

# Improvements of the search for a light CP-odd Higgs boson decaying to a pair of tau leptons using artificial neural networks

Masterarbeit  
zur Erlangung des Hochschulgrades  
Master of Science  
im Masterstudiengang Physik

vorgelegt von

Hannah Jacobi  
geboren am 20.02.1999 in Gehrden

Institut für Kern- und Teilchenphysik  
Fakultät Physik  
Bereich Mathematik und Naturwissenschaften  
Technische Universität Dresden  
2023

Eingereicht am 13. November 2023

1. Gutachter: Prof. Dr. Arno Straessner
2. Gutachter: Prof. Dr. Dominik Stöckinger

## Abstract

Despite the Standard Model's (SM) many achievements, there are still some open questions in particle physics. For example, there is a significant deviation between theoretically predicted and experimentally observed value of the anomalous magnetic moment of the muon  $a_\mu$ . This problem can be solved by the two-Higgs-doublet Model (2HDM). This theory predicts, among other new particles, a CP-odd Higgs boson  $A$ . In order to explain the measured value  $a_\mu$  the  $A$  boson needs to have a light mass and couple strongly to leptons and up-type quarks. This thesis presents an analysis searching for such a particle in the mass range from 20 GeV to 110 GeV using  $140.1 \text{ fb}^{-1}$  of data recorded by the ATLAS detector between 2015 and 2018. The  $A$  boson is expected to almost exclusively decay into a pair of  $\tau$ -leptons that are investigated in their leptonic decay channel.

Two techniques to improve the modelling of background processes compared to data are presented in the context of a cut-based analysis. The focus of this thesis lies on the extension of this analysis by introducing a parameterised neural network to classify events into signal or background. In the end it was possible to reach an expected cross section times branching ratio limit that is roughly comparable to the one of the cut-based analysis.

## Zusammenfassung

Trotz aller Errungenschaften des Standardmodells (SM) gibt es noch unbeantwortete Fragen in der Teilchenphysik. Das Zwei-Higgs-Dublett-Modell (2HDM) ist eine einfache Erweiterung des SM, das unter anderem in der Lage ist, die signifikante Abweichung zwischen theoretisch vorhergesagten und experimentell gemessenen Wert des anomalen magnetischen Moment des Myons  $a_\mu$ , zu erklären. Diese Theorie sagt unter anderem die Existenz eines CP-ungeraden Higgs-Bosons  $A$ . Um den gemessenen Wert von  $a_\mu$  zu erklären, muss es eine geringe Masse und eine starke Kopplung zu Leptonen und Quarks vom up-Typ besitzen. Die vorliegende Arbeit stellt eine Analyse vor, die nach einem derartigen Teilchen mit einer Masse zwischen 20 GeV und 110 GeV sucht. Hierfür werden  $140.1 \text{ fb}^{-1}$  von Messdaten genutzt, die zwischen 2015 und 2018 vom ATLAS Detektor aufgezeichnet wurden. Es ist zu erwarten, dass das beschriebene  $A$ -Boson fast ausschließlich in Paare von  $\tau$  Leptonen zerfällt, die hier im rein leptonischen Zerfallskanal untersucht werden. Im Rahmen einer Cut-basierten Analyse werden zwei datengetriebene Methoden vorgestellt, um die Modellierung von Hintergrundprozessen bezüglich der Verteilung von gemessenen Daten zu verbessern. Der Fokus liegt jedoch darin, diese Basisanalyse durch die Einführung eines parametrisierten Neuronalen Netzes zur Klassifizierung von Ereignissen als Signal- oder Hintergrundprozess, zu erweitern. Letztlich war es möglich eine statistische obere Grenze auf den Wirkungsquerschnitt der Erzeugung des  $A$ -Bosons multipliziert mit der Zerfallsbreite in den Endzustand zu bestimmen. Diese ist in etwa vergleichbar mit der zugrundeliegenden Analyse.

# Contents

<b>1. Introduction</b>	<b>1</b>
<b>2. Theoretical Foundations</b>	<b>2</b>
2.1. The Standard Model of Particle Physics . . . . .	2
2.1.1. Phenomenology and Particle Content . . . . .	3
2.1.2. Discrete Symmetries in the Standard Model . . . . .	4
2.1.3. Electroweak Unification . . . . .	5
2.1.4. Higgs Mechanism . . . . .	7
2.1.5. Deficiencies of the Standard Model . . . . .	9
2.2. The Two-Higgs-Doublet Model . . . . .	11
2.2.1. The Flavour-Aligned Two-Higgs Doublet Model . . . . .	12
2.2.2. Parameter Constraints . . . . .	13
2.3. Statistics . . . . .	16
2.3.1. Theory of Limit Setting . . . . .	16
2.3.2. Expected Limits . . . . .	19
<b>3. The ATLAS Experiment at the LHC</b>	<b>20</b>
3.1. The Large Hadron Collider . . . . .	20
3.2. The ATLAS Detector . . . . .	21
3.2.1. The ATLAS Coordinate System . . . . .	22
3.2.2. The Inner Detector . . . . .	23
3.2.3. The Calorimeter System . . . . .	24
3.2.4. The Muon System . . . . .	25
3.2.5. The Trigger System . . . . .	25
<b>4. Base Analysis</b>	<b>27</b>
4.1. Physics Object Identification and Reconstruction . . . . .	27
4.1.1. Electrons . . . . .	27
4.1.2. Muons . . . . .	28
4.1.3. Jets . . . . .	28
4.1.4. B-Tagging . . . . .	28
4.1.5. Overlap Removal . . . . .	28
4.1.6. Missing Transverse Energy . . . . .	29
4.1.7. Missing Mass Calculator and Other Mass Variables . . . . .	29
4.2. Monte Carlo and Data Samples . . . . .	30
4.2.1. Generation of Monte Carlo Samples . . . . .	31
4.2.2. Signal Process . . . . .	31
4.2.3. Background Processes . . . . .	33
4.3. Event Selection . . . . .	37
4.3.1. Trigger . . . . .	37
4.3.2. Baseline Selection . . . . .	37

4.3.3. Common Selection . . . . .	38
4.4. Data-Driven Estimation of Fake Background Using the Matrix Method	39
4.4.1. On the Matrix Method in General . . . . .	40
4.4.2. Calculation of Matrix Method Efficiencies . . . . .	42
4.4.3. Application of the Matrix Method . . . . .	44
4.5. Reweighting Procedure for the $Z \rightarrow \tau\tau$ -Sample . . . . .	45
4.6. Results of the Cut-Based Analysis . . . . .	50
<b>5. Introduction to Neural Networks</b>	<b>52</b>
5.1. Neural Networks as an Extension to High-Energy Physics Analyses . .	52
5.2. Neural Networks . . . . .	52
5.3. Training of Neural Networks . . . . .	55
5.4. $K$ -Fold Cross Validation . . . . .	58
5.5. Parameterised Neural Networks . . . . .	59
5.6. Comparing Classification Performance . . . . .	60
5.7. Optimisation of Hyperparameters . . . . .	62
<b>6. Multivariate Extension to the Base Analysis</b>	<b>64</b>
6.1. Training Data and Event Weights . . . . .	64
6.2. Hyperparameter Optimisation . . . . .	67
6.3. Input Variables . . . . .	68
6.4. Architecture of the Neural Network . . . . .	70
6.5. Training of the Neural Network . . . . .	70
6.6. Application of the Neural Network . . . . .	72
6.7. Scores and Limits . . . . .	74
6.8. Additional Studies . . . . .	77
6.8.1. Event Weights . . . . .	77
6.8.2. Interpolation Between Signal Masses . . . . .	80
6.8.3. Differences Between $K$ -Folds . . . . .	81
<b>7. Summary and Outlook</b>	<b>84</b>
<b>Bibliography</b>	<b>86</b>
<b>Acronyms</b>	<b>95</b>
<b>List of Figures</b>	<b>99</b>
<b>List of Tables</b>	<b>101</b>
<b>A. Full Sample Selection</b>	<b>102</b>
A.1. Data Samples . . . . .	102
A.2. MC Samples . . . . .	102
A.2.1. mc16a . . . . .	102
A.2.2. mc16d . . . . .	106
A.2.3. mc16e . . . . .	110
A.3. Full List of Triggers . . . . .	114
<b>B. Dependence of MM Efficiencies on the <math>Z \rightarrow \tau\tau</math> Reweighting</b>	<b>115</b>

---

---

<b>C. Mismodelling and Reweighting Studies on the <math>Z \rightarrow \tau\tau</math>-background</b>	<b>117</b>
<b>D. Neural Network Extension to the Analysis</b>	<b>123</b>
D.1. Example Configuration of the Hyperparameter Optimisation Framework	123
D.2. Distribution of All Input Variables to the NN . . . . .	127
D.2.1. Distributions in the mvaVR . . . . .	127
D.2.2. Distributions in the ZVR . . . . .	132
D.3. Overfitting in a Longer Training . . . . .	136
D.4. Distribution of All Score Variables Calculated by the NN . . . . .	137
D.4.1. Score Distributions in the mvaVR . . . . .	137
D.4.2. Score Distributions in the TVR . . . . .	140
D.4.3. Score Distributions in the ZVR . . . . .	142
<b>E. Limit Studies</b>	<b>144</b>
E.1. Influence of the Input-Binning on Limits . . . . .	144
E.2. Neural Network Limits Calculated in the Signal Regions . . . . .	146
E.3. Statistics-Only Final Limit . . . . .	147

---





# 1. Introduction

Particle physics studies the fundamental building blocks of our universe and their interactions. This field of physics was initialised by the discovery of the electron in 1897 by J.J. Thomson [1]. Since then, our knowledge of the subatomic world grew dramatically and culminated in the Standard Model of particle physics (SM). It combines the description of the electroweak sector, the quantum chromodynamics (QCD) and the Higgs mechanism into one theory. The SM is one of the most thoroughly tested theories in physics and is able to describe all experimentally observed particles, their properties and interactions. With the discovery of the Higgs boson at the Large Hadron Collider (LHC) in 2012 [2, 3], all particles predicted by the SM have been experimentally observed.

However, it is known that the SM is not a complete theory. One example would be that the fourth fundamental force, gravity, is not included in it. There are also several phenomena which have been observed in experiments but can't be explained using the SM. These phenomena include the existence of dark matter, the matter-antimatter asymmetry in the universe as well as the deviation between the measured and predicted value of the anomalous magnetic moment of the muon  $a_\mu = (g_\mu - 2)/2$  [4].

Many different theories have been developed that extend the Standard Model in order to explain these discrepancies. One simple way is introducing a second Higgs doublet to achieve a two-Higgs-doublet model (2HDM). This leads to a total of four additional Higgs bosons  $H$ ,  $H^\pm$  and  $A$ . Specific value choices for parameters of this model are able to explain the deviation of the anomalous magnetic moment of the muon while the model remains compatible with other experimental measurements. These parameter values suggest a light CP-odd  $A$  boson with relatively strong couplings to leptons and up-type quarks [5].

This thesis presents an analysis searching for a particle with properties like this  $A$  boson. It uses data recorded by the ATLAS detector during Run 2 of the LHC with a centre-of-mass energy of  $\sqrt{s} = 13$  TeV. It builds on previous work by P. Moder [6], T. Krefe [7], X. Sonntag [8], C. Schmidt [9] and J. Friese [10]. The focus lies on the improvement of a neural network that is used as an extension to the otherwise cut-based analysis as well as the estimation of a fake lepton background and a data-driven reweighting procedure for a single Monte Carlo simulated background process.

This thesis is structured as follows: Chapter 2 gives an introduction to the theoretical foundations of the analysis. This includes the SM and the two-Higgs-doublet model as a possible expansion. It also touches the topic of statistics, which is an integral part of any particle physics analysis. The LHC and the ATLAS detector in particular are described in Chapter 3. The base analysis is outlined in Chapter 4. This section introduces the signal and background processes as well as the data samples and event selection used. Chapter 5 and 6 focuses on neural network (NN). It contains an introduction to the general concepts of NN and describes their integration into the underlying base analysis. Finally, Chapter 7 closes this thesis with a summary of the results and an outlook on possible future work.

## 2. Theoretical Foundations

This chapter covers the theoretical principles that are needed to understand the goal of this thesis and its reasoning. As the CP-odd  $A$  boson is part of the 2HDM, which is an extension to the SM, both the SM and the 2HDM will be introduced here. After that an overview on the statistical methods used in this analysis is given.

### 2.1. The Standard Model of Particle Physics

The Standard Model is very good at theoretically describing phenomena observed in particle physics experiments. It is a very extensive theory that combines different theoretical descriptions, that were developed individually over the course of the 20th century, into one.

The development of the SM started with A. Einstein's theory of relativity in 1916 [11]. In the 1940s, quantum electrodynamics (QED) was formulated by S. Tomonaga [12], J. Schwinger [13] and R. P. Feynman [14]. It is a fundamental quantum theory describing electromagnetism in a fully covariant manner. A similar formalism for the theory of weak interactions was introduced in the 1960s by S. L. Glashow [15], A. Salam [16] and S. Weinberg [17]. It predicted the existence of the  $W^\pm$  and  $Z^0$  bosons which were then experimentally observed by the UA1 and UA2 experiments [18, 19] 20 years later. However, this theory was not able to explain the existence of mass. Especially the gauge bosons were predicted to be massless, which is contradictory to many measurements [20, 21]. The problem of mass was later solved in 1964 by the introduction of the Higgs mechanism. It was independently developed by P. Higgs [22] as well as F. Englert and R. Brout [23]. The Higgs mechanism combines the QED and the weak interaction into the unified theory (electroweak unification). It is based on the existence of an additional field, the Higgs field, and is able to explain particle masses via interactions with it. It also describes a new particle, the Higgs boson, as the massive excitation of the Higgs field. With QCD the last piece of the Standard Model was introduced by M. Gell-Mann and G. Zweig also in 1964 [24–27]. The theory of QCD describes the strong interaction between colour-charged quarks and gluons.

All in all, the Standard Model is a relativistic quantum field theory. All kinematics and interactions can be derived from one Lagrangian which is invariant under local transformation of the gauge group  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . In this theory, Dirac spinor fields are used to describe fermions and the gauge bosons are derived from the gauge fields. The  $SU(3)_C$  symmetry describes the QCD while the  $SU(2)_L \times U(1)_Y$  symmetry is associated with the electroweak sector. The electroweak symmetry breaking reduces this to  $U(1)_Q$  and at the same time creates the masses for the gauge bosons  $W^\pm$  and  $Z^0$ .

The following subsection will give a more phenomenological introduction to the particles of the SM, their properties and interactions. Later subsections will dive deeper into the theoretical formulation of the electroweak sector and the Higgs mechanism.

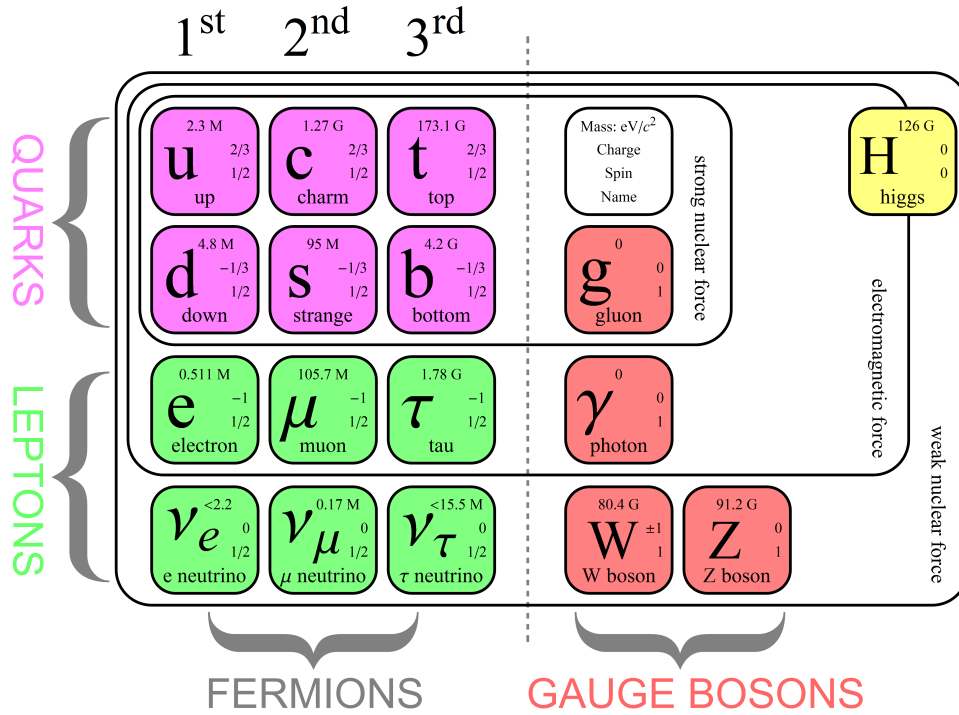


Figure 2.1.: Overview on the particle content of the SM [28]. Displayed are the fermions and bosons together with their mass, electric charge, spin and the interactions they participate in. Updates on the current experimental values for the masses are summarised by the Particle Data Group [29].

### 2.1.1. Phenomenology and Particle Content

All particles included in the Standard Model can be categorised into different groups, which can be seen in Figure 2.1. Half-integer spin particles are called *fermions* while those with an integer spin are *bosons*. The SM distinguishes between twelve fermions, which all have spin  $1/2$ , and five kinds of bosons.

Depending on their charges, fermions can be arranged further in subgroups. All fermions interact via the weak force but only some of them via the electromagnetic or strong force. *Quarks* are the only fermions that carry a colour charge  $C$  (red, blue or green) and therefore can interact using the strong force. Depending on their weak isospin  $I_W^3$  (see Subsection 2.1.3) they are further categorised into *up-type* (up  $u$ , charm  $c$  and top  $t$ ) and *down-type* quarks (down  $d$ , strange  $s$  and bottom  $b$ ). Up-type quarks also have an electrical charge of  $q = +\frac{2}{3}$ , while the down-type ones have  $q = -\frac{1}{3}$ . This means both up- and down-type quarks can be categorised into three *generations*. The only difference between particles of the same kind in different generations is their mass. The concept of *confinement* of the strong force means that no single quarks can be observed. They will always form colour-neutral bound states.

*Leptons* are fermions, which do not carry a colour charge. They can be further separated by their electrical charge and weak isospin. The three neutrinos (electron neutrino  $\nu_e$ , muon neutrino  $\nu_\mu$  and tau neutrino  $\nu_\tau$ ) are the only particles that interact exclusively via the weak force as they do not carry an electrical or colour charge. In the SM neutrinos are considered massless. This is, however, contrary to experimental observations. For example, the Super-Kamiokande detector [30] was able to

measure neutrino oscillations, a phenomenon that is only possible if neutrinos have non-vanishing masses. The three charged particles electron  $e$ , muon  $\mu$  and  $\tau$ -leptons have an electrical charge  $q = -1$ . Both charged and neutral leptons can be classified into one of three generations, just like the quarks.

All fermions have an antimatter partner. These antiparticles have the same properties, like mass and spin, as their matter partners but with opposite charges. This is predicted by the Dirac equation for charged massive particles. As the origin of neutrino masses is not yet understood and they are neutral particles, it is unclear if they have antiparticles. If neutrinos are Majorana particles [31] instead of Dirac fermions can be identified as their own antiparticles.

The spin 1 bosons on the other hand are carriers of the fundamental forces. This includes the massless *photon*, which does not have an electric charge and is associated with the electromagnetic force. The  $W^\pm$  and  $Z^0$  bosons are the carriers of the weak force. The  $W$  bosons are electrically charged with  $q = \pm 1$  while the  $Z$  boson is neutral. They only interact with left-handed particles. *Gluons*  $g$  are the relevant bosons for the strong force. They are massless and electrically neutral but come in eight different colour-charge combinations. The last particle of the SM is the *Higgs boson*  $H$ . It differs from the others as it is not the carrier of a fundamental force and has spin 0. It is a massive excitation of the Higgs field, introduced by the Higgs mechanism (more details on this in Subsection 2.1.4). All of these bosons can interact with fermions as well as other bosons, that carry the corresponding charges. All of these interactions must conserve the total charges, momentum and energy.

### 2.1.2. Discrete Symmetries in the Standard Model

There are three important discrete symmetry operations related to the symmetries of space-time: *parity*, *charge conjugation* and *time reversal*.

When applying the parity operator  $\hat{P}$  to a quantum field, all coordinates are inverted. As the name suggests, the charge conjugation operator  $\hat{C}$  transforms all particles into their associated antiparticle. The time reversal operation  $\hat{T}$  flips the direction of time. The three operations can be defined as:

$$\hat{P}\phi(x, y, z, t) = \phi(-x, -y, -z, t), \quad (2.1)$$

$$\hat{C}\phi(x, y, z, t) = \tilde{\phi}(x, y, z, t), \quad (2.2)$$

$$\hat{T}\phi(x, y, z, t) = \phi(x, y, z, -t). \quad (2.3)$$

Applying the same operator twice should result in exactly the same state as before. This means that all three operations must have eigenvalues of  $\lambda_{\hat{P}, \hat{C}, \hat{T}} = \pm 1$ .

Particles can be assigned an intrinsic parity. A particle has *even* parity if it has  $\lambda_{\hat{P}} = +1$  and *odd* parity if it has  $\lambda_{\hat{P}} = -1$ . Fermions with spin 1/2 are defined to have even parity while antifermions have odd parity. Bosons with spin 0 can either be *scalar* particles with parity  $\lambda_{\hat{P}} = +1$  or *pseudoscalar* with  $\lambda_{\hat{P}} = -1$ . Bosons with spin 1 can also have even or odd parity and are then called *pseudo-vector bosons* or *vector bosons*, respectively. The total parity of a system of multiple particles depends not only on their intrinsic parities. For example, a system of two particles  $a$  and  $b$  has the parity  $\lambda_{ab} = \lambda_a \lambda_b (-1)^L$  with the relative angular momentum  $L$  [32].

In the Standard Model, the electromagnetic and the strong interaction conserve parity. The weak interaction, however, maximally breaks parity conservation by only

coupling to left-handed particles (definition see Subsection 2.1.3) and right-handed antiparticles. Equally, charge conjugation is conserved in the electromagnetic and strong but not the weak interaction. There are, for example, no left-handed antineutrinos in the SM which would be the charge conjugate of left-handed neutrinos [32].

At first, it seems like the combination of parity and charge conjugation would lead to a  $\hat{C}P$  symmetry of the weak interaction. Due to mixing of quark mass eigenstates via the CKM-matrix to eigenstates of the weak interaction,  $\hat{C}P$  symmetry is also broken. This was discovered by J. Cronin and V. Fitch in 1964, who looked at the decay of neutral kaons [33]. A  $CP$  even particle has an intrinsic eigenvalue of  $\lambda_{\hat{C}P} = +1$  and a  $CP$  odd one a negative  $\lambda_{\hat{C}P}$ .

The CPT theorem states that the SM should be invariant under the simultaneous combination of all three transformations. This should be true for all fundamental interactions also including the weak interaction [34]. So far, no violation of this symmetry has been observed [35].

### 2.1.3. Electroweak Unification

The electroweak sector of the Standard Model can be described by the unified gauge group  $SU(2)_L \times U(1)_Y$  as mentioned before. The Noether theorem [36] states, that each continuous symmetry leads to a conserved quantity. In this case the conserved quantities are the weak isospin  $I_W$  and the weak hypercharge  $Y$ . They are related to the electrical charge

$$Q = Y + I_W^3 \quad (2.4)$$

with  $I_W^3$  being the third component of the weak isospin.

The  $SU(2)_L$  group has three generators  $T_a$  ( $a = 1, 2, 3$ ) with two-dimensional representations, which are given by

$$T_a = \frac{1}{2}\sigma_a \quad (2.5)$$

with the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (2.6)$$

The  $U(1)_Y$  gauge group on the other hand has only one generator with a one-dimensional representation: the hypercharge  $Y$ .

The Wu experiment [37] was able to show that the weak interaction violates parity as it only affects left-handed fermions. The left- and right-handed parts  $\Psi_{L/R}$  of a fermion field  $\Psi$  are defined using the fifth Dirac matrix  $\gamma_5^a$

$$\Psi_{L/R} = \frac{1}{2}(1 \mp \gamma_5)\Psi. \quad (2.7)$$

The fact that only left-handed particles interact weakly is incorporated into the theory by describing fermions as having weak isospin  $|I_W| = \frac{1}{2}$  and being grouped into doublets of  $I_W$ . This means that the upper component must have a positive third component  $I_W = +\frac{1}{2}$  and the lower one a negative  $I_W = -\frac{1}{2}$ . This groups up- and down-type

---

<sup>a</sup> defined as  $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} 0 & \mathbb{1}_{2 \times 2} \\ \mathbb{1}_{2 \times 2} & 0 \end{pmatrix}$ , with  $\gamma_0 = \begin{pmatrix} \mathbb{1}_{2 \times 2} & 0 \\ 0 & \mathbb{1}_{2 \times 2} \end{pmatrix}$  and  $\gamma_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}$  in the Dirac representation, using the Pauli matrices  $\sigma_i$

---

quarks into one doublet per generation and the same happens with neutrinos and their respective charged leptons. An example of these isospin doublets would be:

$$L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \text{or} \quad Q = \begin{pmatrix} u \\ d \end{pmatrix}_L. \quad (2.8)$$

At the same time right-handed particles have isospin  $|I_W| = 0$  and are organised in individual singlets. They also behave differently under the  $SU(2)_L$  gauge transformation. The local gauge transformations of the  $SU(2)_L \times U(1)_Y$  group for left- and right-handed fermions are defined as:

$$L \rightarrow L' = e^{i\alpha_a(x)T^a} e^{i\beta(x)Y} L, \quad (2.9)$$

$$\Psi_R \rightarrow \Psi'_R = \mathbb{1} e^{i\beta(x)Y} \Psi_R. \quad (2.10)$$

The Standard Model's Lagrangian density  $\mathcal{L}$  is required to be invariant under these transformations. The term describing the kinematics of fermions is given by:

$$\mathcal{L}_{\text{fermion}} = \bar{L}i\gamma_\mu D^\mu L + \bar{e}_R i\gamma_\mu D^\mu e_R + \bar{Q}i\gamma_\mu D^\mu Q + \bar{u}_R i\gamma_\mu D^\mu u_R + \bar{d}_R i\gamma_\mu D^\mu d_R. \quad (2.11)$$

$\bar{\psi}$  is a way of noting an adjoint spinor which is defined as:  $\bar{\psi} = \psi^\dagger \gamma^0$ . In this case the covariant derivative  $D^\mu$  is used instead of the standard derivation  $\partial^\mu$  to ensure the gauge invariance. This  $D^\mu$  is specific for left- and right-handed particles and introduces new gauge fields  $W_a^\mu$  ( $a = 1, 2, 3$ ) and  $B^\mu$ . It is defined to be:

$$D^\mu L = \left( \partial^\mu + \sum_a ig_W T^a W_a^\mu + ig_Y Y B^\mu \right) L, \quad (2.12)$$

$$D^\mu e_R = (\partial^\mu + ig_Y Y B^\mu) e_R. \quad (2.13)$$

$g_{W/Y}$  are simple coupling constants. The newly introduced gauge fields are related to the physical fields  $W^\pm$ ,  $Z$  and  $\gamma$  of the electroweak sector. The connection between these fields will be discussed in Subsection 2.1.4.

The gauge fields themselves also contribute to the SM Lagrangian density with a kinetic term:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} W_{\mu\nu}^a W^{a,\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \quad (2.14)$$

This formulation uses the field strength tensors

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_W \epsilon^{abc} W_{b,\mu} W_{c,\nu}, \quad (2.15)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (2.16)$$

The kinetic term for the gauge bosons not only determines the dynamics of the gauge fields, but also introduces cubic and quartic gauge boson self-interactions.

The introduction of the covariant derivative in Equation 2.11 alone is not sufficient to reach gauge invariance. In order to achieve this, the gauge fields have to be transformed at the same time as follows:

$$W_a^\mu \rightarrow W_a^{\mu'} = W_a^\mu - \frac{1}{g_W} \partial^\mu \alpha_a(x) - \epsilon^{abc} \alpha_b(x) W_c^\mu, \quad (2.17)$$

$$B^\mu \rightarrow B^{\mu'} = B^\mu - \frac{1}{g_Y} \partial^\mu \beta(x). \quad (2.18)$$

The whole Lagrangian density describing the electroweak sector of the Standard Model is the sum of  $\mathcal{L}_{\text{fermion}}$  in Equation 2.11 and  $\mathcal{L}_{\text{gauge}}$  in Equation 2.14.

## 2.1.4. Higgs Mechanism

One problem with the kind of the Standard Model's electroweak sector as described in the previous section is, that all fermions and bosons are still massless. This is contradictory to the experimental results (an overview of the most current experimental values for particle masses can be found in [29]). The mass term for a real gauge field  $\Psi^\mu$  has the form:

$$\mathcal{L}_{\text{mass}} = \frac{1}{2} m_\Psi^2 \Psi_\mu \Psi^\mu. \quad (2.19)$$

When inspecting the Lagrangian densities in Subsection 2.1.3 no such term is present. In fact, it would not be gauge invariant under the transformations introduced in Equation 2.9 and Equation 2.17.

In 1964, P. Higgs [22] as well as F. Englert and R. Brout [23] proposed a way of solving this problem using spontaneous symmetry breaking: the *Higgs mechanism*.

The first necessary step is to introduce two new complex scalar fields, that together form an isospin doublet with  $Y = \frac{1}{2}$ :

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \quad (2.20)$$

These contribute a kinetic and a potential term to the Lagrangian density. This potential plays a crucial role in the context of spontaneous symmetry breaking. It is chosen in a way to be as simple as possible (but non-trivial) while also conserving the gauge symmetry and renormalizability:

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(|\Phi|^2) = (D^\mu \Phi)^\dagger (D_\mu \Phi) - (\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2). \quad (2.21)$$

The  $\lambda$  parameter of the potential needs to be larger than zero to get a metastable ground state. In the case of  $\mu^2 > 0$  it has one trivial minimum at  $|\Phi|^2 = 0$ . However, in the interesting case of  $\mu^2 < 0$  the potential has an infinite set of degenerate non-vanishing minima with:

$$|\Phi_0|^2 = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}. \quad (2.22)$$

This defines the so-called *vacuum expectation value (VEV)*  $v = \sqrt{-\frac{\mu^2}{\lambda}}$ . The form of such a potential can be seen in Figure 2.2. The Lagrangian is still invariant under gauge transformations. However, the system will occupy one of the possible ground states and will thereby spontaneously break the symmetry. Without loss of generality we can assume the system to occupy the state

$$\langle \Phi \rangle_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.23)$$

This ground state remains invariant under the transformation of the  $U(1)_Q$  gauge group which is the remaining symmetry with the generator  $Q$  according to Equation 2.4. It is possible to expand the Higgs field around this minimum to get

$$\Phi(x) \approx \begin{pmatrix} \tilde{\phi}_1(x) + i\tilde{\phi}_2(x) \\ \frac{v+H(x)}{\sqrt{2}} + i\tilde{\phi}_3 \end{pmatrix} \quad (2.24)$$

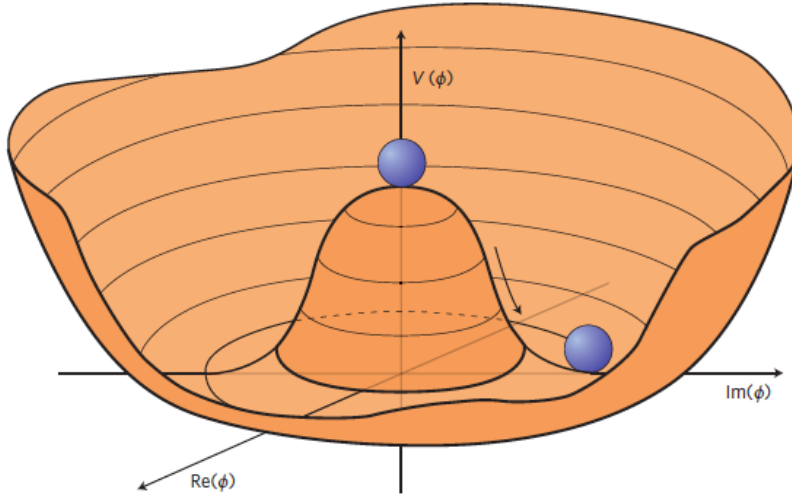


Figure 2.2.: Form of the Higgs potential for the case of  $\mu^2 < 0$  and  $\lambda > 0$  (visualised for only one complex scalar field  $\phi$  instead of two, for clarity) [38].

which can be transformed using gauge transformations to the simpler form

$$\Phi(x) \approx \begin{pmatrix} 0 \\ \frac{v+H(x)}{\sqrt{2}} \end{pmatrix}. \quad (2.25)$$

This version of the Higgs field can be inserted into the Lagrangian Equation 2.21. When using the definition of the covariant derivative in Equation 2.12 this results in:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}} &= |\partial^\mu \Phi|^2 - |\Phi|^2 \left( \frac{1}{4} g_W^2 |W_\mu^1 - iW_\mu^2|^2 + \frac{1}{4} |g_W W_\mu^3 - g_Y B_\mu|^2 \right) - V(|\Phi|^2) \\ &= |\partial^\mu \Phi|^2 - \frac{1}{8} (v + H(x))^2 (g_W^2 |W_\mu^1 - iW_\mu^2|^2 + |g_W W_\mu^3 - g_Y B_\mu|^2) \\ &\quad - (\mu^2 H^2(x) + \lambda v H^3(x) - \frac{1}{4} \lambda H^4(x) + \text{const.}). \end{aligned} \quad (2.26)$$

It is now possible to calculate all individual terms, which will produce interaction terms between the gauge bosons and the Higgs field as well as mass terms for the gauge bosons. However, these mass terms only appear for mixings of the original gauge fields ( $W_1^\mu$ ,  $W_2^\mu$ ,  $W_3^\mu$  and  $B^\mu$ ) which will now be identified as the physical gauge bosons ( $W^{+, \mu}$ ,  $W^{-, \mu}$ ,  $Z^\mu$  and  $A^\mu$ ):

$$W^{\pm, \mu} = \frac{1}{\sqrt{2}} (W_1^\mu \mp iW_2^\mu), \quad (2.27)$$

$$\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B^\mu \\ W_3^\mu \end{pmatrix}. \quad (2.28)$$

In these definitions the *weak mixing angle*  $\theta_W$  is used:

$$\tan \theta_W = \frac{g_W}{g_Y}. \quad (2.29)$$



The masses of the physical bosons can be identified as:

$$m_W = \frac{1}{2}g_W v, \quad (2.30)$$

$$m_Z = \frac{1}{2}\sqrt{g_W^2 + g_Y^2} \cdot v, \quad (2.31)$$

$$m_\gamma = 0. \quad (2.32)$$

As there is no mass term for the field  $A^\mu$ , the photon remains massless. From the Higgs potential a term arises in the Lagrangian density, that describes a mass for the  $H(x)$  field. Therefore, the *Higgs boson* can be identified as a massive excitation of the Higgs field around its ground state. It has a mass of

$$m_H = \sqrt{2} \cdot \mu = \sqrt{2\lambda} \cdot v. \quad (2.33)$$

The introduction of the Higgs field also results in interaction terms between the new Higgs boson and the gauge bosons. This includes interactions between one or two Higgs bosons with two  $Z$  bosons or a  $W^+W^-$  pair. The Higgs potential also results in one  $H^3(x)$  and a  $H^4(x)$  term, which can be interpreted as Higgs self-couplings between three or four Higgs bosons [38].

However, there is still the remaining problem of fermion masses. In contrast to the gauge bosons, fermion masses are not automatically generated by the Higgs mechanism but need to be added by hand. This can be done using so-called *Yukawa interactions* based on *Yukawa couplings*  $\lambda_f$  [39]:

$$\mathcal{L}_{\text{Yukawa}} = -\lambda_e \bar{L}\Phi e_R - \lambda_d \bar{Q}\Phi d_R - \lambda_u \bar{Q}(i\sigma_2\Phi^\dagger)u_R - \dots + \text{h.c.} \quad (2.34)$$

Instead of  $\phi$ , the term  $(i\sigma_2\Phi^\dagger)$  is needed for the calculation of masses of up-type quarks, because the Higgs field was chosen to be  $\phi = \begin{pmatrix} 0 \\ (v + H(x))/\sqrt{2} \end{pmatrix}$  with a non-zero component in the lower position. This leads to terms like this

$$-\frac{\lambda_e}{\sqrt{2}}(v + H(x))(\bar{e}_L e_R + \bar{e}_R e_L), \quad (2.35)$$

which contain *Dirac mass* terms. The masses of the fermions are therefore given by

$$m_f = \frac{\lambda_f v}{\sqrt{2}} \quad (2.36)$$

and dependent on the Yukawa couplings  $\lambda_f$ . As these are free parameters of the SM, it is not possible to make a theoretical prediction of fermion masses.

The masses of neutrinos are still controversial. It might be possible to describe them as Dirac or Majorana masses [31]. The latter would imply, that neutrinos are their own antiparticles. From neutrino oscillation measurements it is known that neutrinos have very small but non-zero masses [30], but the nature of their mass is not yet known.

### 2.1.5. Deficiencies of the Standard Model

The Standard Model is able to explain the majority of observations from particle physics experiments. Still, there are some phenomena which cannot be explained using the SM.

