{miss}}$

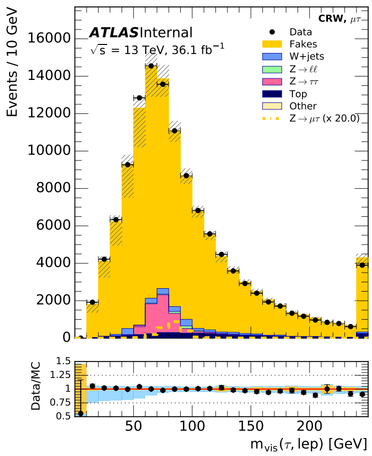
Figure 7.9: Data and Standard Model background distributions of the NN inputs in the W+jets control region of the $\tau - \mu$ channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.



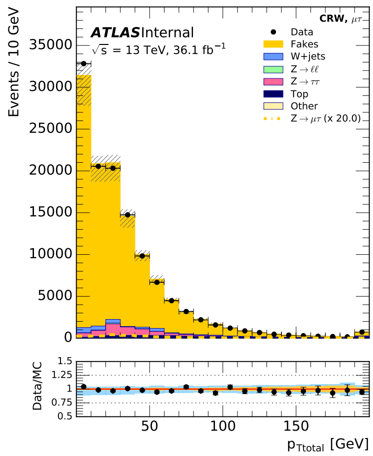
(a) $\Delta\alpha$



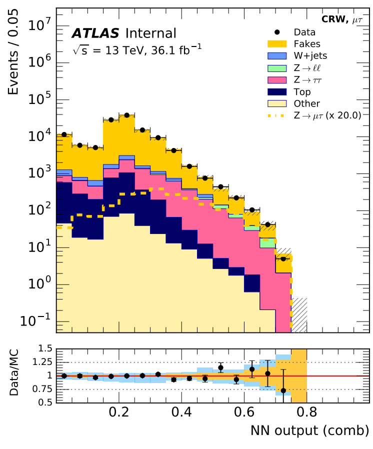
(b) m_{coll}



(c) $m_{\text{vis(lep},\tau_h)}$

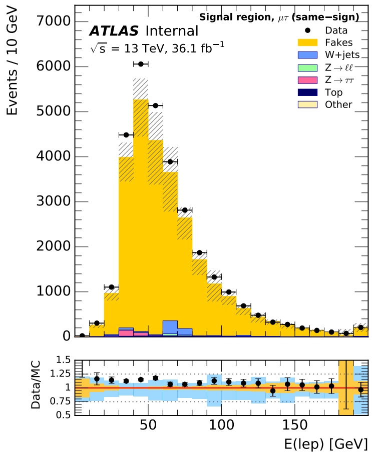


(d) $p_{\text{T}}^{\text{tot}}$

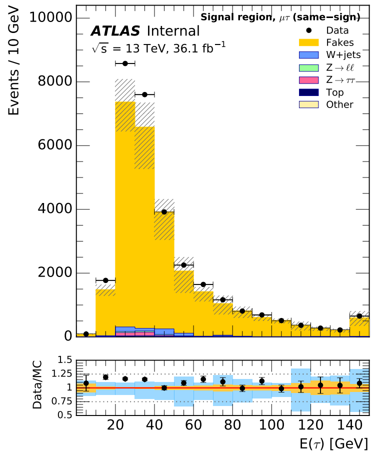


(e) $p_{\text{T}}^{\text{tot}}$

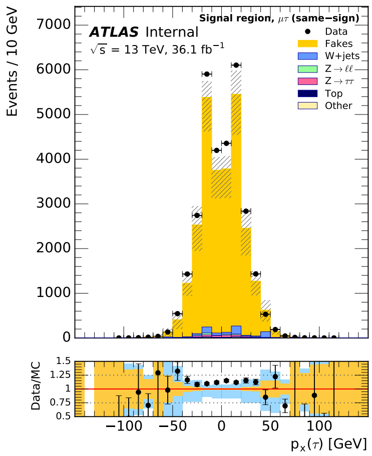
Figure 7.10: Data and Standard Model background distributions of the NN inputs and the combined output in the W+jets control region of the channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.



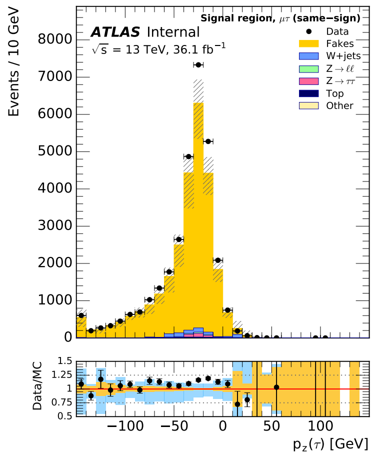
(a) $\hat{E}^{\mathrm{lep}}$



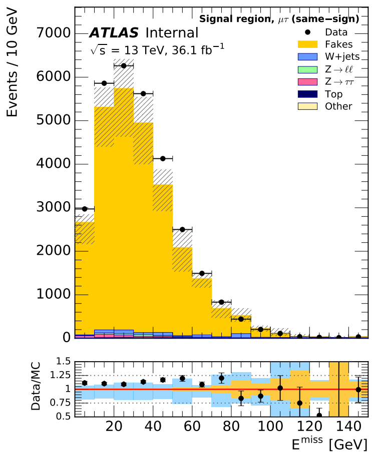
(b) $\hat{E}^{\tau_h}$



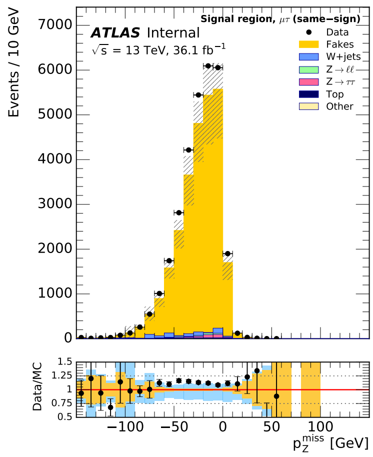
(c) $\hat{p}_x^{\tau_h}$



(d) $\hat{p}_z^{\tau_h}$

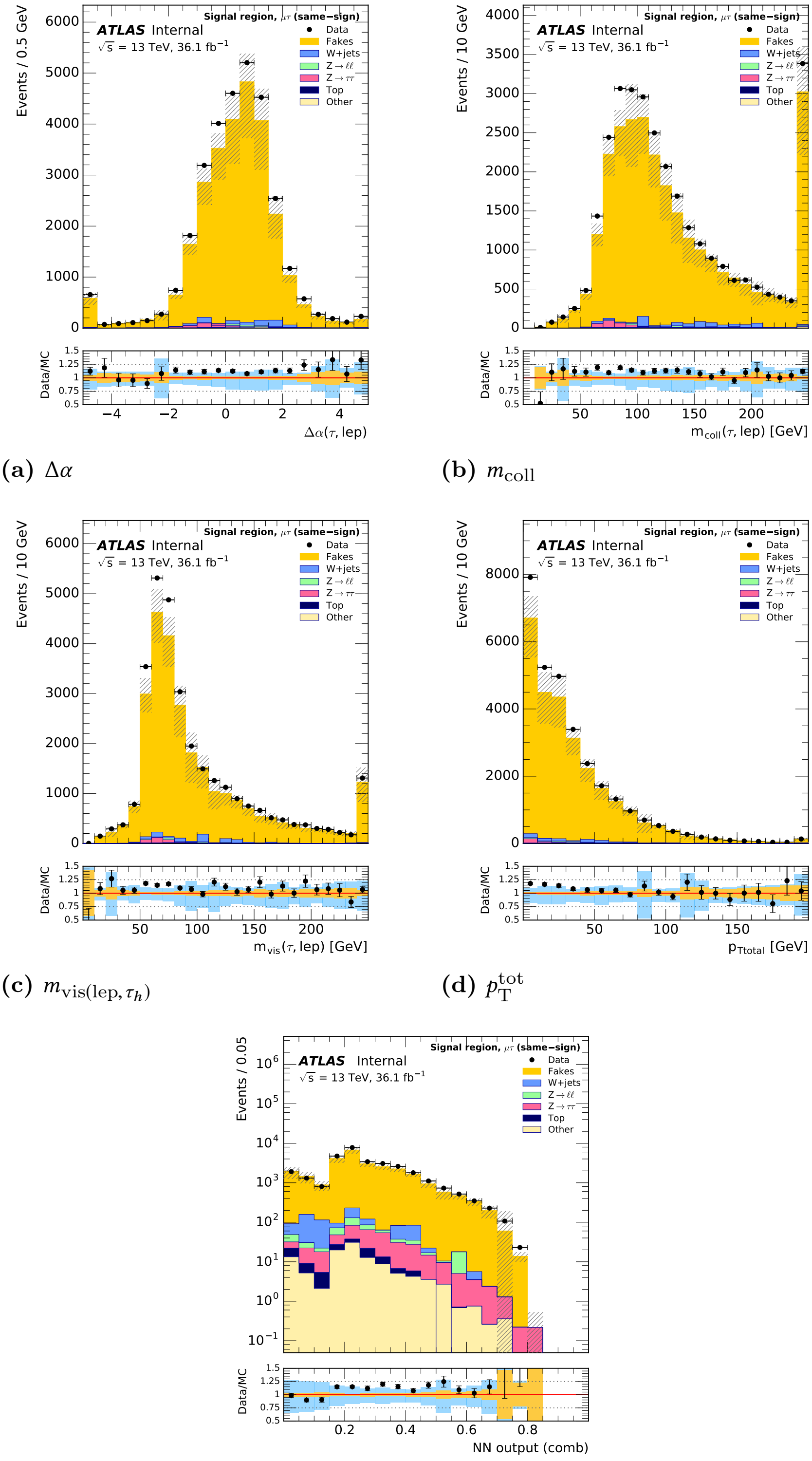


(e) $\hat{E}^{\mathrm{miss}}$



(f) $\hat{p}_z^{\mathrm{miss}}$

Figure 7.11: Data and Standard Model background distributions of the NN inputs in the same-sign signal region of the $\tau-\mu$ channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.



(a) $\Delta\alpha$

(b) m_{coll}

(c) $m_{\text{vis(lep},\tau_h)}$

(d) $p_{\text{T}}^{\text{tot}}$

(e) $p_{\text{T}}^{\text{tot}}$

Figure 7.12: Data and Standard Model background distributions of the NN inputs and the combined output in the same-sign signal region of the $\tau-\mu$ channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of $36.1\,\text{fb}^{-1}$ and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error.

Fit Model and Results

This chapter gives an overview of the used statistical treatment in Section 8.1, followed by the description of the fit model in Section 8.2. and the systematical uncertainties in Section 8.3. The validation of the fit and an unblinded[1] sensitivity test are done using Asimov data which is described in Section 8.4. Section 8.5 describes the sideband fit, done in order to validate the fit model and its predictions in the sideband of the signal region. Finally, the results obtained in the unblinded fit are given in Section 8.6.

8.1 Statistical Treatment

The parameter estimation is based on the *maximum likelihood* method where likelihood is maximized by adjusting the values of the unknown parameters. The likelihood function is describing the joint probability to obtain the observed data, given the model which can be expressed as [49]:

$$\mathrm{L}(x;\theta) = \prod_{\mathrm{i}}^{\mathrm{Nobs}} f(x_i;\theta), \tag{8.1}$$

where θ denotes the unknown parameters which are being estimated and x_i describes the measurement. If the events are grouped into the bins of the measured observable, the binned likelihood fit can be defined as a product of Poissons included per each bin of the discriminating variable which quantifies the probability to obtain the observed number of events, given the expectation from the model.

The likelihood function can depend on several parameters such as the signal branching ratio, the background normalization factors and the parameters associated to the systematical and statistical uncertainties. The parameter one is interested in is called the *parameter of interest*, while the remaining parameters are called the *nuisance parameters* θ. In this analysis, the parameter of interest is the branching ratio of the $Z \to \tau e$ and the $Z \to \tau\mu$ decays, while the nuisance parameters are the background normalization factors and the parameters describing the effect of the systematical uncertainties.

Typically, the parameter of interest is fitted in signal - enriched regions called *signal regions*, while the background normalization factors can be derived in background - enriched *normalization regions* and further extrapolated to the signal regions, under the assumption that the signal and the normalization regions are kinematically similar.

The likelihood function can be built to simultaneously describe multiple signal and nor-

[1] *Unblinded* test refers to a test done without looking at real data

malization regions: [50]:

$$L = \prod_{\text{SRs}} P_{\text{SR}} \times \prod_{\text{NRs}} P_{\text{NR}} \times C_{\text{syst}}. \tag{8.2}$$

where P_{SR} and P_{NR} factors describe the probability to obtain the observed number of events in the signal and normalisation region, given the expected. The C_{syst} factor describes the impact of systematical uncertanties and for independent systematical uncertanties, it is built as a product of probability distributions constructed for each systematical uncertainty.

If the likelihood function depends on multiple parameters μ and θ, it can be maximized by allowing all parameters to float to obtain the *unconditional maximum likelihood* estimates, denoted as $\hat{\mu}, \hat{\theta}$. If one parameter is fixed (typically the parameter of interest μ) while the others are allowed to float, the obtained estimators are called *conditional maximum likelihood* estimators and denoted as $\mu, \hat{\hat{\theta}}$, where the notation of μ is telling that the parameter is not estimated from the fit, but fixed at a given value.

In the analysis, two types of tests are done:

- **Discovery test** attempting to reject the background (null) hypothesis in order to claim a discovery.
- **Exclusion test** attempting to reject the signal+background (null) hypothesis, for a specific branching ratio of the signal process which is proportional to signal strength μ; if the background hypothesis (Discovery test) can not be rejected, the upper limit is set by scanning the branching ratio and by excluding the values for which the null hypothesis can be rejected.

In both tests, the test statistic is the profile log likelihood ratio. In the **discovery test**, the test statistic q_0 is defined as [51]:

$$q_0 = \begin{cases} -2\ln\frac{L(0,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}, & \hat{\mu} \geq 0. \\ 0, & \hat{\mu} < 0. \end{cases} \tag{8.3}$$

where the $L(0, \hat{\hat{\theta}})$ is the likelihood computed for the background hypothesis ($\mu = 0$), while the $L(\hat{\mu}, \hat{\theta})$ is the likelihood computed at the best-fit values for μ and θ. From Equation 8.3, it is seen that if the best-fit value for μ is negative, the test statistic q_0 is set at 0. Based on the observed q_0, the p-value for the null (background) hypothesis is computed as [51]:

$$p_b = \int_{q_{0,\text{obs}}} f(q_0) dq_0 \tag{8.4}$$

where $f(q_0)$ is the distribution of the q_0 test-statistic under the null hypothesis and the direction of the integration is towards the less likely results under the null hypothesis. If p_b is small enough, the background hypothesis can be rejected with certain confidence. In order to reject the background hypothesis, it is customary to require the p-value to be less than $3 \cdot 10^{-7}$ *i.e.* the result to be significant at the 5σ level.

If the background hypothesis can not be rejected, the **exclusion test** is performed in order to set the upper limit on the signal branching ratio. The test statistic q_μ is defined as [51]:

$$q_\mu = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}, & \hat{\mu} \leq \mu. \\ 0, & \hat{\mu} > \mu. \end{cases} \tag{8.5}$$

where $L(\mu, \hat{\hat{\theta}})$ is the likelihood computed at a fixed μ for which the nuisance parameters θ are fitted. The $L(\hat{\mu}, \hat{\theta})$ is the likelihood computed at the best-fit values for μ and θ. If the hypothesized μ is smaller than the best-fit value $\hat{\mu}$, the test-statistic q_μ is set at 0. The test statistic q_μ is used to compute the p-value for a series of hypothesis for the value of μ. When setting the upper limit, it is required that the p-value[2] is less than 0.05 for the μ to be excluded, with the highest excluded value defining the upper limit.

In **exclusion test**, the common approach at ATLAS is to not use the p-value for the null hypothesis, but to use the CL_s-value[52, 53], defined as:

$$\mathrm{CLs} = \frac{CL_{s+b}}{CL_b} \tag{8.6}$$

where $\mathrm{CL_{s+b}}$ describes the probability to obtain the observed or a less likely result, under the signal+background hypothesis *i.e.* the probability for the signal to underfluctuate:

$$\mathrm{CL_{s+b}} \equiv p_{s+b} = \mathrm{prob}(q_{\mu|\mu'} \leq q_\mu^{\mathrm{obs}}) \tag{8.7}$$

The $\mathrm{CL_b}$ is defined as:

$$\mathrm{CL_b} \equiv 1 - p_b = \mathrm{prob}(q_{\mu|\mu=0} \leq q_\mu^{\mathrm{obs}}) \tag{8.8}$$

Hence the $1 - \mathrm{CL_b}$ describes the probability to obtain the observed or a less likely result, under the background hypothesis *i.e.* the probability for the background to overfluctuate. The $\mathrm{CL_{s+b}}$, $\mathrm{CL_b}$ and CLs are illustrated in left part of Figure 8.1. The $\mathrm{CL_b}$ quantifies the sensitivity *i.e.* the separation between the two hypothesis, as illustrated in the right part of Figure 8.1., where it is shown that large values of $\mathrm{CL_b}$ indicate a good sensitivity.

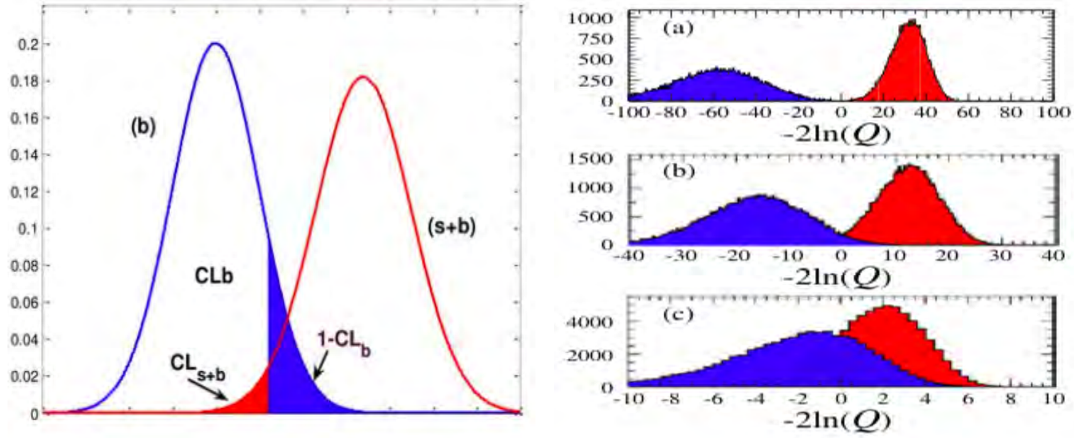
If instead of $\mathrm{CL_s}$, the p_{s+b} is used to decide on the hypothesis, it can be shown that the two approaches give similar results for good sensitivity, while for low sensitivity, the $\mathrm{CL_s}$ is more conservative and a signal which would be excluded using the p_{s+b} does not have to be excluded using the CL_s approach, if the sensitivity to the signal is low.

Further, the p-values for both hypothesis depend on the value of signal strength μ. As the observed value for the test statistic depends of μ, the integration limit and further the both the $\mathrm{CL_b}$ and $\mathrm{CL_{s+b}}$ depend on the scanned μ value.

In both tests, in order to compute the p-values, the distribution of the test-statistic needs to be constructed, either using the toy Monte Carlo method or the assymptotic approximation [51]. In this analysis, the assymptotic approximation which holds for large samples is used.

The statistical treatment in this analysis is done using the HistFitter [50] package, built upon the HistFactory [54] , RooStats [55] and RooFit [56] packages.

[2] In the CLs method, instead of the p-value, used is the CLs value, defined as a ratio of probabilities



(a) (b)

Figure 8.1: Left: The background (blue) and the signal+background (red) hypothesis. The $1 - \mathrm{CL_b}$(shaded blue area) is describing the probability to obtain the observed or less likely result under the background hypothesis, while the $\mathrm{CL_{s+b}}$(shaded red area) is describing the probability to obtain the observed or less likely result under the signal+background hypothesis. Right: $\mathrm{CL_b}$ is used as a measure of the separation of the two hypothesis (sensitivity). The top right plot has the largest $\mathrm{CL_b}$ and the highest sensitivity, while the bottom right plot has the smallest $\mathrm{CL_b}$ and the lowest sensitivity.

8.2 Fit Model

Binned-likelihood fit using the *neural-network score* (NN) as the final discriminating variable is done. The fit is performed separately for the τ-μ and τ-e channels, in the signal region defined in Tables 6.4 and 6.5.

Large backgrounds ($Z \to ee/\mu\mu$, $Z \to \tau\tau$ and fakes) are corrected using the data. The $Z \to ee/\mu\mu$ MC prediction is corrected before the fit, by deriving a scale factor in the $Z \to \mu\mu$ normalization region and extrapolating it to the signal region. In the fit model, the Z-boson production is corrected by applying a normalization $\mu(Z)$ factor to the $Z \to \tau\tau$ and the signal samples. The $Z \to ee/\mu\mu$ is corrected independently since this background is almost entirely formed due to the $e/\mu \to \tau$ fakes which could include additional effects from the real $Z \to \tau\tau$ background.

The normalization of the fakes is allowed to float in the fit, with a separate normalization factor assigned to 1-prong and 3-prong events in the fakes background, due to the independent derivation of fake-factors for 1-prong and 3-prong events. The minor backgrounds are estimated from MC simulation, with the dedicated systematical uncertainties on the cross section and the luminosity.

The $Z \to \tau\tau$ and the fakes backgrounds are normalized in the signal region. Since the discriminating neural-network score distribution contains backgrounds mainly at the low scores and the signal mainly at the large scores, the normalization of the backgrounds is mainly sensitive to the low sideband of the neural-network score distribution. The $Z \to \tau\tau$ normalization in a dedicated $Z \to \tau\tau$ normalization region was attempted, but due to the different kinematical properties with respect to the signal region, the derived scale factor was not providing a good prediction of the $Z \to \tau\tau$ background in the low sideband of the signal region, where the model was validated.

The likelihood L under the signal+background hypothesis is defined as:

$$L = P_{\mathrm{SR}} \times C_{\mathrm{syst}} = \prod_{\mathrm{b} \in \mathrm{bins}} \mathrm{P}_{\mathrm{b}}^{\mathrm{SR}}(n_{\mathrm{b}}^{\mathrm{SR,obs}} | n_{\mathrm{b}}^{\mathrm{SR,exp}}) \times C_{\mathrm{syst}}. \tag{8.9}$$

were P_{SR} is a term describing the discriminating neural-network distribution in the signal region defined as a product of Poisson probabilities to observe $n_{\mathrm{b}}^{\mathrm{SR,obs}}$ in bin b when the expected number of events is $n_{\mathrm{b}}^{\mathrm{SR,exp}}$ and C_{syst} term is describing the effect of systematical uncertanties.

The expected number of events in signal region per bin b of the discriminating neural network score $n_{\mathrm{b}}^{\mathrm{SR,exp}}$ is defined as:

$$n_{\mathrm{b}}^{\mathrm{SR,exp}} = \mu_{\mathrm{sig}} \mu_{\mathrm{Z}} n_{\mathrm{b}}^{\mathrm{exp,sig}} + \mu_{\mathrm{Z}} n_{\mathrm{b}}^{\mathrm{exp},Z\to\tau\tau} + \mu_{\mathrm{F,1p}} n_{\mathrm{b}}^{\mathrm{exp},F,1p} + \mu_{\mathrm{F,3p}} n_{\mathrm{b}}^{\mathrm{exp},F,3p} + \text{other MC} \tag{8.10}$$

where the μ_{sig} is the signal strength defined as the factor multiplying the signal sample, initially normalized to the branching ratio of 10^{-5}. This means that the fitted signal strength of 1.00 relates to the branching ratio of 10^{-5}. The μ_{Z}, $\mu_{\mathrm{F,1p}}$ and $\mu_{\mathrm{F,3p}}$ are normalization factors related to $Z \to \tau\tau$, 1-prong fakes and 3-prong fakes backgrounds.The μ_{Z} factor, derived from the normalization of the $Z \to \tau\tau$ background is applied to the signal sample as well in order to correct the Z production. Other MC backgrounds, including the corrected $Z \to ee/\mu\mu$ are included by the textother MC term.

8.3 Systematical Uncertainties

The systematical uncertainties arise due to:

- Detector simulation: uncertainties due to the calibration, resolutions and efficiency of the reconstruction and identification of objects,
- Theoretical uncertainties: uncertainties due to the factorization and renormalization scale and due to the choice of the MC generator,
- Monte Carlo statistical uncertainties.

They are typically incorporated into the likelihood function, using the Gaussian parametrization:

$$C_{\mathrm{syst}}(\theta^0, \theta) = G(\theta^0 - \theta) \tag{8.11}$$

where θ^0 is fixed at the pre-fit value, while θ is the floating nuisance parameter.

An overview of systematical uncertanties, used in the analysis, is given in Tables 8.2 and 8.1. Shown are the uncertainties due to the simulated objects which are applied to Monte Carlo samples, the systematics applied to data-driven *Fakes* and the luminosity and the theoretical uncertainties on the cross sections, applied to the smaller backgrounds which are not normalized in the fit.

Uncertainties on scale factors of simulated objects are applied to simulated samples. To account for the uncertainty on the $e \to \tau$ fake rate, the recommended 1-prong uncertainty is applied, while since the 3-prong uncertainty and the uncertainty on $\mu \to \tau$ fake rate have not been derived, an *ad-hoc* conservative $\pm 50\%$ uncertainty is applied.

Statistical uncertainties of the *Fakes* background are treated as systematical upward and downward fluctuations. An uncertainty of ±50% on the fractional contribution of the largest fractional fake background, W+jets, is applied to the fakes. The ±2.1% uncertainty on luminosity and the uncertainties on cross sections are applied to the samples whose normalization is not derived from the fit.

Further, systematical uncertanties resulting in the relative variation of the normalization factor of less than 0.1% and systematics with a small effect on the shape are removed from the fit.

Table 8.1: Sources of systematical uncertainties in the analysis.

Label	Description
α_PRW_DATASF	Pileup reweighting SF
α_BTAG_SF_MVX_FT_EFF_B	b-jet uncertainty on b-tagging SF efficiency
α_BTAG_SF_MVX_FT_EFF_C	c-jet uncertainty on b-tagging SF efficiency
α_BTAG_SF_MVX_FT_EFF_Light	Light-jet uncertainty on b-tagging SF efficiency
α_BTAG_SF_MVX_FT_EFF_extrapolation	Extrapolation uncertainty on b-tagging SF efficiency
α_BTAG_SF_MVX_FT_EFF_extrapolation_from_charm	Extrapolation from charm uncertainty on b-tagging SF efficiency
α_EG_RESOLUTION_ALL	EGamma resolution uncertainty
α_EG_SCALE_ALL	EGamma scale uncertainty
α_EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	Total uncertainty on electron ID SF efficiency
α_EL_EFF_ISO_TOTAL_1NPCOR_PLUS_UNCOR	Total uncertainty on electron isolation SF efficiency
α_EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	Total uncertainty on electron reconstruction SF efficiency
α_EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR	Total uncertainty on electron trigger SF efficiency
α_FAKEFACTORS_*P_PTBIN*_TRKPTBIN*_STAT	Bin statistical uncertainty on fake factors
α_FAKEFACTORS_*P_R_W	$R(Wjets)$ uncertainty on fake factors
α_JET_EtaIntercalibration_NonClosure	JES eta intercalibration non-closure uncertainty
α_JET_GroupedNP_1	JES grouped NP 1 uncertainty
α_JET_GroupedNP_2	JES grouped NP 2 uncertainty
α_JET_GroupedNP_3	JES grouped NP 3 uncertainty
α_JET_JER_SINGLE_NP	JER single NP uncertainty
α_LUMI	Luminosity uncertainty for minor samples
α_MET_SoftTrk_Scale	Missing Energy soft track scale
α_MET_SoftTrk_ResoPara	Missing Energy soft track parallel resolution
α_MET_SoftTrk_ResoPerp	Missing Energy soft track perpendicular resolution

Table 8.2: Sources of systematical uncertainties in the analysis.

Label	Description
α_OTHER_THEORY	Cross section uncertanties for minor samples
α_WJETS_THEORY	Cross section uncertanties for W+jets
α_Z_REWEIGHTING_SIG_STAT	Statistical uncertainty on signal reweighting
α_Zll_REWEIGHTING_SIG_STAT	Statistical uncertainty on Zll reweighting
α_TAUS_TRUEELECTRON_3P_EVETO_ADHOC	Ad-hoc 3-prong e-veto uncertainty for taus originated from true electrons
α_TRUEELECTRON_EFF_ELEOLR_TOTAL	Electron overlap removal total uncertainty on true electron tau ID SF efficiency
α_TAUS_TRUEHADTAU_3P_EVETO_ADHOC	Ad-hoc 3-prong e-veto uncertainty for taus originated from true taus
α_TAUS_TRUEHADTAU_EFF_ELEOLR_TOTAL	Electron overlap removal total uncertainty on true hadronic tau ID SF efficiency
α_TAUS_TRUEHADTAU_EFF_JETID_TOTAL	Jet ID total uncertainty on true hadronic tau ID SF efficiency
α_TAUS_TRUEHADTAU_EFF_RECO_TOTAL	Reconstruction total uncertainty on true hadronic tau ID SF efficiency
α_TAUS_TRUEHADTAU_SME_TES_DETECTOR	TES detector uncertainty on true hadronic tau smearing
α_TAUS_TRUEHADTAU_SME_TES_INSITU	TES in-situ uncertainty on true hadronic tau smearing
α_TAUS_TRUEHADTAU_SME_TES_MODEL	TES model uncertainty on true hadronic tau smearing
α_MUON_ID	Muon ID uncertainty
α_MUON_MS	Muon MS uncertainty
α_MUON_SCALE	Muon scale uncertainty
α_MUON_EFF_TrigStatUncertainty	Trigger statistical uncertainty on muon global trigger SF efficiency
α_MUON_EFF_TrigSystUncertainty	Trigger systematic uncertainty on muon global trigger SF efficiency
α_MUON_EFF_STAT	Statistical uncertainty on muon reconstruction SF efficiency
α_MUON_EFF_SYS	Systematic uncertainty on muon reconstruction SF efficiency
α_MUON_TTVA_STAT	Statistical uncertainty on muon reconstruction SF TTVA
α_MUON_TTVA_SYS	Systematic uncertainty on muon reconstruction SF TTVA

8.4 Asimov Fit

The Asimov fit is done in order to estimate the sensitivity of the analysis, before looking at the data in the full signal region in the process called *unblinding*. In the Asimov fit, the data is replaced with the pre-fit background expectation referred to as *Asimov data*. In the signal+background model fit to the Asimov data the upper limit which would be set if the observed data would be equal to the background prediction is obtained.

The results for the branching ratio and the background normalization parameters for the τ-e and the τ-μ channels are shown in Tables 8.3 and 8.4. The branching ratios for the $Z \to \tau e$ and the $Z \to \tau\mu$ processes are fitted at $(0.04 \pm 3.29) \cdot 10^{-5}$ and $(0.00 \pm 1.11) \cdot 10^{-5}$ while the background normalization factors are all fitted at 1.00 as the central value. The results are further validating the fit machinery, as the backgrounds are expected to be fitted at 1.00, while the signal yield is expected to be negligible after the fit. The full list of the post-fit parameters with the uncertainties is shown in Figure 8.2.

The nuisance parameters ranked by their impact on the signal strength are shown in Figure 8.3. Pre-fit and post-fit NN output distributions in the signal region, for the τ-e and the τ-μ channels are given in Appendix C.

The sensitivity scan is shown in Figure 8.4, where the signal strength is scanned in the predetermined range and the $\mathrm{CL_b}$, $\mathrm{CL_{s+b}}$ and $\mathrm{CL_s}$ values are shown. The branching ratio resulting in the $\mathrm{CL_s} = 0.05$ value is used to determine the upper limit. The observed upper limit is obtained the fit to the Asimov data *i.e.* to the pre-fit background prediction. The expected upper limit is obtained in the fit to the post-fit background expectation. The upper limits on branching ratios are summarized in Table 8.5.

Table 8.3: Initial and the fitted values of the $Z \to \tau e$ branching ratio BR($Z \to \tau e$), Z normalization factor μ(Z) and the 1-prong and 3-prong fakes normalization factor μ(fakes,1P) and μ(fakes,3P), obtained in the Asimov fit in the τ-e channel. In the fit, the signal+background model is fitted to the Asimov data, defined as the Standard Model background expectation. The uncertainties include both statistical and systematic contributions.

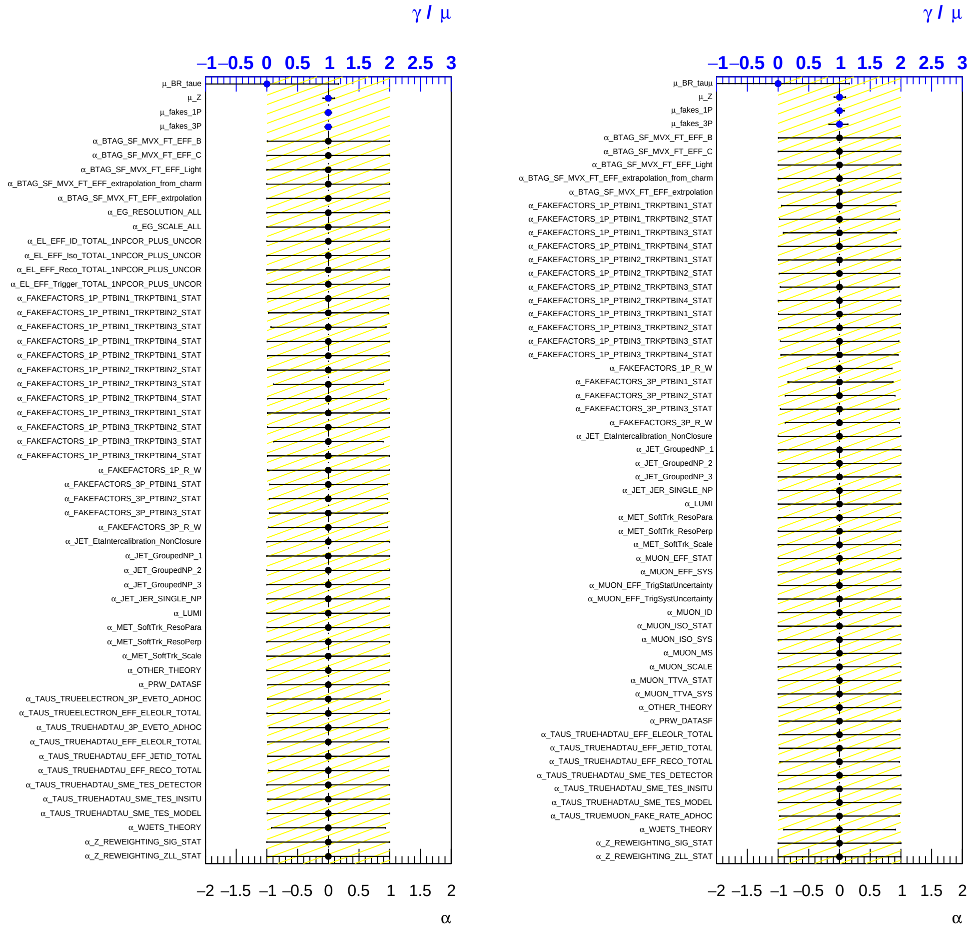
Parameter	Initial value	Fitted value
BR($Z \to \tau e$)	1.00 ± 0.00	$(-0.00 \pm 1.15) \times 10^{-5}$
μ(Z)	1.00 ± 0.00	1.00 ± 0.09
μ(fakes,1P)	1.00 ± 0.00	1.00 ± 0.05
μ(fakes,3P)	1.00 ± 0.00	1.00 ± 0.05

Table 8.4: Initial and the fitted values of the $Z \to \tau\mu$ branching ratio BR($Z \to \tau\mu$), Z normalization factor μ(Z) and the 1-prong and 3-prong fakes normalization factor μ(fakes,1P) and μ(fakes,3P), obtained in the Asimov fit in the τ-μ channel. In the fit, the signal+background model is fitted to the Asimov data, defined as the Standard Model background expectation. The uncertainties include both statistical and systematic contributions.

Parameter	Initial value	Fitted value
BR($Z \to \tau\mu$)	1.00 ± 0.00	$(0.00 \pm 1.11) \times 10^{-5}$
μ(Z)	1.00 ± 0.00	1.00 ± 0.09
μ(fakes,1P)	1.00 ± 0.00	1.00 ± 0.07
μ(fakes,3P)	1.00 ± 0.00	1.00 ± 0.15

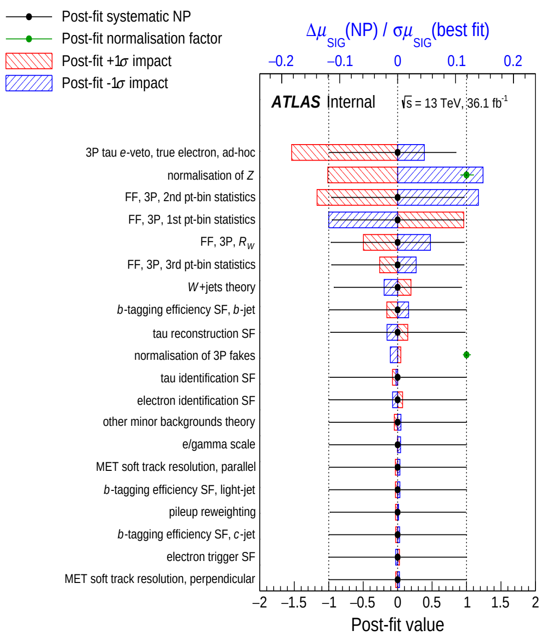
Table 8.5: Expected and observed upper limit on branching ratio in the τ-e and the τ-μ channels. The observed upper limit is obtained in the signal+background model fit to the Asimov data, defined as the pre-fit Standard Model background expectation, while the expected upper limit is obtained in the signal+background fit to the post-fit background expectation *i.e.* the background model, with the post-fit values for nuisance parameters.

Parameter	Expected	Observed
BR upper limit at 95% CL (τ-e)	2.34×10^{-5}	2.34×10^{-5}
BR upper limit at 95% CL (τ-μ)	2.37×10^{-5}	2.37×10^{-5}

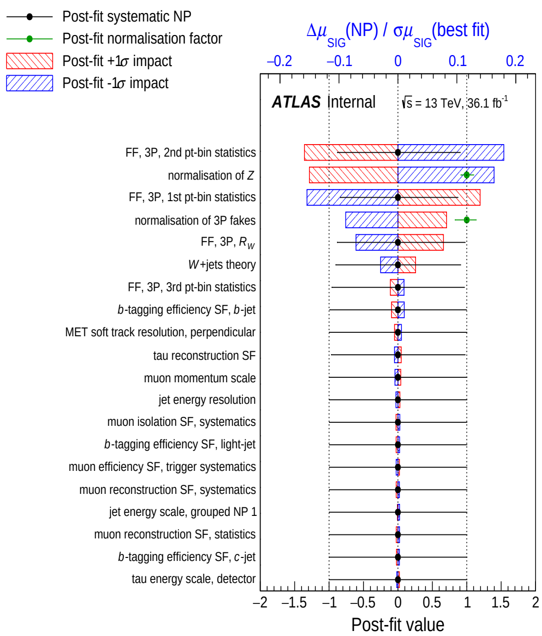


(a) τ-e **(b)** τ-μ

Figure 8.2: Post-fit values for fit parameters obtained in the Asimov fit for the (a) τ-e and the (b) τ-μ channels. The branching ratio and the background normalization parameters are presented by its post-fit values and the post-fit uncertainties, while the nuisance parameters showing the effect systematical errors are shown as the difference between the post-fit and the pre-fit central value, divided by the pre-fit uncertainty. Systematical uncertainties are labeled with α and described in Tables 8.2 and 8.1.