The nature of neutrinos and their masses is still unknown, as stated in the previous section. Another example is that the Standard Model only covers three out of the four fundamental forces with gravity missing from the description. Although gravity can be neglected for energy scales, that are relevant in particle physics, it would be desirable to have one theory describing all the fundamental forces. Especially, because gravity needs to be considered for energies close to the Planck scale at about  $10 \times 10^{19}$  GeV [40].

Another problem is the existence of *dark matter*. From the velocity of stars rotating around the centre of their galaxies [41] as well as from gravitational lenses [42] it is apparent that there has to be a lot more matter interacting via gravitation than expected. This means that there must be a kind of matter that interacts predominantly via gravitation and is unlikely to interact via any other force, if any. The lack of these other kinds of interactions would explain why it was not possible to detect this kind of matter in a direct way before. An idea to incorporate this phenomenon into the SM are massive neutrinos. However, it is unlikely that the full amount of dark matter can be explained using neutrinos alone [43].

In the universe, matter is much more prevalent compared to antimatter. This is called *baryon asymmetry*. There are three conditions that were formulated by A. Sakharov [44] that need to be fulfilled in order to explain the baryon asymmetry. One condition is CP violation, but the CP violating processes in the SM have too small of an influence to explain this [45]. The other two conditions are the existence of baryon number violating interactions and the existence of a stage of thermal non-equilibrium during in the universe.

Next to these concrete deficiencies of the SM, there are also arguments regarding the elegance of the SM. For example, its 19 free parameters are fully unmotivated. Also the Higgs mass is very small (electroweak scale) compared to its loop corrections in the order of the Planck scale which therefore need to be *fine-tuned* in order to lead to such a small total mass (*hierarchy problem*) [46].

This thesis focuses on one specific case where the experimentally observed value of a parameter differs significantly from the value predicted by the SM. This parameter is the *anomalous magnetic moment* of the muon  $a_\mu$ . The muon is a spin 1/2 particle and therefore has an intrinsic magnetic moment given by:

$$|\vec{\mu}| = g_\mu \frac{e}{4m_\mu}. \quad (2.37)$$

Its strength is determined by the g-factor  $g_\mu$  and is also influenced by the mass of the muon  $m_\mu$  and the electric elementary charge  $e$ . It is possible to calculate a theoretical prediction for  $g$  using the Dirac equation. At tree-level this leads to a prediction of  $g = 2$ . However, this is not the true value as loop corrections also need to be taken into account, which change the value slightly. This deviation between the prediction at tree-level and the prediction using higher orders is called anomalous magnetic moment:

$$a_\mu = \frac{g - 2}{2}. \quad (2.38)$$

There are several experiments measuring this value including the Brookhaven National Laboratory (BNL) [47] and the Fermi National Accelerator Laboratory (FNAL) [48, 49]. Combining all measured values into one experimental average [48] results in:

$$a_\mu^{exp} = 116\,592\,059(22) \times 10^{-11} \quad (2.39)$$

while the current Standard Model prediction [50] is:

$$a_\mu^{SM} = 116\,591\,810(43) \times 10^{-11}. \quad (2.40)$$

The comparison of the two leads to a disagreement of  $5.0\sigma$  [48]. This is a strong indicator that the Standard Model is an incomplete theory and needs to be expanded. However, there is another theoretical prediction for  $a_\mu$  based on lattice calculations, which is dramatically deviating from the other SM prediction and is actually more compatible with the experimental values [51]. This relativises the deviation between the SM prediction and the experiments. At the moment, the discussion around the anomalous magnetic moment of the muon is unresolved.

All of these problems suggest that the Standard Model is not a complete theory. As it already describes many phenomena very successful, it is common to simply expand this theory to fill in the blanks instead of creating a whole new description. When doing so one needs to ensure that these extended models are still compatible with all observations, the regular SM can explain. There are many theories trying to describe physics beyond the Standard Model (BSM). An example would be the 2HDM, which introduces a second Higgs doublet and is the basis for Minimal Supersymmetric Standard Models (MSSM) which extend the SM by supersymmetry. The 2HDM is able to explain the deviation of  $a_\mu$  by introducing additional loop diagrams, that involve new Higgs bosons. This is taken as a motivation to search for one of these additional Higgs bosons. The following Section 2.2 will give an introduction to the two-Higgs-doublet model and its parameter constraints in order to explain the measured  $a_\mu$ .

## 2.2. The Two-Higgs-Doublet Model

The two-Higgs-doublet model is a relatively simple extension of the regular SM, as only a second Higgs doublet is added. It is able to explain the deviation of the muon's anomalous magnetic moment, described in the preceding section.

This kind of Higgs sector with two doublets, is needed in supersymmetric theories like the MSSM. In the Standard Model, the adjoint Higgs field  $\Phi^\dagger$ , which is the complex conjugate of the transposed matrix, is necessary to describe the masses of up-type quarks. However, in supersymmetric theories interactions are derived from a superpotential, which cannot depend on the complex conjugate of the fields, but only the fields themselves. Therefore, the second Higgs doublet is used instead of the adjoint field to describe the masses of up-type fermions [39]. The two Higgs fields must have opposite signs of hypercharge in order to explain the couplings to up- or down-type fermions.

The 2HDM will be discussed here from a more general point of view without supersymmetry in mind. Both complex scalar doublets are therefore chosen to have a weak hypercharge of  $Y = +\frac{1}{2}$ . This leads to the two doublets:

$$\Phi_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix}, \quad \Phi_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix}. \quad (2.41)$$

The Higgs potential has to take both doublets into account, which makes it more

complicated. The most general potential is given by [52]:

$$\begin{aligned}
V(\Phi_1, \Phi_2) = & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - \left( m_{12}^2 \Phi_1^\dagger \Phi_2 + \text{h.c.} \right) + \frac{\lambda_1}{2} \left( \Phi_1^\dagger \Phi_1 \right)^2 + \frac{\lambda_2}{2} \left( \Phi_2^\dagger \Phi_2 \right)^2 \\
& + \lambda_3 \left( \Phi_1^\dagger \Phi_1 \right) \left( \Phi_2^\dagger \Phi_2 \right) + \lambda_4 \left( \Phi_1^\dagger \Phi_2 \right) \left( \Phi_2^\dagger \Phi_1 \right) + \frac{1}{2} \left( \lambda_5 (\Phi_1^\dagger \Phi_2)^2 + \text{h.c.} \right) \\
& + \left( [\lambda_6 (\Phi_1^\dagger \Phi_1) + \lambda_7 (\Phi_2^\dagger \Phi_2)] \Phi_1^\dagger \Phi_2 + \text{h.c.} \right). \tag{2.42}
\end{aligned}$$

Depending on the values of the parameters  $\lambda_i$  ( $i = 1, \dots, 7$ ), the potential can contain CP-violating terms. In the minimum of this potential the two Higgs doublets obtain the VEVs  $v_{1,2}$ :

$$\langle \Phi_1 \rangle_0 = \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad \langle \Phi_2 \rangle_0 = \begin{pmatrix} 0 \\ v_2 \end{pmatrix}. \tag{2.43}$$

Their ratio defines the parameter  $\tan \beta = \frac{v_2}{v_1}$ . Following the argumentation in Subsection 2.1.4, we can again expand  $\phi_{1/2}$  around this newly found ground state and get:

$$\Phi_1(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} H_1^+(x) \\ v_1 + \rho_1(x) + iP_1^0 \end{pmatrix}, \quad \Phi_2(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} H_2^+(x) \\ v_2 + \rho_2(x) + iP_2^0 \end{pmatrix}. \tag{2.44}$$

This can be inserted into the Lagrangian and will describe a total of five Higgs bosons after diagonalizing the mass matrix (by rotating it by the angle  $\alpha$ ). The other three degrees of freedom that come from four (two per doublet) complex fields are absorbed by the masses of the gauge bosons. The second doublet with  $v_2 = 0$  generates two CP-even and charged bosons  $H^\pm$  as well as one neutral but CP-odd boson  $A$ . From a mixing of  $\rho_1$  and  $\rho_2$  of both doublets, two neutral Higgs  $h$  and  $H$  are created. By convention,  $h$  denotes the one with the lighter mass and is usually identified as the SM-like Higgs boson  $H_{\text{SM}}$ , which is experimentally observed [52].

It is possible to transform  $\Phi_{1,2}$  into the so-called *Higgs basis* by rotating the doublets by the angle  $\beta$ :

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix}. \tag{2.45}$$

The Higgs basis is defined in a way so that one of the doublets has a vacuum expectation value of  $v'_1 = v_{\text{SM}} = \sqrt{v_1^2 + v_2^2}$ , which is equal to the value from the SM. The other doublet has a VEV of zero [53].

### 2.2.1. The Flavour-Aligned Two-Higgs Doublet Model

The fermion masses are generated by Yukawa couplings, just like in the SM. However, as there are now two doublets the most general version of the 2HDM contains couplings between the Higgs fields and fermions that are non-diagonal in flavour. This means that there will occur *flavour-changing neutral currents (FCNCs)* at tree level. Experiments were able to find very tight constraints on FCNC [54]. Therefore, couplings that lead to FCNC in the 2HDM need to be suppressed in order to get a theoretical description that is compatible with experiments [55].

There are multiple ways to ensure the suppression of FCNC. A common method is to impose that a given fermion field only couples to one of the two scalar Higgs fields. There are four different ways to distribute the fermion couplings between the

Table 2.1.: Overview on the couplings between fermion fields and Higgs doublets for the four different types of 2HDM where only one of the Higgs doublets couples to a given type of fermion in order to suppress FCNC [55].

	Type-I	Type-II	Type-X	Type-Y
Up-type quarks	$\Phi_2$	$\Phi_2$	$\Phi_2$	$\Phi_2$
Down-type quarks	$\Phi_2$	$\Phi_1$	$\Phi_2$	$\Phi_1$
Charged leptons	$\Phi_2$	$\Phi_1$	$\Phi_1$	$\Phi_2$

Table 2.2.: Special coupling parameters of the flavour-aligned 2HDM that results in the four types of  $\mathcal{Z}_2$  symmetric models [55].

	Type-I	Type-II	Type-X	Type-Y
$\zeta_u$	$\cot \beta$	$\cot \beta$	$\cot \beta$	$\cot \beta$
$\zeta_d$	$\cot \beta$	$-\tan \beta$	$\cot \beta$	$-\tan \beta$
$\zeta_l$	$\cot \beta$	$-\tan \beta$	$-\tan \beta$	$\cot \beta$

two Higgs doublets. They are denoted as type-I, type-II, type-X and type-Y. Table 2.1 gives an overview on the couplings between fermions and the Higgs fields for the four types. This kind of model can be implemented by requiring a discrete  $\mathcal{Z}_2$  symmetry, where the two Higgs doublets transform in the following way  $\Phi_1 \rightarrow \Phi_1$ ,  $\Phi_2 \rightarrow -\Phi_2$ .

Another softer method is *aligning* the Yukawa coupling matrices in flavour space. This is done by requiring that  $\Phi_1$  has SM-like Yukawa-couplings and that the coupling of the  $\Phi_2$  to fermions is proportional to the ones of  $\Phi_1$ . These proportionalities are described by three factors:  $\zeta_u$ ,  $\zeta_d$  and  $\zeta_l$  which correspond to interactions with up-type and down-type quarks as well as with charged leptons. The four types described before can be identified as special cases of flavour-aligned models with specific values for the three proportionality factors. The values for the four types are given in Table 2.2.

### 2.2.2. Parameter Constraints

The flavour-aligned 2HDM is able to explain the experimentally measured value of the anomalous magnetic moment of the muon. This is, however, not the case for all possible parameter values. In fact there is only a limited parameter space which is able to explain the deviation and is still compatible with other measurements. This section will explore these constraints on the parameters of the 2HDM, including the masses of the Higgs bosons, the Higgs potential as well as Yukawa couplings.

Theoretical considerations lead to the general result that small masses of the  $A$  boson  $m_A$  in combination with high values for couplings to up-type quarks and leptons ( $\zeta_u$ ,  $\zeta_l$ ) are most important in order to explain the deviation of  $a_\mu$ . Meanwhile, the masses of the other Higgs bosons  $m_h$ ,  $m_H$  and  $m_{H^\pm}$  as well as the coupling  $\zeta_d$  have a smaller influence [5].

These very general constraints are already sufficient to exclude some of the special cases of  $\mathcal{Z}_2$  symmetric models discussed in the previous section. Type-II and X don't allow  $\zeta_u$  and  $\zeta_l$  to have large values at the same time as they are given by  $\cot \beta$  and

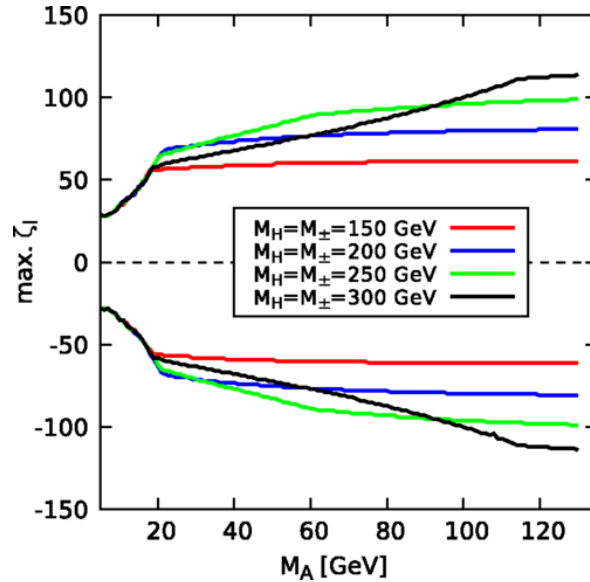


Figure 2.3.: Upper limits on possible values for  $\zeta_l$  as a function of the mass of the  $A$  boson and for multiple possible masses of the other Higgs bosons  $H$  and  $H^\pm$  [5].

$-\tan\beta$ . Therefore, instead of looking at these special cases the general flavour-aligned 2HDM is investigated and its parameters are tuned. The rest of the chapter will discuss a selection of constraints, especially on  $\zeta_u$  and  $\zeta_l$ . A deeper investigation on this can be found in [5].

The coupling to leptons of the second Higgs doublet  $\zeta_l$  has the loosest constraints. For slightly bigger masses of the  $A$  boson  $m_A > 20$  GeV the upper limit on  $\zeta_l$  is dominated by the decay of  $\tau$  leptons via  $\tau \rightarrow \mu\nu_\mu\nu_\tau$  and the decay of  $Z$  bosons. In higher orders, this decay can contain  $A$  or  $H^\pm$  processes, which are enhanced by  $\zeta_l$  and would lead to a disagreement with the experimental observations. For smaller masses  $m_A < 20$  GeV other processes determine the upper limit, namely  $ee \rightarrow \tau\tau(A) \rightarrow \tau\tau(\tau\tau)$ . The production of  $\tau$ -pairs in electron collisions is a process that was measured by the DELPHI collaboration at the Large Electron-Positron collider (LEP) [56]. The allowed maximal values of  $\zeta_l$  depending on the value of  $m_A$  are shown in Figure 2.3 for different mass hypotheses  $m_H$  and  $m_{H^\pm}$ . In general, the maximal possible (absolute) value for  $\zeta_l$  is somewhere between 50 and 100, depending on the other parameters.

The parameter  $\zeta_u$ , which describes the coupling of  $\Phi_2$  to up-type quarks, is constrained mainly by measurements of B physics and processes which involve a neutral Higgs boson as an intermediate particle. B physics describes processes involving hadrons which contain  $b$  quarks. The most important processes here are  $B_s \rightarrow \mu^+\mu^-$  and  $b \rightarrow s\gamma$ . These interactions depend on the couplings to other kinds of fermions  $\zeta_l$  and  $\zeta_d$  as well as the masses of the Higgs bosons  $m_A$  and  $m_{H^\pm}$ . This means that the maximal possible value for  $\zeta_u$  can only be calculated in dependence on these parameters. The continuous lines in Figure 2.4 show this. In general the upper limit dictated by B physics lies at  $\zeta_u < 0.5$  or  $\zeta_u < 0.7$ . Meanwhile,  $\zeta_u$  is constrained by measurements of processes including a neutral Higgs boson as intermediate particle. An example of such a process would be  $pp \rightarrow A/H \rightarrow \tau\tau$ , that can be measured at the LHC. Depending on the masses of the involved particles, especially  $m_A$ ,  $m_Z$  and  $m_H$ , different processes can contribute to measurements.

In the case of the  $A$  boson being heavier than the  $Z$  boson,  $pp \rightarrow A \rightarrow \tau\tau$  is most relevant. As  $A$  is produced via gluon fusion and a top-loop at the LHC, its production cross section is strongly governed by its coupling to up-type quarks  $\zeta_u$ . Therefore, measurements of this process can be used to calculate general upper limits for the case of  $m_A > 100$  GeV and  $\zeta_u < 0.2$ . If the process  $H \rightarrow AA$  is kinematically forbidden (if  $m_A > m_H/2$ ), the cross section of  $pp \rightarrow H \rightarrow \tau\tau$  is determined by  $\zeta_u$  and can therefore, be used to get upper limits. In the case that  $H \rightarrow AA$  is kinematically allowed (if  $m_A < m_H/2$ ), not only  $\zeta_u$  determines the cross section but also the triple Higgs coupling  $c_{HAA}$  as the  $H$  can decay either via a  $\tau$ -pair or an  $A$ -pair instead. This strongly influences the upper limits on  $\zeta_u$  in these parameter regions. The dashed lines in Figure 2.4 show the dependence of upper limits on  $\zeta_u$  calculated based on these LHC processes alone.

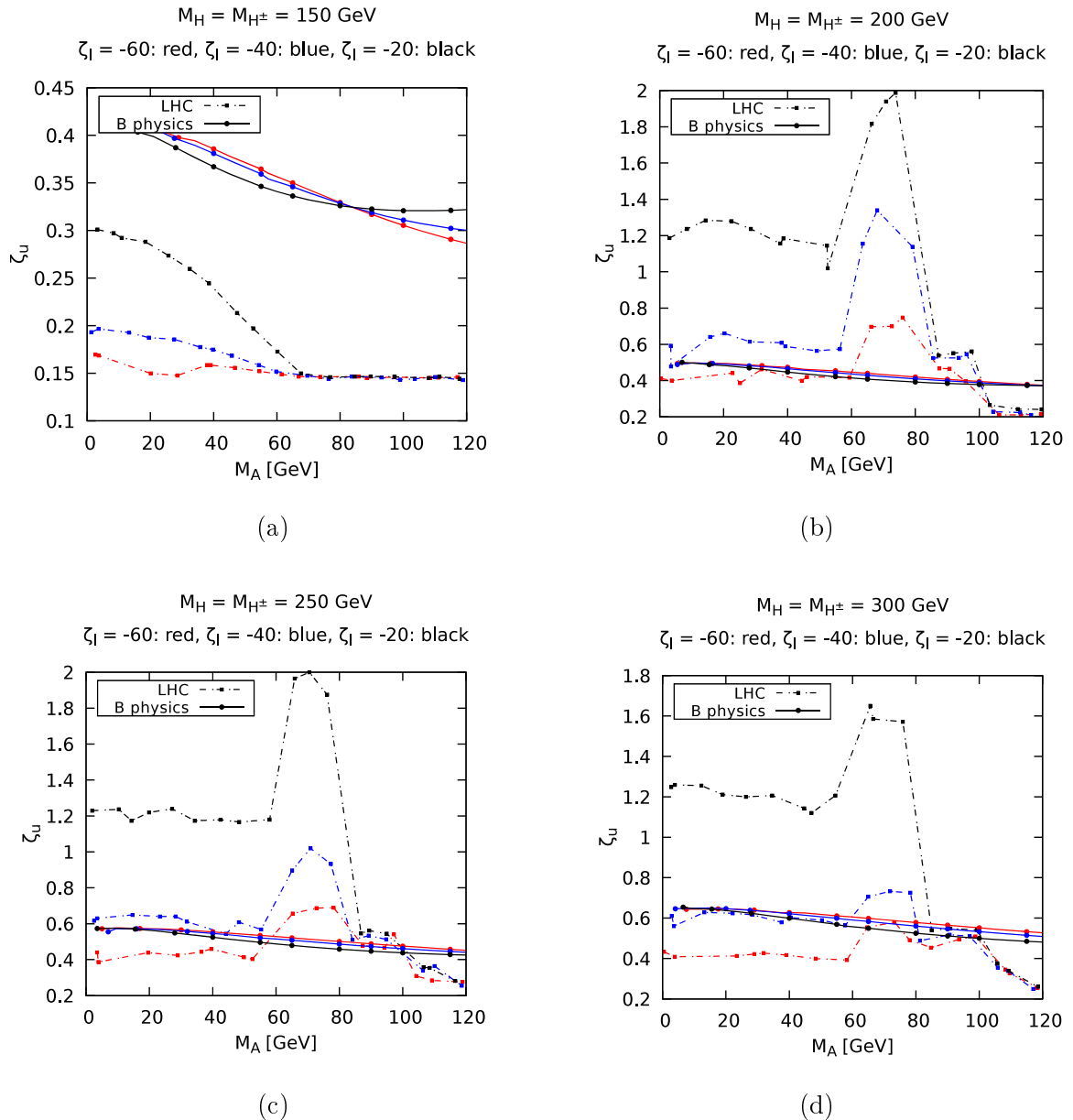


Figure 2.4.: Upper limits on possible values for  $\zeta_u$  as a function of the  $A$  boson mass for different hypotheses for  $m_H$  and  $m_{H^\pm}$  [5].

## 2.3. Statistics

The majority of high energy physics analyses relies on a very large amount of data with many different types of uncertainties. This means that sophisticated statistical methods need to be used in order to make any statement about the possible confirmation or rejection of a theoretical prediction. This section is based on the paper [57] and will give an introduction into these methods.

In the search for new particles, two statistical hypotheses are of importance. The *null hypothesis*  $H_0$  describes all known processes, the background. In the context of this analysis these background processes are defined by the SM predictions and are further explained in Subsection 4.2.3. Meanwhile, hypothesis  $H_1$  describes the case of all known background processes plus a new signal. In this thesis, the light CP-odd  $A$  boson described by the 2HDM is the signal process (see Subsection 4.2.2 for more details).

In the end of the statistical analysis, the goal is to reject the background-only hypothesis and still being compatible with  $H_1$  in order to discover the new signal. This is usually done by considering the so-called *p-value* which will be introduced in the following section in more detail. Generally speaking, it is a measure for the probability to measure the observed result or one that is even less compatible with the prediction under the assumption that the given hypothesis is true. A hypothesis is now discarded, if the p-value is below a defined threshold. A common choice is  $p < 0.05$  to reject a hypothesis, which corresponds to a 95% confidence level (CL) and is used in this case for the signal hypothesis. When the goal is to discover a new particle, the threshold is usually chosen more conservatively at  $p < 5.7 \times 10^{-7}$  to reject the background-only hypothesis.

In most cases, the theoretical predictions are dependent on many parameters. If it is possible to reject the background-only hypothesis, the parameters of the signal plus background hypothesis should be investigated further. For example, one could investigate the minimal or maximal values of these parameters that still lead to a theory that is compatible with the measurements. This process is called *limit setting* and will be discussed in more detail in the following section.

### 2.3.1. Theory of Limit Setting

The process of calculating exclusion limits is based on profile likelihood ratio tests. As mentioned above, the difference between the two hypotheses  $H_0$  and  $H_1$  lies in the presence of a signal, while the description of background processes is the same for both. This can be parametrised by introducing the *signal strength*  $\mu$ , which is used to scale the nominal signal contribution. A value of  $\mu = 0$  is therefore equivalent to the background-only hypothesis  $H_0$  while  $\mu = 1$  is associated with  $H_1$ .

The statistical considerations start at the binned distribution of one variable with a total of  $N$  bins. The number of experimentally observed events in the different bins is given by  $\mathbf{n} = (n_1, \dots, n_N)$ . Based on the hypotheses, the expected number of events per bin  $E_\mu[n_i]$  can be expressed as:

$$E_\mu[n_i] = s_i + b_i. \quad (2.46)$$

Here  $s_i$  describes the signal prediction of  $H_1$  for bin  $i$  and  $b_i$  the corresponding background prediction. Assuming an underlying Poisson distribution for the number of

---



events in each individual bin, the likelihood function can be expressed as:

$$L(\mu) = \prod_{j=1}^N \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)}. \quad (2.47)$$

This formula, however, ignores any uncertainties on the theoretical prediction. In reality  $s_i$  and  $b_i$  depend on a total of  $k$  so-called *nuisance parameters*  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ . They are assumed to follow a normal distribution and represent independent and uncorrelated statistical and systematic uncertainties coming, for example, from the normalization and shape of the  $s_i$  and  $b_i$  distributions.

In order to constrain these nuisance parameters, additional control measurements  $\mathbf{m} = (m_1, \dots, m_M)$  are performed. This could, for example, be done by choosing a sample which is expected to mainly contain background events. The expectation values of these measurements are given by:

$$E[m_j] = u_j(\boldsymbol{\theta}). \quad (2.48)$$

Here  $u_j$  are quantities that depend on the total set of nuisance parameters  $\boldsymbol{\theta}$ . Often these control measurements are constructed in a way to gain information on the background normalization or shape. The additional measurements need to be taken into account when one calculates the likelihood function, which yields:

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^N \frac{(\mu s_j(\boldsymbol{\theta}) + b_j(\boldsymbol{\theta}))^{n_j}}{n_j!} e^{-(\mu s_j(\boldsymbol{\theta}) + b_j(\boldsymbol{\theta}))} \times \prod_{k=1}^M \frac{u_k(\boldsymbol{\theta})^{m_k}}{m_k!} e^{-u_k(\boldsymbol{\theta})}. \quad (2.49)$$

On the basis of this likelihood function one can construct the profile likelihood ratio:

$$\lambda_\mu = \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}. \quad (2.50)$$

This makes use of  $\hat{\mu}$  and  $\hat{\boldsymbol{\theta}}$ , which are the global maximum-likelihood estimator (MLE), and  $\hat{\boldsymbol{\theta}}(\mu)$ , which is the conditional maximum likelihood estimator that maximises  $L$  for a fixed value of  $\mu$ . In the context of searching for a new particle having a negative signal strength  $\mu$  is unphysical. However, it is possible that the estimated  $\hat{\mu}$  becomes negative. To suppress this, a modified profile likelihood ratio  $\tilde{\lambda}_\mu$  is defined as:

$$\tilde{\lambda}_\mu = \begin{cases} \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})}, & \hat{\mu} \geq 0 \\ \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(0, \hat{\boldsymbol{\theta}}(0))}, & \hat{\mu} < 0 \end{cases} \quad (2.51)$$

where  $\hat{\mu}$  is set to 0 by hand if it normally would become negative. If there is a good agreement between the measured data and the prediction  $\tilde{\lambda}_\mu$  converges towards 1. Meanwhile, a bad agreement leads to  $\tilde{\lambda}_\mu$  approaching 0.

Based on this quantity, a new and more convenient test statistic  $\tilde{t}_\mu$  is defined as:

$$\tilde{t}_\mu = -2 \ln \tilde{\lambda}(\mu). \quad (2.52)$$

High values of this test statistic correspond with an incompatibility between measurement and prediction. This test statistic needs further adjusting in the context of

limit setting in a signal search.  $\hat{\mu} > \mu$  indicates an even stronger signal than predicted by  $H_1$  which cannot be interpreted as less compatible than  $\hat{\mu} = \mu$ . This case is, therefore, treated separately in the definition of the new test statistic  $\tilde{q}_\mu$ :

$$\begin{aligned} \tilde{q}_\mu &= \begin{cases} \tilde{t}_\mu, & \hat{\mu} \leq \mu \\ 0, & \hat{\mu} > \mu \end{cases} \\ &= \begin{cases} -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))}, & \hat{\mu} < 0 \\ -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})}, & 0 \leq \hat{\mu} \leq \mu \\ 0, & \hat{\mu} > \mu \end{cases} . \end{aligned} \quad (2.53)$$

The p-value can be calculated for a given test statistic  $\tilde{q}_\mu$  (or the before mentioned  $\tilde{t}_\mu$ ) if the probability density function (PDF)  $f(\tilde{q}_\mu|\mu)$  for the distribution of the test statistic is known for a fixed signal strength  $\mu$ . Its measured value is denoted as  $\tilde{q}_{\mu, \text{obs}}$  and is needed for the definition of the p-value:

$$p_\mu = \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu) d\tilde{q}_\mu . \quad (2.54)$$

The problem, however, is that usually the PDF is unknown and has to be determined first. One way of estimating a PDF is to generate toy Monte Carlo events according to the hypothesis with the chosen signal strength  $\mu$ . Especially when one calculates the p-value for a range of different  $\mu$ , this procedure needs a lot of computing power in order to get sufficient statistics for the toy sample.

An alternative approach is based on Wilk's theorem [58] and Wald's approximation for the logarithm of likelihood ratios [59]. In this case a simpler approximated version of the test statistic and its PDF can be used for the calculation of the p-value if the sample size is sufficiently large. The simplified forms of  $\tilde{q}_\mu$  and  $\tilde{t}_\mu$  are given by:

$$\tilde{t}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2}, & \hat{\mu} < 0 \\ \frac{(\mu - \hat{\mu})^2}{\sigma^2}, & 0 \leq \hat{\mu} \end{cases} , \quad (2.55)$$

$$\tilde{q}_\mu = \begin{cases} \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2}, & \hat{\mu} < 0 \\ \frac{(\mu - \hat{\mu})^2}{\sigma^2}, & 0 \leq \hat{\mu} \leq \mu \\ 0, & \mu < \hat{\mu} \end{cases} . \quad (2.56)$$

In both cases the estimator  $\hat{\mu}$  is assumed to follow a Gaussian distribution with mean  $\mu'$  and standard deviation  $\sigma$ . The PDFs describing these processes are:

$$f(\tilde{t}_\mu|\mu) = \begin{cases} \frac{1}{\sqrt{2\pi\tilde{t}_\mu}} e^{-\tilde{t}_\mu/2}, & \tilde{t}_\mu \leq \mu^2/\sigma^2 \\ \frac{1}{2\sqrt{2\pi\tilde{t}_\mu}} e^{-\tilde{t}_\mu/2} + \frac{1}{\sqrt{2\pi}(2\mu/\sigma)} \exp\left(-\frac{1}{2} \frac{(\tilde{t}_\mu + \mu^2/\sigma^2)^2}{(2\mu/\sigma)^2}\right), & \mu^2/\sigma < \tilde{t}_\mu \end{cases} , \quad (2.57)$$

$$f(\tilde{q}_\mu|\mu) = \frac{1}{2} \delta(\tilde{q}_\mu) + \begin{cases} \frac{1}{2} \frac{1}{\sqrt{2\pi\tilde{q}_\mu}} e^{-\tilde{q}_\mu/2}, & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2 \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)} \exp\left(-\frac{1}{2} \frac{(\tilde{q}_\mu + \mu^2/\sigma^2)^2}{(2\mu/\sigma)^2}\right), & \mu^2/\sigma < \tilde{q}_\mu \end{cases} . \quad (2.58)$$

Here the definition of  $f(\tilde{q}_\mu|\mu)$  uses a  $\delta$ -distribution  $\delta(\tilde{q}_\mu)$ . This method is faster than generating pseudo data as the values for  $\mu'$  and  $\sigma$  can be calculated directly from data. It only requires the data set to be large enough so that the approximations are valid.

When the goal is to discover a new particle, one usually uses the p-value calculated based on  $\tilde{t}_\mu$  and requires  $p < 5.7 \times 10^{-7}$  to reject the background-only hypothesis. At the same time, one often wants to know the biggest value of  $\mu$  that leads to a theory that is still compatible with measurements. To formulate such an *upper limit* one looks for the biggest  $\mu^{\text{up}}$  so that the p-value is still compatible with a defined threshold  $\alpha$ :

$$p_\mu^{\text{up}} < \alpha. \quad (2.59)$$

In the context of upper limits, it is common to use  $\tilde{q}_\mu$  and choose the confidence level at 95%, which is equivalent to  $\alpha = 0.05$ .

### 2.3.2. Expected Limits

The method described in the previous section assumes that there is a measurement that can be used in order to calculate the so-called *observed limit*. In order to be able to make statements on the sensitivity of the measurement before actual data is available, it is common to calculate *expected limits*. They can provide information on the limit and its uncertainties coming from statistical fluctuations on the background alone.

Expected limits are calculated based on simulated background and signal distributions. These can be provided by the *Asimov dataset* which is defined in a way to ensure that one always gets the true values of (nuisance) parameters when calculating the ML estimators. This defines the Asimov data  $n_{i,A}$  and  $m_{i,A}$  (for the regular and control measurements):

$$n_{i,A} = E[n_i] = \mu' s_i + b_i, \quad (2.60)$$

$$m_{i,A} = E[m_i] = u_i(\boldsymbol{\theta}). \quad (2.61)$$

The Asimov dataset also allows the estimation of the standard deviation of the estimator  $\hat{\mu}$  which is assumed to be Gaussian distributed around the mean  $\mu'$ :

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{\tilde{q}_{\mu,A}}. \quad (2.62)$$

With these tools on hand, one can calculate expected upper limits on  $\mu$  in a similar way as described for observed limits above.

# 3. The ATLAS Experiment at the LHC

The analysis presented in this thesis uses experimental data acquired by the **A Toroidal LHC ApparatuS** (ATLAS) detector at the LHC. As it is important to know the general experimental setup to interpret the results, this chapter gives a short overview on the LHC and the ATLAS detector in particular.

## 3.1. The Large Hadron Collider

The LHC is the biggest and most powerful particle accelerator of the world. It is located at the French-Swiss border near Geneva and is operated by the European Organization for Nuclear Research (CERN)<sup>a</sup>. The LHC was constructed between 2000 and 2008 in a 26.7 km long tunnel, that originally was built for the LEP [60].

It is a circular particle accelerator and collider. This means that particle bunches are accelerated by radio frequency cavities in two separate beams in opposite directions. Before they can be injected into the LHC, the beams have to be pre-accelerated by several smaller accelerators at the facility. A schematic view of the whole accelerator complex at CERN can be seen in Figure 3.1.

The LHC was intended to accelerate protons up to energies of 7 TeV, however, this design beam energy has not been reached yet. The Run 2 data that is used in this analysis was taken between 2015 and 2018 using beam energies of 6.5 TeV, which leads to a centre-of-mass energy of  $\sqrt{s} = 13$  TeV. The original goal of Run 3, which is ongoing since 2022, was to reach the design centre-of-mass energy of 14 TeV. But in the end it was only possible to utilise  $\sqrt{s} = 13.6$  TeV. The LHC can also accelerate lead ions up to 2.76 TeV per nucleon.

The particle beams are made of 2808 bunches which consist of around  $2 \times 10^{11}$  particles each. This means that bunch crossings happen every 25 ns and there are always multiple proton collisions happening at the same bunch crossing which is called pile-up. These bunch crossings of the two opposing beams can happen at four points of the LHC. Where the two beams cross, the four major experiments ATLAS [62], CMS [63], LHCb [64] and ALICE [65] are located. ATLAS and CMS are high-luminosity general purpose detectors. LHCb and ALICE are more specialised with LHCb focusing on physics of bottom quarks and ALICE investigating quark gluon plasma using heavy ion collisions. As the LHC is a circular collider, the particle beams need to be forced onto bent trajectories. This is done using 1232 superconducting dipole magnets. They are cooled down to 1.9 K in order to reach magnetic fields up to 8.3 T. Quadrupole and higher order magnets are used to focus the particle beam.

Because data from experiments at the LHC is used to investigate very rare processes, it is important to detect as many of these events as possible. The number of a specific

---

<sup>a</sup>based on: **Conseil Européen pour la Recherche Eucléaire**

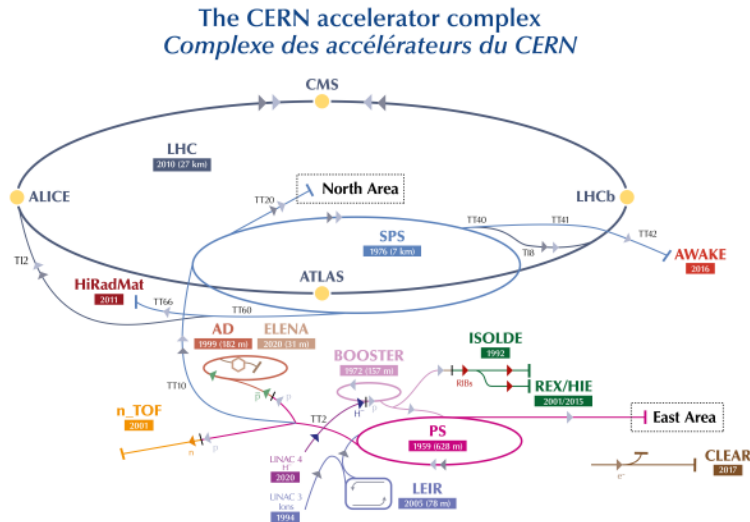


Figure 3.1.: Schematic view of the CERN accelerator complex. The LHC can be seen in dark blue/grey at the top [61].

kind of event  $N_{\text{event}}$  only depends on the machine luminosity  $L$  and the cross section of the processes of interest  $\sigma_{\text{event}}$ :

$$N_{\text{event}} = L \cdot \sigma_{\text{event}}. \quad (3.1)$$

The cross section is an intrinsic property that is a measure for the probability of this specific kind of event to happen. The machine luminosity only depends on beam parameters like the number of bunches per beam and the number of particles per bunch or the beam emittance. During the whole Run 2 a total integrated luminosity of  $140.1 \text{ fb}^{-1}$  has been recorded at the ATLAS detector for  $pp$  collisions [66].