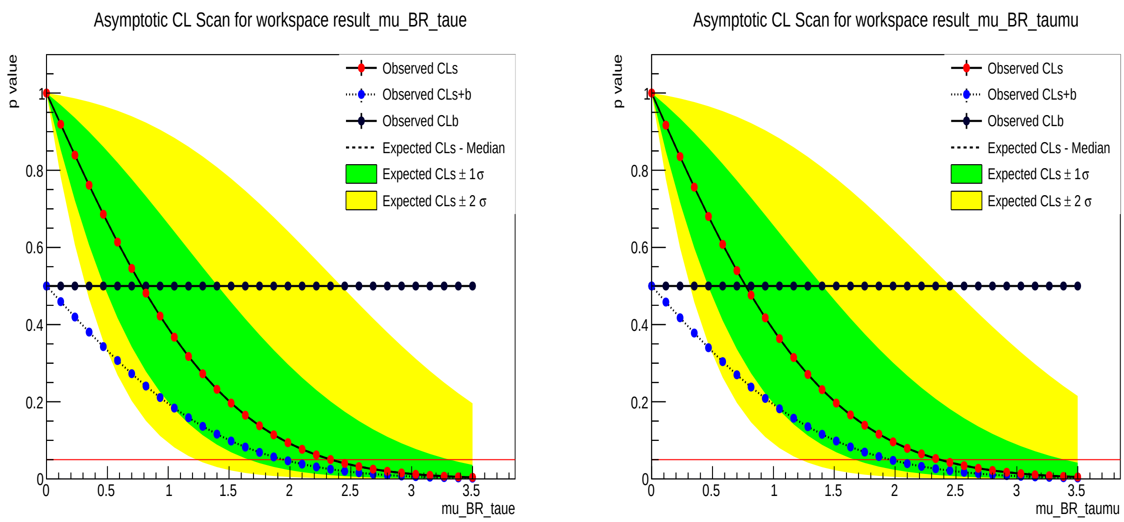


(a) τ-e



(b) τ-μ

Figure 8.3: Nuisance parameters, ranked by their impact on the signal strength, in the Asimov fit for the (a) τ-e and the (b) τ-μ channels. The black lines are showing their post-fit values with post-fit uncertanties expressed in units of pre-fit uncertanties. The impact of a nuisance parameter at the signal strength is measured by fixing the central value of the nuisance parameter at its up/down variation and redoing the fit: the difference between the signal strength and the best-fit value signal strength divided by the post-fit uncertainty of the best-fit signal strength is shown by the hatched boxes for the up (red) and the down (blue) variation.



(a) τ-e **(b)** τ-μ

Figure 8.4: Sensitivity scan in the (a) τ-e and the (b) τ-μ channels, obtained using the signal+background model fitted to the Asimov data, defined as the pre-fit Standard Model background expectation. The branching ratio in units of 10^{-5} is shown in the x-axis, while the p-values for the signal+background (CLs+b), background (1-CLb) and the CLs = $\frac{CL_{s+b}}{CL_b}$ are shown in the y-axis. The expected and the observed limits are obtained as the values resulting in the CLs-value of 0.05, using the ATLAS default CLs method. The observed (=expected) upper limit are set at 2.34×10^{-5} for the τ-e and at 2.37×10^{-5} for the τ-μ channel.

8.5 Sideband Fit

Before the unblinding, the fit model and its predictions are validated in a region that is kinematically similar to the signal region, but with a negligible signal yield. In this analysis, the validation region is constructed using the signal region selection as defined in Tables 6.4 and 6.5 with an additional selection applied on the neural network score NN< 0.4. As the neural network distribution was optimized to mainly contain the backgrounds at the low scores and the signal at the high scores, the low sideband is expected to mainly consists of the backgrounds, while being kinematically similar to the full distribution.

The fit is using a signal+backgrounds model fit and is expected to result in a reliable fit of the backgrounds to the data, with a negligible fitted signal. The results for the branching ratio and the background normalization parameters for the τ-e and the τ-μ channels are shown in Tables 8.6 and 8.7. The branching ratio of the $Z \to \tau e$ process is fitted at $(3.33 \pm 5.26) \times 10^{-5}$ and is containing zero within one standard deviation, while the branching ratio for the $Z \to \tau\mu$ process is fitted at $(0.00 \pm 0.95) \times 10^{-5}$, with the zero at the central value. The full list of the post-fit parameters with the uncertainties is shown in Figure 8.5.

The nuisance parameters ranked by their impact on the signal strength are shown in Figure 8.6. Pre-fit and post-fit NN output distributions in the sideband, for the τ-e and the τ-μ channels are shown in Figures 8.10 and 8.11. The distributions are separately shown for 1-prong and 3-prong events, due to different Fakes normalization factors. The post-fit distributions are showing a good agreement between the model and the data.

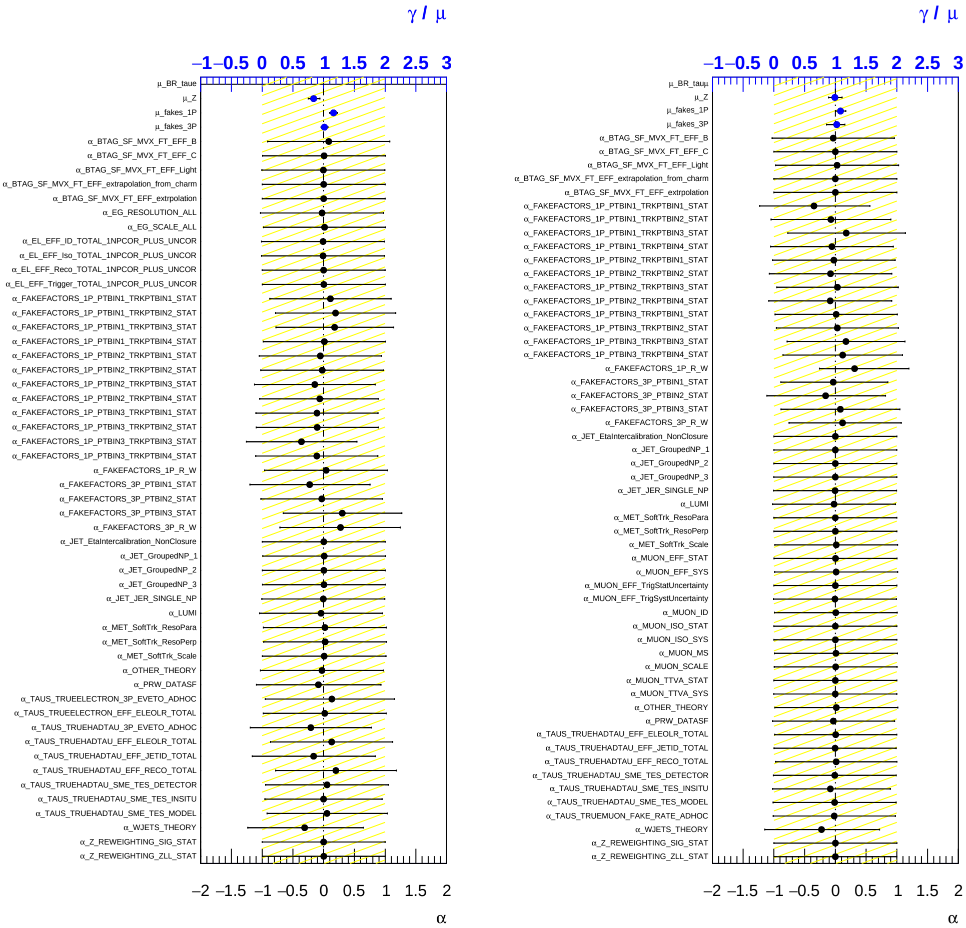
The pre-fit and post-fit NN output distribution in the sideband fit are shown in Figures 8.10 and 8.11 for the τ-e and the τ-μ channel.

Table 8.6: Initial and the fitted values of the $Z \to \tau e$ branching ratio BR($Z \to \tau e$), Z normalization factor μ(Z) and the 1-prong and 3-prong fakes normalization factor μ(fakes,1P) and μ(fakes,3P), obtained in the sideband fit in the τ-μ channel. In the fit, the signal+background model is fitted to the data in the NN sideband (NN<0.4). The uncertainties include both statistical and systematic contributions.

Parameter	Initial value	Fitted value
BR($Z \to \tau e$)	1.00 ± 0.00	$(3.32 \pm 5.41) \times 10^{-5}$
μ(Z)	1.00 ± 0.00	0.84 ± 0.10
μ(fakes,1P)	1.00 ± 0.00	1.16 ± 0.06
μ(fakes,3P)	1.00 ± 0.00	1.01 ± 0.06

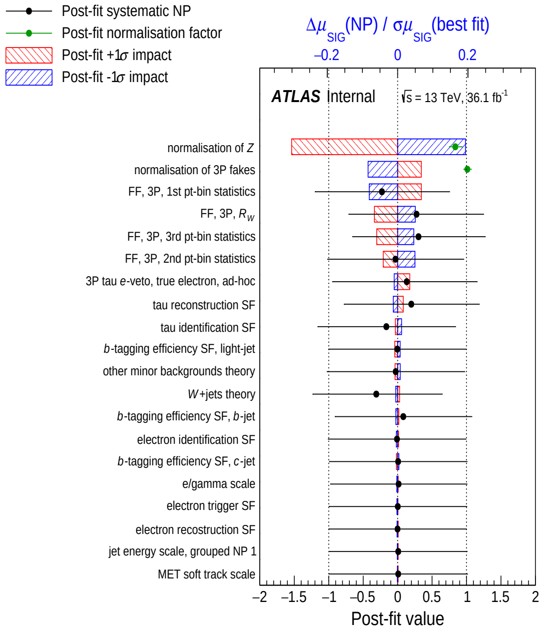
Table 8.7: Initial and the fitted values of the $Z \to \tau\mu$ branching ratio BR($Z \to \tau\mu$), Z normalization factor μ(Z) and the 1-prong and 3-prong fakes normalization factor μ(fakes,1P) and μ(fakes,3P), obtained in the sideband fit in the τ-μ channel. In the fit, the signal+background model is fitted to the data in the NN sideband (NN<0.4). The uncertainties include both statistical and systematic contributions.

Parameter	Initial value	Fitted value
BR($Z \to \tau\mu$)	1.00 ± 0.00	$(-13.43 \pm 3.95) \times 10^{-5}$
μ(Z)	1.00 ± 0.00	0.99 ± 0.11
μ(fakes,1P)	1.00 ± 0.00	1.08 ± 0.08
μ(fakes,3P)	1.00 ± 0.00	1.02 ± 0.14

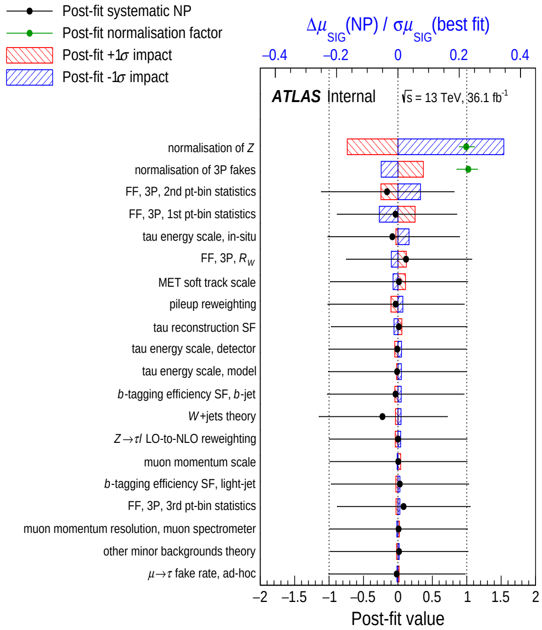


(a) τ-e **(b)** τ-μ

Figure 8.5: Post-fit values for fit parameters obtained in the sideband fit for the (a) τ-e and the τ-μ channels. The branching ratio and the background normalization parameters are presented by its post-fit values and the post-fit uncertainties, while the nuisance parameters which showing the effect of statistical and systematical errors are shown as the difference between the post-fit and the pre-fit central value, divided by the pre-fit uncertainty. Systematical uncertainties are labeled with α and described in Tables 8.2 and 8.1.

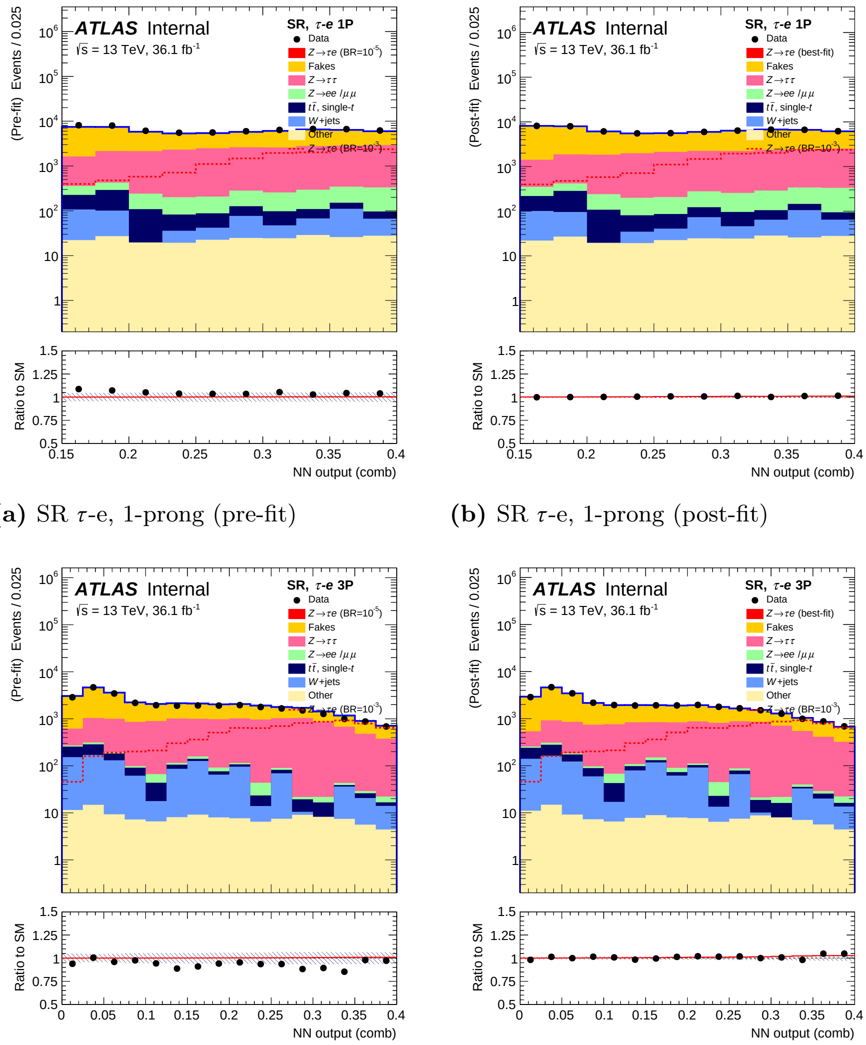


(a) τ-e



(b) τ-μ

Figure 8.6: Nuisance parameters, ranked by their impact on the signal strength, in the side-band fit for the (a) τ-e and the (b) τ-μ channels. The black lines are showing their post-fit values with post-fit uncertanties expresses in units of pre-fit uncertanties. The impact of a nuisance parameter at the signal strength is measured by fixing the central value of the nuisance parameter at its up/down variation and redoing the fit: the difference between the signal strength obtained in this way and its best-fit value divided by the post-fit uncertainty of the best-fit signal strength is shown by the hatched boxes for the up (red) and the down (blue) variation.



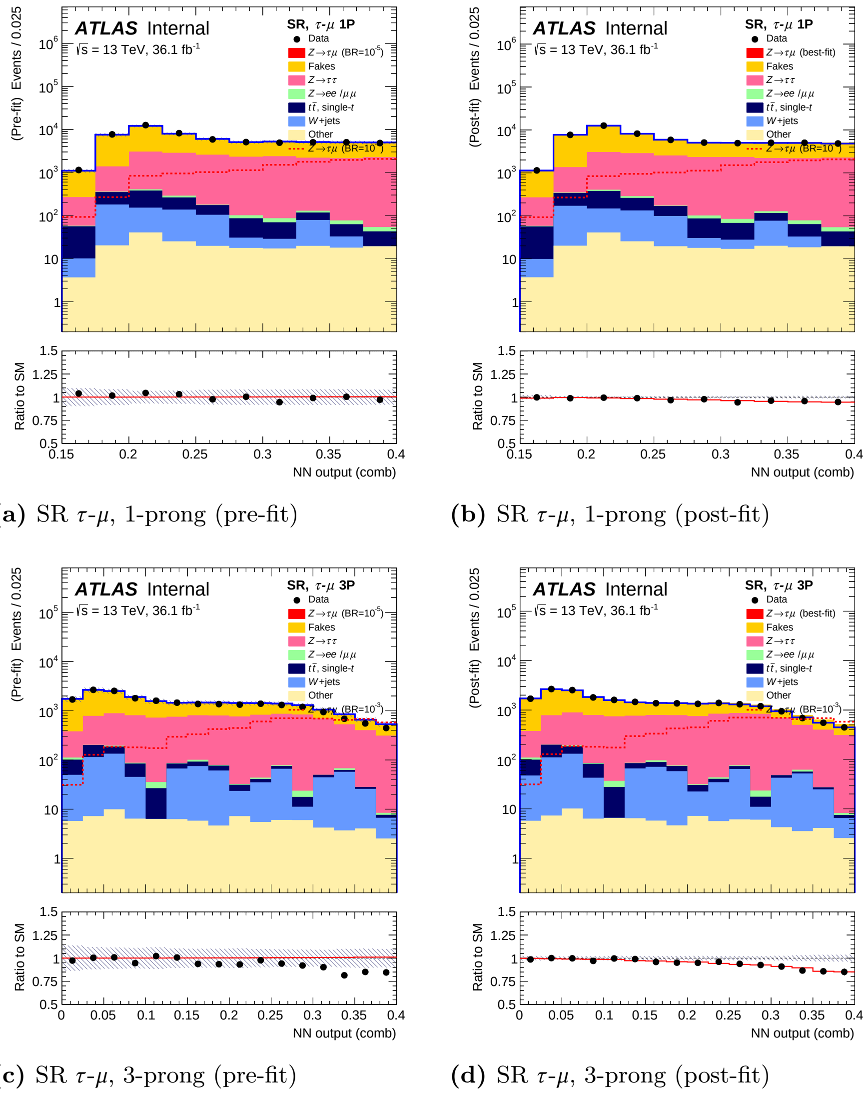
(a) SR τ-e, 1-prong (pre-fit)

(b) SR τ-e, 1-prong (post-fit)

(c) SR τ-e, 3-prong (pre-fit)

(d) SR τ-e, 3-prong (post-fit)

Figure 8.7: Pre-fit and post-fit distribution of the NN output in the low sideband (NN<0.4) of the signal region, for the τ-e channel.



(a) SR τ-μ, 1-prong (pre-fit)

(b) SR τ-μ, 1-prong (post-fit)

(c) SR τ-μ, 3-prong (pre-fit)

(d) SR τ-μ, 3-prong (post-fit)

Figure 8.8: Pre-fit and post-fit distribution of the NN output in the low sideband (NN<0.4) of the signal region, for the τ-μ channel.

8.6 Unblinded Fit

The *unblinded* fit is done using the data in the full range of the neural network score distribution. The results for the free floating parameters obtained using the signal+background model are shown in Tables 8.8 and 8.9. The full list, including the nuisance parameters is shown in Figure 8.9.

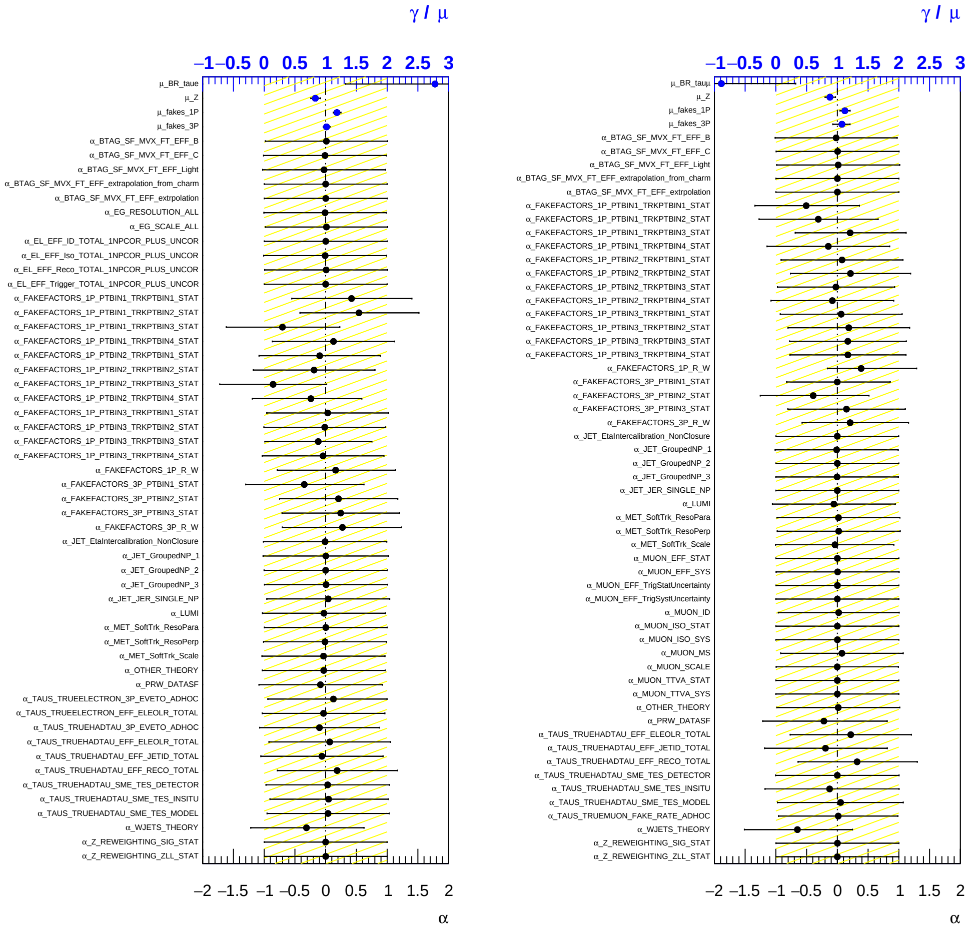
The discriminating unblinded neural network score distributions in the signal region are given in Figures 8.10 and 8.11, for the τ-e and the τ-μ channels. The fit validation is done using the NN-inputs whose post-fit distributions in the signal region of the τ-e channel, for 1-prong and 3-prong events are shown in Figures 8.12 and 8.13. The post-fit NN-inputs in the τ-μ channel are shown in Figures 8.14 and 8.15.

Table 8.8: Initial and the fitted values of the $Z \to \tau e$ branching ratio BR($Z \to \tau e$), Z normalization factor μ(Z) and the 1-prong and 3-prong fakes normalization factor μ(fakes,1P) and μ(fakes,3P), obtained in the unblinded fit in the τ-e channel. The uncertainties include both statistical and systematic contributions.

Parameter	Initial value	Fitted value
BR($Z \to \tau e$)	1.00 ± 0.00	$(2.78 \pm 1.48) \times 10^{-5}$
μ(Z)	1.00 ± 0.00	0.83 ± 0.08
μ(fakes,1P)	1.00 ± 0.00	1.18 ± 0.06
μ(fakes,3P)	1.00 ± 0.00	1.01 ± 0.06

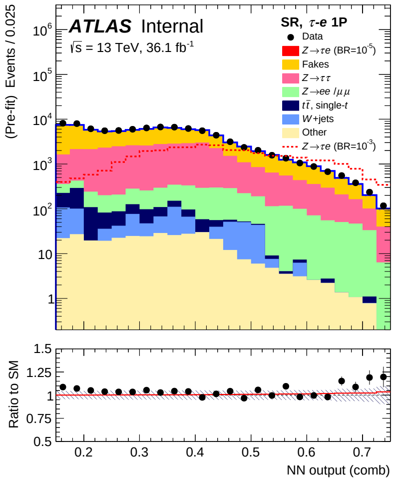
Table 8.9: Initial and the fitted values of the $Z \to \tau e$ branching ratio BR($Z \to \tau\mu$), Z normalization factor μ(Z) and the 1-prong and 3-prong fakes normalization factor μ(fakes,1P) and μ(fakes,3P), obtained in the unblinded fit in the τ-e channel. The uncertainties include both statistical and systematic contributions.

Parameter	Initial value	Fitted value
BR($Z \to \tau\mu$)	1.00 ± 0.00	$(-0.89 \pm 1.17) \times 10^{-5}$
μ(Z)	1.00 ± 0.00	0.88 ± 0.09
μ(fakes,1P)	1.00 ± 0.00	1.12 ± 0.08
μ(fakes,3P)	1.00 ± 0.00	1.07 ± 0.13

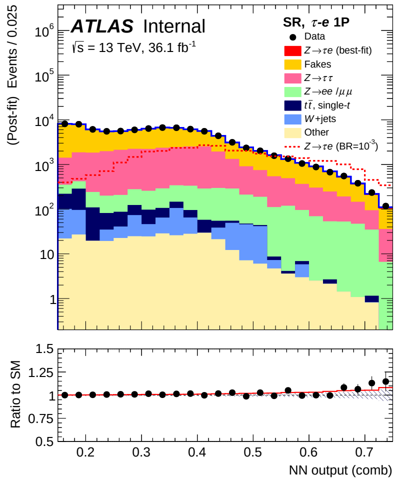


(a) τ-e **(b)** τ-μ

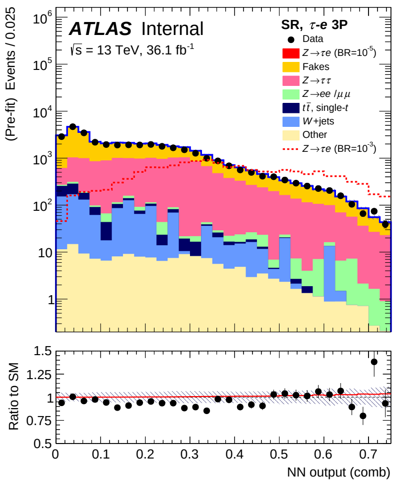
Figure 8.9: Post-fit values for fit parameters obtained in the unblinded fit for the (a) τ-e and the (b) τ-μ channels. The branching ratio and the background normalization parameters are presented by its post-fit values and the post-fit uncertainties, while the nuisance parameters showing the effect systematical errors are shown as the difference between the post-fit and the pre-fit central value, divided by the pre-fit uncertainty. Systematical uncertainties are labeled with α and described in Tables 8.2 and 8.1.



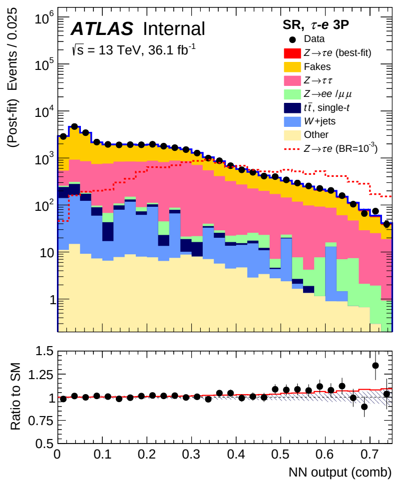
(a) SR τ-e, 1-prong (pre-fit)



(b) SR τ-e, 1-prong (post-fit)

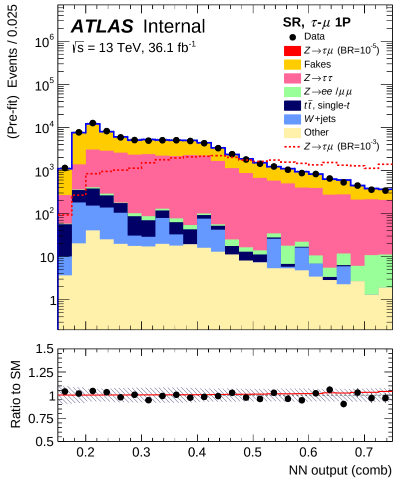


(c) SR τ-e, 3-prong (pre-fit)

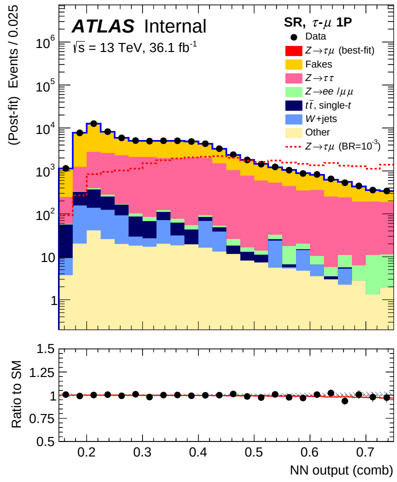


(d) SR τ-e, 3-prong (post-fit)

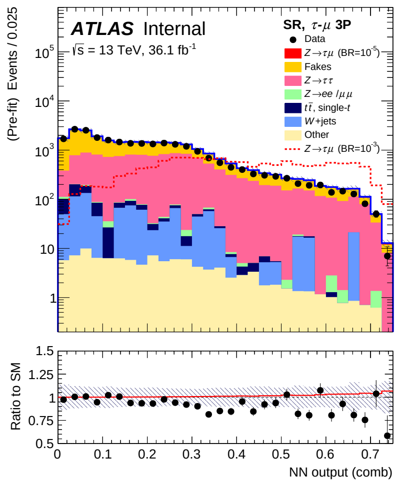
Figure 8.10: Pre-fit and post-fit distribution of the NN output in the unblinded fit in the signal region, for the τ-e channel.



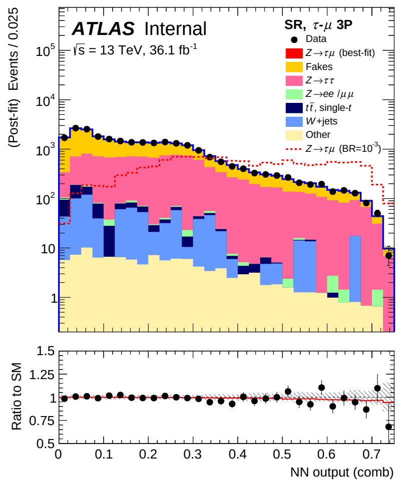
(a) SR τ-μ, 1-prong (pre-fit)



(b) SR τ-μ, 1-prong (post-fit)

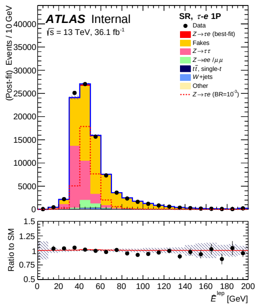


(c) SR τ-μ, 3-prong (pre-fit)

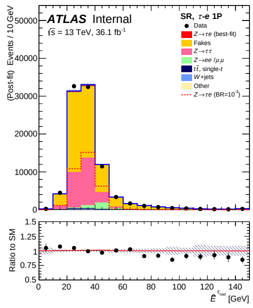


(d) SR τ-μ, 3-prong (post-fit)

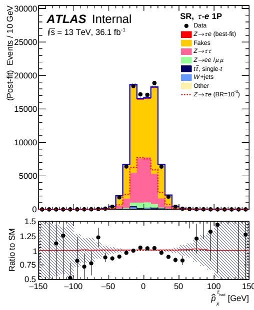
Figure 8.11: Pre-fit and post-fit distribution of the NN output in the unblinded fit in the signal region, for the τ-μ channel.



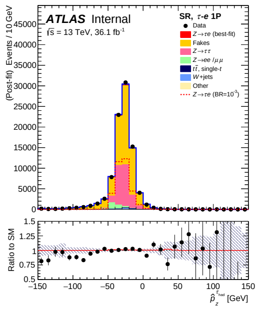
(a) $\hat{E}^{\text{lep}}$



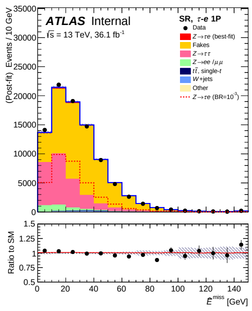
(b) $\hat{E}^{\tau_h}$



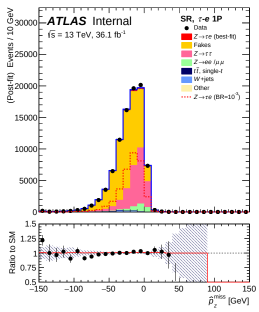
(c) $\hat{p}_x^{\tau_h}$



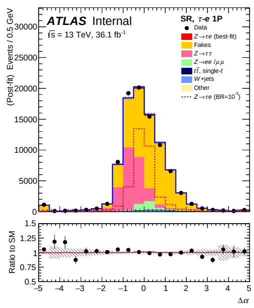
(d) $\hat{p}_z^{\tau_h}$



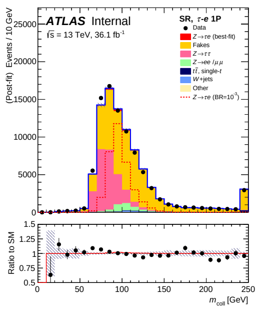
(e) $\hat{E}^{\text{miss}}$



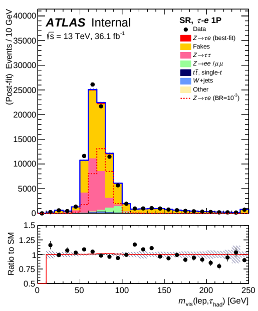
(f) $\hat{p}_z^{\text{miss}}$



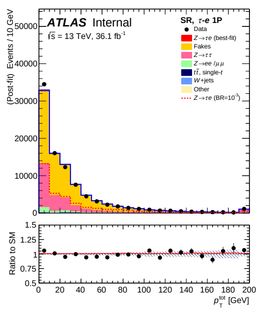
(g) $\Delta\alpha$



(h) m_{coll}

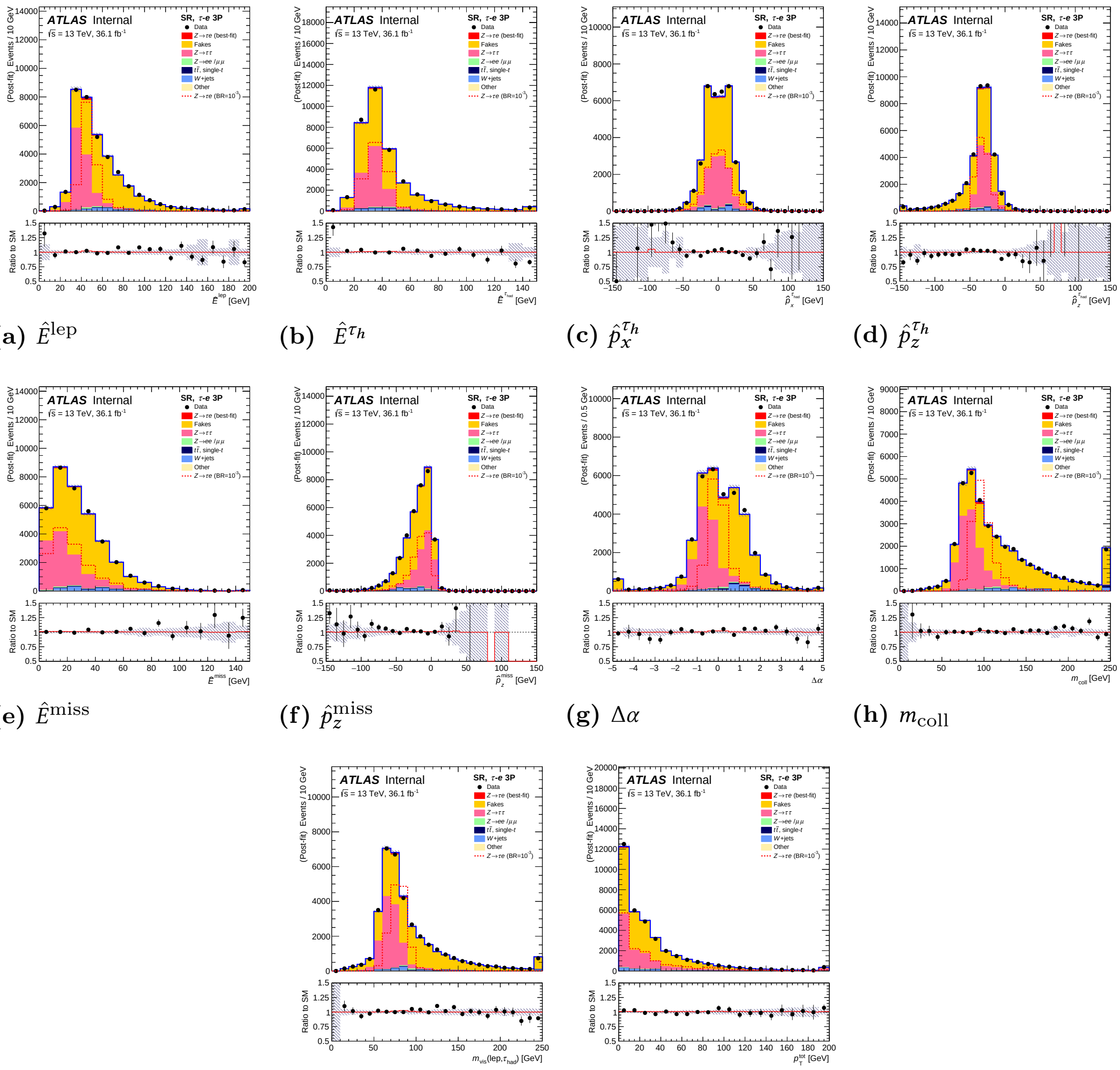


(i) $m_{\text{vis(lep},\tau_h)}$



(j) $p_{\text{T}}^{\text{tot}}$

Figure 8.12: Post-fit NN-input distributions for 1-prong events in the signal region of the τ-e channel.