More details on the LHC can be found in its design report [67].

## 3.2. The ATLAS Detector

The ATLAS detector is one of the four big experiments of the LHC. It is a high-luminosity general purpose detector, which uses cutting-edge technologies and is used for high precision measurements of Standard Model properties as well as the search for new beyond SM physics. It has a cylindrical shape and measures 25 m in height, 44 m in length and weighs around 7000 t. The detector can generally be split into two parts: the barrel region and the end-caps. The barrel is build in concentric layers around the beam-axis with the interaction point in the middle. In the two end-caps the detector parts are assembled in large disks on either side of the barrel perpendicular to the beam. This structure ensures a forward-backward symmetry.

The ATLAS detector includes a system of magnets and consists of three major detector components: the Inner Detector, the calorimetry system and the muon spectrometer. A schematic view can be seen in Figure 3.2. The next sections focus on the coordinate system used to describe events recorded at ATLAS and the three different detector systems. A more thorough description of the ATLAS detector can be found in its design report [62].

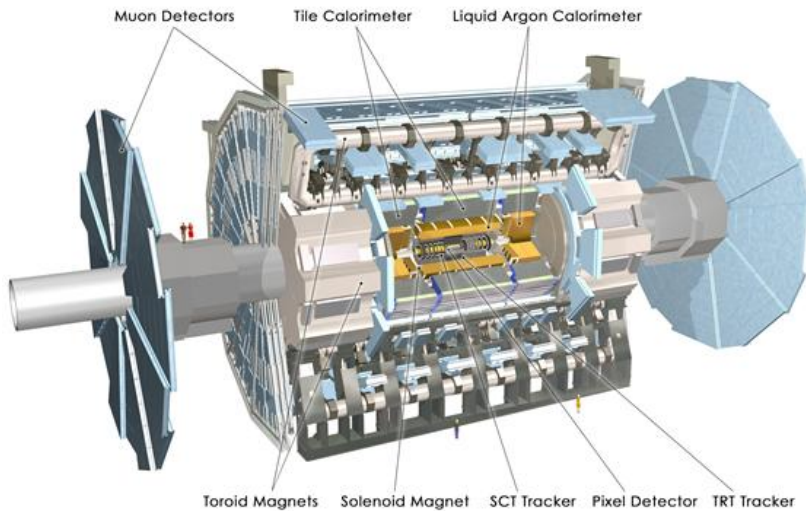


Figure 3.2.: Cut-away view of the ATLAS detector [62].

### 3.2.1. The ATLAS Coordinate System

A right-handed Cartesian coordinate system is used to describe positions inside the ATLAS detector. The origin is chosen to be the interaction point. The  $z$ -axis is parallel to the beam axis, the  $x$ -axis points towards the centre of the LHC and the  $y$ -axis consequently points upwards. Alternatively, spherical coordinates can be used. In this case the radius  $r$ , the polar angle  $\theta$  in respect to the  $z$ -axis and the azimuthal angle  $\phi$  around the  $z$ -axis are the relevant coordinates.

In proton-proton-collisions the total momentum of collision products along the beam axis  $p_z$  is unknown. In reality the partons of the protons are the collision partners. While the total momentum of the protons is fixed, it is not possible to determine the energy fractions that the colliding partons carry. Because of this, momentum conservation can only be applied in the transverse  $x$ - $y$ -plane to the so-called *transverse momentum*  $p_T$ :

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (3.2)$$

As the total transverse momentum is known to be zero, the quantity of *missing transverse energy*  $E_T^{\text{miss}}$  is defined as:

$$E_T^{\text{miss}} = \left| \sum \vec{p}_T \right|. \quad (3.3)$$

The unknown momentum component  $p_z$  leads to a boost of the centre-of-mass frame along the  $z$ -axis. Therefore, observables that are invariant under such Lorentz boosts are desirable. Differences in the so-called *rapidity*  $Y$  fulfil this condition. Rapidity is defined as:

$$Y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right). \quad (3.4)$$

For massless particles, this quantity is equal to the so-called *pseudo-rapidity*  $\eta$ , which is defined only by geometrical quantities as

$$\eta = -\ln \tan \left( \frac{\theta}{2} \right). \quad (3.5)$$

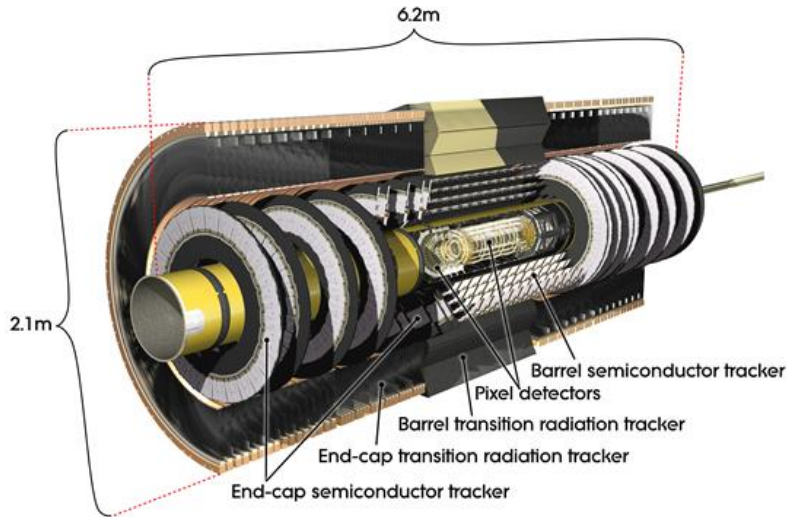


Figure 3.3.: Cut-away view of the ATLAS Inner Detector [62].

and is often used instead of the angle  $\theta$ . The pseudo-rapidity equals zero for particles perpendicular to the  $z$ -axis and  $\pm\infty$  for those moving parallel to the beam axis. Using the pseudo rapidity, the *angular distance*  $\Delta R$  can be defined as:

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \quad (3.6)$$

### 3.2.2. The Inner Detector

The Inner Detector (ID) is the part of ATLAS that is closest to the collision point. The main goal of the ID is to measure the momentum and vertex of charged particles as well as the identification of electrons. It measures 6.2m in length, 2.1m in diameter and covers the region  $|\eta| < 2.5$ . The ID is enclosed by superconducting solenoid magnets that produce a 2T magnetic field. Charged particles experience Lorentz force due to this field and are forced onto curved tracks. This curvature allows measuring the charge and  $p_T$  of those particles.

As the Inner Detector is closest to the collision point, the track density is the highest with approximately 1000 particles emerging from each collision every 25 ns. This means that the ID needs to have a very high resolution, which is ensured using three different kinds of detector module: the pixel detector with the highest granularity is closest to the interaction point followed by Semiconductor Tracker (SCT) and Transition Radiation Tracker (TRT). The structure of the Inner Detector can be seen in Figure 3.3.

The pixel detector is a silicon semiconductor detector with over 90 million individual pixels of the size  $(R - \phi) \times z = 50 \times 400 \mu\text{m}^2$ . A charged particle passing through will create an electron-hole pair and thereby a measurable current. In the barrel region it is organized into four cylindrical layers and each end-cap region has three pixel detector disks perpendicular to the beam [68].

The SCT is also a semiconductor tracker, but in the form of strips that are arranged in four cylindrical layers in the barrel region and a total of 18 disks for the two end-caps.

The TRT consists of gas-filled drift tubes. Charged particles can ionise the gas inside and thereby create an electrical signal. It is the outermost layer of the Inner Detector

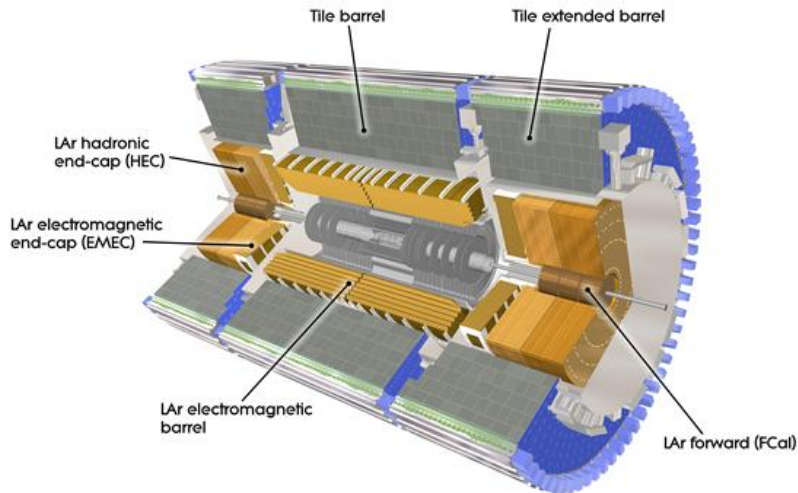


Figure 3.4.: Cut-away view of the ATLAS calorimeter system [62].

and covers the region  $|\eta| < 2$ . As each tube has a diameter of 4 mm, they feature a lower resolution than the other two. They are arranged as 73 straw planes for the barrel and 160 for each end-cap. This means that each charged particle will trigger multiple hits per track, which improves the accuracy of the momentum measurement.

### 3.2.3. The Calorimeter System

The calorimeter is installed around the Inner Detector and the solenoid magnets. Its main task is the precise measurement of particle energies and their identification in a wide region  $|\eta| < 4.9$ . There are two kinds of calorimeters used that are specialised in the energy measurement of different particles. The electromagnetic calorimeter (ECal) located closer to the beam is used for electrons and photons, while the hadronic calorimeter is specialised in strongly interacting particles. By using this combination, the energy of almost all secondary particles can be measured, except for neutrinos and high-energetic muons. The calorimeter system is divided into a barrel and an end-cap region. A schematic view of its structure can be seen in Figure 3.4.

The electromagnetic (EM) calorimeter uses the effect, that electrons emit bremsstrahlung in an absorber material. The produced photons undergo pair production and the emitted electrons and positrons in turn produce bremsstrahlung. This process is called *electromagnetic shower*. If the secondary particles are sufficiently low-energetic, they deposit their energy in the detector, which can then be added up for all particles of one shower and give a measure for the total energy of the mother particle. For this process, it is important that showers are stopped inside this part of the ATLAS detector, which is ensured by choosing 22 times the radiation length as thickness for the EM calorimeter. In the electromagnetic calorimeter, lead plates with stainless steel sheets are used as absorber material to initiate electromagnetic showers and liquid argon (LAr) detectors measure the deposited energy. Its barrel region covers  $|\eta| < 1.48$  while the end-caps extend over  $1.38 < |\eta| < 3.2$ .

The hadronic calorimeter is used to measure the energy of jets. It uses similar principles as EM calorimeter, but initiates hadronic jets instead of electromagnetic



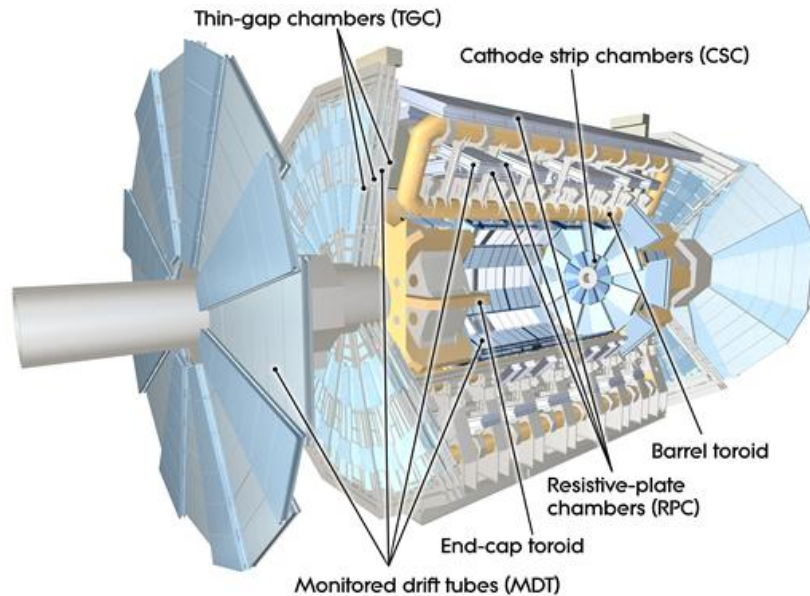


Figure 3.5.: Cut-away view of the ATLAS muon spectrometer [62].

ones. In the barrel region, a tile calorimeter with steel as absorber material and scintillators for the measurement of the energy is used. It covers  $|\eta| < 1.7$ . In the end-cap region at  $1.5 < |\eta| < 3.2$  LAr hadronic end-cap calorimeter (HEC) apply another technology, with copper as absorber and LAr as the active medium.

For the really high  $\eta$  regions  $3.1 < |\eta| < 4.9$ , the LAr forward calorimeter (FCal) uses LAr technology. Its innermost part uses copper optimised for electromagnetic measurements, while the absorber material of the two other modules are made of tungsten.

### 3.2.4. The Muon System

The emission of bremsstrahlung by high energetic muons (above several GeV) is suppressed due to their high mass. For this reason, they are not stopped inside the calorimeter system. To be able to identify and measure the momentum of muons anyway, the Muon Spectrometer (MS) is the outermost layer of the ATLAS detector which can be seen in Figure 3.5. For this momentum measurement the muon system is placed inside the magnetic field of one barrel and two end-cap toroid magnets.

The muon spectrometer consists of four different detector modules. Monitored Drift Tube Chambers (MDT) use cylindrical drift chambers with a gas mixture of carbon dioxide and argon to offer a precise measurement of the track coordinates for a large  $\eta$  region. Cathode Strip Chambers (CSC), that can handle a higher particle rate, replace MDTs in the forward region  $2 < |\eta| < 2.7$ . The other two detectors are used for the trigger system, namely the Resistive Plate Chambers (RPC) in the barrel region and Thin Gap Chambers (TGC) for the end-caps.

### 3.2.5. The Trigger System

At the LHC, collisions happen at a rate of 40 MHz. As every event needs around 1.3 MB to be stored, this would lead to a data rate of around  $50 \text{ TB s}^{-1}$ . This is impossible

and unnecessary, because many of those events don't contain any interesting processes. To reduce this data flow a two level trigger system is deployed to filter the actually relevant events.

The Level-1 Trigger (L1) is hardware-based and uses only a limited amount of available detector information. This is done to limit the processing time so that each decision takes less than 2.5  $\mu\text{s}$ . Events that include particles with high transverse momentum or a large amount of missing transverse energy are more likely to pass the L1 trigger. It is able to reduce the rate of interesting events down to 100 kHz. It also assigns one or more Region-of-Interest (RoI) in  $\eta$  and  $\phi$  for each event, where promising features have been identified. This RoI information is then used by the High Level Trigger (HLT). It is software based and can use the whole detector data available. The final rate of selected events goes down to 1 kHz [69] after applying all triggers.

---

# 4. Base Analysis

This section will give an overview on the cut-based analysis that is the base for the neural network extension described in Chapter 6. At first, basic principles of object reconstruction and identification at the ATLAS detector are introduced. A description of Monte Carlo and data samples analysed will be following, including a discussion of the relevant background processes. After the event selection and triggers are specified, the data-driven estimation of the fake background using the Matrix Method will be introduced. Following the description of a reweighting procedure for one of the backgrounds, the final results of the base analysis will be presented.

## 4.1. Physics Object Identification and Reconstruction

A short overview will be given here on strategies used to recognise, identify and reconstruct physics objects by the signatures they leave inside the ATLAS detector. Relevant objects for this analysis are electrons, muons, ( $b$ -)jets and missing transverse energy  $E_T^{\text{miss}}$ , as well as the mass, reconstructed by the Missing Mass Calculator.

### 4.1.1. Electrons

In the ATLAS detector, electrons pass the Inner Detector (ID) and are then stopped in the electromagnetic calorimeter. They leave a track in the ID and form an EM shower in the calorimeter. The information of ID and ECal can therefore be used to reconstruct and identify them. In the case of electrons this is done in three steps. First, clusters of energy deposition are identified in the ECals, as well as charged tracks in the ID. The last step is to match the information of both detector components.

A sliding-window algorithm is used to find localised energy deposits in the ECal of above  $E_T > 2.5 \text{ GeV}$ . To do this, the  $(\eta \times \phi)$  space is divided into a grid before scanning it. The entries in the different layers (but in the same grid cell) are then summed up, building a so-called tower. In the ID, tracks are reconstructed from multiple individual hits in the tracking layers, using three-dimensional information. In the next step, pattern recognition is used to select those tracks, which are likely to come from a particle which loses its energy while traversing the matter of the detector. In the last step, the  $(\eta \times \phi)$  information of the clusters in the ECal and the tracks in the ID have to match in order to be considered as one electron candidate.

The identification of electrons is done using a multivariate likelihood algorithm. Its input comes from measurements of the two detector systems as well as quantities which combine information from both. There are four available working points: **Very Loose**, **Loose**, **Medium** and **Tight**. They increase in their background rejection in this order [70].

### 4.1.2. Muons

Muons behave similarly to electrons in the ID. They are, however, not stopped in the calorimeter system and reach the MS, where they also leave a signature. Muons are therefore reconstructed independently in the ID and the MS at first. The candidates from both detector parts are then matched in the end.

The track reconstruction in the ID is done similar to the one for electrons. In the MS, tracks are reconstructed by fitting a straight line to hits in different segments in different layers together. The final candidates from this reconstruction are selected by requiring certain fitting criteria. There are several algorithms used to finally combine the information of the ID and MS.

In the case of muons, the different identification working points are based on cuts that are applied to suppress background. These cuts are for example requiring a minimum number of hits in certain detector systems or a high fit quality. The available working points are: **Loose**, **Medium**, **Tight** and **High- $p_T$** . This last working point is used to maximise momentum resolution for tracks with high  $p_T > 100$  GeV [71]. As this thesis focuses on events with relatively low lepton momenta, this working point is not important.

### 4.1.3. Jets

The term *jets* is used for hadronic showers, independent of the specific particle that initiated them. This kind of process is very likely to occur in proton-proton collisions.

The identification of jets is done using the anti- $k_t$  algorithm with the distance parameter  $R = 0.4$  [72] and the particle flow algorithm [73]. Calorimeter measurements, more specifically topologically clustered calorimeter cells, are used as input for the algorithms. Generally speaking the anti- $k_t$  starts with very high energetic deposits and further groups adjacent clusters to form jets [74].

### 4.1.4. B-Tagging

This thesis also utilises a specific kind of jets: *b*-jets. These are jets that originate from the decay of a hadron which contains a bottom-quark.

*b*-jets can be distinguished from others because hadrons containing a *b*-quark are relatively stable with a mean lifetime of roughly  $1.5 \times 10^{-12}$  s [29], depending on the specific *b*-hadron. This means they can travel a significant distance of a few millimetres from the primary collision vertex before decaying. This leads to a secondary vertex which can be identified by interpolating the tracks of charged particles in the ID.

A deep learning neural network *b*-tagging algorithm is used to classify jets as *b*-jets. More specifically, the DL1r is used for this thesis [75]. *b*-jets are selected by choosing a fixed cut on the network discriminant. This cut is based on requiring a *b*-tagging efficiency of 85%.

### 4.1.5. Overlap Removal

The reconstruction of different kinds of physics objects happens independently. This means it is possible that one detector signature gets assigned to multiple different objects. This is not representative of the physics behind it. Therefore, it is necessary

Table 4.1.: Overlap removal applied in this thesis. An object is removed if it is located in a cone of a given  $\Delta R$  around the kept object. The column "step" defines the order in which the different overlap removals are performed.

Step	Removed object	Kept object	Angular separation $\Delta R$
1	electron	muon	0.2
2	jet	muon	0.4
3	jet	electron	0.4

to remove some of the physics objects that have low angular separation  $\Delta R$ . This process is called *overlap removal*.

The decision to remove some objects and keep others is based on the signal purity. For example, muons are favoured over electrons, as additional information from the MS is used during their reconstruction. Also, leptons are generally favoured over jets.

The specific parameters of overlap removal used in this thesis are given in Table 4.1.

#### 4.1.6. Missing Transverse Energy

A very important physics object, that is reconstructed after the others, is the missing transverse energy  $E_T^{\text{miss}}$ .

This quantity is related to neutrinos or potential other BSM particles that were involved in the process. Usually, neutrinos leave the detector without interacting with it and therefore produce no direct signal. Because of momentum conservation in the transverse plane, the total transverse momentum, that was not recorded by the detector, can be reconstructed. It is described by  $E_T^{\text{miss}}$  and is defined as the negative sum of all measured transverse momenta (see Equation 3.3). This quantity is strongly related to the neutrinos, that were not directly measured. This sum includes so-called *hard terms*, which come from the momentum of physics objects that were directly reconstructed in the earlier steps. *Soft terms* are also added. They come from track information which was not assigned to any object [74].

#### 4.1.7. Missing Mass Calculator and Other Mass Variables

The ultimate goal would be to reconstruct the true invariant mass of the mother particle. This is, however, not possible, because of information loss from neutrinos leaving the detector without interacting. Additionally, momentum conservation can only be exploited in the transverse plane due to unknown initial momentum in the  $z$ -direction. There are several relevant definitions of mass variables in high-energy particle physics, that handle this problem in different ways. This section presents three different ways of defining a mass variable.

**Dilepton Mass** The dilepton mass  $m_U$  is defined as the invariant mass of the dilepton system in the transverse plane:

$$m_U = \sqrt{(E_T^e + E_T^\mu)^2 - (\vec{p}_T^e + \vec{p}_T^\mu)^2}. \quad (4.1)$$

As it does not contain any information on neutrinos from  $E_T^{\text{miss}}$ , it is not a good estimate for the invariant mass of the mother particle in this analysis, because we expect neutrinos from the decay of  $\tau$  leptons. It is typically shifted toward masses lower than the true value.

**Total Transverse Mass** The total transverse mass  $m_T^{\text{tot}}$  is a way of incorporating this information into the calculation of a mass variable. It is defined as:

$$m_T^{\text{tot}} = \sqrt{(E_T^e + E_T^\mu + E_T^{\text{miss}})^2 - \left(\vec{p}_T^e + \vec{p}_T^\mu + \vec{E}_T^{\text{miss}}\right)^2}. \quad (4.2)$$

This definition provides an improved reconstruction compared to the dilepton mass. It is, however, still only an estimation of the true invariant mass of the initial particle. Analogously to the total transverse mass, the momentum of the mother particle can be defined as:

$$p_T^{\text{Higgs}} = p_T^e + p_T^\mu + p_T^{\text{miss}}. \quad (4.3)$$

**Missing Mass Calculator** The *Missing Mass Calculator (MMC)* [76] is a tool to calculate an estimate for the invariant mass of a di- $\tau$  system. It is based on a likelihood method and uses additional information on  $\tau$ -lepton decay kinematics for the calculation. The MMC relies on the assumption that only neutrinos from the decay of the two  $\tau$ -leptons are present and build up  $E_T^{\text{miss}}$ .

The additional information regarding the decay of di- $\tau$  systems are based on  $Z \rightarrow \tau\tau$  Monte Carlo (MC) samples. The algorithm can be applied to other sources for the  $\tau$ -lepton pairs, like the  $A$  boson in this thesis, if the polarization of the  $\tau$  leptons can be neglected [76].

As it uses additional information and not only directly measured detector signatures, it is superior to other mass variables like  $m_T^{\text{tot}}$  in resolution and separation power of signal and background.

## 4.2. Monte Carlo and Data Samples

This analysis is based on ATLAS data from the whole Run 2 of the LHC. These events were recorded at proton-proton-collisions in the time between 2015 and 2018 with a centre-of-mass energy  $\sqrt{s} = 13$  TeV. During this time, a total integrated luminosity of  $140.1 \text{ fb}^{-1}$  was recorded [66].

It is crucial for the analysis to have a good estimation for the abundance and behaviour of background and signal events during the measurement. It is the base for formulating the background-only and the signal-plus-background hypothesis, which are needed to statistically examine the existence of a new particle.

In an ideal case one would be able to analytically calculate expected cross sections based directly on the SM predictions. Simulating events is, however, an essential part of any particle physics analysis, as it is for example impossible to directly calculate the needed cross sections based on SM equations in most cases. These simulations are usually done using *Monte Carlo (MC)* simulations which use a probabilistic approach to find approximate solutions to these high-dimensional problems.

The following section will introduce the general principles of MC generators. Subsection 4.2.2 and Subsection 4.2.3 will describe the simulated physics processes used

in this analysis. A complete list of all Monte Carlo samples used in this thesis can be found in Section A.2.

### 4.2.1. Generation of Monte Carlo Samples

The simulated events have to match the actual recorded data as well as possible in regards to their physical behaviour and their output format, in order to allow for a meaningful statistical comparison later on. This leads to the necessity to well simulate the different possible particle physics processes that contribute to the analysis, including their direct decay products as well as their interactions with the detector material and finally the produced signals [77].

For the first part of simulating the actual physics processes, (MC) Event Generators are used. They separate one event in multiple stages from the *hard process*, *parton shower (PS)* and *hadronization*, the *underlying event* to finally the *unstable particle decays* in the end.

The hard process uses *parton density functions (PDFs)* as input, which describe the probability of the partons of the proton to carry a certain fraction of the total momentum. In this step, the *matrix element (ME)* of the physics process of interest is evaluated. PS are used to describe the hadronic showers that are initiated by particles which carry a colour-charge. Hadronization models take the confinement of quarks into account. It is very likely that other processes will happen simultaneously with the hard process in one collision, which is described by the underlying event step. Finally, particle decays of longer-lived but still unstable particles need to be taken into account [78].

A selection of different event generators is used to simulate these steps for the processes that are relevant to this analysis. They are namely PYTHIA [79], POWHEG-BOX [80] and SHERPA [81]. The input PDFs are available through the LHAPDF library [82].

To simulate the detector response, GEANT4 [83] is used. All samples used for the background estimation use the full ATLAS simulation. On the other hand, signal samples are produced using the ATLAS fast simulation ATLFAST-II which utilises a simplified calorimeter simulation [77, 84].

An important additional step is applying so-called *pile-up* corrections. This describes recording the products of multiple proton-collisions at the same time. They can either originate from the same bunch-crossing (*in-time* pile-up) or from different bunch-crossings (*out-of-time* pile-up). The latter is caused by finite readout times of the detector components combined with very small time differences of 25 ns between individual bunch-crossings [77]. There are multiple ways of handling pile-up that are described in reference [85].

### 4.2.2. Signal Process

The goal of this analysis is to search for a light CP-odd Higgs boson as described in Section 2.2. Its couplings to fermions are determined by the three parameters  $\zeta_u$ ,  $\zeta_d$  and  $\zeta_l$ , as well as the fermion masses. As described in Subsection 2.2.2, the parameter values that are able to explain the deviation of the anomalous magnetic moment of the muon  $a_\mu$  lead to a very strong coupling to up-type quarks and leptons. Due to

Table 4.2.: Different decay modes of the  $\tau$ -lepton with their BR [29]. The BR are the same for the decay of  $\tau^+$ -leptons to the corresponding anti-particles of the decay products, shown here.

Decay mode	Branching ratio
Leptonic	
$e^- \bar{\nu}_e \nu_\tau$	17.82 %
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.39 %
Hadronic	
$h^- \nu_\tau$ (1-prong, no neutrals)	11.51 %
$h^- n^0 \nu_\tau$ (1-prong, with neutrals)	37.01 %
$h^- h^- h^+ \nu_\tau$ (3-prong)	14.55 %
others	1.81 %

their heavy masses this means that  $\tau$ -leptons and top-quarks are preferred coupling partners of the  $A$  boson.

This thesis uses data from proton-proton-collisions at the ATLAS detector. Therefore the most prominent production mode in this case is via gluon fusion and a top-loop. On the other hand, the  $A$  boson is predicted to be relatively light, so that the decay via top-quarks is kinematically suppressed. For this reason, it decays almost exclusively over the production of  $\tau$ -pairs, as it is electrically neutral.

$\tau$ -leptons cannot be measured directly in the ATLAS detector as their lifetime of  $2.90 \times 10^{-13}$  s is too short, this means they have to be reconstructed indirectly from their decay products. They can either decay leptonically by emitting a lepton and two neutrinos, or hadronically, producing mainly pions or kaons with associated neutrinos. The branching ratio (BR) of the different decay modes are listed in Table 4.2.

Although in general hadronic decays are dominating with a total BR of around 65 %, it was decided to focus on leptonic decays instead. The reasoning behind this is the fact that the trigger thresholds for hadronic  $\tau$ -leptons are relatively high, with the smallest one starting at  $p_T^\tau = 30$  GeV. A full description of the available triggers on hadronic  $\tau$ -leptons can be found in [86]. This is done to reduce the background from hadronic jets from QCD events, which are dominant in proton-proton collisions, especially in these low  $p_T$  regions.

Leptonic triggers, on the other hand, do not need to suppress this kind of low  $p_T$  background to such an extent and therefore have lower thresholds. As the  $A$  boson is predicted to be light, the decay products are expected to likely have relatively small momenta. This is better matched with the leptonic triggers. The triggers that are utilised in this thesis can be found in Section 4.3.

The final signal process is further restricted to a final state with exactly one electron and one muon (plus neutrinos). This is done to reduce the most prominent background of leptonically decaying  $Z$  bosons and will be discussed in more detail in Subsection 4.2.3. A Feynman diagram of the full signal process can be seen in Figure 4.1.

The signal  $A \rightarrow \tau\tau$  events were generated using PYTHIA v8.2 [79] and the signal



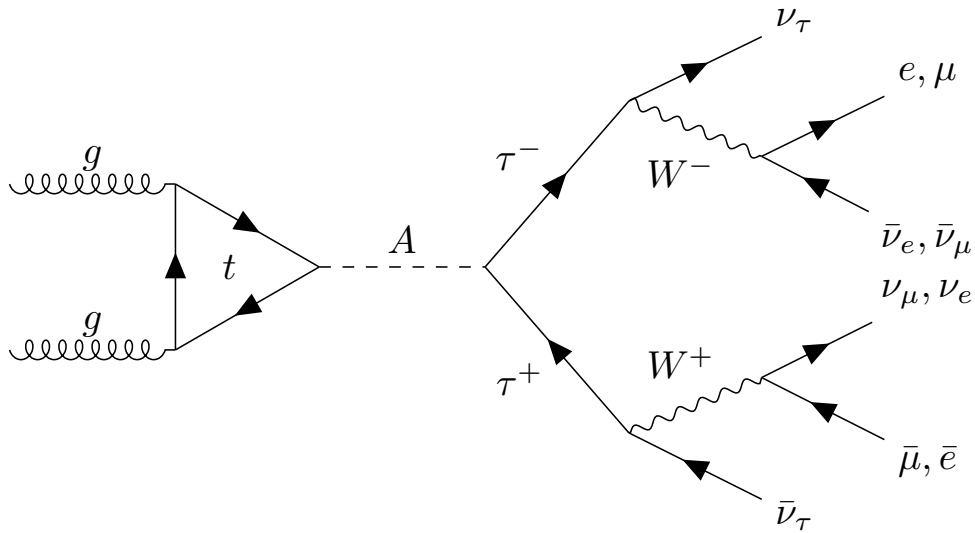


Figure 4.1.: Feynman diagram of the full signal process. The  $A$  boson is produced via gluon fusion and a top quark loop and decays into a  $\tau$ -lepton pair, which then decay leptonically [6].

cross sections were calculated with `ggHiggs` v 4.0 [87] up to  $N^3LO^a$ .

### 4.2.3. Background Processes

This section gives an overview on the most prominent background processes that lead to detector signatures similar to the  $A \rightarrow \tau\tau$  process. It focuses specifically on the processes which can be modelled using Monte Carlo simulation. Section 4.4 will complete the description of relevant backgrounds with the introduction of the so-called *fake background*, which is based on a data-driven method, instead of simulations.

**$Z \rightarrow \tau\tau$  Background** The background process, that is most similar to the signal, is the decay of a  $Z$  boson into a  $\tau$ -pair. A Feynman diagram of this decay can be seen in Figure 4.2. In fact, the final state of these two is exactly the same. The branching ratio of this decay of the  $Z$  boson is 3.4% [29], but as the production of  $Z$  bosons is relatively high, it is a very important background process in this analysis. Because it shares the same final state with the signal process, it is impossible to fully suppress it.

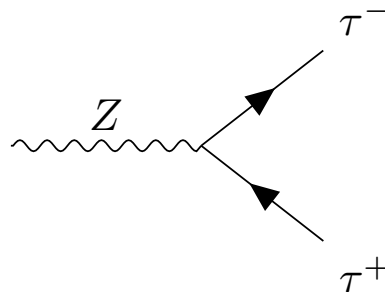


Figure 4.2.: Feynman diagram of the decay of a  $Z$  boson into a  $\tau$ -lepton pair [6].

<sup>a</sup> next-to-next-to-next-to-leading order

Nevertheless, there are two main strategies to suppress it. The first one is to look at reconstructed mass variables that also consider  $E_T^{\text{miss}}$ . These should correlate to the invariant mass of the original particle. This, however, has the problem that due to the unknown momenta of neutrinos there is only a limited mass resolution available. In addition to that, the masses of the  $Z$  and the  $A$  boson might be relatively close to each other, especially for the higher mass hypothesis, which makes them hard to distinguish.

Luckily the two particles are not similar in all their properties. The  $A$  boson has spin 1 and a negative CP eigenvalue, while the  $Z$  boson has spin 0 and is not a CP-eigenstate. This has an influence on the angle between the decay planes of two  $\tau$  leptons. In the case of a CP-odd particle the visible decay products (non-neutrinos) are more often parallel, while they are more likely to be antiparallel if the mother particle is CP-even. It is not possible to look at the full decay planes as the neutrinos are undetected, however, the angular distance  $\Delta R$  between the leptons still carries some information on this.

The  $Z \rightarrow \tau\tau$  events were generated using SHERPA v2.2.1 [81].

**$Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  Background** A very similar process is the decay of a  $Z$  boson into a pair of electrons or muons, which can be seen in Figure 4.3. The branching ratios for these two processes are 3.4% each [29]. The only difference to the signal final state is the absence of neutrinos. As mentioned before, looking at  $E_T^{\text{miss}}$  alone is not a reliable way to distinguish them.

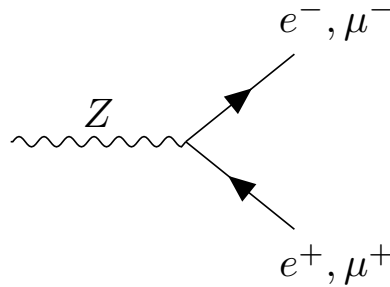


Figure 4.3.: Feynman diagram of the decay of a  $Z$  boson into an electron or muon pair [6].

A very effective way to suppress this background is to require exactly one electron and one muon in the final state. This reduces the impact of this background to only few events which include particles, that were misidentified during reconstruction. A disadvantage of this method is, however, that it also impacts the signal strength significantly, which is roughly cut in half. As this procedure can suppress almost all  $Z \rightarrow ee, \mu\mu$  decays, only  $Z \rightarrow e\mu$  decays are still considered, which make up 2.92%<sup>b</sup> of leptonic  $Z$  boson decays.

The  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  events were generated using SHERPA v2.2.1 [81].

**Top Quark Processes** This category includes all processes that involve top quarks like  $t\bar{t}$  or single top processes. The most prominent decay mode for top quarks is

<sup>b</sup>  $\frac{\mathcal{B}(Z \rightarrow e\mu)}{\mathcal{B}(Z \rightarrow e\mu, ee, \mu\mu)} = \frac{2\mathcal{B}(Z \rightarrow \tau\tau) \cdot \mathcal{B}(\tau \rightarrow e\nu_e\nu_\tau) \cdot \mathcal{B}(\tau \rightarrow \mu\nu_\mu\nu_\tau)}{\mathcal{B}(Z \rightarrow ee) + \mathcal{B}(Z \rightarrow \mu\mu) + \mathcal{B}(Z \rightarrow \tau\tau) \cdot (\mathcal{B}(\tau \rightarrow e\nu_e\nu_\tau) + \mathcal{B}(\tau \rightarrow \mu\nu_\mu\nu_\tau))} = 2.92\%$ , values taken from [29].

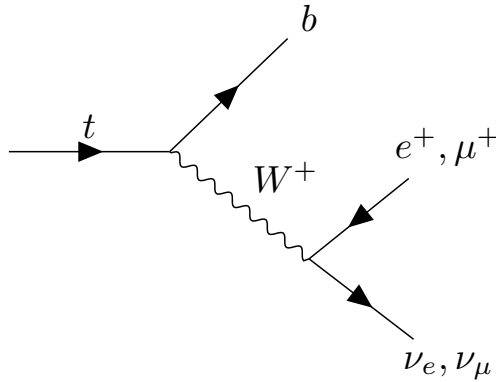


Figure 4.4.: Feynman diagram of the decay of a top quark to a bottom quark and a  $W$  boson, which then decays leptonically [6].

into a bottom quark and a  $W^+$  boson, with a branching ratio of 99.7% [29]. The vector boson then decays further, either leptonically by emitting a lepton and its corresponding neutrino or hadronically. Figure 4.4 shows the decay of a top quark with the  $W$  boson decaying leptonically.

In the case of a  $t\bar{t}$  process, this can lead to a final state similar to the signal process, if both  $W$  bosons decay leptonically, one into an electron and the other into a muon (plus respective neutrinos). This is similar to the signal final state, but not exactly the same, as it contains additional  $b$  quarks and fewer neutrinos.

Single top events can also look like the signal, when the  $W$  boson decays leptonically and there is another lepton, that comes either from misidentification or an underlying process.