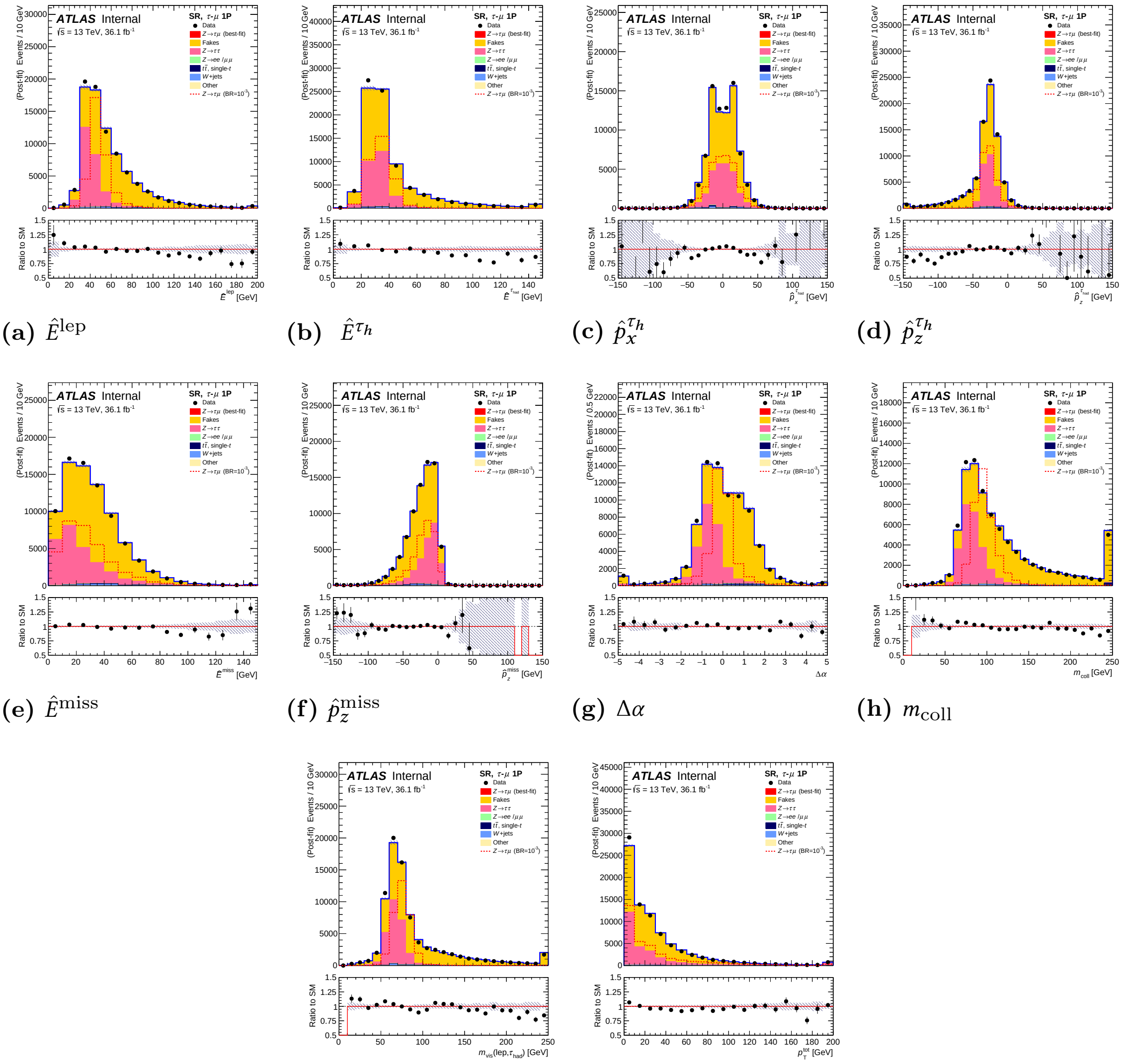


(a) $\hat{E}^{\text{lep}}$ **(b)** $\hat{E}^{\tau_h}$ **(c)** $\hat{p}_x^{\tau_h}$ **(d)** $\hat{p}_z^{\tau_h}$

(e) $\hat{E}^{\text{miss}}$ **(f)** $\hat{p}_z^{\text{miss}}$ **(g)** $\Delta\alpha$ **(h)** m_{coll}

(i) $m_{\text{vis(lep},\tau_h)}$ **(j)** $p_{\text{T}}^{\text{tot}}$

Figure 8.13: Post-fit NN-input distributions for 3-prong events in the signal region of the τ-e channel.

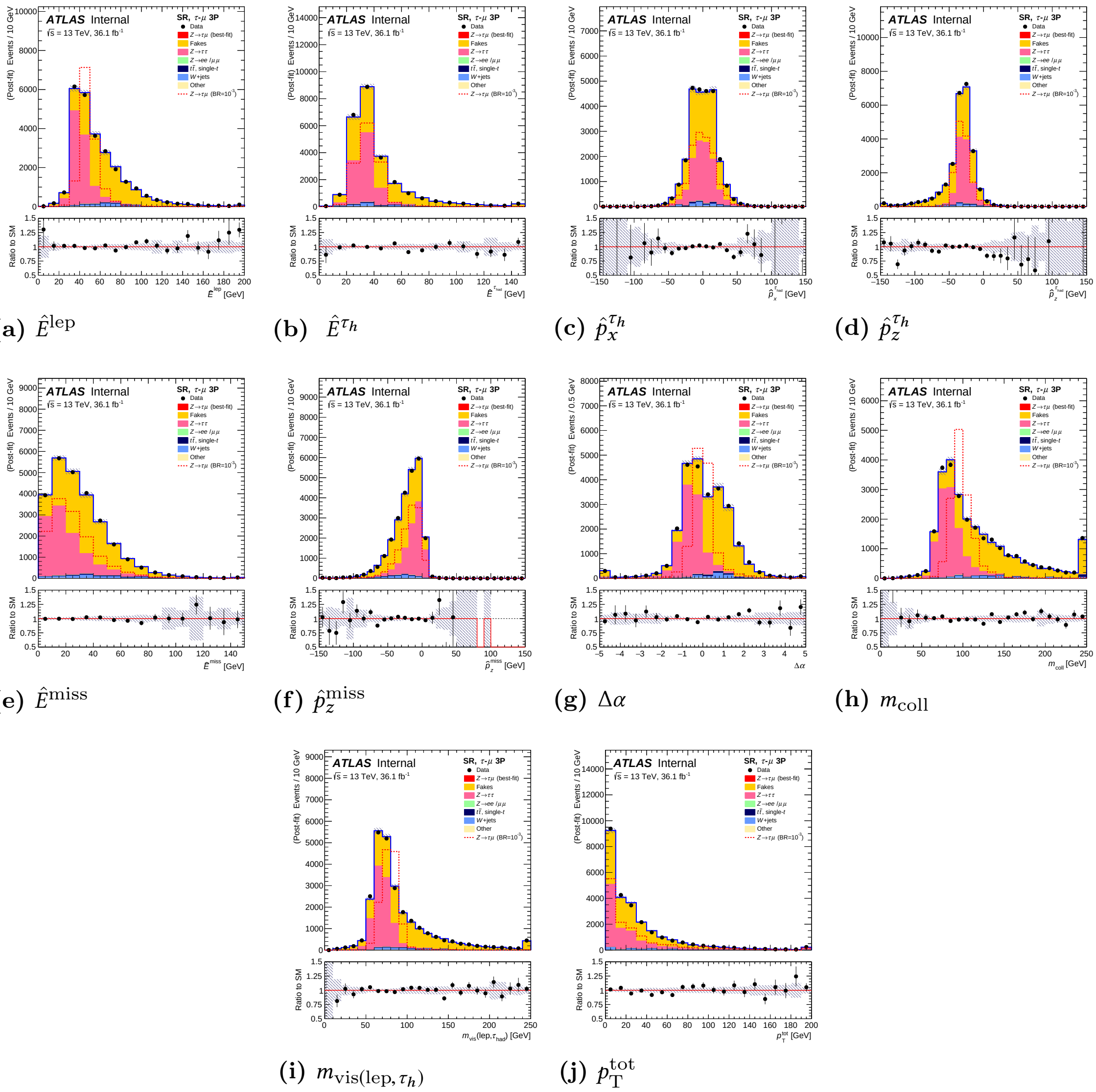


(a) $\hat{E}^{\mathrm{lep}}$ **(b)** $\hat{E}^{\tau_h}$ **(c)** $\hat{p}_x^{\tau_h}$ **(d)** $\hat{p}_z^{\tau_h}$

(e) $\hat{E}^{\mathrm{miss}}$ **(f)** $\hat{p}_z^{\mathrm{miss}}$ **(g)** $\Delta\alpha$ **(h)** m_{coll}

(i) $m_{\mathrm{vis(lep},\tau_h)}$ **(j)** $p_{\mathrm{T}}^{\mathrm{tot}}$

Figure 8.14: Post-fit NN-input distributions for 1-prong events in the signal region of the τ-μ channel.



(a) $\hat{E}^{\text{lep}}$ **(b)** $\hat{E}^{\tau_h}$ **(c)** $\hat{p}_x^{\tau_h}$ **(d)** $\hat{p}_z^{\tau_h}$

(e) $\hat{E}^{\text{miss}}$ **(f)** $\hat{p}_z^{\text{miss}}$ **(g)** $\Delta\alpha$ **(h)** m_{coll}

(i) $m_{\text{vis(lep},\tau_h)}$ **(j)** $p_{\text{T}}^{\text{tot}}$

Figure 8.15: Post-fit NN-input distributions for 3-prong events in the signal region of the τ-μ channel.

8.6.1 Testing the discovery hypothesis

The **discovery test**, testing the background against the signal+background hypothesis is done using the test statistic q_0 defined in Equation 8.3. The p-value for the null hypothesis is characterizing the probability to obtain the observed or the more extreme value for q_0 under the background hypothesis. The p-value is computed from the q_0 distribution under the null hypothesis which is constructed using the asymptotic approximation. The null hypothesis is rejected is the computed p-value is less than $3 \cdot 10^{-7}$, corresponding to the significance of 5σ. The hypothesis testing is done independently for 1-prong and 3-events of the τ-e and the τ-μ channels, with the results being summarized in Table 8.10. No significant excess in observed in the τ-μ channel, while the τ-e channel shows an excess of 2.29σ for 3-prong events.

Table 8.10: P-values, characterizing the probability to obtain the observed data, under the background hypothesis for 1-prong and 3-prong events of the τ-e and the τ-μ channel. No significant excess in observed in the τ-μ channel, while an excess of 2.29σ is observed for 3-prong events, in the τ-e channel.

Channel	P-value	Significance
τ-e (1-prong)	0.60	-0.25
τ-e (3-prong)	0.01	2.29
τ-μ (1-prong)	0.87	-1.12
τ-μ (3-prong)	0.21	0.78

8.6.2 Setting the Upper Limit

The **exclusion test** testing the null +background hypothesis at signal strength μ, is done using the test statistic defined in Equation 8.5. The hypothesis is rejected if the the CLs is computed to be less than 0.05. The highest rejected value of μ defines the upper limit on the signal strength.

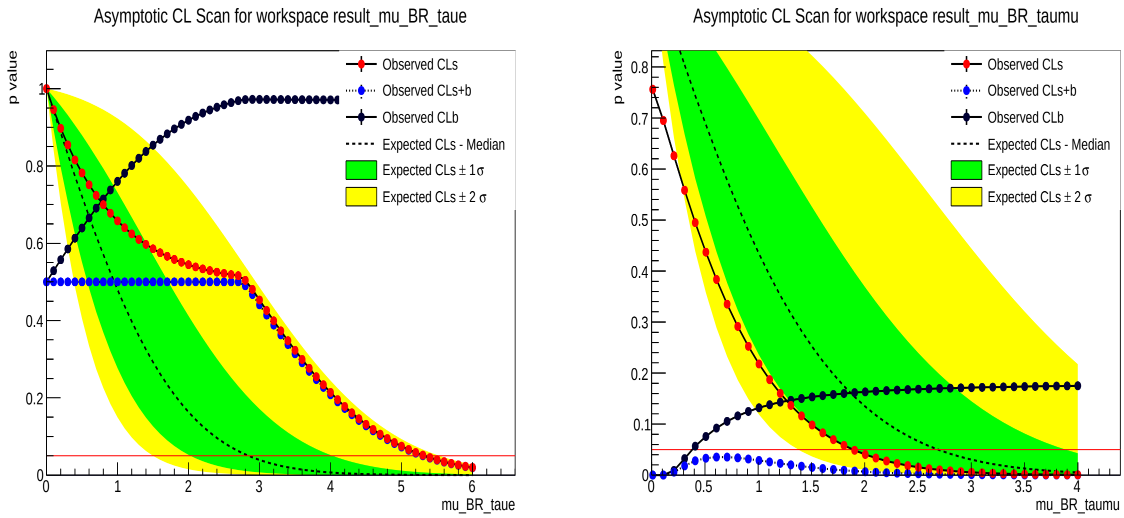
Figure 8.16 shows the p-values for the signal+background (CLs+b), background (1-CLb) and the CLs = $\frac{CL_{s+b}}{CL_b}$. Shown are the observed upper limit, computed using the observed data, and the expected upper limit, computed using the post-fit backgrounds instead of data.

At μ = 0, the two hypothesis are equal causing the $\mathrm{CL_b}$ and $\mathrm{CL_{s+b}}$ to have the same starting value. Further, the $\mathrm{CL_b}$ which characterizes the sensitivity, is increasing with signal strength, especially in the τ-e channel.

In the τ-e channel, the $\mathrm{CL_{s+b}}$ *i.e.* the probability for the underfluctuation of the signal under the signal+ background hypothesis, is significantly larger than in the τ-μ channel and is decreasing more sharply for larger values. For μ larger than ≈ 3, the sensitivity is very good ($CL_b \approx 1$), leading to a small difference between the $\mathrm{CL_{s+b}}$ and $\mathrm{CL_s}$ method.

In the τ-μ, the $\mathrm{CL_{s+b}}$ is continuously small which would lead, using the $\mathrm{CL_{s+b}}$, lead to the exclusion of a smaller value *i.e.* to a better limit. Due to the relatively low sensitivity characterized by small $\mathrm{CL_b}$, the more conservative $\mathrm{CL_s}$ approach leads to a larger upper limit.

Finally, the observed(expected) upper limits are set at $5.32 \times 10^{-5}(2.84 \times 10^{-5})$ for the τ-e and $2.69 \cdot 10^{-5}(1.89 \cdot 10^{-5})$ for the τ-μ channel and summarized in Table 8.11.



(a) τ-e **(b)** τ-μ

Figure 8.16: Upper limit scan obtained in the signal+background model fit to the data for the (a) τ-e and the (b) τ-μ channels. The branching ratio in units of 10^{-5} is shown in the x-axis, while the p-values for the signal+background (CLs+b), background (1-CLb) and the CLs = $\frac{CL_{s+b}}{CL_b}$ are shown in the y-axis. The expected and the observed limits are obtained as the values resulting in the CLs-value of 0.05, using the ATLAS default CLs method. The observed (expected) upper limit are set at for the τ-e and at for the τ-μ channel.

Table 8.11: Expected and observed upper limit on branching ratio in the τ-e and the τ-μ channels. The observed upper limit is obtained in the signal+background model fit to the data, while the expected upper limit is obtained in the signal+background fit to the post-fit background expectation *i.e.* the background model, with the post-fit values for nuisance parameters.

Parameter	Expected	Observed
BR upper limit at 95% CL (τ-e)	2.84×10^{-5}	5.32×10^{-5}
BR upper limit at 95% CL (τ-μ)	$1.89 \cdot 10^{-5}$	$2.69 \cdot 10^{-5}$

Conclusion

Lepton flavour numbers are conserved in the original Standard Model with massless neutrinos, but violated in experimentally discovered neutrino oscillations. Other lepton-flavour violating can occur at the loop level, via neutrino oscillations, at tiny rates. As several Beyond the Standard Model predict significantly larger rates, an observation of such a process would be a sign of new physics.

In this thesis, searches for $Z \to \tau e$ and $Z \to \tau\mu$ lepton-flavour violating processes with hadronic τ-leptons in the final state is performed using the $36.1\,\mathrm{fb}^{-1}$ of proton-proton collision, recorded by the ATLAS detector at the centre-of-mass energy $\sqrt{s} = 13\,\mathrm{TeV}$.

The background model is formed using a data-driven approach combined with the Monte Carlo simulation. The backgrounds are suppressed by applying dedicated selections to define a signal-enriched region where the final fit is done using the neural network score as the discriminating variable.

No significant excess is found and the upper limits on the branching ratios, at the 95% CL, are set at 5.32×10^{-5} for the $Z \to \tau e$ and $2.69 \cdot 10^{-5}$ for the $Z \to \tau\mu$ channel.

Appendices

List of Monte Carlo Samples

The k- factor is defined as the ratio of the cross section at the next leading order and the cross section at the leading order:

$$k = \frac{\sigma^{NLO}}{\sigma^{LO}} \tag{A.1}$$

It is applied as a correction to the cross section at the leading order in order to obtain the cross section at the next-to-leading order.

The generator efficiency ϵ is defined as the ratio of the number of events after a generator filter is applied to the initial number of events:

$$\epsilon = \frac{N_{filter}}{N_{gen}} \tag{A.2}$$

A.1 Signal Samples

Table A.1: Summary of the $Z \to \tau e$ and $Z \to \tau\mu$ signal samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section multiplied with the generator filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for the process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	**Name**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	$\Sigma\, w$	$\mathcal{L}$ [fb^{-1}]
303779	$Z \to \tau e$	4.53×10^{-1}	1.00	345 801	4.98×10^{5}	1.10×10^{3}
303778	$Z \to \tau\mu$	4.57×10^{-1}	1.00	176 620	5.00×10^{5}	1.09×10^{3}

A.2 Background Samples

A.2.1 Z+jets production

Table A.2: Summary of the $Z \to ee$ and $Z \to \mu\mu$ background samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section multiplied with the generator filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for the process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	**Process**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	$\Sigma\, w$	$\mathcal{L}$ [fb^{-1}]
361106	$Z \to ee$	1.90×10^3	1.03	51 003 442	1.50×10^{11}	7.70×10^4
361107	$Z \to \mu\mu$	1.90×10^3	1.03	9 928 580	1.47×10^8	7.55×10^1

Table A.3: Summary of $Z \to \tau\tau$ background samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section multiplied with the generator filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for the process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	max$(H_{\mathrm{T}}, p_{\mathrm{T}}(Z))$	**Filter**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	$\Sigma\, w$	$\mathcal{L}$ [fb^{-1}]
344772	0,70	l13l7	5.01×10^1	0.98	4 677 877	5.41×10^6	1.11×10^2
344774	0,70	l15h20	1.06×10^2	0.98	7 398 490	5.36×10^6	5.21×10^1
344776	70,140	l13l7	4.88×10^{-1}	0.98	1 130 860	7.09×10^5	1.49×10^3
344778	70,140	l15h20	7.95	0.98	1 907 548	6.90×10^5	8.90×10^1
344780	140,280	l13l7	2.04×10^{-1}	0.98	616 399	5.84×10^5	2.94×10^3
344781	140,280	l15h20	3.80	0.98	778 105	5.69×10^5	1.53×10^2
364137	280,500	CBVeto	4.88	0.98	454 306	1.66×10^6	3.48×10^2
364138	280,500	CFilter	2.30	0.98	236 436	9.05×10^5	4.03×10^2
364139	280,500	BFilter	1.53	0.98	539 482	1.85×10^6	1.24×10^3
364140	500,1000	-	1.81	0.98	776 104	2.92×10^6	1.66×10^3
364141	1000,$\mathrm{E}_{\mathrm{CMS}}$	-	1.48×10^{-1}	0.98	295 386	9.98×10^5	6.90×10^3

A.2.2 W+jets

Table A.4: Summary of $W \to e\nu_e$ samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	max($\boldsymbol{H_\mathrm{T}}, \boldsymbol{p_\mathrm{T}(W)}$)	**Filter**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	Σw	$\mathcal{L}$ [fb^{-1}]
364170	0,70	CBVeto	1.58×10^4	0.97	10 568 945	1.66×10^7	1.09
364171	0,70	CFilter	2.49×10^3	0.97	4 695 467	5.65×10^6	2.34
364172	0,70	BFilter	8.45×10^2	0.97	7 521 640	1.04×10^7	1.27×10^1
364173	70,140	CBVeto	6.30×10^2	0.97	7 753 133	5.36×10^6	8.76
364174	70,140	CFilter	2.15×10^2	0.97	5 659 470	3.69×10^6	1.76×10^1
364175	70,140	BFilter	9.77×10^1	0.97	5 142 091	3.98×10^6	4.20×10^1
364176	140,280	CBVeto	2.03×10^2	0.97	5 574 577	6.16×10^6	3.13×10^1
364177	140,280	CFilter	9.84×10^1	0.97	4 510 655	5.26×10^6	5.51×10^1
364178	140,280	BFilter	3.70×10^1	0.97	5 516 665	7.33×10^6	2.04×10^2
364179	280,500	CBVeto	3.92×10^1	0.97	2 886 502	4.31×10^6	1.13×10^2
364180	280,500	CFilter	2.28×10^1	0.97	1 829 267	2.78×10^6	1.25×10^2
364181	280,500	BFilter	9.66	0.97	1 726 534	2.84×10^6	3.03×10^2
364182	500,1000	-	1.52×10^1	0.97	3 598 359	6.00×10^6	4.06×10^2
364183	1000,$\mathrm{E_{CMS}}$	-	1.23	0.97	2 493 629	4.08×10^6	3.41×10^3

Table A.5: Summary of $W \to \mu\nu_\mu$ samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	$\max(\boldsymbol{H_\mathrm{T}}, \boldsymbol{p_\mathrm{T}(W)})$	**Filter**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_gen	$\Sigma\, w$	$\mathcal{L}$ [fb^{-1}]
364156	0,70	CBVeto	1.58×10^4	0.97	2 445 074	1.66×10^7	1.09
364157	0,70	CFilter	2.49×10^3	0.97	1 374 323	5.64×10^6	2.33
364158	0,70	BFilter	8.44×10^2	0.97	2 092 248	1.04×10^7	1.27×10^1
364159	70,140	CBVeto	6.37×10^2	0.97	3 837 225	5.42×10^6	8.76
364160	70,140	CFilter	2.20×10^2	0.97	2 927 817	3.69×10^6	1.73×10^1
364161	70,140	BFilter	7.15×10^1	0.97	5 290 793	7.99×10^6	1.15×10^2
364162	140,280	CBVeto	2.13×10^2	0.97	2 800 816	6.16×10^6	2.98×10^1
364163	140,280	CFilter	9.84×10^1	0.97	2 457 040	5.26×10^6	5.51×10^1
364164	140,280	BFilter	3.69×10^1	0.97	3 038 979	7.27×10^6	2.03×10^2
364165	280,500	CBVeto	3.94×10^1	0.97	1 535 913	4.33×10^6	1.13×10^2
364166	280,500	CFilter	2.29×10^1	0.97	1 066 498	2.78×10^6	1.25×10^2
364167	280,500	BFilter	9.61	0.97	1 036 941	2.84×10^6	3.04×10^2
364168	500,1000	-	1.50×10^1	0.97	2 045 587	5.94×10^6	4.08×10^2
364169	1000,$\mathrm{E_{CMS}}$	-	1.23	0.97	1 410 827	4.07×10^6	3.40×10^3

Table A.6: Summary of $W \to \tau\nu_\tau$ samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	max$(\boldsymbol{H_\mathrm{T}}, \boldsymbol{p_\mathrm{T}(W)})$	**Filter**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_gen	$\Sigma\, w$	$\mathcal{L}$ [fb^{-1}]
364184	0,70	CBVeto	1.58×10^4	0.97	708 621	1.67×10^7	1.09
364185	0,70	CFilter	2.48×10^3	0.97	373 956	5.67×10^6	2.36
364186	0,70	BFilter	8.55×10^2	0.97	564 352	1.05×10^7	1.27×10^1
364187	70,140	CBVeto	6.39×10^2	0.97	1 001 819	5.43×10^6	8.76
364188	70,140	CFilter	2.10×10^2	0.97	805 742	3.72×10^6	1.82×10^1
364189	70,140	BFilter	9.81×10^1	0.97	719 749	3.97×10^6	4.17×10^1
364190	140,280	CBVeto	2.02×10^2	0.97	862 743	6.17×10^6	3.14×10^1
364192	140,280	BFilter	4.01×10^1	0.97	1 008 004	7.29×10^6	1.88×10^2
364193	280,500	CBVeto	3.93×10^1	0.92	506 245	4.32×10^6	1.19×10^2
364194	280,500	CFilter	2.28×10^1	0.97	364 555	2.77×10^6	1.25×10^2
364195	280,500	BFilter	9.67	0.97	394 585	2.83×10^6	3.02×10^2
364196	500,1000	-	1.50×10^1	0.97	778 316	5.98×10^6	4.10×10^2
364197	1000,$\mathrm{E_{CMS}}$	-	1.23	0.97	647 500	4.06×10^6	3.39×10^3

A.2.3 Top

Table A.7: Summary of $t\bar{t}$, single-top and Wt samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	**Process**	**Filter**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	Σw	$\mathcal{L}$ [fb^{-1}]
410000	ttbar	nonallhad	3.78×10^{2}	1.19	25 185 913	4.94×10^{7}	1.09×10^{2}
410011	single t,tchan	lep	4.37×10^{1}	1.01	1 712 900	0.00	4.94×10^{-6}
410012	single $\bar{t}$,tchan	lep $\bar{t}$	2.58×10^{1}	1.02	1 807 572	0.00	4.90×10^{-6}
410013	Wt	-	3.40×10^{1}	1.05	1 512 264	4.99×10^{6}	1.39×10^{2}
410014	$W\bar{t}$	-	3.40×10^{1}	1.05	1 511 422	4.99×10^{6}	1.39×10^{2}
410025	single t,schan	nonallhad	2.05	1.00	328 996	0.00	9.94×10^{-7}
410026	single $\bar{t}$,schan	nonallhad	1.26	1.02	338 361	0.00	9.75×10^{-7}

A.2.4 Diboson

Table A.8: Summary of diboson samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

DSID	**Process**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	Σw	$\mathcal{L}$ [fb^{-1}]
361063	llll	1.28×10^{1}	0.91	588 489	2.13×10^{6}	1.82×10^{2}
361064	lllvSFMinus	1.84	0.91	105 360	4.40×10^{5}	2.62×10^{2}
361065	lllvOFMinus	3.62	0.91	200 526	8.78×10^{5}	2.66×10^{2}
361066	lllvSFPlus	2.57	0.91	139 606	5.88×10^{5}	2.52×10^{2}
361067	lllvOFPlus	5.02	0.91	267 174	1.18×10^{6}	2.59×10^{2}
361088	lvvv	3.40	0.91	516 001	2.28×10^{6}	7.37×10^{2}
361089	vvvv	6.60×10^{-1}	0.91	1942	1.12×10^{6}	1.86×10^{3}
361091	WplvWmqq_SHv21_improved	2.49×10^{1}	0.91	1 288 858	9.64×10^{5}	4.26×10^{1}
361092	WpqqWmlv_SHv21_improved	2.49×10^{1}	0.91	1 277 730	9.64×10^{5}	4.26×10^{1}
361093	WlvZqq_SHv21_improved	1.15×10^{1}	0.91	1 324 320	1.01×10^{6}	9.70×10^{1}
361094	WqqZll_SHv21_improved	3.42	0.91	1 878 560	3.02×10^{5}	9.70×10^{1}
361095	WqqZvv_SHv21_improved	6.78	0.91	49 043	7.43×10^{5}	1.21×10^{2}
361096	ZqqZll_SHv21_improved	2.35	0.91	1 889 745	2.66×10^{6}	1.24×10^{3}
361097	ZqqZvv_SHv21_improved	4.63	0.91	56 718	2.99×10^{6}	7.08×10^{2}

A.2.5 Higgs

Table A.9: Summary of Higgs samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

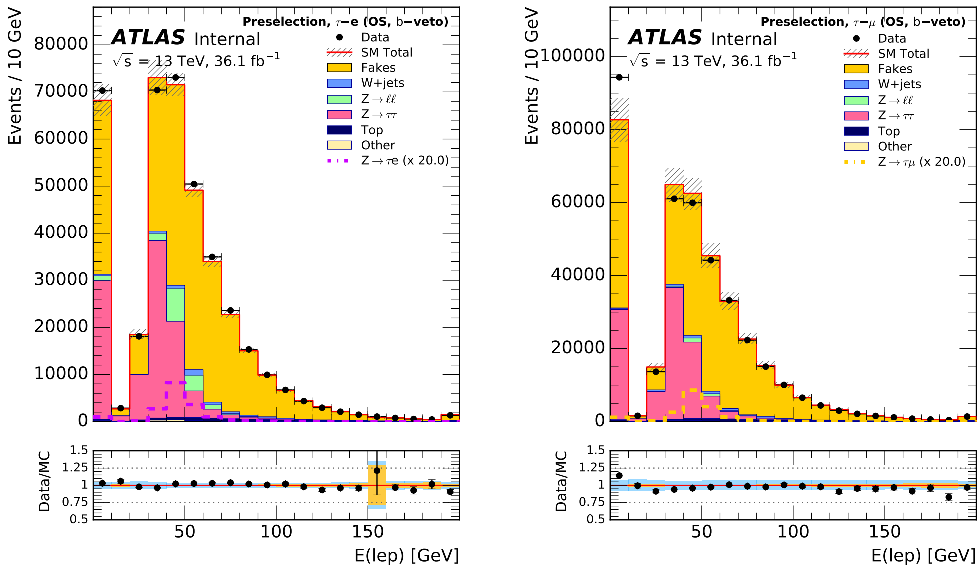
DSID	**Filter**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	Σw	$\mathcal{L}$ [fb^{-1}]
341079	ggH125_WWlvlv_EF_15_5	4.87×10^{-1}	1.00	238 434	4.79×10^{5}	9.83×10^{2}
341080	VBFH125_WWlvlv_EF_15_5	4.33×10^{-2}	1.00	142 983	2.50×10^{5}	5.77×10^{3}
341122	ggH125_tautaull	2.34×10^{-1}	1.45	517 489	4.60×10^{7}	1.35×10^{5}
341123	ggH125_tautaulh	8.68×10^{-1}	1.45	510 247	4.65×10^{7}	3.69×10^{4}
341124	ggH125_tautauhh	8.05×10^{-1}	1.45	32 056	4.66×10^{7}	3.98×10^{4}
341155	VBFH125_tautaull	2.97×10^{-2}	0.98	907 031	2.08×10^{6}	7.15×10^{4}
341156	VBFH125_tautaulh	1.10×10^{-1}	0.98	814 637	2.09×10^{6}	1.94×10^{4}
341157	VBFH125_tautauhh	1.02×10^{-1}	0.98	64 410	2.09×10^{6}	2.09×10^{4}
341195	ggH125_mumu	6.61×10^{-3}	1.45	203 888	2.89×10^{7}	3.00×10^{6}
341206	VBFH125_mumu	8.53×10^{-4}	0.96	314 518	3.89×10^{6}	4.74×10^{6}
342178	ggH125_ee	1.55×10^{-7}	1.45	56 982	1.99×10^{6}	8.86×10^{9}
342189	VBFH125_ee	1.96×10^{-8}	0.98	90 157	3.83×10^{5}	2.00×10^{10}

A.2.6 Drell-Yan

Table A.10: Summary of Drell-Yan samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample.

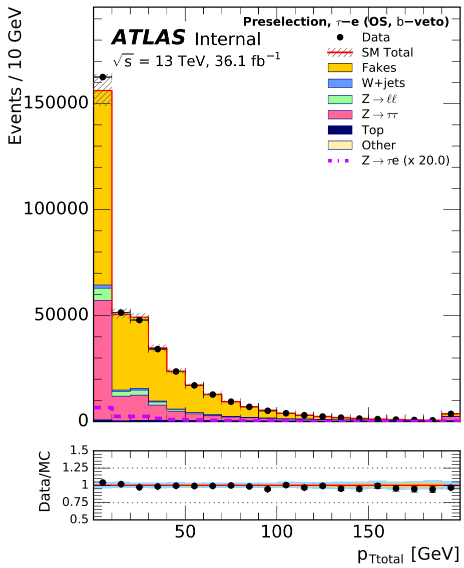
DSID	**Process**	**Filter**	$\sigma \cdot \epsilon$ [pb]	k-factor	N_{gen}	Σw	$\mathcal{L}$ [fb^{-1}]
361665	DYee	10M60	1.76×10^{3}	1.00	646 309	6.75×10^{10}	3.83×10^{4}
361667	DYmumu	10M60	1.81×10^{3}	1.00	244 237	6.75×10^{10}	3.73×10^{4}

Additional Discriminating Variables

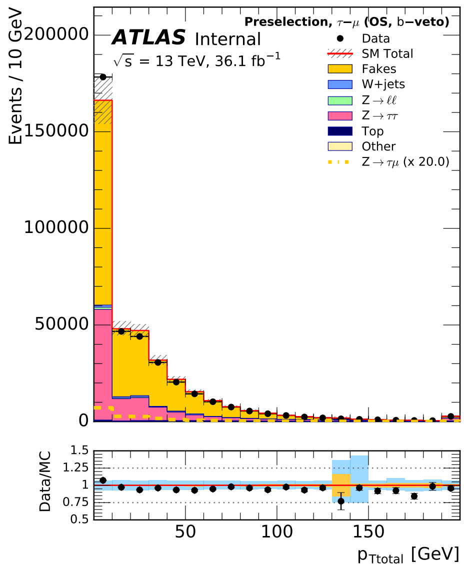


(a) **(b)**

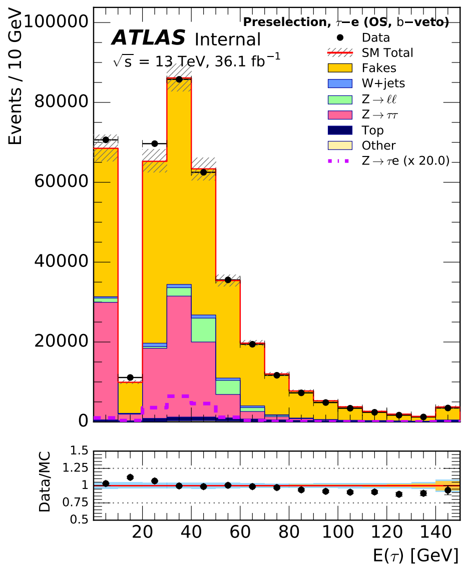
Figure B.1: The $\hat{E}^{\text{lep}}$ distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin.



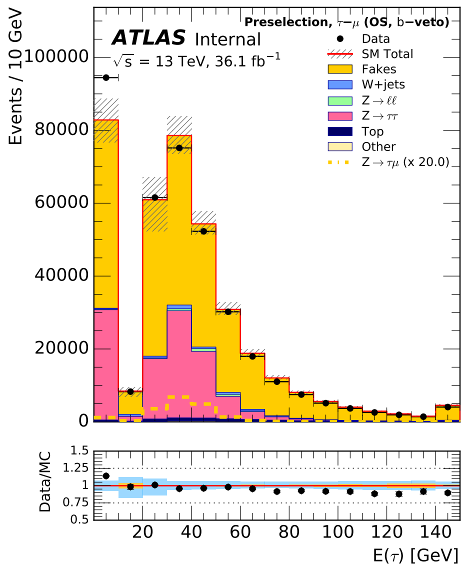
(a)



(b)



(c)



(d)

Figure B.2: The $p_{\mathrm{T}}^{\mathrm{tot}}$ distribution in the (a) τ-e and the (b) τ-μ and the $\hat{E}^{\tau_{\text{had-vis}}}$ in the (c) τ-e and the (d) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin.

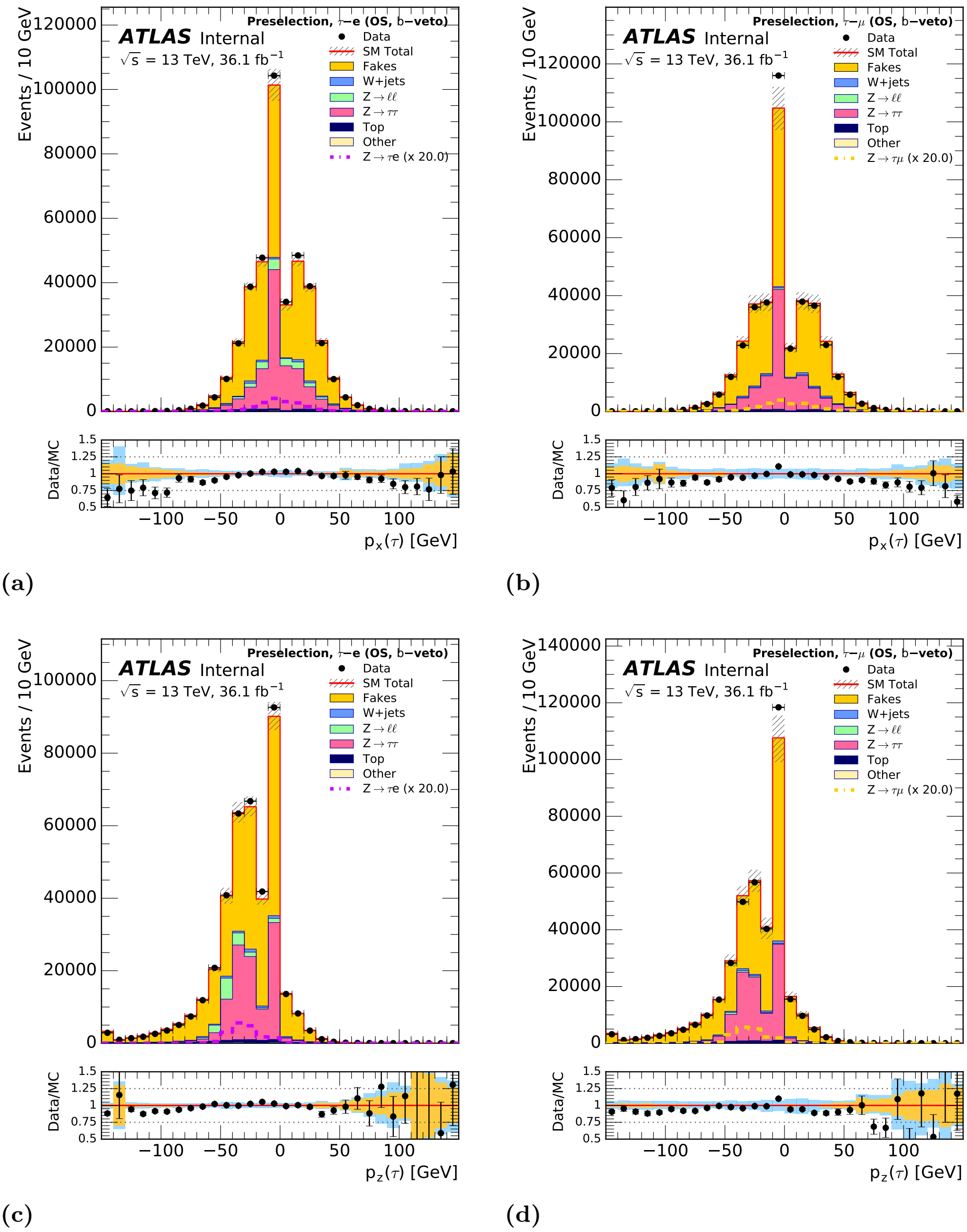
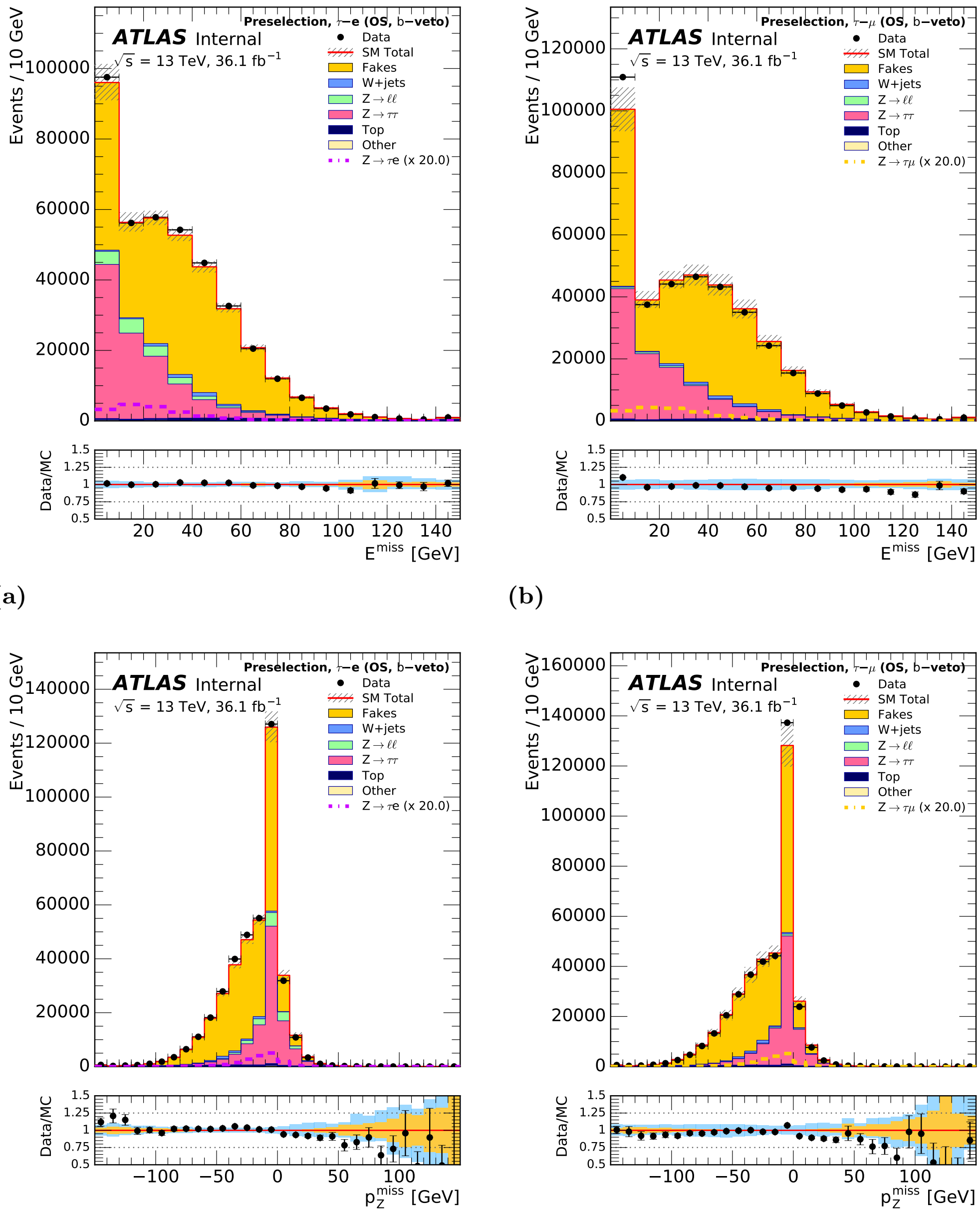