A very good way to suppress top events is to use a  $b$ -tagging algorithm and only look at events which do not contain any  $b$ -quarks. This strategy is, however, not perfect, as the  $b$ -tagging methods only have limited efficiency (85% in this analysis). Additionally, while the decay of a top quark into a bottom quark is dominating, there are also decay modes possible which lead to events without  $b$ -quarks.

The ME for single top processes in the s- [88], t- [88] and  $Wt$ -channel [89] as well as  $t\bar{t}$ -events [90] were all simulated using POWHEG-BOX v2. The generation of PS was done using PYTHIA v8.2 [79].

**Standard Model Higgs Processes** The Standard Model Higgs  $H_{125}$  can also directly decay into a  $\tau$ -lepton pair, just like the  $A$  and  $Z$  boson. This process has a small BR of 6.27% and a relatively low cross section of 56 pb at  $\sqrt{s} = 13$  TeV proton-proton collisions [29]. Another decay mode of the  $H_{125}$  that can lead to a similar final state is the process  $H_{125} \rightarrow W^+W^- \rightarrow e + \mu + 2\nu$ , which also has a low BR.

As the  $H_{125}$  is a CP-even particle, measures to suppress the influence of  $Z \rightarrow \tau\tau$  will also reduce the impact of this process. As its mass is above the range of examined signal hypotheses, it is easier to separate it from the signal using mass variables than in the case of the  $Z$  boson. Due to these methods and the lower cross sections, this process has less of an impact on this analysis than the  $Z$  decays. However, as it is so similar to the signal process it is not possible to fully suppress it.

The ME of SM Higgs processes were generated with POWHEG-BOX v2 [91, 92] while PYTHIA v8.2 [79] was used for the PS.

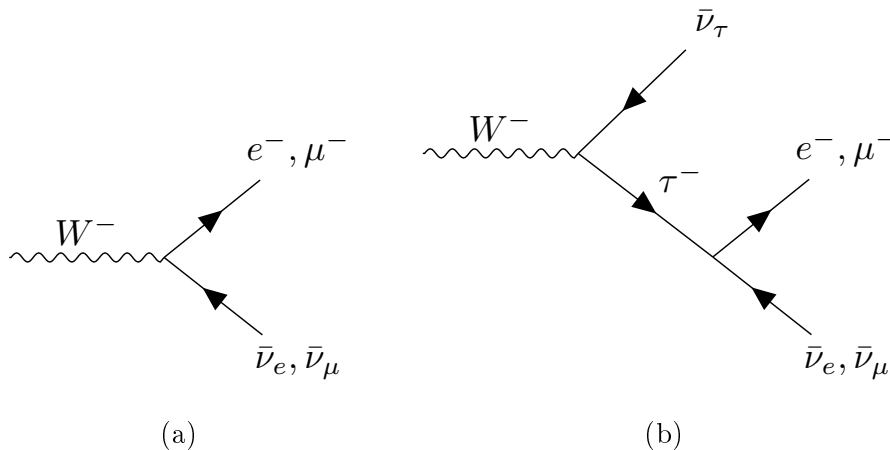


Figure 4.5.: Feynman diagrams of the leptonic decay modes of a  $W$  boson that lead to one electron or muon in the final state. In (a) the  $W$  boson decays directly into an electron or muon, while in (b) it decays first into a  $\tau$  lepton which decays further to the investigated final state [6].

**$W \rightarrow l\nu$  and  $W \rightarrow \tau\nu$  Processes** In proton-collisions,  $W$  bosons are often present from quark-antiquark annihilation. These  $W$  bosons can then decay leptonically via one lepton and its corresponding neutrino. This lepton can already be an electron or muon (see Figure 4.5a) or it is a  $\tau$ -lepton, which then decays further (Figure 4.5b). If this coincides with the production of an electron or muon from an underlying process or the misidentification of a physics object as a lepton, this can look like the signal process.

This process is expected to play a minor role in this analysis, because of good background rejection for the identification of electrons and muons. This is especially the case when the Matrix Method is used for estimating the fake background (see Section 4.4) as this requires a truth cut on all MC events. This cut is used to select simulated events that truly come from the described process. This will suppress all  $W \rightarrow l\nu$  events, which contain leptons from non-prompt sources, like misidentification. SHERPA v2.2.1 [81] was used for the full simulation of  $W \rightarrow l\nu$  and  $W \rightarrow \tau\nu$  processes.

**Diboson Background** The category *diboson* background includes all processes which include two vector bosons, meaning  $WW$ ,  $WZ$  and  $ZZ$ . This definition allows for a wide range of possible final states. Some of these are very different to the signal and therefore, do not need to be suppressed individually. An example for this would be the decay of two  $Z$  bosons to four leptons  $ZZ \rightarrow 4l$ .

Meanwhile, there is also a variety of processes which need to be considered in this thesis. If two  $W$  bosons decay leptonically, one of them to an electron and the other to a muon (plus neutrinos), it leads to a final state similar to the signal. The best way to distinguish this process from signal, is the use of mass variables as they are usually higher for the diboson state.

Some diboson processes, that also need to be taken into account, contain the misidentification of other physics objects as electron or muon, or leptons not being recognised as such.

While semi-leptonic diboson processes were simulated with SHERPA v2.2.1, SHERPA v2.2.2 [81] was used for fully leptonic ones.

## 4.3. Event Selection

This section will provide information on the strategy of selecting events used in this thesis. It will start with the description of the used triggers and then continue to illustrate further requirements that all events considered in any form need to pass, the *baseline selection*. The section will close with a summary of different selections used to define the signal region, as well as a variety of validation regions.

This thesis uses `DAOD_HIGG3D1` derivations as input and the `AtautauAnalysis` framework [93] for the production of NTuples. The base for this analysis is the "Dresden framework" `ELCore` [94] with the underlying `AnalysisBase 21.2.202`. In this step the trigger and baseline selections are performed.

### 4.3.1. Trigger

This thesis uses combined electron-muon trigger for the initial selection of events. For each data-taking period there are three triggers available, that differ in their  $p_T$  thresholds and are called here: `e7mu24`, `e17mu14` and `e26mu8`. A list of their full names can be found in Section A.3.

All of these require the two triggering leptons to be the ones with the highest  $p_T$  values. The two numbers in the trigger names represent the trigger thresholds for the leptons in GeV. Because there are no trigger efficiencies available fully down to the thresholds, it was decided to apply a cut that only allows events with  $p_T$  of the leptons at least 1 GeV above the corresponding thresholds. For the two triggers, which allow for very low lepton  $p_T$ , this cut is increased to 10 GeV (equivalent to 2 GeV and 3 GeV above the thresholds).

In the  $p_T$  regions, where the different triggers overlap, `e17mu14` is preferred over the other two. This way the three triggers are strictly separated. A schematic view of this trigger setup, including the  $p_T$  regions where the triggers are available and used in reality, can be seen in Figure 4.6.

### 4.3.2. Baseline Selection

All events that are considered in this thesis (also those used in the calculation of efficiencies for the Matrix Method, see Section 4.4) need to pass certain requirements. Generally only events with exactly one electron and one muon are considered, which is motivated by enhancing the signal-to-background ratio (see Subsection 4.2.3). The other requirements are used to guarantee a good reconstruction quality, while others ensure a kinematic parameter space with good detector resolution.

Electrons are required to pass the `LooseBLayer` identification working point and `Loose` isolation. They need to have a transverse momentum above 7 GeV and  $|\eta_e| < 2.47$ . Additionally, events that contain electrons in the crack region  $1.37 < |\eta_e| < 1.52$  are vetoed, as well as ones with the `DFCommonCrackVetoCleaning` flag.

Muons on the other hand need to pass the `Loose` identification and the `Loose_VarRad` isolation working point. They also must have  $p_T^\mu > 7$  GeV. Their allowed  $\eta$ -range is a bit wider than for electrons, with  $|\eta_\mu| < 2.7$ . If muons carry the `isBadMuon` flag, they are also sorted out.

If the event contains jets, they need to have  $p_T^{\text{jet}} > 20$  GeV and  $|\eta_{\text{jet}}| < 4.5$ .

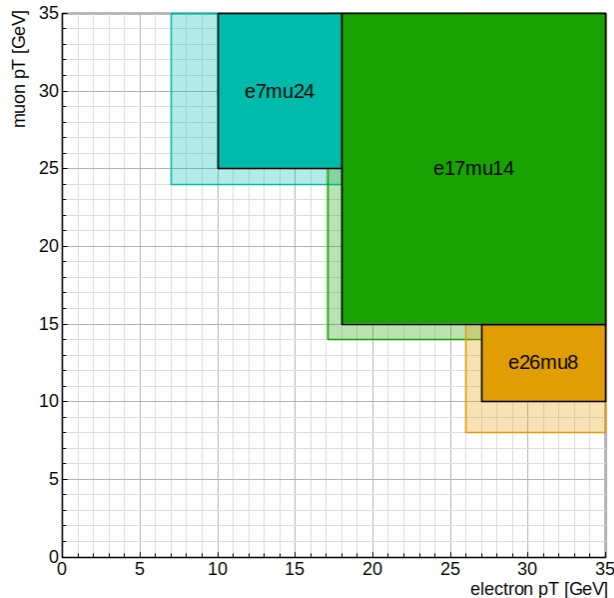


Figure 4.6.: Schematic view of the trigger setup used in this thesis, in dependence of the  $p_T$  of the electron and muon. The transparent regions mark parameter space that is available from the triggers in theory but is not used in this thesis. The opaque regions mark the actually used parameter space [7].

### 4.3.3. Common Selection

The purpose of some general cuts is to further select the events with the highest lepton reconstruction quality. Namely, only electrons that pass the **Medium** identification and **Tight** isolation working point are selected. Simultaneously muons are required to pass **Medium** identification and **Tight\_VarRad** isolation.

The analysis described in this section of the thesis is cut based. This means cuts on kinematic variables are used to improve the signal-to-background ratio in the final selection, the so-called *Signal Region (SR)*. The specific cuts are physically motivated and use the differences in behaviour between the signal and the background processes, as described in Subsection 4.2.3. As there is a relatively large range of mass hypotheses of 20 GeV to 110 GeV investigated, a change of behaviour in some variables like  $m_{\text{MMC}}$  can be observed between the different signal hypotheses. For this reason it was decided to create two (non-orthogonal) SRs, that are specialised on different mass regions: the low-mass signal region (LMSR), for masses up to 80 GeV and the high-mass signal region (HMSR), for mass of 80 GeV and above. An additional mvaVR is defined that is very similar to the two standard SR but further loosens their requirements. Only the very basic cuts and the ones used to suppress the  $Z \rightarrow \tau\tau$  and top background are kept. This region is not used in the base analysis, but is important for the neural network extension that is described in Chapter 6. All regions contain a cut  $m_{\text{MMC}} > 0$ , to simply suppress events with  $m_{\text{MMC}} = -1$ , where the Missing Mass Calculator was unable to converge.

On the other hand it is also interesting to look at parameter regions, where specific background processes are enhanced and the signal is suppressed. These regions are called *validation region (VR)*. A VR can be defined by inverting the cut that is used to suppress one background process. VRs can be used to examine the modelling of the

Table 4.3.: Different cuts that are applied in the signal and validation regions. See Subsection 4.1.7 for the definition of  $E_T^{\text{miss}}$  and  $m_T^{\text{tot}}$ .

	LMSR	HMSR	mvaVR	ZVR	SS ZVR	TVR
$E_T^{\text{miss}}$	$> 50 \text{ GeV}$	$> 30 \text{ GeV}$	-	-	-	$> 30 \text{ GeV}$
$m_T^{\text{tot}}$	$< 45 \text{ GeV}$	$< 65 \text{ GeV}$	$< 65 \text{ GeV}$	$< 65 \text{ GeV}$	$< 65 \text{ GeV}$	$< 65 \text{ GeV}$
$\Delta R_{ll}$	$< 0.7$	$< 1.0$	$< 1.4$	$> 1.4$	$> 1.4$	$< 1.0$
$m_{\text{MMC}}$	$> 0 \text{ GeV}$	$> 35 \text{ GeV} \ \& \ < 130 \text{ GeV}$	$> 0 \text{ GeV}$	$> 0 \text{ GeV}$	$> 0 \text{ GeV}$	$> 0 \text{ GeV}$
$q_e \cdot q_\mu$	-1	-1	-1	-1	1	-1
$n_{b\text{-jets}}$	0	0	0	0	0	$> 0$

different background processes. This thesis utilises the so-called  $Z \rightarrow \tau\tau$  Validation Region (ZVR), with an enhanced  $Z \rightarrow \tau\tau$ -background by flipping the  $\Delta R_{ll}$  cut. A same-sign version of it, the SS ZVR, enhances the impact from misidentified fake events. There is also the so-called Top Validation Region (TVR). This region is created by flipping the  $n_{b\text{-jets}}$ -cut which strengthens the background from top-related events. The TVR is only of minor importance for this thesis and is mostly used to compare results from different VR. The specific cuts that are used to define the different regions are given in Table 4.3.

The ATLAS analysis this thesis is based on is still blinded. This means, it is not allowed to look at data in parameter regions, where a lot of signal is expected. This has several implications for this thesis. Firstly, all variable distributions in the SR, including HMSR, LMSR and mvaVR, are only allowed to show the background expectation using MC samples and estimated fake events from the Matrix Method (MM) (see following section), but **no** distribution of measured data events. This also means that no data-to-background ratio is given in this case. The blinding procedure also has implications on distributions in the VRs. Bins with an expected discovery significance  $Z = \frac{s}{\sqrt{b}}$  and  $Z = \frac{s}{\sigma_{\text{MCstat}}}$  greater than 3.0 are blinded. This can lead to a distribution where the background expectation is displayed in all bins, while data points are only shown in some of them.

## 4.4. Data-Driven Estimation of Fake Background Using the Matrix Method

In Figure 4.7 the exemplary distribution of  $m_{\text{MMC}}$  in the ZVR can be seen. In this diagram, only MC samples, as described in Subsection 4.2.3, are used to form the background. It is apparent that there is background missing in order to explain the data distribution.

This disagreement can be explained by events which contain physics objects that were falsely reconstructed as prompt leptons. This kind of background is called *fake background*. There is a big variety of physics objects that can be misidentified as prompt lepton. Possible sources are hadronic jets, the decay of heavy quarks and photon conversion or charge flip events. A fraction of fake events is already included

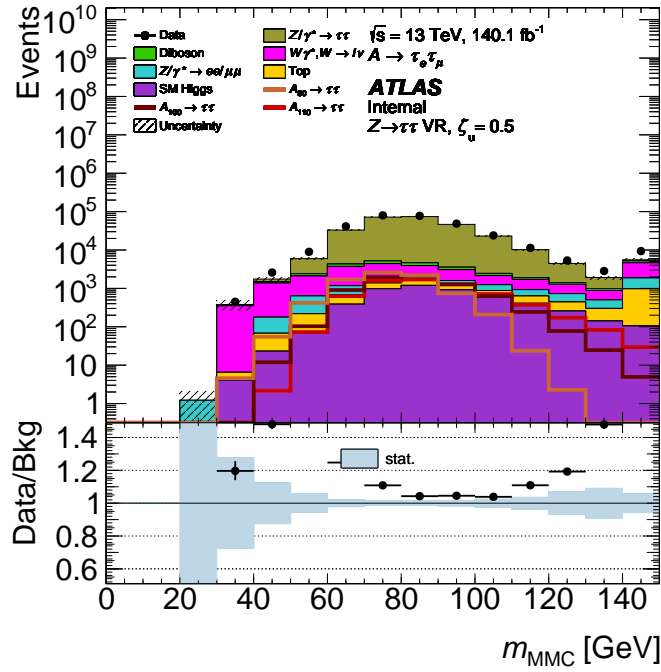


Figure 4.7.: Distribution of  $m_{\text{MMC}}$  using only Monte Carlo generated background samples, as presented in Subsection 4.2.3, alone.

in the MC samples introduced in the section before. These are, however, not sufficient to describe the full fake background, especially not events with both leptons from fake processes. The goal of strict identification and isolation working points is to reduce these misidentifications. Although this filtering of events is very good, it is not perfect. As the sources for fake events can have quite big cross sections in proton-collisions, the fake background is not negligible.

The fact that fake events originate from many different sources with possibly high cross sections, makes it hard to simulate this background with Monte Carlo methods. All sub-processes would need to be well-modelled and a lot of data would need to be simulated to account for the high cross sections. In addition to that, the probability for misidentification would have to be determined with a high accuracy.

Due to these problems with an MC generated fake background, it is common practice to use data-driven approaches instead. There is a variety of slightly different algorithms available. This thesis uses the so-called *Matrix Method (MM)* [95]. It is especially well suited for this thesis, as it naturally incorporates the case of multiple (different) particles in the final state that might come from misidentification.

This section will introduce the general idea of the MM and the specific way it was implemented in this thesis.

#### 4.4.1. On the Matrix Method in General

The general idea of the matrix method is to use *real* and *fake efficiencies* to find a relation between the realness of an event and its reconstructed tightness. This correlation can then be used to get an estimate for the number of fake events based on the information on tight and non-tight events, which can be measured easily.



Very important terms in the context of MM are the definitions of a *tight* and a *loose* event. In this analysis, loose leptons need to meet the baseline identification and isolation criteria as described in Subsection 4.3.2 (**Loose ID** and **Loose isolation**). Tight leptons are required to pass **Medium ID** and **Tight isolation** (see Subsection 4.3.3).

The real efficiency  $r_{e/\mu}$  describes the probability that a prompt lepton, that passes the baseline selection, is also tight. The fake efficiency  $f_{e/\mu}$  is defined in the same way for fake leptons. The two kinds of efficiencies must be calculated separately for electrons and muons. In an ideal case of almost no misidentification, the real efficiencies would be very close to one, while the fake efficiencies would converge towards zero. This is unfortunately not the case in reality.

In this thesis the following convention will be used to describe the realness and tightness of events: upper indices always describe if we look at tight (T), loose (L) or non-tight ( $\bar{T}$ ) events. In this case loose is given by the sum of tight and non-tight events ( $L = T + \bar{T}$ ). The lower index denotes if the leptons are real (R) or fake (F). An X means, that the realness of the lepton is irrelevant in this case ( $X = R + L$ ). When discussing the two-lepton case, the first upper and lower index is used to describe the electron and the second one describes the muon.

Assuming the real and fake efficiencies are known, the matrix method can be used to relate the number of tight/non-tight leptons with the number of real/fake ones. In the simple one-lepton case this is given by:

$$\begin{pmatrix} N_{XX}^T \\ N_{X\bar{T}}^T \end{pmatrix} = \begin{pmatrix} r & f \\ \bar{r} & \bar{f} \end{pmatrix} \cdot \begin{pmatrix} N_R^L \\ N_F^L \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} N_R^L \\ N_F^L \end{pmatrix}. \quad (4.4)$$

Here  $\bar{r}$  is defined as  $\bar{r} = (1 - r)$  which is the probability that a prompt lepton is not reconstructed as tight.  $\bar{f}$  is defined analogously.

This relation can be inverted to get

$$\begin{pmatrix} N_R^L \\ N_F^L \end{pmatrix} = \mathbf{M}^{-1} \cdot \begin{pmatrix} N_{XX}^T \\ N_{X\bar{T}}^T \end{pmatrix}. \quad (4.5)$$

This way one gets a description for the number of loose fake and real leptons depending on the number of tight, non-tight ones. This is desirable because the number of tight and non-tight leptons can easily be measured. The last step is to calculate the number of events with a tight fake lepton, which is given by:

$$N_F^T = f \cdot N_F^L. \quad (4.6)$$

This number can be used to describe the fake background in the one-lepton case. In this thesis, however, a final state with one electron and one muon is considered. Therefore the formulas from above get more complex, but the strategy remains exactly the same.

The equivalent to the basic Equation 4.4 in the two lepton case is:

$$\begin{pmatrix} N_{XX}^{TT} \\ N_{XX}^{T\bar{T}} \\ N_{XX}^{\bar{T}T} \\ N_{XX}^{\bar{T}\bar{T}} \end{pmatrix} = \begin{pmatrix} r_e r_\mu & r_e f_\mu & f_e r_\mu & f_e f_\mu \\ r_e \bar{r}_\mu & r_e \bar{f}_\mu & f_e \bar{r}_\mu & f_e \bar{f}_\mu \\ \bar{r}_e r_\mu & \bar{r}_e f_\mu & \bar{f}_e r_\mu & \bar{f}_e f_\mu \\ \bar{r}_e \bar{r}_\mu & \bar{r}_e \bar{f}_\mu & \bar{f}_e \bar{r}_\mu & \bar{f}_e \bar{f}_\mu \end{pmatrix} \cdot \begin{pmatrix} N_{RR}^{LL} \\ N_{RF}^{LL} \\ N_{FR}^{LL} \\ N_{FF}^{LL} \end{pmatrix} = \mathbf{M} \cdot \begin{pmatrix} N_{RR}^{LL} \\ N_{RF}^{LL} \\ N_{FR}^{LL} \\ N_{FF}^{LL} \end{pmatrix}. \quad (4.7)$$

Inverting the matrix is just the same, but when calculating the final number of tight fake events, one needs to consider the three kinds of fake background now: single

electron, single muon and double fake background. This influences the way  $N_{\text{F}}^{\text{TT}}$  needs to be calculated:

$$N_{\text{TT}}^{\text{F}} = r_e f_{\mu} N_{\text{RF}}^{\text{LL}} + f_e r_{\mu} N_{\text{FR}}^{\text{LL}} + f_e f_{\mu} N_{\text{FF}}^{\text{LL}}. \quad (4.8)$$

There are two minor problems with the matrix method that one needs to keep in mind. Firstly, due to the definition of the single and double fake backgrounds, it is possible that their contribution might become negative<sup>c</sup>. This depends on the values of the efficiencies and the background composition. It is usually just a problem with low statistic bins but should be prevented if possible.

The other thing one needs to think about is, that using the Matrix Method indirectly unblinds the data, because the number of tight/non-tight events are taken from data. Also, the calculation of fake efficiencies relies on data (but only from VRs), which will be discussed in the following section.

#### 4.4.2. Calculation of Matrix Method Efficiencies

Once the MM efficiencies are calculated, the MM is a straight-forward way of estimating the fake background. More challenging is the actual calculation of the efficiencies beforehand. It is desirable to perform these in a region which is as close to the SR as possible. This is important because the efficiencies will differ slightly for the different processes that contribute to the fake background. When switching to another region, the composition of fake events might change, which will have an impact on the measured effective efficiencies.

To determine the real efficiencies one needs a way of selecting real events with as less contamination from fakes as possible. One can then calculate the fraction of events which contain electrons that not only pass the baseline selection but also the tight conditions:  $r_e = N_{\text{RX}}^{\text{TL}}/N_{\text{RX}}^{\text{LL}}$  in order to calculate a real, electron efficiency.

In practice, the selection of only real events can be achieved when looking on MC events only and applying a cut on their truth information. By doing so, one trusts in the modelling of real efficiencies of the MC simulation<sup>d</sup>. This thesis uses the `TruthClassification` tool [96], that is provided by the Isolation and Fake Forum (IFF) of the ATLAS collaboration, for this selection. To select real electrons the `IFF class 2`, which stands for prompt electrons, is used. Prompt muons in MC events are assigned `IFF class 4` and are investigated for muon efficiencies here. As the real efficiencies are based solely on MC generated events, they can be calculated directly in the signal regions, without unblinding data there. For the low- and high-mass SR individual real efficiencies are calculated. Due to statistical limitations, the LMSR needs to be widened a bit by dropping the  $E_{\text{T}}^{\text{miss}}$  and MMC cuts.

For the calculation of fake efficiencies one needs a way of selecting almost exclusively fake events. This cannot be done using the truth information of MC events as the data-driven estimation of fakes is only necessary because fake events are not well-modelled in MC. Instead, one chooses a parameter region which contains an enhanced contribution from fake events. In the context of proton-collisions, this can be achieved by looking at regions with two leptons, that have charges of the same sign. Processes which lead to

---

<sup>c</sup>Negative contributions might also happen for MC generated backgrounds, as they can have negative event weights.

<sup>d</sup>Small corrections of the MC modelling, based on data, are already applied by using recommended ATLAS scale factors. These are, however, calculated in very different regions.

---

Table 4.4.: Regions, where the different MM efficiencies are calculated. The region is always the same for electron and muon. LMSR\* represents a wider LMSR without the  $E_T^{\text{miss}}$  and MMC cut. See Table 4.3 for the definition of all regions.

efficiency	2 <sup>nd</sup> lepton	applied in	calculated in
real	tight	HMSR	HMSR
real	non-tight	HMSR	ZVR
fake	tight	HMSR	SS ZVR
fake	non-tight	HMSR	SS ZVR
real	tight	LMSR	LMSR*
real	non-tight	LMSR	ZVR
fake	tight	LMSR	SS ZVR
fake	non-tight	LMSR	SS ZVR

final states with two same-sign leptons are relatively rare. This means the probability is high that one can find events with at least one fake lepton there. It was found that the SS ZVR is the same-sign region that is as close as possible to the SR but still provides sufficient statistic for the calculation of fake efficiencies. To further suppress the contribution of real events, the fake events are calculated by subtracting prompt MC events (using truth information) from data. The calculation of fake efficiencies is then performed analogously to the real efficiencies:

$$f_e = \frac{N_{\text{data}}^{\text{TL}} - N_{\text{RX}}^{\text{TL}}}{N_{\text{data}}^{\text{LL}} - N_{\text{RX}}^{\text{LL}}}. \quad (4.9)$$

All efficiencies depend strongly on the transverse momentum of the lepton. For example, the reconstruction of a lepton with higher momentum is more likely to be accurate. This means the real efficiencies go towards 1 for higher momenta, while fake efficiencies decrease. Therefore, the electron efficiencies are binned in  $p_T^e$  and muon efficiencies in  $p_T^\mu$ .

A fundamental assumption of the MM is, that the two leptons and their probability to be reconstructed as tight or not is totally independent of each other. It was shown, that for the here presented analysis this is not entirely the case, probably due to the use of combined electron-muon triggers. It was observed that the modelling of fake background improves drastically, when an additional parametrization for the tightness of the other lepton is introduced. This means, for example, that the electron efficiencies change slightly if the muon is reconstructed as tight or non-tight. In the following, efficiencies will be calculated separately depending on the tightness of the second lepton. This also needs to be considered in the application of the Matrix Method. Due to a lack of non-tight events in the SR the real efficiencies are calculated in the ZVR in this case. A full overview on the regions where the different efficiencies are calculated is given in Table 4.4. The final efficiencies can be seen in Figure 4.8.

Using the MM introduces a range of new systematic uncertainties. These are applied using variations for the different efficiencies. One needs, for example, to consider the statistical and systematic uncertainties on the distributions that are used to calculate

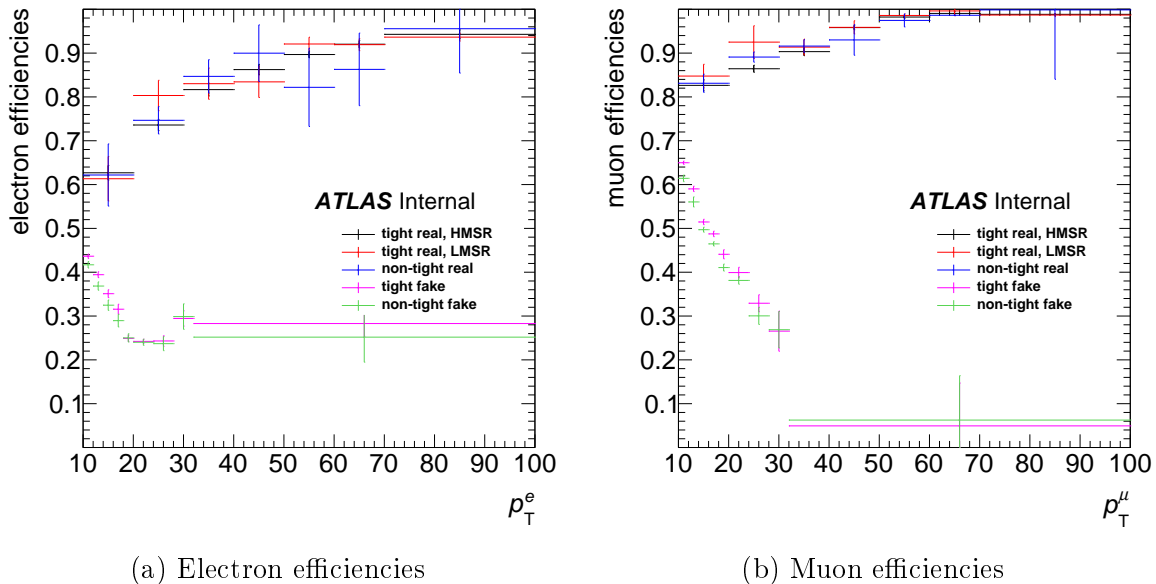


Figure 4.8.: Distribution of the final electron and muon efficiencies in dependence of transverse momentum  $p_T$  of the electron and muon, respectively. This includes real and fake efficiencies in the different tight and non-tight cases, as well as a separation of HMSR and LMSR for the tight real efficiencies.

the efficiencies. As described above it would be best to calculate everything in the SR. This is not possible due to a range of reasons and can be considered by including a composition uncertainty. Lastly, the efficiencies also depend on other parameters than the lepton's momentum, like on its value for  $\eta$ . Ideally, it would be possible to use a two-dimensional  $p_T - \eta$  distribution as parametrisation. This procedure would introduce more bins. Because of that the number of events, used to calculate efficiencies in the individual bins, can get relatively small, which means that statistical uncertainties will become very dominant. To prevent this behaviour it was decided to only use a parametrisation in the lepton's momenta and cover the missing  $\eta$ -dependence by introducing a corresponding systematic uncertainty.

In my work I focused on the calculation of efficiencies. It was advised to use the `FakeEfficiencyTool`, that was developed by the IFF [97]. Due to a very different code base of the framework that was used for this thesis and the one that was used in the `FakeEfficiencyTool`, this was not possible for the most part. It was therefore decided to only use the `RatePlotter.cxx` which takes ROOT histograms of the distributions for different tightness working points and performs the actual calculation of the efficiencies based on that. The input histograms for the efficiency calculations are extracted from NTuples, that contain all data and MC events relevant for this analysis, using a custom script.

### 4.4.3. Application of the Matrix Method

In practice, the MM is applied in this analysis using event weights. These are assigned to all data events that pass the baseline selection and depend on the tightness and momenta of both leptons. The calculation of these weights is performed by the `FakeBkgTool` [98], that is also provided by the IFF. The fake background is then

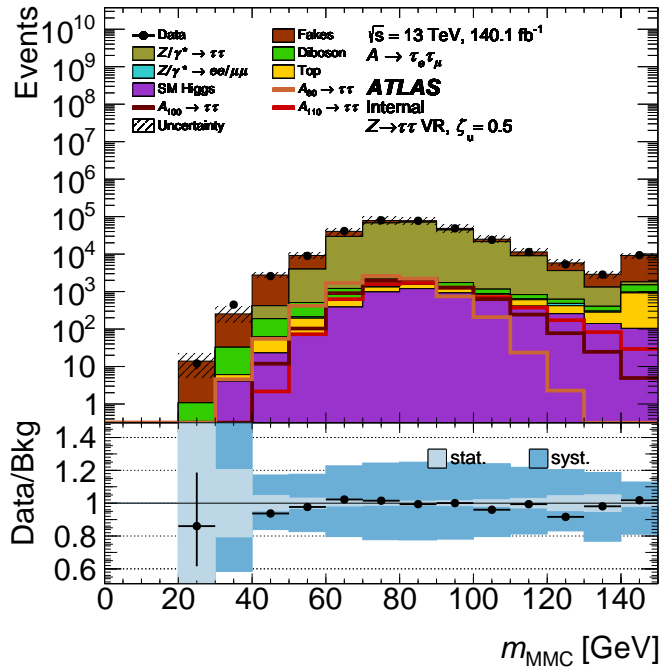


Figure 4.9.: Distribution of  $m_{\text{MMC}}$  using Monte Carlo generated background samples, as well as a fake background estimated using the Matrix Method. The reweighting of the  $Z \rightarrow \tau\tau$  background as presented in the following section is already applied.

calculated as the weighted sum over all data events that pass the baseline selection.

The MM is used to estimate both single and double fake events. Many MC generated backgrounds also include fake events to some extent. To prevent double counting, an additional cut is applied to all MC events to select only those with two real leptons (cut on IFF class 2 for prompt electrons and IFF class 4 for prompt muons).

Figure 4.9 shows the distribution, that was used in the beginning of this section to motivate the need for a fake background. It now includes fake events, which drastically improves the background modelling. In the end the fake background makes up 9.1% of all background events in the HMSR, 18.6% in the ZVR and a total of 97.5% in the SS ZVR.

## 4.5. Reweighting Procedure for the $Z \rightarrow \tau\tau$ -Sample

After all the background processes, including the estimation of fake events, was finalised, there was still a mismodelling of some variables visible. This was apparent in different distributions in the  $Z \rightarrow \tau\tau$  validation region. A selection of mismodelled variables can be seen here in Figure 4.10 and additional ones can be found in the appendix Figure C.1. There seems to be an excess of estimated background events in comparison to data especially for  $E_{\text{T}}^{\text{miss}}$  values between 30 GeV and 120 GeV. At the same time, the total data-to-background ratio of 97.8% in the ZVR is relatively good.

In the ZVR the  $Z \rightarrow \tau\tau$  background makes up 77.8% of all background events. In the HMSR this value is at 67.9%, which makes  $Z \rightarrow \tau\tau$  the most relevant process there, as well. Therefore, it is likely, that the mismodelling observed in the ZVR will

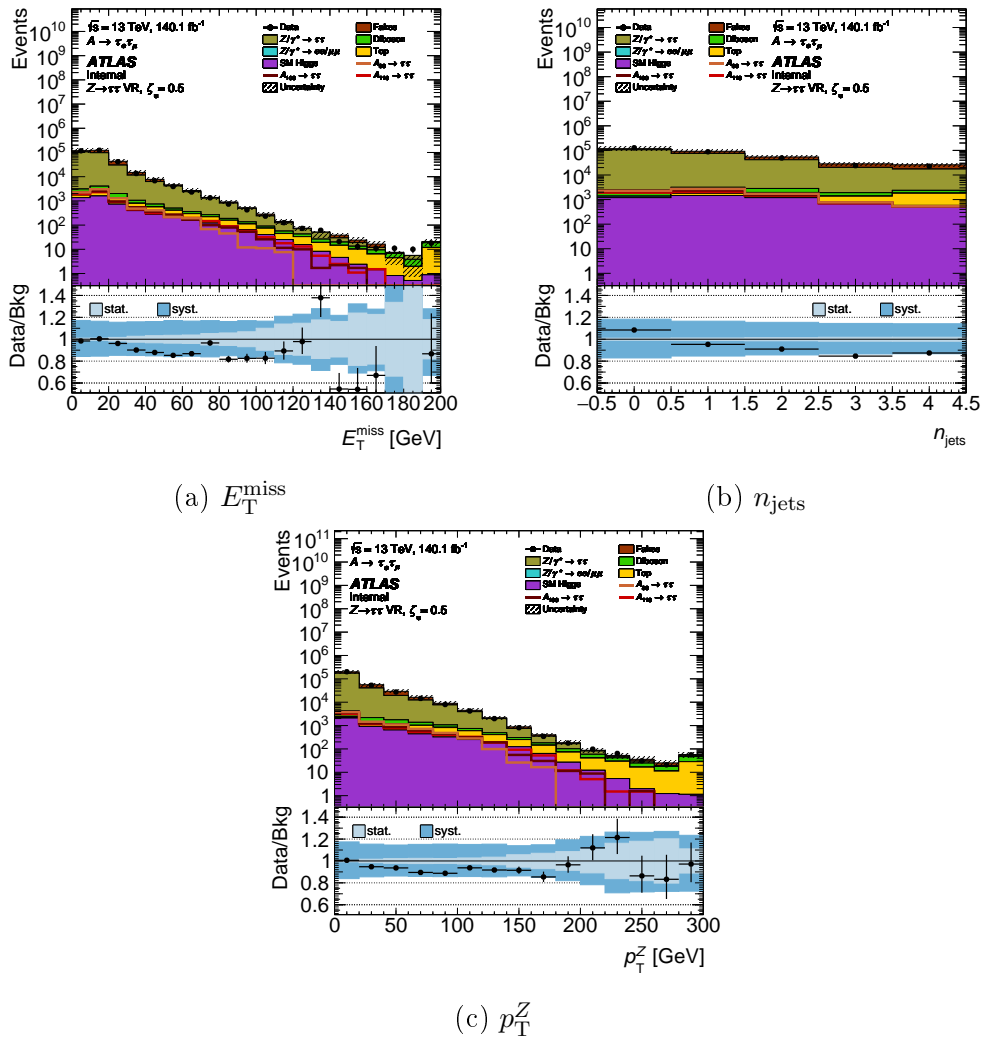


Figure 4.10.: Distributions in the ZVR that seem to be mismodelled (the  $Z \rightarrow \tau\tau$  reweighting procedure is not applied here). The shown distributions are  $E_T^{\text{miss}}$  (a),  $n_{\text{jets}}$  (b) and  $p_T^Z$  (c).

translate to the SR. The mismodelling is especially pronounced in the distribution of  $E_T^{\text{miss}}$  in the region above 40 GeV. The definition of high- and low-mass SR include cuts on  $E_T^{\text{miss}}$  ( $> 50$  GeV,  $> 30$  GeV for LMSR, HMSR) that explicitly exclude the  $E_T^{\text{miss}}$  region where the modelling seems to be fine. This might even worsen the mismodelling in the SR.