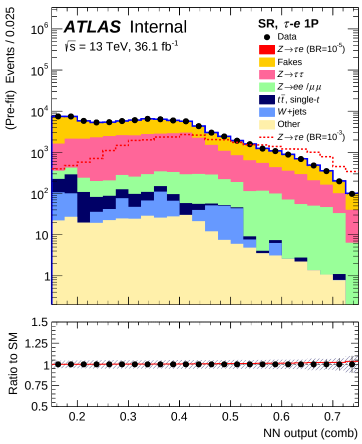
Figure B.3: The $\hat{p}_x^{\tau_{\text{had-vis}}}$ distribution in the (a)τ-e and the (b) τ-μ, $\hat{p}_z^{\tau_{\text{had-vis}}}$ in the (c)τ-e and (d) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin.



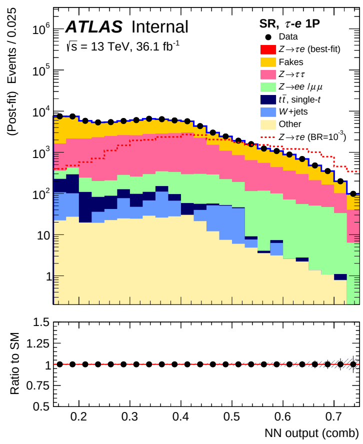
(a) (b) (c) (d)

Figure B.4: The $\hat{E}^{\text{miss}}$ distribution in the (a) τ-e and the (b) τ-μ and $\hat{p}_z^{\text{miss}}$ in the (c) τ-e and (d) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin.

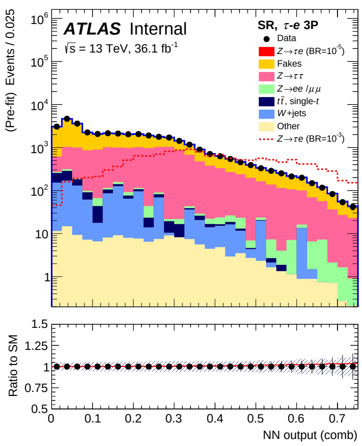
Asimov Fit Distributions



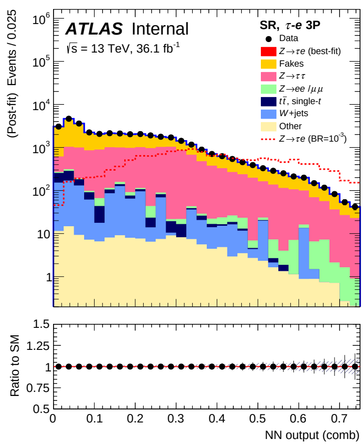
(a) SR τ-e, 1-prong (pre-fit)



(b) SR τ-e, 1-prong (post-fit)

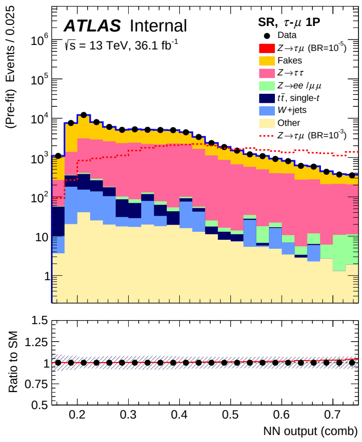


(c) SR τ-e, 3-prong (pre-fit)

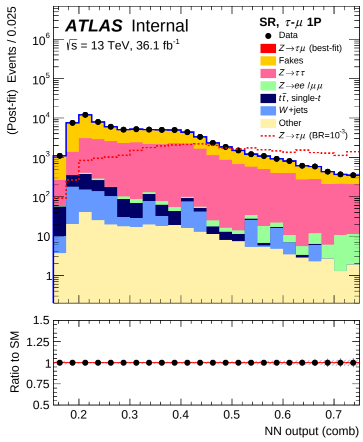


(d) SR τ-e, 3-prong (post-fit)

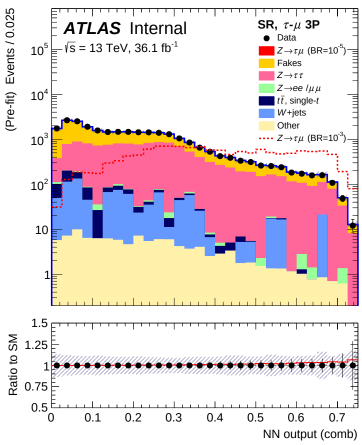
Figure C.1: Pre-fit and post-fit distribution of the NN output in the Asimov fit in the signal region, for the τ-e channel.



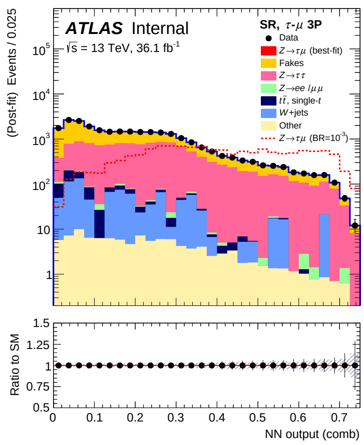
(a) SR τ-μ, 1-prong (pre-fit)



(b) SR τ-μ, 1-prong (post-fit)



(c) SR τ-μ, 3-prong (pre-fit)



(d) SR τ-μ, 3-prong (post-fit)

Figure C.2: Pre-fit and post-fit distribution of the NN output in the Asimov fit in the signal region, for the τ-μ channel.

List of Figures

1.1 An overview of the Standard Model particles. 1
1.2 A neutrino produced as an electron neutrino further interacting as a muon neutrino, due to the change of flavour. 6
1.3 Lepton flavour-changing decay of the Z-boson, due to neutrino oscillations. . . 7
1.4 Feynman diagrams contributing to flavour-changing decays of the Z-boson, in seesaw sypersymmetric theories [2]. 7
1.5 Feynman diagrams contributing to flavour-violating decays of the Z-boson, in two Higgs doublet models [4]. 8

2.1 A scheme of the acceleration chain: the proton-acceleration chain starts with the Linear accelerator 2 (LINAC 2), proceeding with the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS) before being injected into the LHC. Similarly, the acceleration of heavy ions starts with the Linear accelerator 3 (LINAC 3) and proceeds with Low Energy Ion Ring (LEIR), the PS and the SPS. 10
2.2 A scheme of the ATLAS detector. 11
2.3 A scheme of the ATLAS inner detector, consisting of the pixel detector, the semiconductor tracker and the transition radiation tracker in the barrel and the two end-caps. 12
2.4 A scheme of the ATLAS calorimeter system: the electromagnetic calorimeter consists of the LAr based barrel and the two LAr based end-caps, while the hadronic calorimeter consists of the central and the two extended Tile barrel parts and the two LAr based end-caps. The LAr forward calorimeter consists of one module optimized for measurements due to the electromagnetic and two modules optimized for measurements due to the strong interaction. 13
2.5 A scheme of the ATLAS toroid magnet. 13
2.6 A scheme of the ATLAS trigger system [20] 14
2.7 The LHC delivered and the ATLAS recorded cumulative luminosity of the (a) 2015 and (b) 2016 proton-proton collision data collected at the centre-of-mass energy $\sqrt{s} = 13$ TeV[21]. 16
2.8 Luminosity-weighted distribution of the mean number of interactions per bunch crossing for proton-proton collisions at the center-of-mass energy $\sqrt{s} = 13$ TeV, delivered by the LHC during a) 2015 and b) 2016 [21]. 16

3.1 The ATLAS simulation data flow: the output of the MC generator is written to HepMC format [24]. At this stage, a particle filter can be applied to filter events of interest. A detailed output of the event generation is stored in the form of truth particles fwhich are then propagated trough the detector simulation to generate the hits in the detector. At this stage, a simulation of pile up hits can be overlayed with the hits from the hard process to produce the merged hits. The detector hits are further digitized to produce *Raw Data Object* (RAW) files. In the final stage, the reconstruction algorithms are run to form the physics objects. 18
3.2 Feynman diagram for the leptonic (a) and hadronic (b) decay of the τ-lepton in the Standard Model. Labels q_i and q_j refer to different quark flavours. . . . 20
3.3 Examples of leading order Feynman diagrams for $Z/\gamma*+1$ jet production. . . 22
3.4 Drell-Yan process: $q/\bar{q}$ annihilates into a $Z/\gamma*$ which further decays into a lepton pair. 23
3.5 Examples of leading order Feynman diagrams for W+1 jet production. Labels q_i and q_j refer to different quark flavours. 23
3.6 Examples of leading order Feynman diagrams for $t\bar{t}$ pair production via gluon scattering in the s-channel (a) and the t-channel (b) and quark scattering in the s-channel (c). 24
3.7 Examples of leading order Feynman diagrams for single-top production in the s-channel (a) and the t-channel (b) and the associated production with a W-boson in the s-channel (c) and the t-channel (d). Labels q_i and q_j refer to different quark flavours. 24
3.8 Examples of leading order Feynman diagrams for diboson production in the s-channel with the TGC vertex (a) and the t-channel(b). Allowed TGC vertices in SM are WWZ and $WW\gamma$. 25
3.9 Examples of leading order Feynman diagrams for dijet production in the s-channel via q-$\bar{q}$ annihilation (a) triple gluon interaction (b) and in the t-channel via gluon (c) and q-$\bar{q}$ scattering (d). 26
3.10 The dominant channels for the Higgs production at the LHC: gluon fusion (a) and vector boson fusion (b). 27

5.1 The $\Delta\alpha(\tau, \ell)$ distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of $36.1\,\text{fb}^{-1}$ and the signal is multiplied by a factor of 20 to be visible. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. The underflow and the overflow are merged into the first and the last bin. . . 37
5.2 The collinear mass distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of $36.1\,\text{fb}^{-1}$ and the signal is multiplied by a factor of 20 to be visible. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. The underflow and the overflow are merged into the first and the last bin. 39

5.3 The visible mass distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20 to be visible. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. The underflow and the overflow are merged into the first and the last bin. 40

5.4 The transverse mass distribution of the ($\tau_{\text{had-vis}}$,$E_{\mathrm{T}}^{\mathrm{miss}}$) versus the (e,$E_{\mathrm{T}}^{\mathrm{miss}}$) system in the τ-e channel for the: (a) $Z \rightarrow \tau e$, (b) $Z \rightarrow \tau\tau$+jets, (c) $Z \rightarrow ee/\mu\mu$, (d) Fakes, (e) W+jets and (f) top after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20 to be visible. The underflow and the overflow are merged into the first and the last bin. The color axis shows the number of events per bin. 42

5.5 The transverse mass distribution of the ($\tau_{\text{had-vis}}$,$E_{\mathrm{T}}^{\mathrm{miss}}$) versus the (μ,$E_{\mathrm{T}}^{\mathrm{miss}}$) system in the τ-μ channel for the: (a) $Z \rightarrow \tau\mu$, (b) $Z \rightarrow \tau\tau$+jets, (c) $Z \rightarrow ee/\mu\mu$, (d) Fakes, (e) W+jets and (f) top after the trigger, the requirement for one one pair of leptons with opposite charge and requesting no b-quark flavour initiated jets. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20 to be visible. The underflow and the overflow are merged into the first and the last bin. The color axis shows the number of events per bin. 43

5.6 MC distributions of $m(\tau_{\mathrm{trk}}, e/\mu)$ versus $m_{\mathrm{vis}}(\tau, e/\mu)$ in signal (left) and $Z \rightarrow \ell\ell$ (right) events with 1-prong τ candidate in the $\tau - \mu$ (top) and $\tau - e$ (bottom) channel after the signal region selection except for the cuts on those two variables (shown in Tables 6.4 and 6.5) and is used to motivate the selection on these two variables. Shown $Z \rightarrow \ell\ell$ is combined $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ background. The histogram is normalized to 1. 44

6.1 A scheme of neural network architecture. The first layer is called the input layer. It is followed by two hidden layers and the output layer. Neurons from different layes are connected to each of the neurons from the previous and the next layer. 45

6.2 Distributions of the input features for $\mathrm{NN}_{Z\tau\tau}$ in the signal (signal) and $Z \rightarrow \tau\tau$(background) events used in the training for the $e\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48] 50

6.3 Distributions of the input features for $\mathrm{NN}_{Z\tau\tau}$ in the signal (signal) and $Z \rightarrow \tau\tau$(background) events used in the training for the $\mu\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48]. 51

6.4 Distributions of the input features for $\mathrm{NN}_{Z\ell\ell}$ in the signal (signal) and $Z \rightarrow ee/\mu\mu$ (background) events used in the training for the $e\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48] 52

6.5 Distributions of the input features for $\mathrm{NN}_{Z\ell\ell}$ the signal (signal) and $Z \to ee/\mu\mu$ (background) events used in the training for the $\mu\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48] 53
6.6 Distributions of the input features for NN_W in the signal (signal) and W+jets (background) events used in the training for the $e\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48] 54
6.7 Distributions of the input features for NN_W in the signal (signal) and W+jets (background) events used in the training for the $\mu\tau$ channel. The plots on the diagonal show the one-dimensional distributions. The plot off the diagonal show the correlations among each pair of input variables.[48] 55
6.8 Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - e$ channel. 56
6.9 Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - e$ channel. 57
6.10 Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - \mu$ channel. 58
6.11 Data and Standard Model background distributions of neural network inputs at the training stage for the $\tau - \mu$ channel. 59
6.12 Monte Carlo and data distributions of the neural network scores $\mathrm{NN}_W, \mathrm{NN}_{Z\tau\tau}$ and $\mathrm{NN}_{Z\ell\ell}$ in the signal region. To test if the modelling is good, the data and the Monte Carlo predictions are compared in the low sideband of the distributions ($\mathrm{NN}_i > 0.4$), done in order to not unblind the data where the signal is expected, before the full statistical model is built. 63
6.13 Monte Carlo and data distributions of the combined neural network scores $\mathrm{NN}_{\mathrm{comb}}$ in the signal region. To test if the modelling is good, the data and the Monte Carlo predictions are compared in the low sideband of the distributions ($\mathrm{NN}_{\mathrm{comb}} > 0.4$), done in order to not unblind the data where the signal is expected, before the full statistical model is built. 64

7.1 Data and Standard Model background distributions of the $p_{\mathrm{T}}(\tau)$ and $p_{\mathrm{T}}(\tau$-track) in the CRW and CRZll regions of the τ-e channel. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The errors and statistical. 68
7.2 Data and Standard Model background distributions of the $p_{\mathrm{T}}(\tau)$ and $p_{\mathrm{T}}(\tau$-track) in the CRT and CRQ regions of the τ-e channel. The multijet events in CRQ are shown as the difference between the data and the MC prediction for other backgrounds. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The errors and statistical. 69
7.3 Data and Standard Model background distributions of the $p_{\mathrm{T}}(\tau)$ and $p_{\mathrm{T}}(\tau$-track) in the CRW and CRZll regions of the τ-μ channel. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The errors and statistical. 70

7.4 Data and Standard Model background distributions of the $p_{\mathrm{T}}(\tau)$ and $p_{\mathrm{T}}(\tau$-track) in the CRT and CRQ regions of the τ-μ channel.The multijet events in CRQ are shown as the difference between the data and the MC prediction for other backgrounds. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20. The errors and statistical. 71

7.5 Data and Standard Model background distributions of the NN inputs in the W+jets control region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 77

7.6 Data and Standard Model background distributions of the NN inputs and the combined output in the W+jets control region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 78

7.7 Data and Standard Model background distributions of the NN inputs in the same-sign signal region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 79

7.8 Data and Standard Model background distributions of the NN inputs and the combined output in the same-sign signal region of the τ-e channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 80

7.9 Data and Standard Model background distributions of the NN inputs in the W+jets control region of the $\tau-\mu$ channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of $36.1\,\mathrm{fb}^{-1}$ and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 81

7.10 Data and Standard Model background distributions of the NN inputs and the combined output in the W+jets control region of the channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 82
7.11 Data and Standard Model background distributions of the NN inputs in the same-sign signal region of the $\tau-\mu$ channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 83
7.12 Data and Standard Model background distributions of the NN inputs and the combined output in the same-sign signal region of the $\tau-\mu$ channel. *Fakes* show backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The ratio plot in the bottom shows the statistical (yellow) and systematical error (blue) while the dashed band in the upper plot shows the combined error. 84

8.1 Left: The background (blue) and the signal+background (red) hypothesis. The $1-\mathrm{CL_b}$(shaded blue area) is describing the probability to obtain the observed or less likely result under the background hypothesis, while the $\mathrm{CL_{s+b}}$(shaded red area) is describing the probability to obtain the observed or less likely result under the signal+background hypothesis. Right: $\mathrm{CL_b}$ is used as a measure of the separation of the two hypothesis (sensitivity). The top right plot has the largest $\mathrm{CL_b}$ and the highest sensitivity, while the bottom right plot has the smallest $\mathrm{CL_b}$ and the lowest sensitivity. 88
8.2 Post-fit values for fit parameters obtained in the Asimov fit for the (a) τ-e and the (b) τ-μ channels. The branching ratio and the background normalization parameters are presented by its post-fit values and the post-fit uncertainties, while the nuisance parameters showing the effect systematical errors are shown as the difference between the post-fit and the pre-fit central value, divided by the pre-fit uncertainty. Systematical uncertainties are labeled with α and described in Tables 8.2 and 8.1. 94
8.3 Nuisance parameters, ranked by their impact on the signal strength, in the Asimov fit for the (a) τ-e and the (b) τ-μ channels. The black lines are showing their post-fit values with post-fit uncertanties expressed in units of pre-fit uncertanties. The impact of a nuisance parameter at the signal strength is measured by fixing the central value of the nuisance parameter at its up/down variation and redoing the fit: the difference between the signal strength and the best-fit value signal strength divided by the post-fit uncertainty of the best-fit signal strength is shown by the hatched boxes for the up (red) and the down (blue) variation. 95

8.4 Sensitivity scan in the (a) τ-e and the (b) τ-μ channels, obtained using the signal+background model fitted to the Asimov data, defined as the pre-fit Standard Model background expectation. The branching ratio in units of 10^{-5} is shown in the x-axis, while the p-values for the signal+background (CLs+b), background (1-CLb) and the CLs = $\frac{CL_{s+b}}{CL_b}$ are shown in the y-axis. The expected and the observed limits are obtained as the values resulting in the CLs-value of 0.05, using the ATLAS default CLs method. The observed (=expected) upper limit are set at 2.34×10^{-5} for the τ-e and at 2.37×10^{-5} for the τ-μ channel. . 96
8.5 Post-fit values for fit parameters obtained in the sideband fit for the (a) τ-e and the τ-μ channels. The branching ratio and the background normalization parameters are presented by its post-fit values and the post-fit uncertainties, while the nuisance parameters which showing the effect of statistical and systematical errors are shown as the difference between the post-fit and the pre-fit central value, divided by the pre-fit uncertainty. Systematical uncertainties are labeled with α and described in Tables 8.2 and 8.1. 98
8.6 Nuisance parameters, ranked by their impact on the signal strength, in the sideband fit for the (a) τ-e and the (b) τ-μ channels. The black lines are showing their post-fit values with post-fit uncertanties expresses in units of pre-fit uncertanties. The impact of a nuisance parameter at the signal strength is measured by fixing the central value of the nuisance parameter at its up/down variation and redoing the fit: the difference between the signal strength obtained in this way and its best-fit value divided by the post-fit uncertainty of the best-fit signal strength is shown by the hatched boxes for the up (red) and the down (blue) variation. 99
8.7 Pre-fit and post-fit distribution of the NN output in the low sideband (NN<0.4) of the signal region, for the τ-e channel. 100
8.8 Pre-fit and post-fit distribution of the NN output in the low sideband (NN<0.4) of the signal region, for the τ-μ channel. 101
8.9 Post-fit values for fit parameters obtained in the unblinded fit for the (a) τ-e and the (b) τ-μ channels. The branching ratio and the background normalization parameters are presented by its post-fit values and the post-fit uncertainties, while the nuisance parameters showing the effect systematical errors are shown as the difference between the post-fit and the pre-fit central value, divided by the pre-fit uncertainty. Systematical uncertainties are labeled with α and described in Tables 8.2 and 8.1. 103
8.10 Pre-fit and post-fit distribution of the NN output in the unblinded fit in the signal region, for the τ-e channel. 104
8.11 Pre-fit and post-fit distribution of the NN output in the unblinded fit in the signal region, for the τ-μ channel. 105
8.12 Post-fit NN-input distributions for 1-prong events in the signal region of the τ-e channel. 106
8.13 Post-fit NN-input distributions for 3-prong events in the signal region of the τ-e channel. 107
8.14 Post-fit NN-input distributions for 1-prong events in the signal region of the τ-μ channel. 108
8.15 Post-fit NN-input distributions for 3-prong events in the signal region of the τ-μ channel. 109

8.16 Upper limit scan obtained in the signal+background model fit to the data for the (a) τ-e and the (b) τ-μ channels. The branching ratio in units of 10^{-5} is shown in the x-axis, while the p-values for the signal+background (CLs+b), background (1-CLb) and the CLs = $\frac{CL_{s+b}}{CL_b}$ are shown in the y-axis. The expected and the observed limits are obtained as the values resulting in the CLs-value of 0.05, using the ATLAS default CLs method. The observed (expected) upper limit are set at for the τ-e and at for the τ-μ channel. 111

B.1 The $\hat{E}^{\text{lep}}$ distribution in the (a) τ-e and the (b) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin. . . . 121

B.2 The $p_{\mathrm{T}}^{\text{tot}}$ distribution in the (a) τ-e and the (b) τ-μ and the $\hat{E}^{\tau_{\text{had-vis}}}$ in the (c) τ-e and the (d) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin. 122

B.3 The $\hat{p}_x^{\tau_{\text{had-vis}}}$ distribution in the (a)τ-e and the (b) τ-μ, $\hat{p}_z^{\tau_{\text{had-vis}}}$ in the (c)τ-e and (d) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin. 123

B.4 The $\hat{E}^{\text{miss}}$ distribution in the (a) τ-e and the (b) τ-μ and $\hat{p}_z^{\text{miss}}$ in the (c) τ-e and (d) τ-μ channel after the trigger, the requirement for one OS(τ, e/μ) pair and the b-jet veto. *Fakes* are backgrounds in which the τ is faked by a jet, while other backgrounds show MC events where the τ is matched to either a truth τ,e or μ. The distributions are normalized to the luminosity of 36.1 fb^{-1} and the signal is multiplied by a factor of 20. The underflow and the overflow are merged in with the first and the last bin. 124

C.1 Pre-fit and post-fit distribution of the NN output in the Asimov fit in the signal region, for the τ-e channel. 125

C.2 Pre-fit and post-fit distribution of the NN output in the Asimov fit in the signal region, for the τ-μ channel. 126

List of Tables

3.1 Monte Carlo generators used in the simulation of the signal and the Standard Model background processes and the order at which the cross section was computed. 20

4.1 The list of lowest unprescaled single-electron/muon triggers used in the analysis. The HLT trigger name is typically formed by the name of the object $(e/\mu), followedbythe p_{\mathrm{T}}\ cut (e.g. e24 means p_{\mathrm{T}} > 24\,\mathrm{GeV} and the identification requirement (e.g. lhmedium = likelihood medium). In addition,$ HLT_e24_lhmedium_L1EM20VH $indicates an unusual requirement on the ener$

4.2 The event cleaning efficiency for MC and data events. The table shows efficiencies with respect to the number of processed events. The GRL and Detector cleaning are applied to data only, while other selections are applied to both the data and the MC. 29

4.3 An overview of the object definitions used in the analysis. 33

6.1 Lists of kinematical variables given as inputs to neural networks. The scores NN_W, $\mathrm{NN}_{Z\tau\tau}$and $\mathrm{NN}_{Z\ell\ell}$refer to different networks trained to discriminate between the signal and the W+jets, $Z \to \tau\tau$ and $Z\ell\ell$ processes. 48

6.2 Selection applied to the events used for training: the common selection is applied in both channels, while the additional selections are applied per a specific channel. 48

6.3 Number of Monte Carlo events used for the training of the neural-network classifiers in the $e\tau$ and $\mu\tau$ channels. 49

6.4 Cutflow for τ-e channel, normalised to 36.1fb^{-1}. The signal branching ratio is assumed to be 1.2×10^{-5}. 61

6.5 Cutflow for τ-μ channel(s), normalised to 36.1 fb^{-1}. The signal branching ratio is assumed to be 1.2×10^{-5}. 62

7.1 Contol regions used to derive fake-factors. Differences from the loose SR selection are listed together with the purities for the target processes 67

7.2 Fake-factor per background computed in the dedicated control regions of the τe channel. 72

7.3 Fake-factor per background computed in the dedicated control regions of the $\tau\mu$ channel. 73

7.4 MC correction k-factor per background computed in the dedicated control regions of the τe channel. 73

7.5 MC correction k-factor per background computed in the dedicated control regions of the $\tau\mu$ channel. 74

7.6 R-factor in the signal region of the τe channel. 74

7.7 R-factor in the signal region of the $\tau\mu$ channel. 75

7.8 Fake-factor in the signal region for the τe and $\tau\mu$ channels. 75

8.1 Systematics Table I 90
8.2 Systematics Table II 91
8.3 Asimov fit results in the τ-e channel 92
8.4 Asimov fit results in the τ-μ channel 93
8.5 Asimov upper limit results 93
8.6 Sideband fit results in the τ-e channel 97
8.7 Sideband fit results in the τ-μ channel 98
8.8 Unblinded fit results in the τ-e channel 102
8.9 Unblinded fit results in the τ-μ channel 102
8.10 Discovery test results 110
8.11 Unblinded upper limit results 111

A.1 Summary of the $Z \to \tau e$ and $Z \to \tau\mu$ signal samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section multiplied with the generator filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for the process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 114
A.2 Summary of the $Z \to ee$ and $Z \to \mu\mu$ background samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section multiplied with the generator filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for the process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 115
A.3 Summary of $Z \to \tau\tau$ background samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section multiplied with the generator filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for the process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. . 115
A.4 Summary of $W \to e\nu_e$ samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 116
A.5 Summary of $W \to \mu\nu_\mu$ samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 117
A.6 Summary of $W \to \tau\nu_\tau$ samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 118

A.7 Summary of $t\bar{t}$, single-top and Wt samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 119
A.8 Summary of diboson samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 119
A.9 Summary of Higgs samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 120
A.10 Summary of Drell-Yan samples used in the analysis. $\sigma \cdot \epsilon$ corresponds to the cross-section provided by the Monte Carlo generator multiplied with a possible truth-level filter efficiency. The k-factor is used to normalise the generator cross-section to the best known for a process. $\sum w$ includes the generator-level weights and the pileup reweighting effect. $\mathcal{L}_{int}$ is the generated equivalent integrated luminosity for the sample. 120

Bibliography

[1] Anna Rossi. "Lepton flavor violation in supersymmetric models". In: *15th Conference on High Energy Physics (IFAE 2003) Lecce, Italy, April 23-26, 2003*. 2003. arXiv: hep-ph/0311320 [hep-ph].

[2] Jin Min Yang. "Lepton flavor violating Z-boson decays at GigaZ as a probe of supersymmetry". In: *Sci. China Phys. Mech. Astron.* 53 (2010), pp. 1949–1952. DOI: 10.1007/s11433-010-4146-3. arXiv: 1006.2594 [hep-ph].

[3] Andreas Crivellin, Christoph Greub, and Ahmet Kokulu. "Flavour-violation in two-Higgs-doublet models". In: *PoS* EPS-HEP2013 (2013), p. 338. arXiv: 1309.4806 [hep-ph].

[4] R. Benbrik, Chuan-Hung Chen, and Takaaki Nomura. "$h, Z \to \ell_i \bar{\ell}_j$, Δa_μ, $\tau \to (3\mu, \mu\gamma)$ in generic two-Higgs-doublet models". In: *Phys. Rev.* D93.9 (2016), p. 095004. DOI: 10.1103/PhysRevD.93.095004. arXiv: 1511.08544 [hep-ph].

[5] Ferruccio Feruglio, Paride Paradisi, and Andrea Pattori. "Lepton Flavour Violation in Composite Higgs Models". In: *Eur. Phys. J.* C75.12 (2015), p. 579. DOI: 10.1140/epjc/s10052-015-3807-9. arXiv: 1509.03241 [hep-ph].

[6] Abhishek M Iyer and Sudhir K Vempati. "Lepton Masses and Flavor Violation in Randall Sundrum Model". In: *Phys. Rev.* D86 (2012), p. 056005. DOI: 10.1103/PhysRevD.86.056005. arXiv: 1206.4383 [hep-ph].

[7] Takaaki Kajita. "Discovery of atmospheric neutrino oscillations". In: *Annalen Phys.* 528 (2016), pp. 459–468. DOI: 10.1002/andp.201600086.

[8] OPAL Collaboration. "A search for lepton flavour violating Z^0 decays". In: *Zeitschrift für Physik C Particles and Fields* 67.4 (1995), pp. 555–563. DOI: 10.1007/BF01553981.

[9] DELPHI Collaboration. "Search for lepton flavour number violating Z^0- decays". In: *Zeitschrift für Physik C Particles and Fields* 73.2 (1997), pp. 243–251. DOI: 10.1007/s002880050313.

[10] Georges Aad et al. "Search for lepton-flavour-violating decays of the Higgs and Z bosons with the ATLAS detector". In: *Eur. Phys. J.* C77.2 (2017), p. 70. DOI: 10.1140/epjc/s10052-017-4624-0. arXiv: 1604.07730 [hep-ex].

[11] Lyndon Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[12] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[13] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[14] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[15] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[16] G. Latino. “The TOTEM Experiment at the LHC”. In: *QCD and high energy interactions. Proceedings, 44th Rencontres de Moriond, La Thuile, Italy, March 14-21, 2009*. 2009, pp. 357–360. arXiv: 0905.2936 [hep-ex]. URL: http://inspirehep.net/record/820761/files/arXiv:0905.2936.pdf.

[17] B. Acharya et al. “The Physics Programme Of The MoEDAL Experiment At The LHC”. In: *Int. J. Mod. Phys.* A29 (2014), p. 1430050. DOI: 10.1142/S0217751X14300506. arXiv: 1405.7662 [hep-ph].

[18] M. del Prete et al. “LHCf experiment: forward physics at LHC for cosmic rays study”. In: *EPJ Web Conf.* 126 (2016), p. 04014. DOI: 10.1051/epjconf/201612604014.

[19] A. Airapetian et al. “ATLAS: Detector and physics performance technical design report. Volume 1”. In: (1999).

[20] Morad Aaboud et al. “Performance of the ATLAS Trigger System in 2015”. In: *Eur. Phys. J.* C77.5 (2017), p. 317. DOI: 10.1140/epjc/s10052-017-4852-3. arXiv: 1611.09661 [hep-ex].

[21] *ATLAS Luminosity PublicResults in Run-2*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_AN4.

[22] *Athena Framework*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasComputing/AthenaFramework.

[23] G. Aad et al. “The ATLAS Simulation Infrastructure”. In: *Eur. Phys. J.* C70 (2010), pp. 823–874. DOI: 10.1140/epjc/s10052-010-1429-9. arXiv: 1005.4568 [physics.ins-det].

[24] *HepMC 2: a C++ Event Record for Monte Carlo Generators*. URL: http://lcgapp.cern.ch/project/simu/HepMC/20400/HepMC2_user_manual.pdf.

[25] Andy Buckley et al. “General-purpose event generators for LHC physics”. In: *Phys. Rept.* 504 (2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].

[26] Torbjorn Sjostrand. “Monte Carlo Generators”. In: *High-energy physics. Proceedings, European School, Aronsborg, Sweden, June 18-July 1, 2006*. 2006, pp. 51–74. arXiv: hep-ph/0611247 [hep-ph]. URL: http://weblib.cern.ch/abstract?CERN-LCGAPP-2006-06.

[27] S. Agostinelli et al. “GEANT4: A Simulation toolkit”. In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[28] W. Lukas. “Fast Simulation for ATLAS: Atlfast-II and ISF”. In: 2012. URL: http://cdsweb.cern.ch/record/1458503/files/ATL-SOFT-PROC-2012-065.pdf.

[29] William Buttinger. *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*. Tech. rep. ATL-COM-SOFT-2015-119. Geneva: CERN, May 2015. URL: https://cds.cern.ch/record/2014726%22.

[30] *Pythia 6*. URL: https://pythia6.hepforge.org.

[31] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. “A Brief Introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[32] Carlo Oleari. “The POWHEG-BOX”. In: *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 36–41. DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph].

[33] Peter Zeiler Skands. “Tuning Monte Carlo Generators: The Perugia Tunes”. In: 82 (Sept. 2010), p. 74018.

[34] The ATLAS collaboration. “Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data”. In: (2016).

[35] Georges Aad et al. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV”. In: *Eur. Phys. J.* C76.5 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex].

[36] Felix Friedrich. “Tau Lepton Reconstruction and Identification at ATLAS”. In: (2012). [EPJ Web Conf.28,2007(2012)]. DOI: 10.1051/epjconf/20122812007. arXiv: 1201.5466 [hep-ex].

[37] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Anti-k(t) jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[38] W. Lampl et al. “Calorimeter clustering algorithms: Description and performance”. In: (2008).

[39] *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. ATLAS-CONF-2014-018. Geneva: CERN, May 2014. URL: https://cds.cern.ch/record/1700870.

[40] K G Tomiwa. “Performance of Jet Vertex Tagger in suppression of pileup jets and IMG [http://ej.iop.org/images/1742-6596/802/1/012012/toc$_JPCS_802_{10}12012_ieqn1.gif$]$E_T^{miss}$ in ATLAS detector”. In: *Journal of Physics: Conference Series* 802.1 (2017), p. 012012. URL: http://stacks.iop.org/1742-6596/802/i=1/a=012012.

[41] Marc Lehmacher. “b-Tagging Algorithms and their Performance at ATLAS”. In: *Proceedings, 34th International Conference on High Energy Physics (ICHEP 2008): Philadelphia, Pennsylvania, July 30-August 5, 2008*. 2008. arXiv: 0809.4896 [hep-ex]. URL: http://weblib.cern.ch/abstract?ATL-PHYS-PROC-2008-052.

[42] Claire A Lee. “Missing p T Reconstruction at ATLAS”. In: *Journal of Physics: Conference Series* 645.1 (2015), p. 012012. URL: http://stacks.iop.org/1742-6596/645/i=1/a=012012.

[43] *ATLAS Analysis Harmonisation Group 5: Overlap Removal*. URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/AnalysisHarmonisationGroup5.

[44] Sacha Davidson, Sylvain Lacroix, and Patrice Verdier. “LHC sensitivity to lepton flavour violating Z boson decays”. In: *JHEP* 09 (2012), p. 092. DOI: 10.1007/JHEP09(2012)092. arXiv: 1207.4894 [hep-ph].

[45] A. Elagin et al. “A New Mass Reconstruction Technique for Resonances Decaying to di-tau”. In: *Nucl. Instrum. Meth.* A654 (2011), pp. 481–489. DOI: 10.1016/j.nima.2011.07.009. arXiv: 1012.4686 [hep-ex].

[46] Daniele Zanzi, Hubert Kroha, and Sandra Kortner. “Search for the Standard Model Higgs Boson in Hadronic $\tau^+\tau^-$ Decays with the ATLAS Detector”. Presented 20 05 2014. May 2014. URL: https://cds.cern.ch/record/1746045.

[47] *Convolutional Neural Networks for Visual Recognition*. URL: http://cs231n.github.io.

[48] Morad Aaboud et al. “A search for lepton-flavor-violating decays of the Z boson into a τ-lepton and a light lepton with the ATLAS detector”. In: *Submitted to: Phys. Rev.* (2018). arXiv: 1804.09568 [hep-ex].

[49] Olaf Behnke et al., eds. *Data analysis in high energy physics*. Weinheim, Germany: Wiley-VCH, 2013. ISBN: 9783527410583, 9783527653447, 9783527653430. URL: http://www.wiley-vch.de/publish/dt/books/ISBN3-527-41058-9.

[50] M. Baak et al. “HistFitter software framework for statistical data analysis”. In: *Eur. Phys. J.* C75 (2015), p. 153. DOI: 10.1140/epjc/s10052-015-3327-7. arXiv: 1410.1280 [hep-ex].

[51] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J.* C71 (2011). [Erratum: Eur. Phys. J.C73,2501(2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. arXiv: 1007.1727 [physics.data-an].

[52] Alexander L. Read. “Presentation of search results: The CL(s) technique”. In: *J. Phys.* G28 (2002). [,11(2002)], pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[53] *Statistical Methods for Data Analysis: Upper Limits*. URL: http://people.na.infn.it/~lista/Statistics/slides/06%20-%20upper%20limits.pdf.

[54] Kyle Cranmer et al. “HistFactory: A tool for creating statistical models for use with RooFit and RooStats”. In: *Eur. Phys. J.* (2012).

[55] Lorenzo Moneta et al. “The RooStats Project”. In: *PoS* ACAT2010 (2010), p. 057. arXiv: 1009.1003 [physics.data-an].

[56] Wouter Verkerke and David P. Kirkby. “The RooFit toolkit for data modeling”. In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].