The analysis, this thesis is based on, is still blinded, therefore it is impossible to check the SR explicitly for mismodelling. Also, if there was any mismodelling happening in the SR, this would need to be treated with caution, as it is hard to distinguish a shape that is caused by the presence of a signal, from one that comes from a problem in the background modelling. Therefore, one can just assume that the problem that caused the mismodelling in the ZVR will translate to the SR. This might lead to problems in the discovery of an actual signal in the end.

There were several studies performed to look for the source of this mismodelling. The influence of different MC campaigns/data-taking years, of triggers as well as different lepton identification working points and alternative MC samples were investigated. A

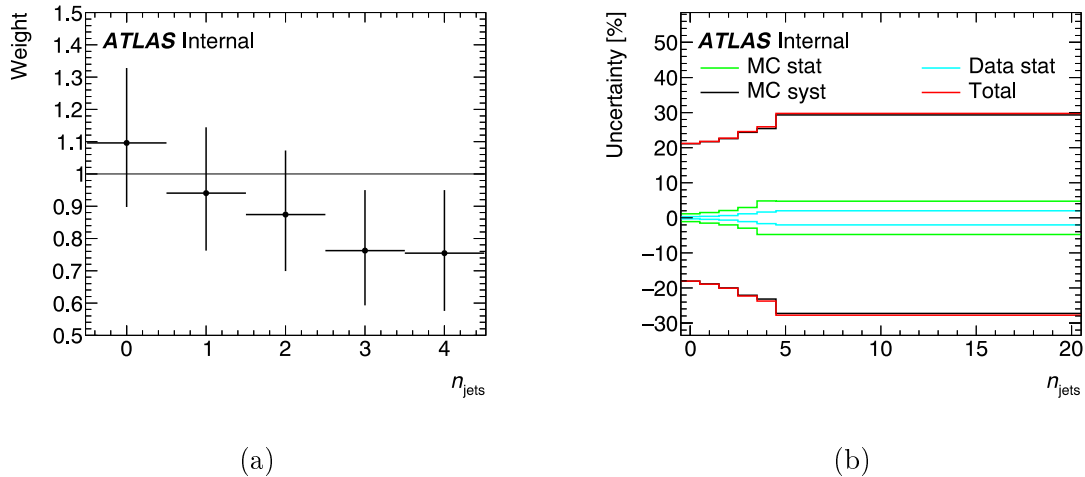


Figure 4.11.: Distribution of  $Z \rightarrow \tau\tau$  weights (a) and their systematic uncertainties (b) as a function of  $n_{\text{jets}}$ . The systematic uncertainties are split depending on their source during calculation: statistical uncertainties on MC and data as well as systematic uncertainties on MC.

more extensive summary of selected studies can be found in Appendix C. None of these studies were able to come to satisfactory conclusions.

It was found that the problem most likely comes from the  $Z \rightarrow \tau\tau$  samples, which seem to be not so good at modelling the transverse momentum of the  $Z$  boson and its interplay with jets. This was concluded because the mismodelling is not observed in the other validation regions, where the  $Z \rightarrow \tau\tau$  background is not as pronounced. An example of distributions in the TVR for comparison can be found in the appendix in Figure C.2.

After the studies were unable to present a possible solution to the problem, it was decided to use a data-driven reweighting method to suppress its influence. A similar procedure was used in [99]. For this, it is assumed that the mismodelling is solely caused by the  $Z \rightarrow \tau\tau$  samples and all other kinds of background are well-modelled. Depending on a selected variable, all  $Z \rightarrow \tau\tau$  events are then assigned an additional weight to correct this behaviour. A common choice for this variable is the momentum of the  $Z$  boson  $p_{\text{T}}^Z$ . This is not an option for this thesis, as the ZVR and the SR are mostly separated by a  $\Delta R_{ll}$  cut (see Table 4.3). This variable is closely correlated with  $p_{\text{T}}^Z$ , which leads to very different  $p_{\text{T}}^Z$ -distributions in the two regions and makes it an unfit choice for the reweighting. Instead, it was decided to use  $n_{\text{jets}}$  as it is relatively independent to other variables that are of importance for the analysis. Additionally it shows a particularly systematic mismodelling, as one can see in Figure 4.10b.

The weights are now calculated for each  $n_{\text{jets}}$  bin individually. They are defined as the ratio of "corrected" number of  $Z \rightarrow \tau\tau$  events in the ZVR over the number of simulated events there. The number of "corrected" events is understood as the difference between data and the sum of all other background processes:

$$\text{weight} = \frac{\text{Data} - \sum_{i \neq Z \rightarrow \tau\tau} \text{Bkg}_i}{\text{Bkg}_{Z \rightarrow \tau\tau}}. \quad (4.10)$$

These weights are then applied to all  $Z \rightarrow \tau\tau$  events, depending on their  $n_{\text{jets}}$ . The same values that were calculated in the ZVR are applied in all regions, assuming the

Table 4.5.: Changes in the number of  $Z \rightarrow \tau\tau$  events in the HMSR and ZVR when applying the  $Z \rightarrow \tau\tau$  reweighting.  $N_{\text{Bkg,tot}}$  describes the total number of background events. The last column is blinded for the SR.

		$N_{Z \rightarrow \tau\tau}$	$N_{Z \rightarrow \tau\tau}/N_{\text{Bkg,tot}}$	$N_{\text{Data}}/N_{\text{Bkg,tot}}$
HMSR	before reweighting	3122.8	0.679	-
HMSR	after reweighting	2570.6	0.635	-
ZVR	before reweighting	248298.8	0.778	0.978
ZVR	after reweighting	241393.0	0.774	1.0

mismodelling is independent of the performed separating cuts. The distribution of the weights can be seen in Figure 4.11a together with their systematic uncertainties.

These are solely based on the uncertainties from the underlying distributions, that were used for the calculation of the weights. Gaussian error propagation is used to determine their influence on the weight calculation. There are different error sources that are considered here: the statistical uncertainties on the data and the different background distributions, as well as systematic uncertainties on the background. For the latter, all different sources of systematics are summed up quadratically to form one single uncertainty.

This calculation relies on a good estimation of all background processes. Therefore, it is essential that the real and fake efficiencies for the Matrix Method are calculated beforehand, so that a fake background is available. These efficiencies, however, should slightly depend on the reweighting itself. Therefore one would ideally iterate the calculation of MM efficiencies and  $Z \rightarrow \tau\tau$  weights until they converge. A study was performed that was able to show that the efficiencies change only marginally when calculated before/after reweighting, the results can be found in Appendix B.

The final limit calculation, which will be discussed in the following section, is slightly influenced by the reweighting. There are two competing effects: improvement of the calculated limits due to a general reduction of background and a worsening of it from the addition of systematic uncertainties. The difference between limits can be seen in Figure C.5 in the appendix. For higher mass hypotheses, there can generally be seen a slight improvement in the calculated limits.

Table 4.5 shows the change in number of  $Z \rightarrow \tau\tau$  events in the ZVR and HMSR. This also includes the change in data-to-background ratio, which is artificially set to 1.0 in the ZVR, due to the definition of the reweighting procedure. Generally this method is able to improve the background modelling in the ZVR. This is not only true for the distribution of  $n_{\text{jets}}$ , but is also beneficial for other variables, see Figure 4.12.



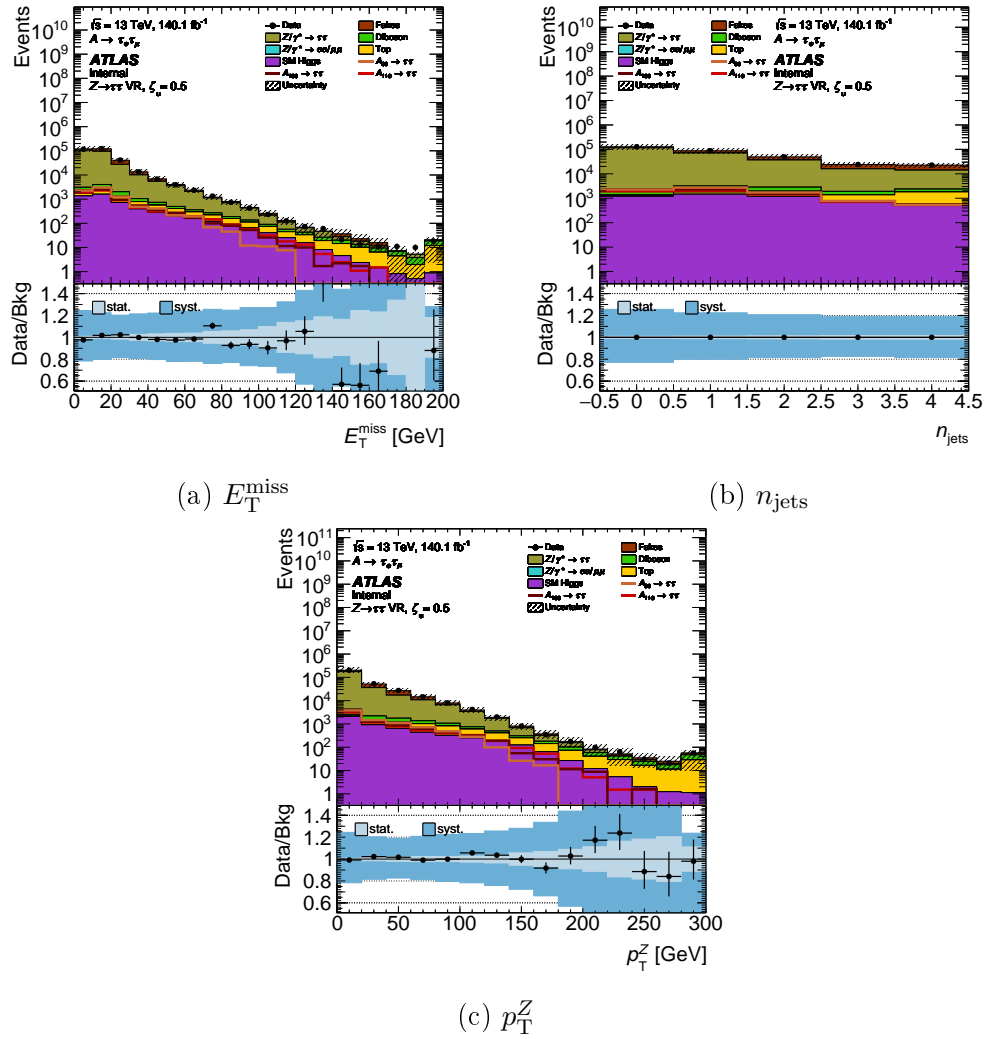


Figure 4.12.: Distributions of formerly mismodelled variables in the ZVR. The  $Z \rightarrow \tau\tau$  reweighting procedure is applied here. The shown distributions are  $E_T^{\text{miss}}$  (a),  $n_{\text{jets}}$  (b) and  $p_T^Z$  (c).

## 4.6. Results of the Cut-Based Analysis

The final goal of the base analysis is setting an upper limit on the production cross section times branching ratio of the  $A$  boson. As the analysis is still blinded, it is only possible to present expected limits here. The statistical analysis is done using the `WSMaker` package [100, 101].

These limits are based on the  $m_{\text{MMC}}$  distributions that are calculated in the two SRs and can be seen in Figure 4.13. Their binnings are chosen in a way, that the statistical uncertainties in each bin do not exceed 10%. At the same time each bin must contain at least 10 events, so that all statistical assumptions remain valid. Also, a minimum bin size of 2 GeV is required to accommodate for a limited mass resolution and suppress the influence of bin fluctuations.

Figure 4.14 shows the final expected upper limit for ten mass hypotheses between 20 GeV and 110 GeV.

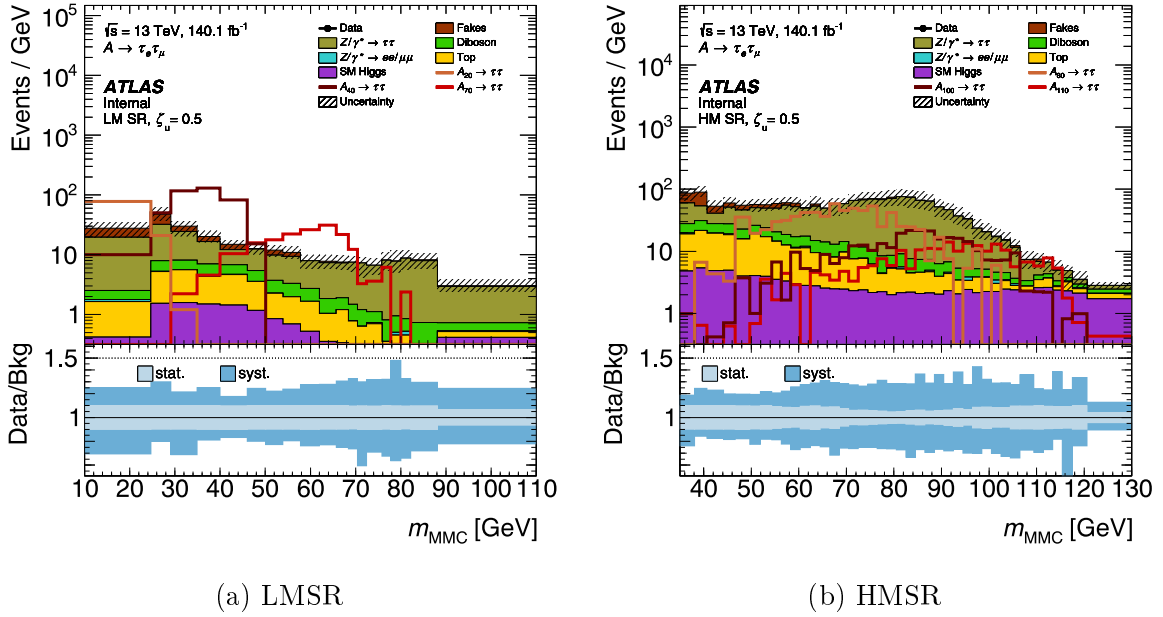


Figure 4.13.: Distribution of  $m_{\text{MMC}}$  in the low-mass and high-mass SR using the binning chosen for the calculation of the expected limit. The y-axis shows the number of events per bin divided by its bin width.

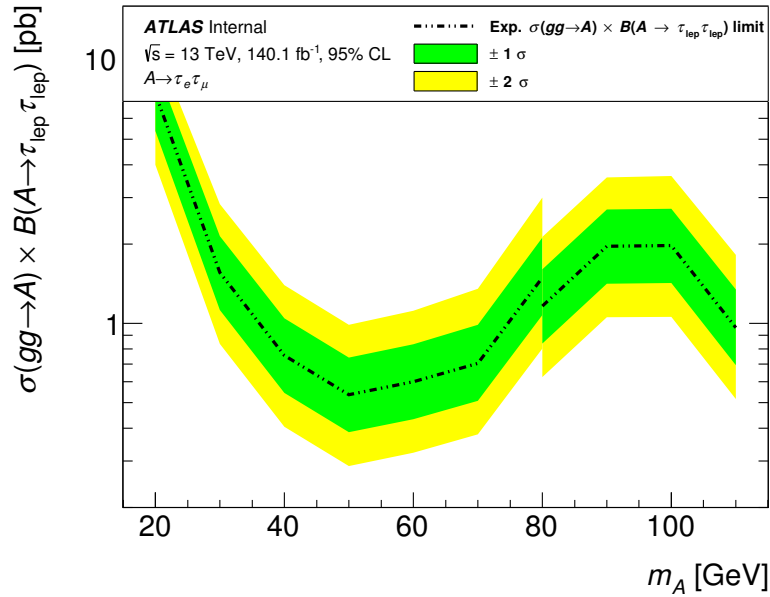


Figure 4.14.: Expected limit of the base analysis on the production cross section for  $gg \rightarrow A$  times the branching ratio for  $A$  decaying into two leptonic  $\tau$ -leptons. The limit is calculated for  $A$  boson masses from 20 to 110 GeV and includes systematic uncertainties. The dashed line shows the limit set by analysing the shape of the  $m_{\text{MMC}}$  distribution. The green area denotes the  $1\sigma$  uncertainty band and the yellow marks the  $2\sigma$  area.

# 5. Introduction to Neural Networks

This section will give a general introduction to artificial neural networks (NNs). It will include basic principles of fully connected, feed-forward networks as well as their training. Additionally, parameterised neural networks and techniques to optimise hyperparameters will be described.

## 5.1. Neural Networks as an Extension to High-Energy Physics Analyses

In the context of high-energy physics experiments, there are many tasks NNs can be applied to. For example, they are used for the reconstruction of tracks, vertices and deposited energy in detectors and for triggering. They can also help in identifying and discriminating certain particles [102].

This thesis focusses on a NN that classifies events. To each event it assigns a score, which can be interpreted as the probability that this event originates from a signal process. Classic cut-based analyses, like the one presented in Chapter 4, rely on the separation power of physically motivated, high-level variables, like reconstructed masses. Applying cuts on these variables can then lead to an improvement of the signal-to-background ratio.

Meanwhile, NNs are not limited to simple cuts alone, but are able to connect information from different input variables in complex ways. Therefore, the score variables can show a very high separation power. This makes them well suited inputs for further statistical analyses. Powerful tools like these can help to discriminate small signal contributions from large backgrounds.

## 5.2. Neural Networks

The general idea of NNs [103] is to model biological neural networks like the human brain, using modern technology. Because of that NNs consist of many *artificial neurons*, where small individual calculations are performed. These neurons are connected with each other in order to allow for an exchange of information between them. Typically, neurons are arranged in different layers. If each neuron has connections to all neurons in its two neighbouring layers, this is called a *fully connected neural network*. It is also *feed-forward* if the direction of information flow between different layers always stays the same, without any loops. A schematic view of such a network is shown in Figure 5.1. Fully-connected, feed-forward networks can mathematically be described as directed graphs, without any cycles.

The first layer of a NN is called the *input layer*. It is used to feed information from the input data into the NN. In the context of a high-energy physics application, this will usually be different (one-dimensional) variables that describe one recorded

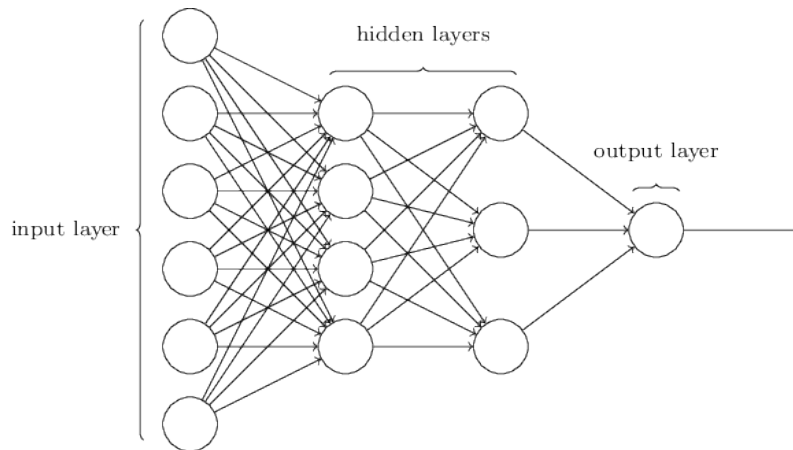


Figure 5.1.: Schematic view of the architecture of a feed-forward, fully-connected neural network. Each circle represents one artificial neuron and the arrows symbolise the flow of information. The input, output and hidden layers of the network are marked [103].

event, like the momentum and direction of involved particles. The number of input neurons depends on the number of input variables. The neurons in the following *hidden layers* perform the main calculation. A NN's *architecture* is defined by the number of hidden layers and neurons per layer. The *output layer* is the last layer of the NN and is responsible for the calculation of final output values. Depending on the task of the NN, this might be just one value, for example, representing the probability that the input event comes from the signal process (binary classification). But also multiple output values are possible, for example, if the classification needs to distinguish between more than two different classes (multi-class classification). The output values always need to be interpreted in dependence on the network's task. The user only interacts with the input and output layers of the NN during application.

All neurons in the hidden layers have the same structure. Each of them has a number of different inputs from the previous layer and one single output that is passed on to the neurons in the following layer. The different inputs are assigned individual *weights*, that are an integral part of the NN, as they are independent of the event to which the network is applied. Each neuron performs one calculation. The first step is to calculate its final input value as the weighted sum of all individual inputs, shifted by a fixed *bias*. These biases behave like the neuron weights in the sense, that they are independent of the applied event and each neuron is assigned its bias individually. If one looks at the  $k$ -th neuron in layer  $l$ , its final input  $i$  is defined by the output values  $o$  of the  $N$  neurons in the  $(l-1)$ -th layer, the weight of their connection  $w$  and its own bias  $b$ :

$$i_k^l = \sum_{j=1}^{N^{l-1}} w_{(k,j)}^l \cdot o_j^{l-1} + b_k^l. \quad (5.1)$$

This calculation can also be expressed as one single entry of the result of a matrix multiplication. In this case, a weight matrix is multiplied with a vector of the output values of the previous layer and a vector of the current bias terms is added to get the total input values for all neurons in the specified layer. This is a calculation that can be performed really fast on modern computers, which makes NNs highly performant

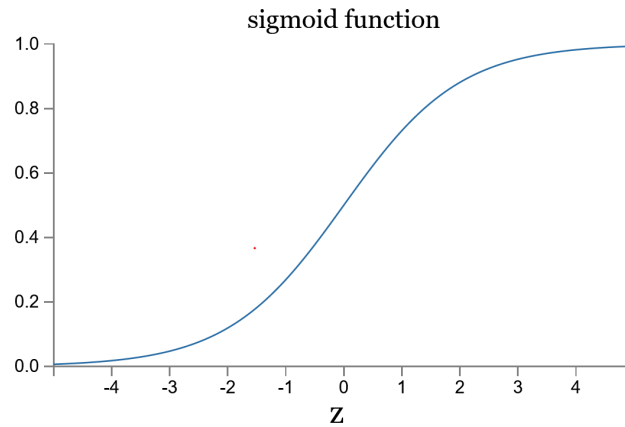


Figure 5.2.: Shape of the logistic sigmoid function from Equation 5.3 [103].

and time efficient [103].

The calculated total input  $i_k^l$  is then processed using an *activation function*  $A$  [104] to calculate the output  $o_k^l$ :

$$o_k^l = A(i_k^l). \quad (5.2)$$

The activation function allows for a non-linear relationship between a neuron's in- and output values, as it can be chosen relatively freely. There is a variety of different activation functions that are regularly used. Popular choices are sigmoid functions like the logistic sigmoid function, which is defined as

$$\text{sig}(x) = \frac{1}{1 + e^{-x}}, \quad (5.3)$$

or the hyperbolic tangent  $\tanh(x)$ . They are motivated by the response of activated biological neurons. The output of the sigmoid function is limited to values between 0 and 1. Negative inputs yield results close to zero and high inputs lead to outputs close to one, with a continuous turn-on effect in between. This can be seen in Figure 5.2.

These kinds of functions show some deficiencies as their output becomes saturated for high and low input values, which is related to the *vanishing gradient problem* [105, 106]. Additionally, their usage results in higher computational complexity than other activation functions like the very commonly used *Rectified Linear Unit (ReLU)* [104]. This alternative activation function is defined as

$$\text{ReLU}(x) = \max(0, x). \quad (5.4)$$

which means that its output values are larger or equal to zero and are unlimited for high input values. One of the downsides of using ReLU is the non-utilization of negative input values.

Usually the same activation function is used for all neurons in the hidden layers. Depending on the task of the network, it might be beneficial to use a different, more suited one for neurons in the output layer. In the case of a classification task, a sigmoid function with an output between zero and one is well suited, as the output can be interpreted as probability.

### 5.3. Training of Neural Networks

The first step of getting a NN that is able to perform its task, is to define its general structure. This includes, for example, choosing input parameters and defining the network's architecture with the number of hidden layers and number of neurons per layer, as well as the kind of activation function to be used. When these *hyperparameters* are fixed, the performance of the NN depends solely on the weights and biases of each neuron. The next step is therefore optimising these parameters. This process is called the *training* of the NN. After the training is finished, the network is ready to be applied.

One possible method are *supervised learning* strategies. Their general idea is to use sets of input values  $\vec{x}$  together with the corresponding expected output value  $y$ . The input values are presented to the NN, which calculates an output  $f(\vec{x})$  based on that. Afterwards, the calculated and expected output is compared and, based on this, the weights and biases are adjusted. This procedure is repeated until the performance of the NN stabilises.

It is necessary to define a measure for how much the expected and the calculated output differ. This is done using a *loss/cost function*  $L(f(\vec{x}), y)$ . The loss function can be chosen according to the NN's task. For a classification application, the binary cross entropy is a well suited option:

$$L(f(\vec{x}), y) = -[y \log(f(\vec{x})) + (1 - y) \log(1 - f(\vec{x}))]. \quad (5.5)$$

In general, loss functions should be differentiable, and small values of  $L$  correspond to a good match between expected and observed loss, by convention. Therefore, training of NN is done in practice by minimising the average loss over all training samples depending on the neurons weights and biases. As deep neural networks become larger and larger, the number of neurons and therefore the number of weights and biases increases. This makes it impossible to solve this minimisation problem analytically. Instead, one relies on iterative, numerical algorithms like *gradient descent (GD)* to find (at least a local) minimum of the loss function.

At first, the weights and biases have to be initialised with some values to get a first (probably bad) guess and a starting point for gradient descent. Typically, a normal distribution around 0 is used to randomly select initial weights, while the biases are simply set to 0. The network's parameters are then updated iteratively according to a training procedure, in order to minimise the loss.

The idea of *batch gradient descent* [107] is to calculate the average loss over all training samples for the current configuration of the NN as well as its gradient. It is usually possible to directly calculate the gradient of the loss function  $\nabla L(\vec{w}, \vec{b})$ , if the average loss function itself and the output of the NN are differentiable. The calculation of partial derivatives  $\partial L / \partial w$  or  $\partial L / \partial b$  is often done using the *back propagation* algorithm. It starts for a given input event with calculating the output of all layers, beginning with the first one and then through the whole network. The next step relies on determining the influence of small changes in the input to neurons on the loss, by looking at the partial derivative  $\delta_j^l = \partial C / \partial i_j^l$ . Here  $i_j^l$  is the final input of a neuron, as defined in Equation 5.1. It is possible to calculate  $\delta_j^l$  for the last layer directly. For other layers  $l$  this quantity can be expressed in dependence of  $\delta_j^{l+1}$  from the following layer. This means the "error"  $\delta$  can be propagated from the last layer, back to the previous layers. Finally, it is possible to relate  $\delta_j^l$  with the interesting partial derivatives  $\partial L / \partial w_{jk}^l$  and  $\partial L / \partial b_j^l$ . The back propagation algorithm strongly relies on the chain

rule for calculating derivatives in the definition of  $\delta^l$ . A more thorough description of the back propagation algorithm can be found in [103].

When all partial derivatives and therefore the gradient of the average loss function is known, the weights and biases are updated by a certain step width  $\alpha$ , called the *learning rate* in the opposite direction of the gradient:

$$(\vec{w}, \vec{b})_{\text{updated}} = (\vec{w}, \vec{b})_{\text{old}} - \alpha \cdot \nabla L(\vec{w}, \vec{b}). \quad (5.6)$$

This should result in weights and biases that lead to a loss closer to the (local) minimum. With the updated parameters, the whole procedure is repeated.

A modified version of this algorithm is *mini-batch gradient descent (mGD)* [107]. The general procedure stays exactly the same, the only difference is the amount of training data that is used to calculate the average loss function. The whole training set is split into multiple *mini-batches* of a fixed number of training events. This means parameter updates depend only on a fraction of all training events and multiple parameter updates can be performed until the whole training set has been seen once. The term *epoch* describes in this context a training session, in which the whole training data has been used once. A whole training usually consists of many epochs.

The size of the mini-batches influences the network's ability to converge towards a good minimum during training. If the batch size is very small, containing only single events in the extreme case, there can be very strong statistical fluctuations of the gradient between different iterations. This prevents the algorithm from approaching a minimum of the loss function in a directed manner. Larger batch sizes, lead to an increasingly smoother progress. An advantage over using the whole dataset (corresponding batch gradient descent) is the greater ability to leave local minima, due to a slightly bigger variance in the parameter updates.

Although mGD is already an improved version of the batch gradient descent, there are some remaining problems. The learning rate, for example, has to be chosen very carefully as it strongly influences the network's ability to converge during training to a (local) minimum of the loss function. If  $\alpha$  is chosen too small, the convergence happens only very slowly, which means that many computation-intensive iterations are necessary and it is hard to find a good local minimum. Meanwhile, if the learning rate is too big this impedes convergence, as updates of the loss function might repeatedly jump from one side of a minimum to the other, without getting significantly closer to it. One answer to this problem is the use of *learning rate schedulers*, which change  $\alpha$  over the course of the training. This way, the step size can be chosen quite large in the beginning to allow leaving local minima. Towards the end of the training,  $\alpha$  can be reduced to make converging towards a minimum possible. Their weakness is, however, that they need to be defined before the training even started and cannot react to the actual current situation. The gradient descent algorithm can also get stuck close to minima or saddle points with very small gradients in the surrounding region [107]. A solution to these problems are more advanced algorithms, like the *adaptive moment estimation optimiser (ADAM) optimiser* [108]. This algorithm is based on gradient descent but additionally takes earlier calculated gradients into account. Using this information the ADAM optimiser is able to scale the learning rate dynamically during training. If the gradient stays very similar over multiple iterations the step size can be increased. On the other hand a strongly fluctuation gradient, which is likely to happen close to a minimum, will lead to a reduction of the learning rate.



This algorithm depends on four scalar constants:  $\alpha$  is the learning as before,  $\beta_1$  and  $\beta_2$  determine the influence of past gradients on the current parameter update and lastly  $\epsilon$  ensures numerical stability. The  $(t+1)$ th parameter update is now defined by:

$$(\vec{w}, \vec{b})(t+1) = (\vec{w}, \vec{b})(t) - \alpha \cdot \frac{m(t)}{\sqrt{v(t) + \epsilon}}. \quad (5.7)$$

The newly introduced quantities  $m(t)$  and  $v(t)$  are recursively defined as:

$$m(t) = \frac{\beta_1 m(t-1) + (1 - \beta_1) \nabla L(t)}{1 - \beta_1^t}, \quad (5.8)$$

$$v(t) = \frac{\beta_2 v(t-1) + (1 - \beta_2) [\nabla L(t)]^2}{1 - \beta_2^t}, \quad (5.9)$$

with  $\beta_i^t$  meaning  $\beta_i$  to the power of  $t$ . Suggested values for the four parameters are:  $\alpha = 0.01$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and  $\epsilon = 10 \times 10^{-8}$  [108].

The combination of mGD or the ADAM optimiser with back propagation is the base for a successful training of NNs, but there are some additional considerations to further improve this.

The first improvement is preparing the data used for training. This includes transforming all input variables so that they are distributed around zero with a standard deviation of 1. This is especially helpful if different input parameters of the network have very different orders of magnitude. If these inputs are not normalised, this can lead to some of the input variables being almost fully ignored by the network. It is important to keep in mind that the samples to which the NN will be applied after the training is completed need to be transformed in the exact same way.

Another way to possibly improve the performance is to introduce event weights to all training events. In this case one will need to calculate the weighted average of the loss function during gradient descent. Event weights can, for example, be used to balance the influence of different classes of training events. This might be necessary for example if there are only a few events of one class, but a lot from another one. In this case the NN will probably be better in handling events from the second class if there are no event weights used. Therefore, it might be beneficial to assign high event weights to events from the first class to strengthen their influence on the training. Meanwhile, it might also be beneficial to use event weights to recreate a given unbalanced distribution of classes.

In general, the selection of training events should be well-thought-out. A high number of training events is crucial for the training's success, as it ensures, that the network can learn from a big variety of different inputs. Usually, however, one has only a limited amount of available input data. Additionally, it is common practice to split this data into two or three orthogonal datasets. One of these will be the *training* dataset, actually used during training. The second one is called the *validation* dataset. It can be used for a continuous check of the training progress. This can be done by repeatedly calculating the average loss of the validation data after a certain number of epochs. The validation loss can then be compared to the training loss. If the validation loss stagnates while the training loss is still improving, this is a sign of *overfitting*. This means the network has stopped to learn general features and now focuses specific details of the training dataset. In this case the training should be stopped as the network will not get better at generalizing from the training data to new inputs.

Table 5.1.: Scheme of the grouping of datasets in the case of a 5-fold method. Each row describes the function of a given dataset and columns represent the five different networks that are trained.

Dataset	NN <sub>1</sub>	NN <sub>2</sub>	NN <sub>3</sub>	NN <sub>4</sub>	NN <sub>5</sub>
1	train	train	train	validate	apply
2	apply	train	train	train	validate
3	validate	apply	train	train	train
4	train	validate	apply	train	train
5	train	train	validate	apply	train

It is possible to use a third dataset, the *test data*, to compare the performance of different networks, for example ones with varying architectures. The test dataset should be totally unbiased regarding the training and tuning of the NN. It is therefore not advisable to use the validation dataset for this, as it is already used, for example, to stop a training when overfitting is starting. For this reason a third and new dataset is necessary to fulfil this task. One has to carefully choose the fraction of available data each of these datasets should contain. Generally speaking, the more events are available during training the better, as it is good for the network to see a big variety of different inputs. At the same time the other datasets should not be neglected, as they also need to be able to produce statistically meaningful results.

## 5.4. $K$ -Fold Cross Validation

*K-fold cross validation* is a tool that is commonly used to test the reliability of a NN's performance. The general idea is to take all available labelled data (meaning there is an expected NN output available) and split it into  $k$  orthogonal datasets of roughly equal size. These can now be grouped together to form a training, validation and application dataset. Usually the training dataset should contain the majority of events, which is achieved by grouping  $(k - 2)$  datasets together, while only one is used for validation and application, each. Up until now this is just the same procedure as described in the section before.

The difference for  $k$ -fold cross validation is now that this procedure of grouping the  $k$  datasets is repeated  $k$  times, always switching the roles a dataset is assigned to, in a cyclical way. This means, for example, that every dataset is used for validation once. For each different data division one NN is trained independently based on a different fraction of all events. The strategy of grouping the  $k$  datasets differently for each  $k$ -fold is visualised in Table 5.1, for the case  $k = 5$ .

This method can be used to test statistical fluctuations between NNs that were trained using the same hyperparameters, by looking at the validation loss for the different  $k$ -folds [109]. In this thesis, the splitting into individual  $k$ -folds is used to address another problem. Generally, NNs are expected to perform better on data they have seen during training than on unknown input. In the context of particle physics analyses, it is likely that there is only a limited amount of simulated (MC) events that can be used for training. As the original process (signal or background) is

known only for simulated events, it is generally not possible to use measured data as training input<sup>a</sup>. However, in the end it is necessary to apply the NN to both MC and data events, to compare the respective final score distributions. If only one network is trained using the majority of MC events, but none of the measured data, the NN will be biased towards the simulated events and might be better at classifying the MC events. This will lead to an unfair comparison between the final score distributions in data and background. This procedure is, therefore, not advisable.

The concept of using multiple networks, trained on different  $k$ -folds, is very helpful to solve this problem. With this approach it is possible to apply each of the NNs only to the events reserved for application, that were therefore not used in any way before. Because of the method of grouping together the  $k$  base datasets for each of the  $k$ -folds, merging all the applied data from the different NN will result in one dataset which contains all available input data. These events are therefore evaluated by different NNs, but each individual evaluation is unbiased.

One disadvantage of the  $k$ -fold technique is the increased computing time, as it uses  $k$  times more resources than training and applying only one NN. The number of splits  $k$  should therefore not be chosen too high in order to limit the necessary computing power. As it allows using a large fraction of events during training, while also providing many unbiased results at the same time,  $k$ -fold is nevertheless a very helpful technique.

## 5.5. Parameterised Neural Networks

The goal of this thesis is to examine ten different mass hypotheses for the  $A$  boson between 20 GeV and 110 GeV. It is expected (and also can be observed in MC simulated events) that all of these behave differently for certain variables, like for example  $m_{\text{MMC}}$ . This can be seen in Figure 4.13. Because of this it is unlikely to successfully train a single regular neural network with the task of classifying all these different mass points as signal compared to background.

An obvious solution is the use of ten different NNs, each of them specialised in the classification of signal events with different masses  $m_A$ . Such networks can be simply trained using all background samples but only one type of signal events during training. This procedure has the disadvantage of needing ten times the computational power than a single NN. Additionally, the different NNs will probably be quite bad at evaluating signal that comes from an  $A$  boson of a different mass than the one it is trained on. This is a problem as it will likely be bad at interpolating between the different mass points. However, it is very likely that the  $A$  boson, if it exists, has a mass somewhere between the investigated masses. If this is the case all ten NNs will be bad at recognizing these kinds of events.

A *parameterised neural network (PNN)* can be a solution to this problem [102]. The idea behind this is adding one adjustable input parameter that represents the different kinds of signal. This parameter can, for example, be chosen to be the signal mass  $m_A$ . There are ten possible values  $m_A = 20, 30, \dots, 110$  GeV in this thesis. The training samples then need to be prepared in the following way: for all signal MC events this additional parameter is set to the true value, here the signal mass. All background

---

<sup>a</sup> an exception can be specially selected data events for the representation of fake events, see Section 6.1.

---

events get assigned a random one of the available values. The last step is to simply train a single NN that also considers this additional parameter as input variable.

The idea behind this is that the network will learn how the behaviour of different signal events change with their  $m_A$ . This will help the network to recognise signal events that are triggered by an  $A$  boson of a different than the considered mass points.

When the NN is applied to experimental data, a value for the newly added parameter  $m_A$  has to be supplied, along with the other input variables. This needs to be done both for MC and data events. It is common practice that each event, is presented to the network multiple times, each time with a different value for  $m_A$ . The values are chosen from the same list as during the training. In this analysis they are  $m_A = 20, 30, \dots, 110$  GeV. In the end each event is assigned a score for all possible values of  $m_A$  (here ten scores are assigned).

It is expected that for signal events these scores will be highest if  $m_A$  is the same or very close to their true value. This means, for example, that if one looks at the score distribution of  $m_A = 20$  GeV the best signal-to-background separation is reached for the signal with the same mass  $m_A = 20$  GeV. This means if one wants to use score distributions for a statistical analysis later on, for example to calculate limits, it is best to always use the corresponding score distribution for a given signal mass.

## 5.6. Comparing Classification Performance

It is important to have tools to assess and compare the performance of different NNs. There is a range of metrics that can be used for this. Many tools that will be introduced in this section can also be used for the assessment of classification problems in general and are not limited to neural network applications [110].

One important measure is given by the already introduced (weighted) loss function. Comparing the development of training and validation loss over the course of a training can be used to identify overfitting. It is also possible to look at the final values of validation loss after a training is finished to get a measure for the network's performance and compare these values.

Other metrics rely on the network's ability to separate signal  $s$  from background  $b$  events based on their assigned score  $x$ . This is quantified by introducing an artificial threshold  $t$  that is varied between the minimal ( $x = 0$ ) and maximal possible ( $x = 1$ ) score value. In the following all events with a score above the threshold ( $x > t$ ) will be classified as signal, and everything below as background. This leads to the definition of *true positive rate (TPR)*, *true negative rate (TNR)*, *false positive rate (FPR)* and *false negative rate (FNR)*:

$$\begin{aligned} \text{tpr}(t) &= \frac{\int_t^1 dx s(x)}{\int_0^1 dx s(x)}, & \text{fnr}(t) &= 1 - \text{tpr}(t) = \frac{\int_0^t dx s(x)}{\int_0^1 dx s(x)}, \\ \text{tnr}(t) &= \frac{\int_0^t dx b(x)}{\int_0^1 dx b(x)}, & \text{fpr}(t) &= 1 - \text{tnr}(t) = \frac{\int_t^1 dx b(x)}{\int_0^1 dx b(x)}. \end{aligned} \quad (5.10)$$

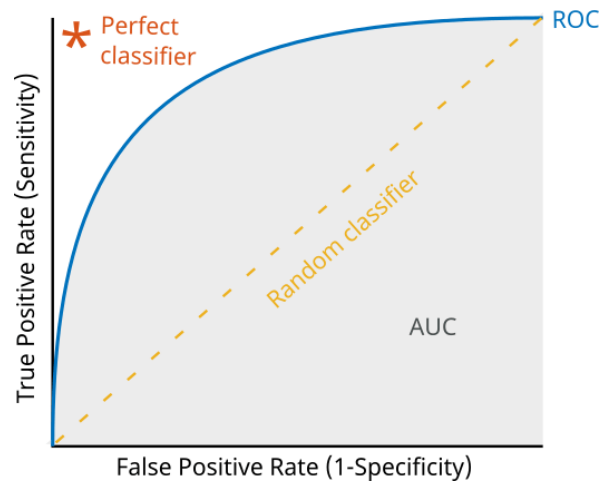


Figure 5.3.: Visualisation of a receiver operating characteristic curve (ROC curve) in the case of a random classifier (yellow, dashed line), a realistic example (blue, line) and a theoretical perfect classifier (red star). The area under curve (AUC) is shown in grey [111].

TPR, for example, is defined as the ratio of signal events which have a score above the threshold  $t$  and are therefore classified correctly. All other signal events that are assigned a score below the threshold and are therefore falsely classified as background contribute to the FNR. All rates have a range of 0.0 to 1.0.

These rates can be used to define the *receiver operating characteristic curve (ROC curve)* [110]. The ROC curve is given by  $\text{tpr}(t)$  as a function of  $\text{fpr}(t)$ , where the threshold  $t$  is varied over its full range from 0 to 1. An example of a ROC curve can be seen in Figure 5.3. In the case of a random classifier the TPR and FPR are expected to be exactly the same for all  $t$  (see yellow, dashed line in Figure 5.3). This comes from the fact that a random classifier is not able to distinguish between signal and background and will therefore treat both of them equally. A theoretical perfect classifier would assign all signal events a score of 1 and all background events 0. Independently of the chosen threshold this would lead to  $\text{tpr}(t) = 1$  and  $\text{fpr}(t) = 0$ . A realistic example of a classifier would lead to a ROC curve similar to the blue line in Figure 5.3. The bigger the deviation of ROC curve from a straight, diagonal line, the better is the network's performance. This behaviour is quantified using the *area under curve (AUC)* of the ROC curve, which is simply defined as the integral of the ROC curve over the full range from 0 to 1. This is marked in the diagram as grey area. A perfect classifier would lead to  $\text{AUC} = 1$  and a random one to  $\text{AUC} = 1/2$ . The closer the value for the AUC is to 1, the better the classification.

Another metric can be the *maximal classification efficiency*  $\epsilon_{\max}$  that can be achieved by applying a threshold  $t$ :

$$\epsilon_{\max} = \max(\epsilon(t)) = \max(\text{tnr}^2(t) + \text{tpr}^2(t)). \quad (5.11)$$

This quantifies the maximal ratio of all events that can be classified correctly.  $\epsilon_{\max}$  contains information on the separation of signal and background, as it is only possible to find a threshold that allows for a high efficiency, if they are assigned very different score values. Additionally, it can be used to determine a final threshold  $t$  that maximises the efficiency, if one actually wants to perform a selection of events depending

on their score.

The introduced rates can also be used to calculate the maximum of *expected signal significance improvement*  $\sigma_{\max}$  that can be obtained by using a threshold  $t$ . The signal significance  $Z$  can be approximated, using the number of signal  $N_{\text{signal}}$  and background events  $N_{\text{background}}$  as:

$$Z = \frac{N_{\text{signal}}}{\sqrt{N_{\text{background}}}}. \quad (5.12)$$

It is important to note that this formula is an approximation assuming an underlying Poisson statistic, that is only valid for high numbers of background events  $N_{\text{background}}$  and neglects any systematic uncertainties. If the investigated dataset has a significance of  $Z_{\text{initial}}$  before any kind of selection is performed, only looking at events with scores above a chosen threshold  $t$  can change this significance by a factor:

$$\sigma(t) = \frac{Z(t)}{Z_{\text{initial}}} = \frac{\int_t^1 dx s(x) / \sqrt{\int_t^1 dx b(x)}}{\int_0^1 dx s(x) / \sqrt{\int_0^1 dx b(x)}} = \frac{\text{tpr}(t)}{\sqrt{\text{fpr}(t)}}. \quad (5.13)$$

It is possible to use  $\max \sigma_t$  as metric in comparing different NNs or, if wanted, to select the value for score cut.

## 5.7. Optimisation of Hyperparameters

The performance of a NN depends strongly on the quality of the training. Other very important factors are the network's hyperparameters, which define general features of the network. These include the architecture of a network, the chosen learning rate, batch size, but also parameters of regularisation methods. All of these strongly influence the ability of a network to be well-trained.

All hyperparameters need to be chosen before the training and are then fixed throughout. The most obvious method of optimising the choice for  $\vec{\theta}$  is trying out different combinations by hand and comparing them. For the evaluation of different networks the metrics, discussed in the preceding section, can be used.

One problem with this method is, however, that the person performing this optimisation relies on their personal experience and understanding of machine learning algorithms and the interplay between hyperparameters and the resulting performance. Additionally, it gets increasingly hard if a bad starting point of hyperparameters is used. Usually there are many hyperparameters which are not independent of each other and need to be optimised simultaneously. Therefore, this method is very inefficient, time-consuming and likely to not yield the best results, especially if performed by relatively inexperienced users [112].

Automated optimisation algorithms are an alternative that can drastically reduce the work that needs to be done by hand. It also means the experience of a user is of minor importance. One disadvantage of these algorithms is that they are likely to need higher computational resources, because they can test more  $\vec{\theta}$  configurations. There is a big variety of different algorithms that can be used for hyperparameter optimisation. There is *grid* and *random search* [113], as well as the *Bayesian approach* [114] to name

a few. All of these use different methods to choose the hyperparameter configurations to test, but they all use these configurations throughout the entirety of one training, each<sup>b</sup>.

Grid and random search are relatively simple algorithms. The way they choose hyperparameters to test makes no assumption on the optimisation problem itself. In the case of grid search, a set of values to test is defined for each hyperparameter beforehand. The optimisation simply performs trials that represent possible combinations of the different parameters and compares them. Random search chooses each value randomly, based on marginal distributions for all  $\theta_i$  (usually uniform or logarithmic distributions are chosen) [112].

Often these methods are outperformed by more sophisticated techniques like the Bayesian approach. The basic idea of this algorithm is to use the results from earlier trials to estimate a probabilistic model  $p(y|\vec{\theta})$  that describes the probability to get a validation loss  $y$  if one trains a network with the hyperparameters  $\vec{\theta}$ . This distribution  $p(y|\vec{\theta})$  can be used to choose the most promising hyperparameter combination, that is expected to minimise  $y$ , for the next trial. Over the course of the optimisation this estimate for  $p(y|\vec{\theta})$  becomes more accurate as it is based on more data-points.

There are multiple ways of estimating  $p(y|\vec{\theta})$ . One is called the *Tree Parzen Estimator (TPE)* [115]. It is based on Bayes theorem, as in fact  $p(\vec{\theta}|y)$  and  $p(y)$  are modelled and then used to calculate  $p(y|\vec{\theta})$ .

The *Asynchronous Successive Halving Algorithm (ASHA)* [116] can be used to save computing resources and is compatible with different optimisation strategies, which usually rely on performing a full training for all trials. It is based on the idea to perform multiple trials at the same time and regularly check their training progress. Poorly performing trials can thereby be identified and aborted.

One special kind of hyperparameter optimisation is the selection of input variables. This requires already roughly optimised hyperparameters to get a meaningful result. The *backwards elimination approach* starts with a fully trained NN that uses all possibly interesting input parameters. Iteratively, variables are removed until a certain condition is met, for example, a certain relative change in the validation loss.

The naive approach would be to retrain the whole NN for different combinations of variables. If one tries to select the variables of a deep neural network that uses a lot of resources during training, another approach, the *shuffle* method, can be beneficial. In this case, one can choose the variable to investigate and shuffle all its values and assign them to other events than they originally came from, which ensures that the distribution of this variable is kept constant. All other input variables are kept the same. It is expected that in all cases the loss should worsen. If the variable is important the increase in loss should be a more drastic change than if the variable is barely used by the network.

---

<sup>b</sup>in contrast to other automated hyperparameter optimisation algorithms like population based training

---

# 6. Multivariate Extension to the Base Analysis

The following section will describe the incorporation of a neural network that acts as a classifier into the base analysis that was described in Chapter 4. It is interesting to investigate the possible improvement this NN can provide over the standard cut-based approach. The chosen network will be characterised, including its training data and the input variables that are used, as well as its architecture. After a summary of the results using the NN compared to the cut-based analysis, the section will close with some additional studies that were performed.

The framework used for the definition, training and application of this NN relies on the Keras library [117], which is based on TensorFlow [118]. It provides many machine-learning functionalities and is written in Python. The source code can be found at [119].

## 6.1. Training Data and Event Weights

The training is based on events from the mvaVR (see Table 4.3). It is a looser version of the signal regions and requires only the most important cuts. This is a compromise between having a larger number of available training data and including a significant fraction of signal events. In the base analysis a cut on truth information is applied to all MC events to make sure that only fake events produced by the MM are taken into account. It was decided to not use this cut for the selection of training events. By doing so, the impact from data on the training can be reduced.

It was decided to use input data that is representative for the physical composition of different types of background in the chosen region. This means in particular, that all available background and signal MC events are used for training. It is common practice to use MC generated events as training data, because they naturally provide information on the underlying process, which makes it easy to classify them into signal or background. Data cannot be used as easily for the training as there is no labelling possible in this case. Additionally, this could result in unblinding the analysis, which is not allowed yet.

As discussed in Section 4.4, it is not possible to model the background from events that include fake leptons using MC generators. They are, however, a relevant source of background in the mvaVR as they contribute almost half of the total background events here. In the base analysis fake events are estimated using the Matrix Method. It is not yet possible to include the elaborate algorithm of the MM into the framework that is used to train the NN. Nevertheless, it is important to present fake events to the network during its training. For this reason, a not so elegant way of estimating fake events is applied. Because MC events are not an option, it was decided to use data events from a special region, which is expected to contain almost only fake events.



The region used here is a `not-Tight` version of the `mvaVR`. This means all data events in the `mvaVR`, that pass the baseline selection, but fail some `Tight` criteria, are included as fake events. More specifically, the muon is required to fail the `Tight` isolation, while a softer condition is applied to electrons. They are required to fail at least one of the `Tight` isolation or identification criteria. This is done to get better statistics. The `not-Tight` region is expected to be strongly dominated by fake events. This is supported as prompt MC events can only explain 0.57% of all data events in this region. Because of the change in isolation and identification criteria, this can be considered as a validation region and one can look at data without having to worry about unblinding. The number of data events in the described region is very different to the weighted number of fake events in the `mvaVR`, this can be seen in Table 6.1. As the goal is to reproduce the distribution in the `mvaVR`, event weights need to be applied to the data-events from the `not-Tight` region, to get a similar contribution. The weighting strategy will be described later on in this section.

As the NN is trained using supervised learning, all training samples need to be labelled. The signal MC events get assigned the value 1.0, while the background events are marked with a 0.0. This means that the final score that the NN will apply to an event can be interpreted as the probability that it comes from a signal process.

The final number of events that can be used as training input are shown in Table 6.1. As mentioned earlier, the goal is to model the weighted distribution of events in the `mvaVR` using the available training data. The weighted contribution of different background sources in this region are listed in the same table.

It is clear that this composition cannot be modelled using the number of events per background alone. One way to change the impact a single background has on the training, is using event weights. These influence the weighted average loss during gradient descent as follows:

$$L = \sum_i^{n_{\text{events}}} w_i \cdot L_i \quad (6.1)$$

with the individual event weights  $w_i$  and the loss of a single event  $L_i$ .

For the MC events, the weights are calculated to resemble the ones that are applied during the base analysis. There are contributions from the process' cross section, luminosity, filter efficiency, so-called k-factor and a general MC weight. The latter can be further split into terms that represent the actual MC, pile-up and jet tagging weights (jet vertex tagging, forward jet vertex tagging and b-tagging). In the case of calculating the weight of the signal, one also needs to consider the dependence of the cross section on the theory parameter  $\zeta_u$  and apply  $\times \zeta_u^2$  to the final weight.

In the base analysis there are additional scale factors that are also applied as individual weights to all MC events. They are used to do small corrections, for example, for different isolation/identification working points or triggers. As all scale factors are close to one for almost all events, they only have a minor impact on the composition of background and are therefore neglected here. This weighting of all samples has one disadvantage. Pure MC weights are used, amongst others, to express positive and negative interference between different contributions in the Lagrangian. Because of this, they can have negative values. This might become a problem, as training a NN is a minimisation problem. These negative terms might therefore lead to the network learning the opposite of what it is supposed to learn, if their influence is too big. A study was performed, comparing this and alternative ways of weighting the training

Table 6.1.: Unweighted number of events per process that are used as training data, as well as their contribution to the mvaVR. The "Other MC" process includes:  $W \rightarrow l\nu$ ,  $Z/\gamma^* \rightarrow ee/\mu\mu$ , SM Higgs. The number of fake training events is given by the number of data events in the not-Tight mvaVR. The (weighted) contribution of different processes to the mvaVR is extracted using the MM, which also includes the cut on truth information for all MC events.

Process	Number of Events	Contribution in mvaVR
$Z \rightarrow \tau\tau$	194 181	10 096.8
Top	19 095	2218.6
Diboson	184 487	1533.9
Other MC	237 059	1069.2
Fakes	26 329	10 316.2
$A \rightarrow \tau\tau$ , $m_A = 20$ GeV	3741	3794.9
$A \rightarrow \tau\tau$ , $m_A = 30$ GeV	4385	5480.6
$A \rightarrow \tau\tau$ , $m_A = 40$ GeV	300	5652.9
$A \rightarrow \tau\tau$ , $m_A = 50$ GeV	583	5806.9
$A \rightarrow \tau\tau$ , $m_A = 60$ GeV	630	4912.0
$A \rightarrow \tau\tau$ , $m_A = 70$ GeV	635	3712.3
$A \rightarrow \tau\tau$ , $m_A = 80$ GeV	626	2996.1
$A \rightarrow \tau\tau$ , $m_A = 90$ GeV	538	2081.9
$A \rightarrow \tau\tau$ , $m_A = 100$ GeV	1022	1488.4
$A \rightarrow \tau\tau$ , $m_A = 110$ GeV	855	1085.8

events. It can be found in Subsection 6.8.1.

The weights for data from the not-Tight mvaVR that is used to represent fake events are calculated differently. For simplicity, it was decided to calculate a single weight  $w_{\text{fake}}$  that is applied to all of these data events. The goal of the weighting strategy is to represent the distribution of background in the normal mvaVR.

It is important to consider that the truth information cut on MC, that is applied in the base analysis in the context of the MM, is not used for the selection of training data. This means that the event number of fake events in the base analysis cannot be used directly as a reference for the calculation of the fake weight  $w_{\text{fake}}$ . The goal is to conserve the total number of weighted events in the mvaVR based on the MM  $N_{\text{tot,MM}}$ . The final contribution of fake events should therefore be equal to the difference between the total number in the base analysis minus the weighted contribution of all MC events, disregarding their truth information  $N_{\text{MC,X}}^a$ . The final weight  $w_{\text{fake}}$  can be calculated by dividing this reference value by the number of data events in the not-Tight mvaVR.

<sup>a</sup>The index  $X$  denotes, that both real and fake events are considered, analogous to the notation in Section 4.4.

This can be expressed as:

$$w_{\text{fake}} = \frac{N_{\text{tot,MM}} - N_{\text{MC,X}}}{\#(\text{data}_{\text{not-Tight}})}. \quad (6.2)$$

The weight was calculated to be  $w_{\text{fake}} = 0.33$ .

One last adjustment is performed on the weights of signal events. Its goal is to balance the network's performance regarding the different signal hypotheses. This means the weighted number of signal events should be equal for all mass points. This can be achieved by scaling all weights of signal events. The factor that needs to be used only depends on the mass hypothesis of an event, not the event itself. It is given by the mean weighted contribution of all signal events divided by the contribution of signal events of the given mass hypothesis. The new weight  $\tilde{w}_i$  of a signal event  $i$  can be defined using original weights  $w_j$  of a signal events, the number of mass points  $n_{\text{signal}}$  (here =10):

$$\tilde{w}_i(m) = w_i \cdot \frac{\sum_j^{\text{all signal}} w_j}{n_{\text{signal}}} \cdot \frac{1}{\sum_k^{\text{signals of mass } m} w_k}, \quad (6.3)$$

$n_{\text{signal}}$  represents the number of different signal hypotheses. This scaling ensures that the overall contribution of signal events does not change, as well as the individual distributions within one mass point.

Unfortunately it was not yet possible to implement the  $Z \rightarrow \tau\tau$  reweighting, that is introduced in Section 4.5, in the NN framework. This should be done in the future, as it is advisory to use only well modelled distributions for the training of the NN.

## 6.2. Hyperparameter Optimisation

As described in Chapter 5, a NN's ability to perform well during and after training strongly depends on a range of general features, the hyperparameters. A goal of this thesis was to improve the choices for these. Section 5.7 introduces some methods for performing such a hyperparameter optimisation.

In this thesis a framework developed by Erik Bachmann [120] is used to perform an automatised hyperparameter optimisation for the here presented NN. This framework relies on the Tune library [121] and is able to perform hyperparameter optimisations for NNs that are based on Keras and implement supervised learning. It was used here to improve the input variable selection, as well as standard hyperparameters like the NN's architecture. The framework's fine-tuning option using population based training was not used here.

The training data that is used for the optimisation and its preparation are exactly the same as for the normal NN. This includes the same event weights and the same preparation of all input variables (see following section), including the addition of the mass parameter and the transformation of all input variables.

The standard hyperparameter optimisation was performed using the Tree Parzen Estimator in combination with the Successive Halving Algorithm. It was decided to try a total of 1000 different hyperparameter combinations and assess them by their weighted binary cross entropy from the validation dataset. The framework was used to optimise the following hyperparameters: number of layers, number of neurons per

layer, batch-size and learning rate. The final most optimised network was chosen based on the "best value" of the weighted validation loss, and not the "best fit" results.

For the selection of input variables the *shuffle* method was used. This procedure is described in Section 5.7. It was used twice with a threshold for minimum relative change in validation loss of 2% and another time with 5%. The results of these two optimisations as well as the original variable selection are described in more detail in the following section. A complete config-file that is needed for the optimisation can be found in the Section D.1.

## 6.3. Input Variables

In Table 6.2 a complete list of all variables, that were considered as input for the NN, is given. They are chosen in a way to contain as much information on the kinematics of each event and combine low-level variables, like  $p_T^e$ , with high-level ones, like  $m_{\text{MMC}}$ .

This broad selection was optimised by investigating the relative change of loss if the values for one variable are shuffled between the events (as described in Section 6.2). A larger change of loss means that the variable has a bigger influence on the output of the NN. One can therefore select the meaningful input variables by defining a threshold of minimum relative change in validation loss. This was performed twice with different thresholds of 2% and 5%. Table 6.2 lists the relative change in validation loss for all variables.

For each of the two variable selections, a further hyperparameter optimisation and training of a NN was performed and their results compared. The validation loss of the network that was trained with a stricter selection of variables showed a larger fluctuation in the validation loss over the course of the training between different k-folds. It was therefore decided to use the milder variable selection, based on a threshold of 2%, for this thesis. The distributions of all final input variables in the mvaVR can be found in Section D.2. They are shown once in the mvaVR, which is used to select the training data, as well as in the ZVR, which can be used to check the variable's modelling.

The modelling of the input variables is important, as this will influence the NN's ability to learn the right relations between different input variables from the simulated events. Only if all variables are well modelled in MC, the NN will be able to interpret data events correctly.

An important step in the context of a parameterised NN is the definition of the additional variable. In this case the difference between signal hypotheses is their mass, because of that, all events are assigned a parameter  $m_A$ . Because high values might lead to numerical instabilities during training it was decided to use  $m'_A = m_A/100$  GeV which leads to a range of  $m'_A \in \{0.2, 0.3, \dots, 1.1\}$ . As described in Section 5.5, signal events will always be assigned the correct  $m_A$  during training, while background events will get a random one out of the possible values.

It is beneficial for the training of the network, when all input parameters have a similar range in their values. Otherwise, it might occur that input parameters with very small values are ignored by the network as they have a small influence on the outputs of neurons.

Table 6.2.: Complete list of all variables investigated as input for the NN. There is a short description for each available and they are classified by the relative change of validation loss  $\Delta L$  that was observed during input selection optimisation (see Section 6.2). The transverse impact parameter significance variable is defined as  $d_0\text{Sig} = d_0/\sigma(d_0)$ .

Variable	$\Delta L$	Description
$p_T^e$	$> 5\%$	transverse momentum of the electron
$p_T^{\text{jet}}$	$> 5\%$	transverse momentum of the leading jet
$p_T^{\text{Higgs}}$	$> 5\%$	transverse momentum of the mother particle
$E_T^{\text{miss}}$	$> 5\%$	missing transverse energy
$m_{\text{MMC}}$	$> 5\%$	mass reconstructed by the Missing Mass Calculator
$m_T^{\text{tot}}$	$> 5\%$	total transverse mass
$m_{ll}$	$> 5\%$	dilepton mass
$m_{\text{jet,jet}}$	$> 5\%$	dijet mass, analogous to the dilepton mass
$\Delta R_{ll}$	$> 5\%$	angular distance between the two leptons
$n_{\text{jets}}$	$> 5\%$	number of jets
$\Delta\phi_{ll}$	$> 5\%$	azimuthal distance between the leptons
$ \Delta z_0^e \cdot \sin\theta^e $	$> 5\%$	longitudinal impact parameter of the electron
$ \Delta z_0^\mu \cdot \sin\theta^\mu $	$> 5\%$	longitudinal impact parameter of the muon
$p_T^\mu$	$> 2\%, < 5\%$	transverse momentum of the muon
$\eta^e$	$> 2\%, < 5\%$	pseudo-rapidity of the electron
$\eta^\mu$	$> 2\%, < 5\%$	pseudo-rapidity of the muon
$\eta^{\text{jet}}$	$> 2\%, < 5\%$	pseudo-rapidity of the leading jet
$\Delta\phi_{\text{jet1}, E_T^{\text{miss}}}$	$> 2\%, < 5\%$	azimuthal distance between leading jet and $E_T^{\text{miss}}$
$\Delta\phi_{\text{lep2}, E_T^{\text{miss}}}$	$> 2\%, < 5\%$	azimuthal distance between subleading lepton and $E_T^{\text{miss}}$
$\phi^e$	$< 2\%$	azimuthal angle of the electron
$\phi^\mu$	$< 2\%$	azimuthal angle of the muon
$\phi^{\text{jet}}$	$< 2\%$	azimuthal angle of the leading jet
$\Delta\phi_{\text{lep1}, E_T^{\text{miss}}}$	$< 2\%$	azimuthal distance between leading lepton and $E_T^{\text{miss}}$
$d_0^e\text{Sig}$	$< 2\%$	transverse impact parameter significance of the electron
$d_0^\mu\text{Sig}$	$< 2\%$	transverse impact parameter significance of the muon

To prevent this, all input parameters  $p$  are transformed in the same way, using their mean value  $\bar{p}_{\text{train}}$  and their standard deviation  $\sigma_{p,\text{train}}$  over all training data:

$$p' = \frac{p - \bar{p}_{\text{train}}}{\sigma_{p,\text{train}}}. \quad (6.4)$$

This ensures, per definition, that all input parameters are distributed around 0 with a standard deviation of 1.

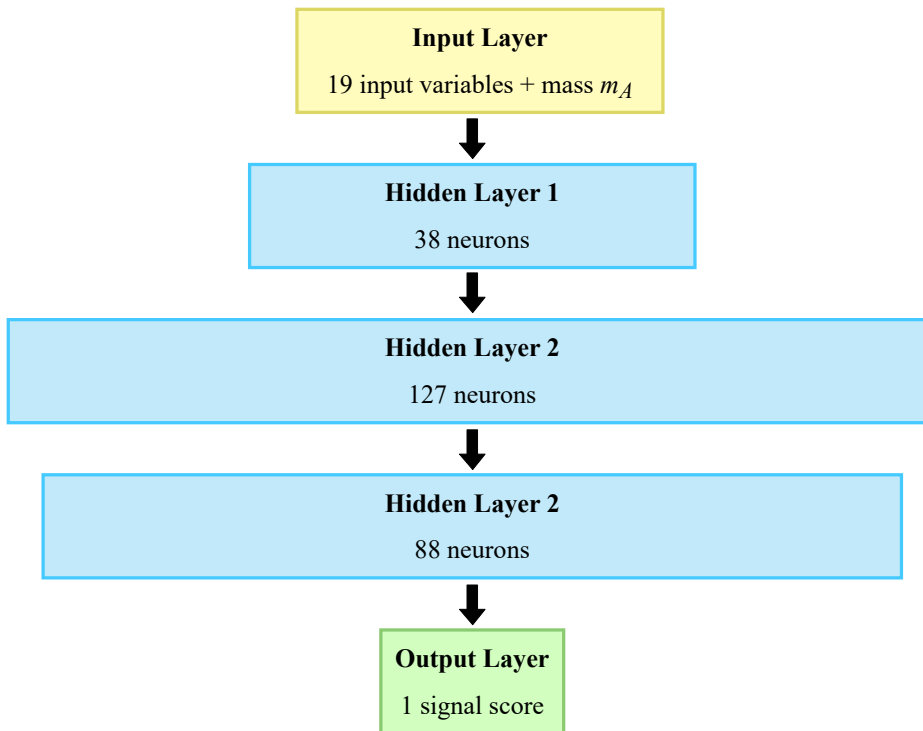


Figure 6.1.: Scheme of the final NN's architecture.

## 6.4. Architecture of the Neural Network

The neural network that is used here to classify events as signal or background is feed-forward and its layers are fully connected. Its final architecture was determined using the hyperparameter optimisation described in Section 6.2. The input layer consists of 20 neurons, one per input parameter (19 standard input variables, see Table 6.2 plus additional  $m_A$  parameter). This is followed by three hidden layers made up by 38, 127 and 88 neurons. Lastly, one output layer consisting of exactly one neuron calculates the final score. This architecture is shown in Figure 6.1. All neurons in the hidden layer, as well as the output layer, use a logistic sigmoid function (Equation 5.3) as activation function.

When comparing the few best trials (based on weighted validation loss) of the hyperparameter optimisation, it was clear that using three hidden layers is strongly preferred. On the other hand, the specific number of neurons per layer was relatively flexible as long as they stayed in the same order of magnitude, the change in validation loss was only marginal.

## 6.5. Training of the Neural Network

The first step of the training is to read in and preprocess all labelled data. The framework at hand can load data from given NTuples and saves only those events, which pass the mvaVR requirements (or the not-Tight mvaVR conditions in the case of data for fake events), into NumPy arrays.

All of these data points are then grouped in datasets according to the k-fold method (see Section 5.4). For this thesis it was decided to use a 5-fold, with five separate

datasets. The split is done based on `EventNr`, a unique identifier for all simulated and measured events. In the following, five independent NNs will be trained. They share the same hyperparameters and only differ in their data split. The training data will always consist of three out of the five available datasets. This leaves exactly one for the validation and application. The mapping, which of the five datasets has what role, is cyclically permuted between the different NN, as shown in Table 5.1.

During the definition of the NN architecture, it is necessary to initialise values for the weights and biases that define the network's behaviour. These initial values are then the starting point for the training. It was decided to use the default Keras initialisers. In the case of weights, this is the *Glorot uniform initialiser* [122], while all biases are simply set to zero. The uniform Glorot algorithm chooses initial weights for a layer  $j$  from a uniform distribution between  $[-x_j, x_j]$ , with the borders defined as  $x_j = \sqrt{\frac{6}{n_j + n_{j-1}}}$ , where  $n_i$  denotes the number of neurons in layer  $i$ .

The NN is designed to perform mini-batch gradient descent. The batch size was determined using the hyperparameter optimisation. For this thesis, a size of 226 was used. This batch size is too small to consider at least one event of each single signal process in each mini-batch. This comes from the very different numbers of events for the various processes (see Table 6.1). The total number of signal events is large enough to ensure that each mini-batch should contain a mean of roughly four signal events.

The hyperparameter optimisation was also applied to the learning rate. Unfortunately the optimisation framework is only able to investigate constant learning rates. The result for this was  $\eta = 1.685 \times 10^{-4}$ . In the past it was observed that using more advanced learning rate schedulers leads to superior results than constant learning rates [10]. Therefore, a similar procedure was chosen. A learning rate scheduler is used to implement a cyclical learning rate that decreases like a cosine over the course of  $t_0 = 20$  epochs and then is set back up to a starting point that is exponentially decreased with a half-time of  $t_{1/2} = 100$  epochs. The initial maximum and minimum of the cosine is given by  $\eta_{\min/\max} = 0.0001/0.001$ . The full definition depending on the training's epoch  $t$  is given by:

$$\eta(t) = e^{-\ln 2 \frac{t}{t_{1/2}}} \cdot \left[ \frac{\eta_{\max} - \eta_{\min}}{2} \cdot \cos\left(\frac{\pi}{t_0 - 1} \cdot (t \bmod t_0 - 1)\right) + \frac{\eta_{\max} + \eta_{\min}}{2} \right]. \quad (6.5)$$

Figure 6.2 shows the change in learning rate over the course of the full training of 50 epochs. To further facilitate good training, the ADAM optimiser is applied using the standard values proposed by the authors, that can be found in Section 5.3.

Each of the five NN is then trained independently. It was decided to use weighted binary cross entropy as loss function. "Weighted" refers here to the use of event weights which were described in Section 6.1. This loss function is well suited for classification problems. The training and validation loss is calculated after every epoch of the training and can be seen in Figure 6.3. As the loss needs to be calculated for each NN from the k-fold method individually, the diagram shows the total range of loss as a transparently coloured area and the mean value over the different NNs as a strong line.

The decision for a total of 50 epochs of training is based on the development of training and validation loss in an analogous training over the course of 100 epochs. In the region above 50 epochs overfitting could be observed, as the training loss kept decreasing, while the validation loss was stagnant. This behaviour is discussed in more

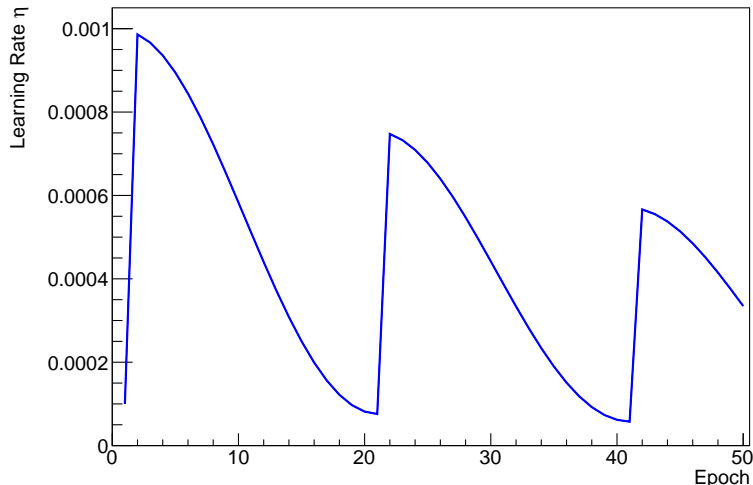


Figure 6.2.: Cyclical change in the learning rate over the course of the training, defined by an exponentially decreasing half-cosine learning rate scheduler as defined in Equation 6.5.

detail in the Section D.3. In the future, more sophisticated approaches could be used to prevent overfitting, like implementing early stopping conditions directly in the Keras model.

## 6.6. Application of the Neural Network

After the training is finished, the NNs can be applied to events to get their final score. As described in Section 5.5, each event gets assigned a total of ten scores, representing a probability that this event originates from one of the considered signal mass hypotheses. It is expected that all NNs that were trained in the context of the k-fold technique, produce slightly different scores for the same event. The final goal, however, is to only have one final score. This is done by always using just one of the scores from the different NNs for an event. Each MC generated event was part of the application dataset of only one of the NNs. That means only this specific network is totally unbiased regarding this event, as it was not used during its training or for validation. For this reason the unbiased NN is chosen for the application of scores for MC events.

In the case of events that none of the NNs has seen during training, like data events that are not in the not-Tight mvaVR, another approach needs to be defined. In earlier versions of this analysis, the mean of all scores from the five networks was used. This, however, lead to a strong mismodelling in the distribution of scores in the ZVR or TVR, between background and data. It was possible to solve this problem, by calculating the final scores for data more similar to the ones for MC events. For each event, that was not used during training, a random NN out of the five available ones is chosen and its score for the event is used.

In practice, the final application is done by looping over all events of input NTuples and saving a full copy of the original NTuples supplemented with ten additional score variables (one per mass hypothesis), that contain the calculated scores for each event.



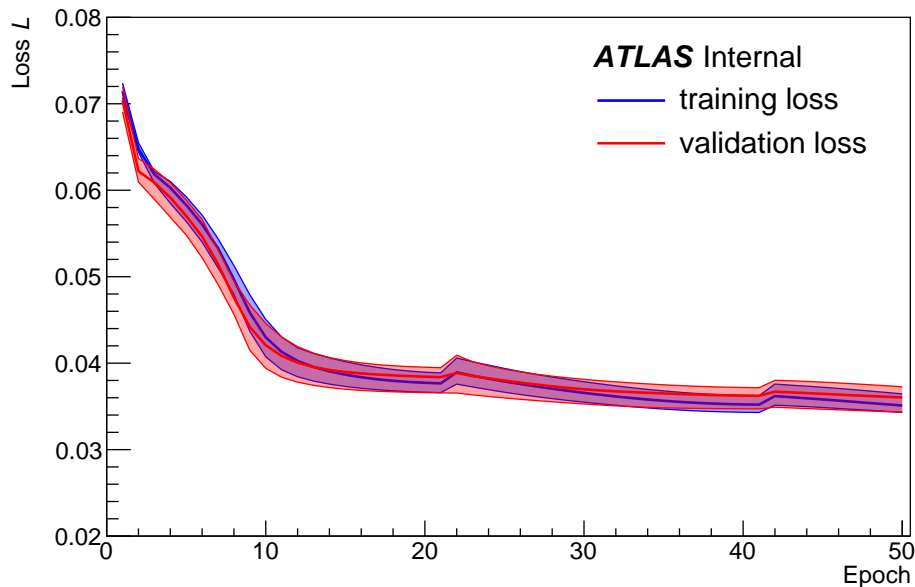


Figure 6.3.: Diagram showing the development of training (blue) and validation (red) loss over the course of the whole training. Weighted binary cross entropy is chosen as loss function. The transparent areas denote the full range of loss values over the different NNs from the k-fold technique and the line shows their mean value for each epoch.

Ideally this is done for MC simulated events as well as for data.

Saving the score variables in NTuples makes it possible to investigate their distributions, as well as comparing them in different regions or between data and background, using the framework that is used in the base analysis. One downside to it is the relatively high amount of computational resources necessary for this step. Therefore in the early stages of comparing the performance of different NNs one has to rely on the metrics determined during the training, instead of ones utilizing score distributions.

These include the validation and training loss, as well as metrics, like the AUC, that are defined in the `model.compile()` function of Keras. It is also possible here to get weighted metrics that use the training event weights. This analysis investigates a PNN, which means that the metrics calculated during training will be averaged over all mass hypotheses. As the individual scores for the different mass points are only calculated during application, the performance of a network regarding individual  $m_A$  can only be investigated after this step. Using the score distributions that can be calculated after application of the NN, the AUC, maximal efficiency and maximal signal significance improvement (see Section 5.6 for definitions) can be determined individually for each mass point. This also has the advantage of not using the same weights as during the training but the event weights from the base analysis which makes it possible to compare the results from NNs that use different weighting strategies (like in the study described in Subsection 6.8.1).

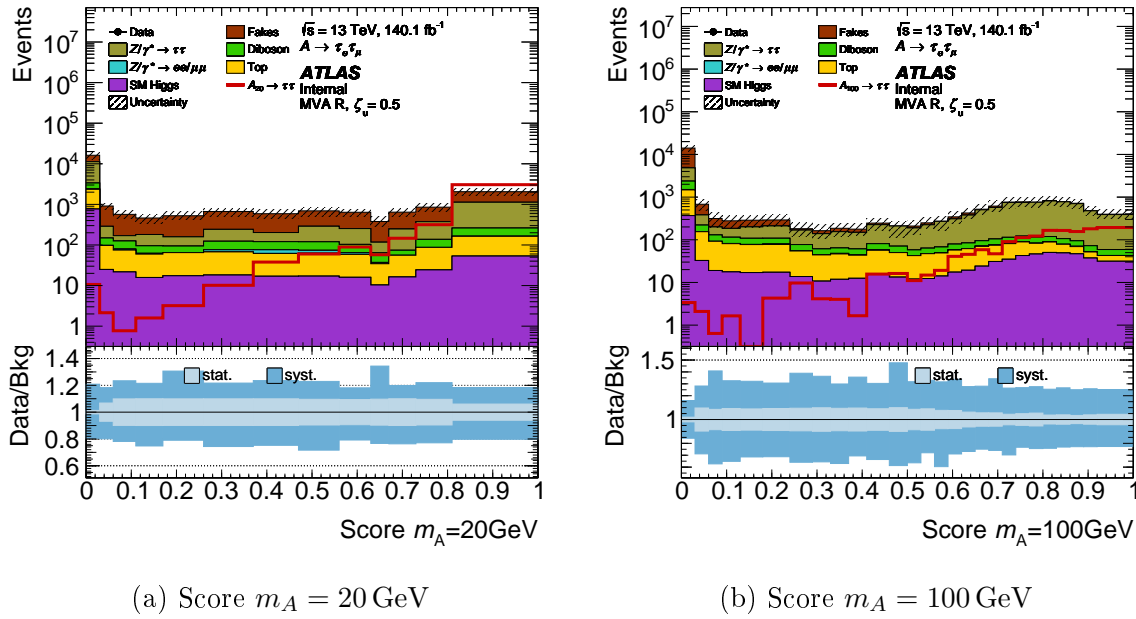


Figure 6.4.: Distribution of score variables for  $m_A = 20$  GeV (left) and  $m_A = 100$  GeV (right) in the mvaVR that were calculated by the neural network described in Chapter 6. The red line represents the distribution of the corresponding signal events. The fake background is estimated using the standard Matrix Method (see Section 4.4) and the  $Z \rightarrow \tau\tau$  reweighting is applied. The binning displayed here, is used as input for the calculation of limits.

## 6.7. Scores and Limits

**Score Distributions** After the final NN is applied to all available data and MC samples it is possible to look at the distributions of newly added score variables. Producing plots of the score variables, is done analogously to the procedure described for the base analysis. This means showing the weighted distribution of all MC simulated and data events, and using the MM for the estimation of the fake background. The final distributions of score  $m_A = 20$  GeV and  $m_A = 100$  GeV in the mvaVR are shown in Figure 6.4. The distributions for other mass hypotheses in the mvaVR, TVR and ZVR can be found in Section D.4. All score distributions in the mvaVR show a relatively similar behaviour in the equidistant binning used in the appendix. There is a strong maximum of background events close to score = 0. This means that the NN correctly assigns a very low score to many of the background events. Starting in the next bin the background contribution stays roughly constant, with a local maximum around a score of 0.8 for higher signal mass scores. This maximum is especially pronounced in the  $Z \rightarrow \tau\tau$  background and is probably related to the strong similarities between higher signal hypotheses and the  $Z \rightarrow \tau\tau$  process, as described in Subsection 4.2.3. The signal distribution also behaves as expected, with only a few ones getting assigned a small score and increasing numbers of events for higher scores. Especially for lower mass hypotheses up to 70 GeV the signal distribution has a maximum in the highest possible bin (in the relatively broad binning used in the appendix) while for higher masses the most frequent score is somewhere between 0.8 and 0.9.

The distributions in the validation regions are allowed to contain data events. There-

fore, in these regions the modelling of the score variables can be checked. The general shapes of score distributions in the TVR is very similar to the ones in the mvaVR, with a large peak of background events close to score = 0 and a maximum of signal events for high values. The modelling between data and background looks quite good with minor fluctuations for all score distributions. In none of the data-to-background ratios a systematic deviation can be observed. The distributions in the ZVR show a slightly different behaviour. This can probably be explained by the fact, that the main difference between mvaVR and ZVR lies in a cut on the  $\Delta R_{ll}$  variable, which is also an input variable to the NN. This means that the NN is solely presented with events that have a value in one input variable that it has never seen during training in the mvaVR. The  $\Delta R_{ll}$  variable seems to be important for the classification of low masses, as the majority of all events in this case, gets assigned very low scores. In the case of very high mass scores, the distributions still show a maximum in background events close to 0, but it is much less pronounced. In general, the network treats signal and background events very similar in the ZVR. Nevertheless, the distributions in the ZVR still show relatively good data-to-background agreement, especially for scores of  $m_A > 60$  GeV.

**Expected Limits** In the following, the score distributions are used as input for the calculation of cross section times branching ratio limits, instead of  $m_{\text{MMC}}$  as it is done in the base analysis. The score distributions are expected to have a good separation power between background and data and are therefore well-suited for this application. To allow this, the statistics framework was slightly adapted to allow for individual input distributions for all mass hypotheses. The binning of the score variables should be chosen carefully, as they strongly influence the final limit. All bins must contain a minimum of ten events, so that all assumptions of the limit calculation framework are fulfilled. Additionally, each bin is required to have a minimum size of 0.03 to reduce the influence from bin migration. Inside these bounds, the bin size is determined by restricting the statistical uncertainty of background events in each bin to be below 10 %.

For the calculation of the expected limit of a given mass point, the distribution of the corresponding  $m_A$  score in the mvaVR is used. This variable is expected to provide the best separation of background and the specific signal. A small study comparing score distributions in the two SRs instead is described in Section E.2. Using all events from the mvaVR proved to give better results.

Figure 6.5 shows the final limit calculated as described. This diagram also contains the limit of the base analysis. Before directly comparing the two limits with each other, it should be mentioned, that the limit strongly depends on the binning of the input distributions. As the two limits are based on completely different variables, with slightly varied binning strategies (regarding the minimum bin size) and are calculated in different regions with different number of events, all direct comparisons should be done with caution. A comparison of two limits calculated using the same input distributions, but with different bin choices can be found in Figure E.1 in the appendix. It is visible that the calculated limit might change by a factor of almost 3, by only changing the bin edges of the input distributions.

The comparison of the two limits in Figure 6.5 shows that the score distributions of the final NN are able to produce a limit that is in the same order of magnitude,

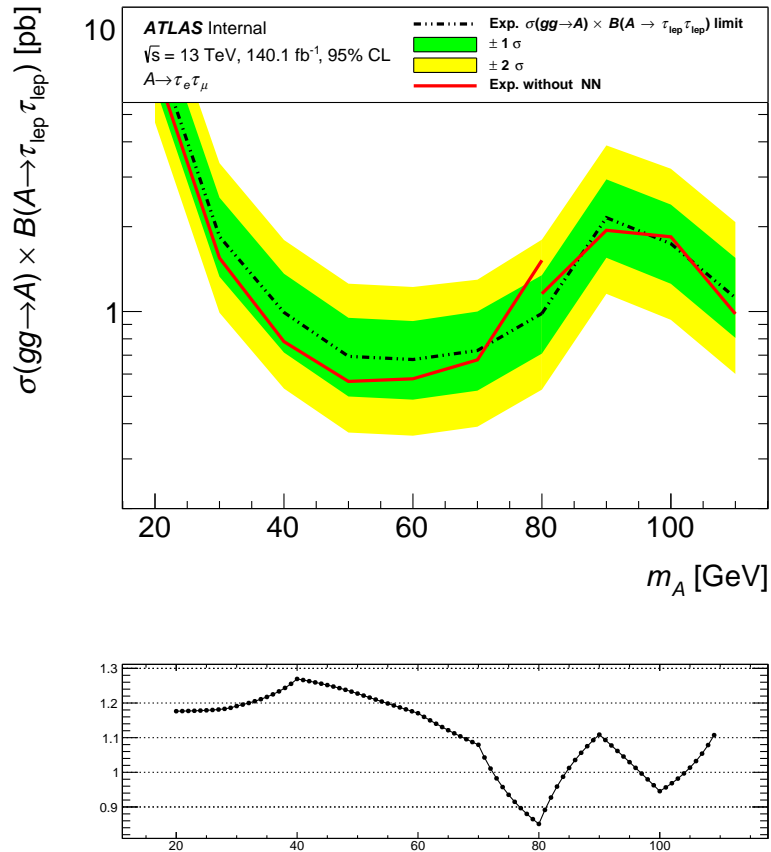


Figure 6.5.: Expected limit on the production cross section for  $gg \rightarrow A$  times the branching ratio for  $A$  decaying into two leptonic  $\tau$ -leptons. The results from the base analysis (red) can be compared to the one based on the score distributions (dashed). The limit is calculated for  $A$  boson masses from 20 to 110 GeV and includes systematic uncertainties in both cases. The green area denotes the  $1\sigma$  uncertainty band for the results using the NN and yellow marks the  $2\sigma$  area.

as the one calculated based on  $m_{\text{MMC}}$ . Except for the nominal LMSR limit for mass point  $m_A = 80 \text{ GeV}$ , all nominal points are compatible with the  $1\sigma$  region around the new NN based limit. It seems like the standard method  $m_{\text{MMC}}$  distribution might be a better fit for this application, especially for lower mass points up to 60 GeV. A small study was performed comparing limits that use only statistical uncertainties for both methods. It is discussed in Section E.3 and shows a slightly better result than the one using also systematic uncertainties, shown here. As mentioned before this needs to be said with caution, as different bin choices for the exact same input distributions can already lead to much larger changes in the limit (see Section E.1). A final comparison should be done using limits that were calculated on unbinned distributions.

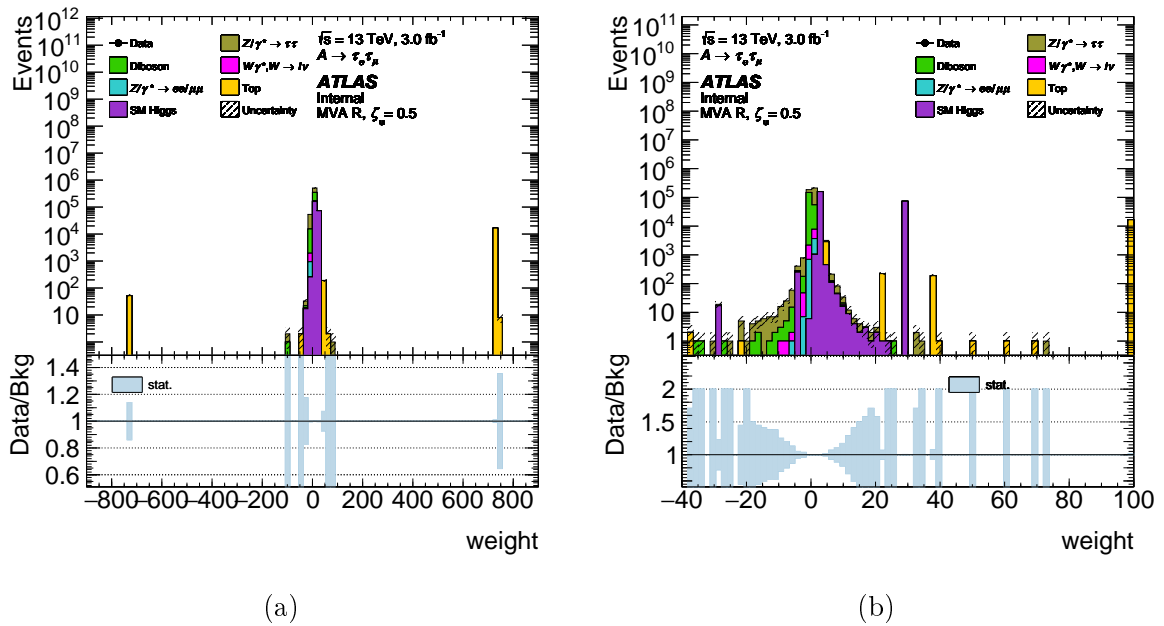


Figure 6.6.: Distribution of event weights in the mvaVR. The full spectrum of different event weights can be seen in (a), while (b) shows only the section around 0. These weights are multiplied with some other terms, the the luminosity, described in Section 6.1, to form the full event weights.

## 6.8. Additional Studies

This section will provide some additional information on studies that were performed around the neural network of this thesis. An important part was the comparison of different methods to calculate event weights that are used for a weighted average loss function during training. Additionally, a study was performed to investigate the parametrised NN’s ability to interpolate between the mass hypotheses that were used for training. Lastly, the difference between scores assigned to the same event by the different networks from the k-fold method are studied.

### 6.8.1. Event Weights

As the available training data is very imbalanced between different event classes, it was decided to use event weights to calculate the weighted average loss during training. The goal is to get a loss function that is representative of the input distribution in the mvaVR. The way of calculating these event weights is done analogously to weights in the base analysis and is described in Section 6.1. These weights also contain a factor that is given by the MC weight. These MC weights can be used to describe negative interference between different processes and therefore can be negative. Depending on the MC sample there is a non-negligible contribution from events with negative MC weights. In total, out of the 620 903 events that are used as training input, 52 038 have a negative weight, which is a portion of roughly 8%. The distribution of MC weights in the mvaVR can be seen in Figure 6.6. Negative weights might lead to problems in the minimisation of the loss function  $L$ , as  $L$  can then become unbound. This can lead to a diverging training [123]. This section investigates the influence of the negative weighted events on the training by comparing neural networks that used

different modifications of the event weights during training. Next to using the weights as they are, four alternative versions are considered. In the first one, only events that naturally have positive weights are used for training, by setting all negative event weights to zero. The second very straight-forward approach uses the absolute value of all weights. The last alternative based on the standard event weights is less intuitive. It always applies the absolute values of event weights and additionally changes the label of events with originally negative weights. This means a background event with negative weight is now considered as a signal event with positive weight. This method is inspired by partial-label learning techniques, here in the special case of only one label [124, 125]. This approach will be called *flip label* in this thesis. The fourth alternative approach is more general and does not utilise weights that are calculated individually for each event. Instead, it relies on weights that are the same for each event of the same process. This means these *process weights* are simply calculated as the contribution of a given type of process in the mvaVR and are then normalised to the number of events available for this process (see Table 6.1). To get process weights that are more in the order of the other methods, this value is multiplied with the mean number of events per process:

$$w_{\text{process } i} = \frac{\text{rel. contribution of } i \text{ in mvaVR}}{\#(\text{events of type } i)} \cdot \frac{\#(\text{all events})}{\#(\text{all processes})}. \quad (6.6)$$

The relative contributions of the different processes in the mvaVR can be found in Table 6.3. It was decided to put emphasis on MC generated events and therefore drop the cut on the IFF class of the leptons that is applied for the MM. As the MM is expected to generate a very good estimation of the total amount of background  $N_{\text{tot,MM}}$ , this number is used as a reference. For all MC simulated processes their relative contribution is calculated as their contribution in the mvaVR without the truth cut, divided by  $N_{\text{tot,MM}}$ . The contribution of fake events is simply chosen to achieve a coverage of 100 % of  $N_{\text{tot,MM}}$  by all backgrounds combined. This procedure results in the values shown in Table 6.3.

With each of the five different weight application methods a NN was trained. This comparison was performed with an earlier optimised configuration than the one presented in preceding sections. Therefore, the hyperparameters are different: The NNs consist of three hidden layers with 40, 39 and 108 neurons. This study was performed before the balancing of the different signal hypotheses as described in Section 6.1 was introduced. This can also be seen as the process weights vary for the different mass points (see Table 6.3.) A batch size of 228 and a constant learning rate of  $1.7 \times 10^{-4}$  was used. This setup was used for all five NNs with the different weighting methods.

Each NN was then applied to MC and data events in order to calculate score distributions, that can be used as input for the calculation of metrics to compare the performance of all networks. These quantities and their meanings are introduced in Section 5.6. It is important to note, that the loss that was calculated over the course of the training can only be compared qualitatively. As the weights are different for all NNs, the weighted loss is intrinsically different for all methods. An overview over the networks' performances can be found in Table 6.4. Each metric is always calculated from a single score distribution, in order to take all scores for the different mass hypotheses into account the mean value is shown here, as well as the standard deviation between them.

Table 6.3.: Contribution of all processes in the mvaVR, relative to the total background contribution. The truth cut on MC is not applied for these values to emphasise the impact of MC samples on the training, the contribution of fake events is chosen to ensure the total background expectation predicted by the MM. The process weights can be calculated based on these values by dividing by the number of events per process.

Process	Rel. Contribution in the mvaVR
$Z \rightarrow \tau\tau$	42.6 %
Top	9.1 %
Diboson	6.1 %
Other MC	19.0 %
Fakes	23.1 %
$A \rightarrow \tau\tau, m_A = 20 \text{ GeV}$	14.3 %
$A \rightarrow \tau\tau, m_A = 30 \text{ GeV}$	20.8 %
$A \rightarrow \tau\tau, m_A = 40 \text{ GeV}$	21.1 %
$A \rightarrow \tau\tau, m_A = 50 \text{ GeV}$	21.9 %
$A \rightarrow \tau\tau, m_A = 60 \text{ GeV}$	18.5 %
$A \rightarrow \tau\tau, m_A = 70 \text{ GeV}$	14.3 %
$A \rightarrow \tau\tau, m_A = 80 \text{ GeV}$	11.3 %
$A \rightarrow \tau\tau, m_A = 90 \text{ GeV}$	7.8 %
$A \rightarrow \tau\tau, m_A = 100 \text{ GeV}$	5.7 %
$A \rightarrow \tau\tau, m_A = 110 \text{ GeV}$	4.1 %

It is important to note that the weights will directly influence the total loss and therefore its gradient. This means that the training conditions are not fully comparable, when the same learning rate is used, like here. It is therefore possible that a weighting strategy, which does not perform as good as others here, might stronger benefit from a hyperparameter optimisation and will outperform the other methods afterwards.

It is clear from Table 6.4, that none of the described weighting methods is superior to the other ones. Using the standard event weights, including the negative ones as they are, seems to perform slightly better than the other methods in regard to the considered averaged metrics. For this reason it was decided to use this strategy throughout the analysis. However, all methods result in a relatively similar performance.

Under the assumption that a good modelling of the input distributions is beneficial to the training, it makes sense that the two methods setting negative weights to zero or using their absolute value, do not perform as good as the standard event weights. The negative weights are used to suppress the contribution of events in a certain phase space. If these weights are set to zero, no suppression is performed and in the case of absolute values they are even enhanced. The flip label method makes it harder for the NN to learn the distinguished features of the signal process as an artificial noise is added to the signal training data. In the case of multi-class classification, the partial

Table 6.4.: Comparing the performance of NNs trained with different event weight methods, based on the AUC, the maximal signal significance improvement  $\sigma_{\max}$  and the maximal classification efficiency  $\epsilon_{\max}$  (see Section 5.6 for their definition). The first value for each NN denotes the mean over all mass hypotheses and the one in brackets their standard deviation.

	AUC	$\sigma_{\max}$	$\epsilon_{\max}$
standard event weights	0.817 (0.073)	1.56 (0.31)	1.18 (0.21)
only pos. event weights	0.810 (0.066)	1.53 (0.28)	1.16 (0.21)
abs. value event weights	0.802 (0.066)	1.49 (0.26)	1.15 (0.20)
flip label event weights	0.791 (0.077)	1.47 (0.31)	1.10 (0.21)
process weights	0.771 (0.071)	1.42 (0.21)	1.08 (0.18)

labelling technique is probably more reasonable than the label flip in this particular case with only one class. Lastly, it seems to negatively impact the training if the information on event distributions inside a given process is lost. This is the case when using process weights, which overall performs worst in this study.

All of the here discussed weighting methods try to reproduce the contribution of events in the mvaVR. It might be an interesting approach to instead try to balance the impact of background and signal events during training. The study presented here suggests that keeping information from MC events is beneficial, therefore one could try an approach, like the one suggested in Equation 6.3 to balance the contribution of different signal mass hypotheses. In this case, this balancing would be done between background and all signal events.

### 6.8.2. Interpolation Between Signal Masses

The decision to use a PNN for the classification of signal in this thesis is based on the fact, that the mass of a potential  $A$  boson cannot be predicted exactly by theory. Therefore, one needs to be sensitive to a range of different mass hypotheses. It is only possible to use a finite set of possible mass points for the training of the network. As it is likely that the  $A$  boson has a mass somewhere between the considered ones (if it exists at all) it is interesting to investigate the NN's ability to interpolate between the trained masses. This is possible by examining the behaviour of neighbour signal hypotheses in a given score distribution, so for example to look at signal  $m_A = 60$  GeV in the score distributions for  $m_A = 50$  GeV and  $m_A = 70$  GeV. The NN should be able to better interpolate to any other signal produced by an  $A$  boson with a mass in between these points, compared to this extreme case.

A qualitative way to do this is simply to look at the distribution of signal events in score distributions of the neighbour mass points. Some examples for this from the mvaVR can be found in Figure 6.7. From these distributions it is apparent, that there is a stronger specialisation for lower signal mass hypotheses. Especially for score  $m_A = 20$  GeV it is very clear that the  $m_A = 30$  GeV signal events generally get assigned lower scores more often than the one from the lower mass point. For higher masses, on the other hand, the signal distributions look very similar over the full range of score values. Especially, as the statistical fluctuations are expected to be higher and



therefore differences are not as significant in bins with smaller number of signal events.

It would be interesting to compare the limit that is calculated for a signal based on the corresponding score distribution with those based on one of the neighbour scores.

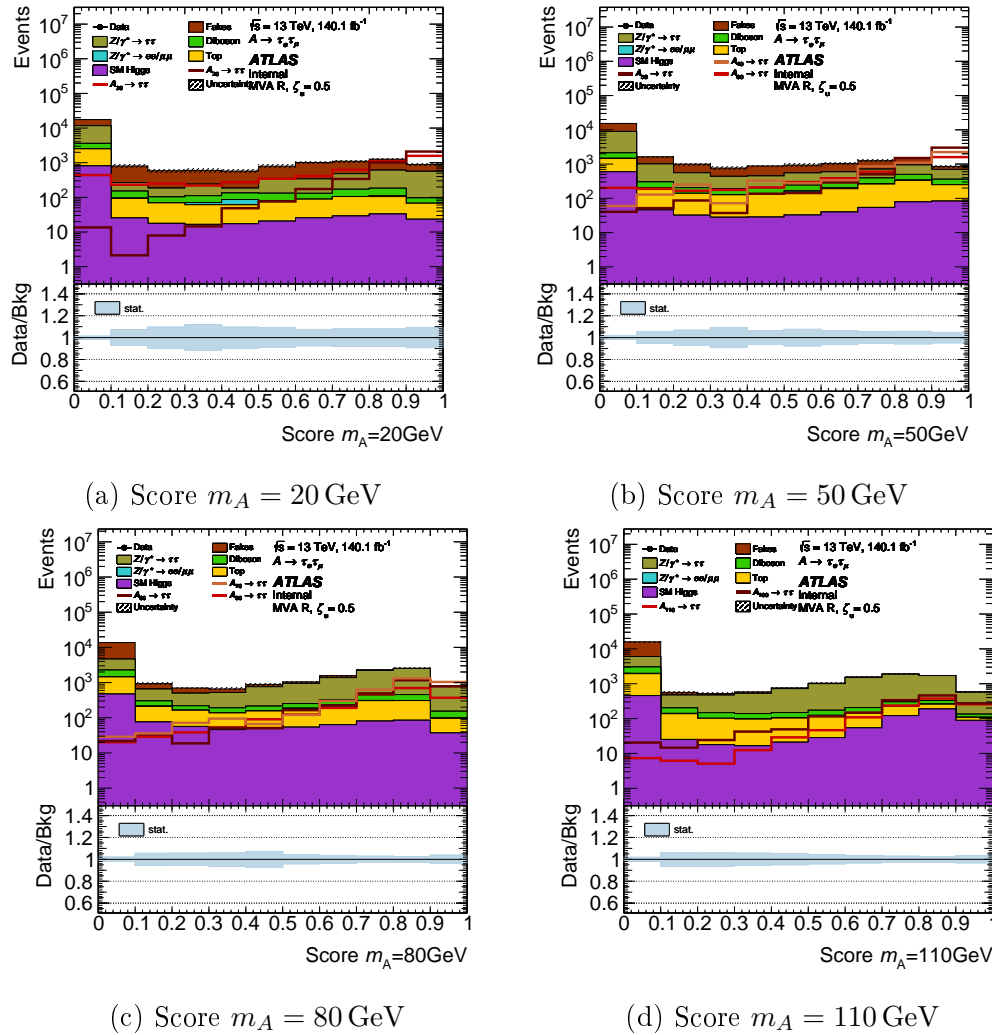


Figure 6.7.: Score distributions from the mvaVR also containing signal events from the corresponding neighbour mass points. The dark brown line always shows the signal distribution that is corresponding to the score and can be used as reference for the signal distributions in orange and red from neighbour masses.

### 6.8.3. Differences Between K-Folds

Lastly, a small study was performed investigating the difference in score that a specific event is assigned by the different NNs from the k-fold procedure. It would be a sign of a model's stability, if the training results depended only loosely on the specific training samples which would lead to small differences between the scores. In an ideal case the scores assigned by different networks would almost be identical for the same event. To investigate this behaviour, it was necessary to slightly change the way the NN is applied to MC and data events. In this case, every network is applied to the same

event and the corresponding score is saved. This means that also events, that were used during the training or validation get assigned a score. This is usually not advised and these events were not used in any way beside this study.

After all neural networks are applied to all available events, the standard deviation between scores  $\sigma_{\text{score}}$  of the different networks regarding the same event is calculated. This is done separately for all different mass hypotheses. These standard deviations can then be investigated. Figure 6.8 shows the distribution of these  $\sigma_{\text{score}}$  in the mvaVR for different mass hypotheses.

It is visible that all distributions of  $\sigma_{\text{score}}$  have a strong peak at zero and then quickly decline with maximum standard deviations of around 0.3. This behaviour seems to be relatively universal for all types of background and signal. For a given mass point, all backgrounds share a relatively similar distribution, which also includes signal events for the most part. Only in the case of signal  $m_A = 20 \text{ GeV}$  the distribution looks significantly different with a smaller peak around zero but a faster decline towards higher deviations. It seems like the different NNs can agree more on scores for higher masses, as these distributions show a steeper decline for higher  $\sigma_{\text{score}}$  values, which also results in smaller maximal values. It should be noted that the bins in Figure 6.8, which seem to be empty but have a large relative statistical uncertainty, actually do contain a few events but with very small weights, so that they do not show up in the logarithmic plot. The general behaviour of all  $\sigma_{\text{score}}$  distributions looks as expected, with a strong peak at deviations close to zero, without any additional maxima. For the future it might be interesting to also have a comparison of only the scores that were assigned by NNs that used the respective event during training. It is possible that they might agree on more similar scores.

This study might be interesting to implement a systematic uncertainty for the application of the neural network in the future. Until now, there is no systematic added for this procedure at all. An uncertainty that is based on the differences in scores assigned by different NNs, that were defined exactly in the same way and were just trained on a different sub-set of training data, should at least cover the influence from different trainings on the final network's performance. To get representative and unbiased uncertainty, it would be necessary to use a slightly different k-fold procedure with more datasets that are reserved for the application/calculation of uncertainties. Only by doing this, it would be possible to compare multiple NN that were trained on different training events and are unbiased regarding specific events at the same time. It is likely that the specific choice of hyperparameters will have a larger impact on the final performance than different trainings of the same network. As there are so many different hyperparameters that can be changed and are additionally highly correlated, it will be challenging to define an uncertainty that covers this dependence fully.

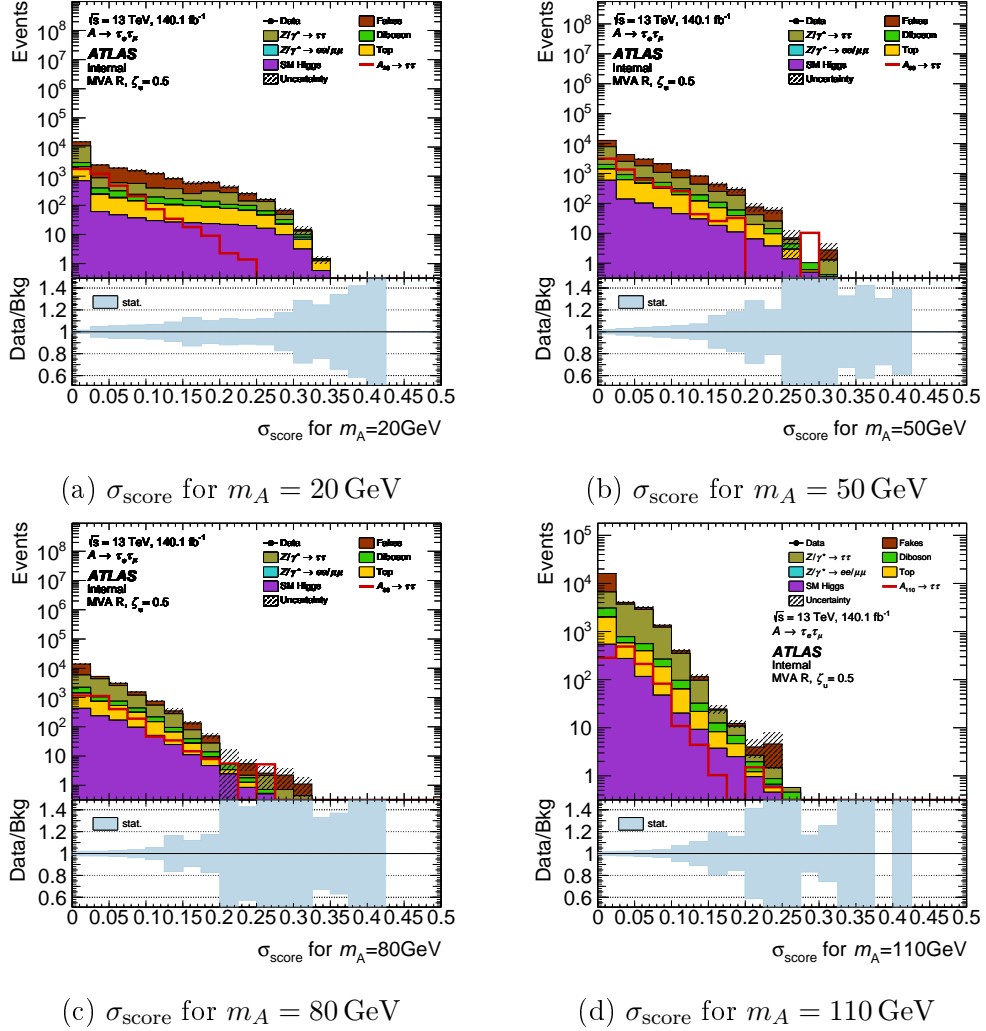


Figure 6.8.: Distribution of standard deviations between scores assigned by the different k-fold NNs to the same events in the mvaVR for different  $m_A$  scores. Bins that seem to be empty but have a large relative statistical uncertainty at the same time actually do contain some events but with very small weights, so that they do not show up in the logarithmic plot.

# 7. Summary and Outlook

This thesis presents the search for a light CP-odd Higgs boson  $A$  that couples strongly to leptons and up-type quarks. The analysis is based on measurements of the ATLAS detector that were recorded between 2015 and 2018, during Run 2 of the LHC. A total luminosity of  $140.1 \text{ fb}^{-1}$  of proton-proton collisions at a centre-of-mass energy of 13 TeV is utilised. The  $A$  boson is expected to almost exclusively decay into a  $\tau$ -lepton pair. For this thesis, it was decided to look into a final state of one electron and one muon, that results from a fully leptonic decay of the two  $\tau$  leptons. As there is no precise theoretical prediction for the mass of the  $A$  boson, a total of ten different mass hypotheses between 20 GeV and 110 GeV is examined.

The first part of this thesis focuses on a cut-based approach to this search. It strongly relies on MC generated background and signal events. However, it is not possible to simulate background events that come from non-prompt sources, the so-called fake background. The Matrix Method is a data-driven way of estimating this type of background and is utilised in this thesis. It is well suited for the described analysis, as it is able to differentiate well between events where the electron, muon or both are reconstructed from non-prompt sources. With all relevant background processes, including the one from fake events, described, there still is a mismodelling in one validation regions observed. This was investigated and after finding no satisfying explanation it was decided to apply a data-driven reweighting procedure to all MC generated events of the problematic  $Z \rightarrow \tau\tau$  process. This is necessary, as  $Z \rightarrow \tau\tau$  is the most prominent background process in the signal regions, which makes it likely that the mismodelling will translate there.

With these techniques it was possible to reach a good agreement between data and estimated background in dedicated validation regions. The base analysis relies on physically motivated kinematic cuts to improve the expected signal significance. It uses distributions in the high-level variable  $m_{\text{MMC}}$ , that is calculated by the Missing Mass Calculator algorithm, as input for the statistical evaluation. In the end, it is able to calculate expected limits on the production cross section times branching ratio of the  $A$  boson. The unblinding of data in the signal region in the near future will allow further insides using observed limits.

It would be beneficial to the analysis to find the cause for the observed mismodelling of the  $Z \rightarrow \tau\tau$  background. The reweighting procedure described here is only an unsatisfactory way to treat it, that introduces significant additional systematic uncertainties and needs to rely on measured data, without any understanding of the underlying problem.

The second part of this thesis focuses on the introduction of a neural network based extension to the earlier described base-analysis. It is trained for the classification of signal and background events. In order to be sensitive to all considered mass hypotheses for the  $A$  boson, a parameterised NN is used that adds an input variable that is representative for the signal mass. It was shown that this allows the network to interpolate between the signal mass points. The training is performed on a selection

of low- and high-level input variables that encapsulate information on the kinematics of the process. MC samples are used as training data for the majority of background and all signal processes. Background from fake events is represented by data events that do not pass certain `Tight` isolation and identification criteria. Because the MC events that are needed for the training are also used for the application later on, a k-fold technique is used to prevent a bias on these events compared to data samples, which the NN has not seen during training. After a finished training, the NN can be applied to data and MC events, meaning it assigns a score for each mass hypothesis to the events. These scores can be interpreted as the probability that this event comes from a signal process of the corresponding mass. The goal of this thesis was to further improve the performance of this classification. To do this, a hyperparameter optimisation framework was used to optimise certain general features. Additionally, different strategies for event weights, that are applied during the training, are compared.

In the end, it is challenging to compare the expected limits that are calculated in the base analysis using the  $m_{\text{MMC}}$  distribution, with the ones based on the score distributions calculated by the NN. In both cases the limits strongly depend on the binning that is used for the input distributions. It is difficult to define binning strategies that are comparable for two distributions in completely different variables and with different numbers of events. Therefore, in the future it would be beneficial to use an unbinned limit calculation, which is more challenging, but will allow for a fair comparison between different strategies. The resulting limits calculated here show a slight degradation in the limit for the one using NN scores compared to the base analysis. However, the final limits are of the same order of magnitude and statistically compatible within the  $1\sigma$  band. This means the NN seems to almost be able to recreate the performance of the base analysis.

Therefore, it can be expected to even exceed the results of the cut-based analysis if the performance of the network can be further increased. Because NNs are highly complex systems, there are many possibilities for further improvements that can be explored in the future. The studies performed in this thesis regarding event weights during training could be further expanded towards balancing the contribution of signal and background during training. To make this possible it would be beneficial to get more signal events that can help to represent the full signal phase space. Another possible improvement is the implementation of a multi-class classification that distinguishes between different kinds of background. This could help the NN to learn distinct features of individual types of background, compared to only learning specific signal features.

In order to use this expansion to the base analysis, it would be important to further investigate the quality of interpolation between different signal masses. This is important because the true mass of the  $A$  boson (if it exists) is likely to be somewhere between the studied mass points. One opportunity could be to calculate expected limits for signal of a given mass based on the score distribution for another mass. Lastly, it would be very important to define systematic uncertainties that are able to cover the ones that arise by the use of the neural network. This will require many additional studies and research, but the ones presented in Subsection 6.8.3 can be seen as a first step.

# Bibliography

- [1] J. J. Thomson, “XL. Cathode Rays”, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44**, 293 (1897).
- [2] G. Aad et al. (The ATLAS Collaboration), “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, Physics Letters B **716**, 1 (2012).
- [3] S. Chatrchyan et al. (The CMS Collaboration), “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, Physics Letters B **716**, 30 (2012).
- [4] B. Abi et al., “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm”, Physical Review Letters **126**, 10.1103/physrevlett.126.141801 (2021).
- [5] A. Cherchiglia et al., “Muon  $g - 2$  in the 2HDM: Maximum results and detailed phenomenology”, Physical Review D **98**, 10.1103/physrevd.98.035001 (2018).
- [6] P. Moder, *Search for a light CP-odd Higgs boson decaying into a  $\tau\tau$  pair*, 2018.
- [7] T. Kreße, *Search for a light CP-odd Higgs boson decaying into a pair of  $\tau$ -leptons with the ATLAS detector*, 2020.
- [8] X.-M. Sonntag, *Optimized search for a light CP-odd Higgs boson decaying into two  $\tau$ -leptons using the ATLAS detector*, 2021.
- [9] C. Schmidt, *Top quark background uncertainty and fake lepton background estimation in the search for a light CP-odd Higgs boson using the ATLAS detector*, 2023.
- [10] J. Friese, *Search for a light CP-odd Higgs boson decaying into a pair of  $\tau$ -leptons*, 2022.
- [11] A. Einstein, “Die Grundlage der allgemeinen Relativitätstheorie”, Annalen der Physik **354**, 769 (1916).
- [12] S. Tomonaga, “On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields\*”, Progress of Theoretical Physics **1**, 27 (1946).
- [13] J. Schwinger, “Quantum Electrodynamics. I. A Covariant Formulation”, Phys. Rev. **74**, 1439 (1948).
- [14] R. P. Feynman, “Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction”, Phys. Rev. **80**, 440 (1950).
- [15] S. L. Glashow, “Partial-symmetries of weak interactions”, Nuclear Physics **22**, 579 (1961).
- [16] A. Salam, “Renormalizability of Gauge Theories”, Phys. Rev. **127**, 331 (1962).
- [17] S. Weinberg, “A Model of Leptons”, Phys. Rev. Lett. **19**, 1264 (1967).

- 
- [18] G. Arnison et al. (The UA1 Collaboration), “Experimental observation of lepton pairs of invariant mass around 95 GeV/c<sup>2</sup> at the CERN SPS collider”, *Physics Letters B* **126**, 398 (1983).
- [19] M. Banner et al. (The UA2 Collaboration), “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider”, *Physics Letters B* **122**, 476 (1983).
- [20] The ATLAS Collaboration, *Improved W boson Mass Measurement using 7 TeV Proton-Proton Collisions with the ATLAS Detector*, ATLAS-CONF-2023-004, 2023.
- [21] P. Achard et al., “Measurement of the Z-boson mass using e+e-→Zγ events at centre-of-mass energies above the Z pole”, *Physics Research Publications* **585**, 10.1016/j.physletb.2004.01.074 (2004).
- [22] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13**, 508 (1964).
- [23] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13**, 321 (1964).
- [24] H. Fritzsch et al., “Advantages of the color octet gluon picture”, *Physics Letters B* **47**, 365 (1973).
- [25] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13**, 321 (1964).
- [26] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Non-Abelian Gauge Theories”, *Phys. Rev. Lett.* **30**, 1343 (1973).
- [27] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30**, 1346 (1973).
- [28] M. Lubej, *Standard Model*, University of Zurich.
- [29] R. L. Workman et al. (Particle Data Group), “Review of Particle Physics”, *PTEP* **2022**, 083C01 (2022).
- [30] Y. Fukuda et al. (The Super-Kamiokande Collaboration), “Evidence for Oscillation of Atmospheric Neutrinos”, *Physical Review Letters* **81**, 1562 (1998).
- [31] E. Majorana, “Theory of the symmetry of electrons and positrons”, *Nuovo Cimento* **14**, 171 (1937).
- [32] D. J. Griffiths, *Introduction to elementary particles*, Physics textbook (Wiley, New York, NY, 2008).
- [33] J. H. Christenson et al., “Evidence for the 2π Decay of the K<sub>2</sub><sup>0</sup> Meson”, *Phys. Rev. Lett.* **13**, 138 (1964).
- [34] G. Lüders, “Proof of the TCP theorem”, *Annals of Physics* **2**, 1 (1957).
- [35] N. E. Mavromatos, *CPT Violation: Theory and Phenomenology*, arXiv:hep-ph/0504143, 2005.
- [36] E. Noether, “Invariante Variationsprobleme”, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918**, 235 (1918).
-

- 
- [37] C. S. Wu et al., “Experimental Test of Parity Conservation in Beta Decay”, *Phys. Rev.* **105**, 1413 (1957).
- [38] J. Ellis, “Higgs Physics”, in 2013 European School of High-Energy Physics (2015), pp. 117–168.
- [39] P. Langacker, *The Standard Model and Beyond*, Series in High Energy Physics, Cosmology and Gravitation (Taylor & Francis, 2017).
- [40] H. Goenner, “Zur Geschichte der einheitlichen Feldtheorie: Einsteins Jahre erfolgloser Forschung”, *Annalen der Physik* **518**, 149 (2005).
- [41] E. Corbelli and P. Salucci, “The extended rotation curve and the dark matter halo of M33”, *Monthly Notices of the Royal Astronomical Society* **311**, 441 (2000).
- [42] D. Clowe et al., “A Direct Empirical Proof of the Existence of Dark Matter\*”, *The Astrophysical Journal* **648**, L109 (2006).
- [43] A. Boyarsky et al., “Sterile neutrino Dark Matter”, *Progress in Particle and Nuclear Physics* **104**, 1 (2019).
- [44] A. D. Sakharov, “Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe”, *Soviet Physics Uspekhi* **34**, 392 (1991).
- [45] M. B. Gavela et al., “Standard Model CP-violation and baryon asymmetry”, *Modern Physics Letters A* **09**, 795 (1994).
- [46] L. Susskind, “Dynamics of spontaneous symmetry breaking in the Weinberg-Salam theory”, *Phys. Rev. D* **20**, 2619 (1979).
- [47] G. W. Bennett et al. (Muon  $g - 2$  Collaboration), “Final report of the E821 muon anomalous magnetic moment measurement at BNL”, *Phys. Rev. D* **73**, 072003 (2006).
- [48] D. P. Aguillard et al. (Muon  $g - 2$  Collaboration), “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm”, *Phys. Rev. Lett.* **131**, 161802 (2023).
- [49] B. Abi et al. (Muon  $g - 2$  Collaboration), “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm”, *Phys. Rev. Lett.* **126**, 141801 (2021).
- [50] T. Aoyama et al., “The anomalous magnetic moment of the muon in the Standard Model”, *Physics Reports* **887**, The anomalous magnetic moment of the muon in the Standard Model, 1 (2020).
- [51] S. Borsanyi et al., “Leading hadronic contribution to the muon magnetic moment from lattice QCD”, *Nature* **593**, 51 (2021).
- [52] G. Branco et al., “Theory and phenomenology of two-Higgs-doublet models”, *Physics Reports* **516**, 1 (2012).
- [53] S. Davidson<sup>1</sup> and H. E. Haber, “Basis-independent methods for the two-Higgs-doublet mode”, *Physical Review D* **72**, 10.1103/PhysRevD.72.035004 (2005).
- [54] M. Sher, “Flavor-changing neutral currents in the Higgs sector”, *Modern Physics Letters A* **37**, 10.1142/s0217732322300117 (2022).
-



- 
- [55] A. Pich and P. Tuzón, “Yukawa alignment in the two-Higgs-doublet model”, *Physical Review D* **80**, 10.1103/physrevd.80.091702 (2009).
- [56] J. Abdallah et al. (The DELPHI Collaboration), “Searches for neutral Higgs bosons in extended models”, *The European Physical Journal C* **38**, 1 (2004).
- [57] G. Cowan et al., “Asymptotic formulae for likelihood-based tests of new physics”, *The European Physical Journal C* **71**, 10.1140/epjc/s10052-011-1554-0 (2011).
- [58] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”, *The Annals of Mathematical Statistics* **9**, 60 (1938).
- [59] A. Wald, “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large”, *Transactions of the American Mathematical Society* **54**, 426 (1943).
- [60] CERN, *LEP design report*, CERN-LEP-84-01, Geneva, 1984.
- [61] E. Mobs, *The CERN accelerator complex in 2019*, CERN-GRAPHICS-2019-002, 2019.
- [62] G. Aad et al. (The ATLAS Collaboration), “The ATLAS Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation* **3**, S08003 (2008).
- [63] S. Chatrchyan et al. (The CMS Collaboration), “The CMS experiment at the CERN LHC”, *Journal of Instrumentation* **3**, S08004 (2008).
- [64] A. A. Alves Jr et al. (The LHCb Collaboration), “The LHCb Detector at the LHC”, *Journal of Instrumentation* **3**, S08005 (2008).
- [65] K. Aamodt et al. (The ALICE Collaboration), “The ALICE experiment at the CERN LHC”, *Journal of Instrumentation* **3**, S08002 (2008).
- [66] The ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*, arXiv:2212.09379, 2022.
- [67] L. Evans and P. Bryant, “LHC Machine”, *Journal of Instrumentation* **3**, S08001 (2008).
- [68] D. Ferrère, “The upgrade of the ATLAS Inner Detector”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **718**, Proceedings of the 12th Pisa Meeting on Advanced Detectors, 30 (2013).
- [69] The ATLAS Collaboration, “Performance of the ATLAS trigger system in 2015”, *The European Physical Journal C* **77**, 10.1140/epjc/s10052-017-4852-3 (2017).
- [70] M. Aaboud et al. (The ATLAS collaboration), “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at  $\sqrt{s} = 13$  TeV”, *The European Physical Journal C* **79**, 10.1140/epjc/s10052-019-7140-6 (2019).
- [71] G. Aad et al. (The ATLAS collaboration), “Muon reconstruction performance of the ATLAS detector in proton-proton collision data at  $\sqrt{s} = 13$  TeV”, *The European Physical Journal C* **76**, 10.1140/epjc/s10052-016-4120-y (2016).
-

- 
- [72] M. Cacciari et al., “The anti- $k_t$  jet clustering algorithm”, *Journal of High Energy Physics* **2008**, 063 (2008).
- [73] M. Aaboud et al. (The ATLAS collaboration), “Jet reconstruction and performance using particle flow with the ATLAS detector”, *The European Physical Journal C* **77**, 10.1140/epjc/s10052-017-5031-2 (2017).
- [74] A. Hrynevich, “ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2”, *Journal of Instrumentation* **12**, C06038 (2017).
- [75] The ATLAS collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2016.
- [76] A. Elagin et al., “A new mass reconstruction technique for resonances decaying to  $\tau\tau$ ”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654**, 481 (2011).
- [77] G. Aad et al. (The ATLAS Collaboration), “The ATLAS simulation infrastructure”, *The European Physical Journal C* **70**, 823 (2010).
- [78] M. H. Seymour and M. Marx, *Monte Carlo Event Generators*, arXiv:1304.6677, 2013.
- [79] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Computer Physics Communications* **191**, 159 (2015).
- [80] S. Alioli et al., “A general framework for implementing NLO calculations in shower monte carlo programs: the POWHEG BOX”, *Journal of High Energy Physics* **2010**, 10.1007/jhep06(2010)043 (2010).
- [81] E. Bothmann et al., “Event generation with Sherpa 2.2”, *SciPost Physics* **7**, 10.21468/scipostphys.7.3.034 (2019).
- [82] A. Buckley et al., “LHAPDF6: parton density access in the LHC precision era”, *The European Physical Journal C* **75**, 10.1140/epjc/s10052-015-3318-8 (2015).
- [83] S. Agostinelli et al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506**, 250 (2003).
- [84] M. Beckingham et al. (The ATLAS Collaboration), *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, Geneva, 2010.
- [85] Z. Marshall (The ATLAS Collaboration), “Simulation of Pile-up in the ATLAS Experiment”, *Journal of Physics: Conference Series* **513**, 022024 (2014).
- [86] The ATLAS collaboration, *The ATLAS Tau Trigger in Run 2*, ATLAS-CONF-2017-061, Geneva, 2017.
- [87] M. Bonvini, *ggHiggs*, <https://www.ge.infn.it/~bonvini/higgs/>, Accessed: 2023-10-10.
- [88] S. Alioli et al., “NLO single-top production matched with shower in POWHEG:s and t-channel contributions”, *Journal of High Energy Physics* **2009**, 111 (2009).
-

- [89] E. Re, “Single-top Wt-channel production matched with parton showers using the POWHEG method”, *The European Physical Journal C* **71**, 10.1140/epjc/s10052-011-1547-z (2011).
- [90] S. Alioli et al., “Hadronic top-quark pair-production with one jet and parton showering”, *Journal of High Energy Physics* **2012**, 10.1007/jhep01(2012)137 (2012).
- [91] S. Alioli et al., “NLO higgs boson production via gluon fusion matched with shower in POWHEG”, *Journal of High Energy Physics* **2009**, 002 (2009).
- [92] P. Nason and C. Oleari, “NLO higgs boson production via vector-boson fusion matched with shower in POWHEG”, *Journal of High Energy Physics* **2010**, 10.1007/jhep02(2010)037 (2010).
- [93] Atautau Analysis Team (The ATLAS collaboration), *AtautauAnalysis*, <https://gitlab.cern.ch/atlas-physics-office/HDBS/ANA-HDBS-2021-08/atautauanalysis>.
- [94] ATLAS Dresden Group (The ATLAS collaboration), *ELCore Documentation*, <https://twiki.cern.ch/twiki/bin/view/Atlas/ELCoreFramework>, Accessed:2023-10-05, 2022.
- [95] ATLAS Collaboration, *Tools for estimating fake/non-prompt lepton backgrounds with the ATLAS detector at the LHC*, arXiv:2211.16178, 2022.
- [96] Isolation and Fake Forum (The ATLAS collaboration), *Truth Classification*, <https://gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/AnalysisCommon/TruthClassification>.
- [97] Isolation and Fake Forum (The ATLAS collaboration), *FakeEfficiencyTool*, <https://gitlab.cern.ch/ATLAS-IFF/FakeEfficiencyTool>.
- [98] Isolation and Fake Forum (The ATLAS collaboration), *FakeBkgTool*, <https://gitlab.cern.ch/atlas/athena/-/tree/main/PhysicsAnalysis/AnalysisCommon/FakeBkgTools>.
- [99] T. A. collaboration, *Study of  $Z \rightarrow ll\gamma$  decays at  $\sqrt{s} = 8$  TeV in the ATLAS experiment at the LHC*, ATLAS-CONF-2022-046, Geneva, 2022.
- [100] ATLAS Higgs Working Group (The ATLAS collaboration), *WsMaker*, <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/WorkspaceMaker>, Accessed: 2023-10-18.
- [101] Atautau Analysis Team (The ATLAS collaboration), *WsMakerAtautau*, <https://gitlab.cern.ch/atlas-physics-office/HDBS/ANA-HDBS-2021-08/WSMakerAtautau>.
- [102] P. Baldi et al., “Parameterized neural networks for high-energy physics”, *The European Physical Journal C* **76**, 10.1140/epjc/s10052-016-4099-4 (2016).
- [103] M. A. Nielsen, *Neural Networks and Deep Learning*, <http://neuralnetworksanddeeplearning.com/>, Accessed:2023-10-20, 2015.
- [104] S. R. Dubey et al., *Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark*, 2022.
- [105] S. Hochreiter, “Untersuchungen zu dynamischen neuronalen Netzen”, (1991).
-

- 
- [106] S. Hochreiter et al., “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies”, (2001).
- [107] S. Ruder, *An overview of gradient descent optimization algorithms*, arXiv:1609.04747, 2017.
- [108] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv:1412.6980, 2017.
- [109] T.-T. Wong, “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation”, *Pattern Recognition* **48**, 2839 (2015).
- [110] V. Bewick et al., “Statistics review 13: Receiver operating characteristic curves”, *Critical Care* **8**, 508 (2004).
- [111] *Compare Deep Learning Models Using ROC Curves*, <https://de.mathworks.com/help/deeplearning/ug/compare-deep-learning-models-using-ROC-curves.html>, Accessed: 2023-10-20.
- [112] I. Goodfellow et al., *Deep Learning* (MIT Press, 2016).
- [113] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization”, *Journal of Machine Learning Research* **13**, 281 (2012).
- [114] E. Brochu et al., *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning*, arXiv:1012.2599, 2010.
- [115] J. Bergstra et al., “Algorithms for Hyper-Parameter Optimization”, in, Vol. 24, edited by J. Shawe-Taylor et al., *Advances in Neural Information Processing Systems* (2011).
- [116] L. Li et al., *A System for Massively Parallel Hyperparameter Tuning*, arXiv:1810.05934, 2020.
- [117] F. Chollet et al., *Keras*, <https://keras.io>, 2015.
- [118] Martín Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, <https://www.tensorflow.org/>, Software available from tensorflow.org, 2015.
- [119] Atautau Analysis Team (The ATLAS collaboration), *MVA Neural Network*, [https://gitlab.cern.ch/atlas-physics-office/HDBS/ANA-HDBS-2021-08/mva\\_neural\\_network](https://gitlab.cern.ch/atlas-physics-office/HDBS/ANA-HDBS-2021-08/mva_neural_network).
- [120] E. Bachmann, *Polarization studies in same-sign W-boson pair production at the LHC with the ATLAS detector*, 2023.
- [121] R. Liaw et al., *Tune: A Research Platform for Distributed Model Selection and Training*, arXiv:1807.05118, 2018.
- [122] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks”, in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, Vol. 9, *Proceedings of Machine Learning Research* (May 2010), pp. 249–256.
- [123] M. Borisyak and N. Kazeev, “Machine learning on sWeighted data”, *Journal of Physics: Conference Series* **1525**, 012088 (2020).
-

- 
- [124] J. Bach et al., *Negative event weights in Deep Neural Network training for heavy Higgs searches*, [https://indico.desy.de/event/38271/contributions/138908/attachments/80694/105559/dpg\\_2023\\_bach\\_rehearsal.pdf](https://indico.desy.de/event/38271/contributions/138908/attachments/80694/105559/dpg_2023_bach_rehearsal.pdf), 2023.
- [125] L. Feng et al., *Provably Consistent Partial-Label Learning*, arXiv:2007.08929, 2020.
-



# Acronyms

**2HDM** two-Higgs-doublet model

**ADAM** adaptive moment estimation optimiser

**ALICE** A Large Ion Collider Experiment

**ASHA** Asynchronous Successive Halving Algorithm

**ATLAS** A Toroidal LHC ApparatuS

**AUC** area under curve

**BNL** Brookhaven National Laboratory

**BR** branching ratio

**BSM** beyond the Standard Model

**CERN** European Organization for Nuclear Research

**CL** confidence level

**CMS** Compact Muon Solenoid

**CSC** Cathode Strip Chambers

**ECal** electromagnetic calorimeter

**EM** electromagnetic

**FCal** LAr forward calorimeter

**FCNC** flavour-changing neutral current

**FNAL** Fermi National Accelerator Laboratory

**FNR** false negative rate

**FPR** false positive rate

**GD** gradient descent

**GRL** Good Runs List

**HEC** LAr hadronic end-cap calorimeter

**HLT** High Level Trigger

**HMSR** high-mass signal region

**ID** Inner Detector

**IFF** Isolation and Fake Forum

**L1** Level-1 Trigger

**LAr** liquid argon

**LEP** Large Electron-Positron collider

**LHC** Large Hadron Collider

**LHCb** Large Hadron Collider beauty

**LMSR** low-mass signal region

**MC** Monte Carlo

**MDT** Monitored Drift Tube Chambers

**ME** matrix element

**mGD** mini-batch gradient descent

**MLE** maximum-likelihood estimator

**MM** Matrix Method

**MMC** Missing Mass Calculator

**MS** Muon Spectrometer

**MSSM** Minimal Supersymmetric Standard Models

**NN** neural network

**PDF** probability density function

**PDF** parton density function

**PNN** parameterised neural network

**PS** parton shower

**QCD** quantum chromodynamics

**QED** quantum electrodynamics

**ReLU** Rectified Linear Unit

---



**ROC curve** receiver operating characteristic curve

**RoI** Region-of-Interest

**RPC** Resistive Plate Chambers

**SCT** Semiconductor Tracker

**SM** Standard Model of particle physics

**SR** Signal Region

**TGC** Thin Gap Chambers

**TNR** true negative rate

**TPE** Tree Parzen Estimator

**TPR** true positive rate

**TRT** Transition Radiation Tracker

**TVR** Top Validation Region

**VEV** vacuum expectation value

**VR** validation region

**ZVR**  $Z \rightarrow \tau\tau$  Validation Region

---



# List of Figures

2.1.	Overview on the particle content of the SM. . . . .	3
2.2.	Form of the Higgs potential for the case of $\mu^2 < 0$ and $\lambda > 0$ . . . . .	8
2.3.	Upper limits on possible values for $\zeta_l$ . . . . .	14
2.4.	Upper limits on possible values for $\zeta_u$ . . . . .	15
3.1.	Schematic view of the CERN accelerator complex. . . . .	21
3.2.	Cut-away view of the ATLAS detector. . . . .	22
3.3.	Cut-away view of the ATLAS Inner Detector. . . . .	23
3.4.	Cut-away view of the ATLAS calorimeter system. . . . .	24
3.5.	Cut-away view of the ATLAS muon spectrometer . . . . .	25
4.1.	Feynman diagram of the full signal process. . . . .	33
4.2.	Feynman diagram of the decay of a $Z$ boson to a $\tau$ -lepton pair. . . . .	33
4.3.	Feynman diagram of the decay of a $Z$ boson to an electron or muon pair . . . . .	34
4.4.	Feynman diagram of the decay of a top quark. . . . .	35
4.5.	Feynman diagrams of the relevant leptonic decay modes of a $W$ boson. . . . .	36
4.6.	Schematic view of the trigger setup. . . . .	38
4.7.	Distribution of $m_{\text{MMC}}$ using only MC background samples. . . . .	40
4.8.	Distributions of the final Matrix Method efficiencies. . . . .	44
4.9.	Distribution of $m_{\text{MMC}}$ using MC samples and fake events, using the MM. . . . .	45
4.10.	Variable distributions in the ZVR that seem to be mismodelled. . . . .	46
4.11.	Distribution of $Z \rightarrow \tau\tau$ reweighting weights and their systematic uncertainties. . . . .	47
4.12.	Distributions of formerly mismodelled variables in the ZVR, using $Z \rightarrow \tau\tau$ reweighting. . . . .	49
4.13.	Distribution of $m_{\text{MMC}}$ in the two SRs binned for the calculation of the expected limit. . . . .	51
4.14.	Expected limit on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow \tau\tau)$ for the base analysis. . . . .	51
5.1.	Schematic view of the architecture of a feed-forward, fully-connected neural network. . . . .	53
5.2.	Shape of the logistic sigmoid function . . . . .	54
5.3.	Receiver operator characteristic curve. . . . .	61
6.1.	Scheme of the final NN's architecture. . . . .	70
6.2.	Cyclical change in the learning rate over the course of the training. . . . .	72
6.3.	Diagram showing the development of training and validation loss over the course of the whole training. . . . .	73
6.4.	Distributions of score variables, calculated by the NN, in the mvaVR. . . . .	74
6.5.	Expected limits on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow \tau\tau)$ for the base analysis and the NN results. . . . .	76
6.6.	Distribution of MC event weights in the mvaVR. . . . .	77

6.7.	Score distributions from the mvaVR also containing signal events from the corresponding neighbour mass points. . . . .	81
6.8.	Distributions of $\sigma_{\text{score}}$ in the mvaVR. . . . .	83
B.1.	Comparing MM efficiencies calculated with and without the $Z \rightarrow \tau\tau$ reweighting. . . . .	115
C.1.	Distribution of different variables in the ZVR without applying the $Z \rightarrow \tau\tau$ reweighting. . . . .	118
C.2.	Distributions of variables in the TVR, which show a mismodelling in the ZVR. . . . .	119
C.3.	Mismodelling of $E_{\text{T}}^{\text{miss}}$ in the ZVR for different data-taking periods and MC campaigns. . . . .	120
C.4.	Mismodelling of $E_{\text{T}}^{\text{miss}}$ in the ZVR for different combined electron-muon triggers. . . . .	121
C.5.	Influence of the $Z \rightarrow \tau\tau$ reweighting on the expected limit on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow \tau\tau)$ . . . . .	122
D.1.	Distributions of NN input variables in the mvaVR. . . . .	128
D.2.	More distributions of NN input variables in the mvaVR. . . . .	129
D.3.	More distributions of NN input variables in the mvaVR. . . . .	130
D.4.	More distributions of NN input variables in the mvaVR. . . . .	131
D.5.	Distributions of NN input variables in the ZVR. . . . .	132
D.6.	More distributions of NN input variables in the ZVR. . . . .	133
D.7.	More distributions of NN input variables in the ZVR. . . . .	134
D.8.	More distributions of NN input variables in the ZVR. . . . .	135
D.9.	Diagram showing the development of training and validation loss over the course of a longer training. . . . .	136
D.10.	Distributions of score variables, calculated by the NN, in the mvaVR. . . . .	138
D.11.	Distributions of score variables, calculated by the NN, in the mvaVR. . . . .	139
D.12.	Distributions of score variables, calculated by the NN, in the TVR. . . . .	140
D.13.	Distributions of score variables, calculated by the NN, in the TVR. . . . .	141
D.14.	Distributions of score variables, calculated by the NN, in the ZVR. . . . .	142
D.15.	Distributions of score variables, calculated by the NN, in the ZVR. . . . .	143
E.1.	Expected limits on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow \tau\tau)$ for the base analysis comparing different input binnings. . . . .	144
E.2.	Expected limits on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow \tau\tau)$ based on score distributions in the SRs and the mvaVR. . . . .	146
E.3.	Expected limits on $\sigma(gg \rightarrow A) \cdot \text{BR}(A \rightarrow \tau\tau)$ for the base analysis and the NN results using statistical uncertainties only. . . . .	147

---

# List of Tables

2.1. Overview on the couplings between fermion fields and Higgs doublets for the four different types of 2HDM. . . . .	13
2.2. Special coupling parameters of the flavour-aligned 2HDM . . . . .	13
4.1. Overlap removal applied in this thesis. . . . .	29
4.2. Different decay modes of the $\tau$ -lepton with their BR. . . . .	32
4.3. Different cuts that are applied in the signal and validation regions. . . . .	39
4.4. Regions, where the different MM efficiencies are calculated. . . . .	43
4.5. Changes in the number of $Z \rightarrow \tau\tau$ events, when applying the $Z \rightarrow \tau\tau$ reweighting. . . . .	48
5.1. Scheme of the grouping of datasets in the case of a 5-fold method. . . . .	58
6.1. Number of training events per process and their contribution to the mvaVR. . . . .	66
6.2. Complete list of all variables investigated as input for the NN. . . . .	69
6.3. Contribution of all processes in the mvaVR, relative to the total background contribution. . . . .	79
6.4. Comparing the performance of NN trained with different event weight methods. . . . .	80
A.1. Triggers for the different data-taking years with their full names. . . . .	114

# A. Full Sample Selection

## A.1. Data Samples

This thesis relies on data from the ATLAS detector at the LHC that was taken during its Run 2 between 2015 and 2018. It does not use all data that was recorded during that time, but only events, which were recorded with all detector systems in good condition. This selection is defined by the *Good Runs List (GRL)*. The following GRL files were used:

```
data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Unknown_PHYS_StandardGRL_All_Good_25ns.xml,
data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml,
data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Unknown_PHYS_StandardGRL_All_Good_25ns_TriggerNo17e33prim.xml,
data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL_All_Good_25ns_TriggerNo17e33prim.xml.
```

## A.2. MC Samples

All Monte Carlo samples that were used in this thesis are listed in this section. They are organised by their MC campaign.

### A.2.1. mc16a

#### Gluon-fusion $A$ production

```
mc16_13TeV.346025.Pythia8EvtGen_A14NNPDF23LO_ggA60_tautau_leplep.deriv.DAOD_HIGG3D1.e7122_a875_r9364_p4164
mc16_13TeV.346026.Pythia8EvtGen_A14NNPDF23LO_ggA70_tautau_leplep.deriv.DAOD_HIGG3D1.e7122_a875_r9364_p4164
mc16_13TeV.346027.Pythia8EvtGen_A14NNPDF23LO_ggA80_tautau_leplep.deriv.DAOD_HIGG3D1.e7122_a875_r9364_p4164
mc16_13TeV.346028.Pythia8EvtGen_A14NNPDF23LO_ggA90_tautau_leplep.deriv.DAOD_HIGG3D1.e7122_a875_r9364_p4164
mc16_13TeV.346375.Pythia8EvtGen_A14NNPDF23LO_ggA50_tautau_leplep.deriv.DAOD_HIGG3D1.e7122_a875_r9364_p4164
mc16_13TeV.451705.Pythia8EvtGen_A14NNPDF23LO_ggA40_tautau_leplep.deriv.DAOD_HIGG3D1.e8139_a875_r9364_p4164
mc16_13TeV.801162.Py8EG_A14NNPDF23LO_ggA20_tautau_l1314.deriv.DAOD_HIGG3D1.e8307_a875_r9364_p4164
mc16_13TeV.801163.Py8EG_A14NNPDF23LO_ggA30_tautau_l1314.deriv.DAOD_HIGG3D1.e8307_a875_r9364_p4164
mc16_13TeV.801827.Py8EG_A14NNPDF23LO_ggA100_tautau_l1314.deriv.DAOD_HIGG3D1.e8307_a875_r9364_p4164
mc16_13TeV.801828.Py8EG_A14NNPDF23LO_ggA110_tautau_l1314.deriv.DAOD_HIGG3D1.e8307_a875_r9364_p4164
```

#### $Z/\gamma^* \rightarrow \tau\tau + \text{jets}$

```
mc16_13TeV.364128.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CVetoBVeto.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
mc16_13TeV.364129.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_CFilterBVeto.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
mc16_13TeV.364130.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_70_BFilter.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
mc16_13TeV.364131.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_140_CVetoBVeto.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
mc16_13TeV.364132.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_140_CFilterBVeto.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
mc16_13TeV.364133.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV0_140_BFilter.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
mc16_13TeV.364134.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CVetoBVeto.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
mc16_13TeV.364135.Sherpa_221_NNPDF30NNLO_Ztautau_MAXHTPTV140_280_CFilterBVeto.deriv.DAOD_HIGG3D1.e5307_s3126_r9364_p4164
```

mc16\_13TeV.364136.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r9364\_p4164  
mc16\_13TeV.364137.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r9364\_p4164  
mc16\_13TeV.364138.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5313\_s3126\_r9364\_p4164  
mc16\_13TeV.364139.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5313\_s3126\_r9364\_p4164  
mc16\_13TeV.364140.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r9364\_p4164  
mc16\_13TeV.364141.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r9364\_p4164  
mc16\_13TeV.364210.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364211.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364212.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV70\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364213.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV70\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364214.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364215.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364282.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r9364\_p4164  
mc16\_13TeV.364360.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r9364\_p4164  
mc16\_13TeV.364363.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r9364\_p4164

## Diboson

mc16\_13TeV.363355.Sherpa\_221\_NNPDF30NNLO\_ZqqZvv.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r9364\_p4164  
mc16\_13TeV.363356.Sherpa\_221\_NNPDF30NNLO\_ZqqZ11.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r9364\_p4164  
mc16\_13TeV.363357.Sherpa\_221\_NNPDF30NNLO\_WqqZvv.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r9364\_p4164  
mc16\_13TeV.363358.Sherpa\_221\_NNPDF30NNLO\_WqqZ11.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r9364\_p4164  
mc16\_13TeV.363359.Sherpa\_221\_NNPDF30NNLO\_WpqqWmlv.deriv.DAOD\_HIGG3D1.e5583\_s3126\_r9364\_p4164  
mc16\_13TeV.363360.Sherpa\_221\_NNPDF30NNLO\_WplvWmqq.deriv.DAOD\_HIGG3D1.e5983\_s3126\_r9364\_p4164  
mc16\_13TeV.363489.Sherpa\_221\_NNPDF30NNLO\_WlvZqq.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r9364\_p4164  
mc16\_13TeV.364250.Sherpa\_222\_NNPDF30NNLO\_l1111.deriv.DAOD\_HIGG3D1.e5894\_s3126\_r9364\_p4164  
mc16\_13TeV.364253.Sherpa\_222\_NNPDF30NNLO\_l111v.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r9364\_p4164  
mc16\_13TeV.364254.Sherpa\_222\_NNPDF30NNLO\_l11vv.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r9364\_p4166  
mc16\_13TeV.364255.Sherpa\_222\_NNPDF30NNLO\_l1vvv.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r9364\_p4164  
mc16\_13TeV.364286.Sherpa\_222\_NNPDF30NNLO\_l1lvjj\_ss\_EW4.deriv.DAOD\_HIGG3D1.e6055\_s3126\_r9364\_p4164  
mc16\_13TeV.364288.Sherpa\_222\_NNPDF30NNLO\_l111\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6096\_s3126\_r9364\_p4164  
mc16\_13TeV.364289.Sherpa\_222\_NNPDF30NNLO\_l11v\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6133\_s3126\_r9364\_p4164  
mc16\_13TeV.364290.Sherpa\_222\_NNPDF30NNLO\_l1vv\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6096\_s3126\_r9364\_p4164

## $W(\rightarrow l\nu) + \text{jets}$

mc16\_13TeV.364156.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364157.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364158.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364159.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364160.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364161.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364162.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364163.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364164.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364165.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364166.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364167.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364168.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164  
mc16\_13TeV.364169.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364170.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364171.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364172.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364173.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364174.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364175.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364176.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364177.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364178.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364179.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364180.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364181.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364182.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364183.Sherpa\_221\_NNPDPF30NNLO\_Wenu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364184.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364185.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364186.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364187.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364188.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364189.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364190.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364191.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364192.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364193.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364194.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364195.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364196.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

mc16\_13TeV.364197.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r9364\_p4164

### $W\gamma^*$

mc16\_13TeV.700015.Sh\_228\_evgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r9364\_p4164

mc16\_13TeV.700016.Sh\_228\_mvgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r9364\_p4164

mc16\_13TeV.700017.Sh\_228\_tvgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r9364\_p4164

### $Z/\gamma^*(\rightarrow \mu\mu) + \text{jets}$

mc16\_13TeV.364100.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364101.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364102.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364103.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364104.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364105.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364106.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364107.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364108.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364109.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364110.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364111.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164

mc16\_13TeV.364112.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164



mc16\_13TeV.364113.Sherpa\_221\_NNPDF30NNLO\_Zmumu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r9364\_p4164  
mc16\_13TeV.364198.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364199.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364200.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV70\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364201.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV70\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364202.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364203.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364204.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364205.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364206.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV70\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364207.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV70\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364208.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364209.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r9364\_p4164  
mc16\_13TeV.364281.Sherpa\_221\_NNPDF30NNLO\_Zmumu\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r9364\_p4164  
mc16\_13TeV.364359.Sherpa\_221\_NNPDF30NNLO\_Zmumu\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r9364\_p4164  
mc16\_13TeV.364362.Sherpa\_221\_NNPDF30NNLO\_Zmumu\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r9364\_p4164

## $Z/\gamma^*(\rightarrow ee) + \text{jets}$

mc16\_13TeV.364114.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364115.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364116.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364117.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364118.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364119.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364120.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364121.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364122.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364123.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364124.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364125.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364126.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364127.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r9364\_p4164  
mc16\_13TeV.364280.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r9364\_p4164  
mc16\_13TeV.364358.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r9364\_p4164  
mc16\_13TeV.364361.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r9364\_p4164

## $t\bar{t}$ , single top

mc16\_13TeV.410470.Py8EG\_A14\_ttbardamp258p75\_nonallhad.deriv.DAOD\_HIGG3D1.e6337\_s3126\_r9364\_p4164  
mc16\_13TeV.410644.PowhegPythia8EvtGen\_A14\_singletop\_schan\_lept\_top.deriv.DAOD\_HIGG3D1.e6527\_s3126\_r9364\_p4164  
mc16\_13TeV.410645.PowhegPythia8EvtGen\_A14\_singletop\_schan\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6527\_s3126\_r9364\_p4164  
mc16\_13TeV.410648.PowhegPythia8EvtGen\_A14\_Wt\_DR\_dilepton\_top.deriv.DAOD\_HIGG3D1.e6615\_s3126\_r9364\_p4164  
mc16\_13TeV.410649.PowhegPythia8EvtGen\_A14\_Wt\_DR\_dilepton\_antitop.deriv.DAOD\_HIGG3D1.e6615\_s3126\_r9364\_p4164  
mc16\_13TeV.410658.Py8EG\_A14\_tchan\_BW50\_lept\_top.deriv.DAOD\_HIGG3D1.e6671\_s3126\_r9364\_p4164  
mc16\_13TeV.410659.Py8EG\_A14\_tchan\_BW50\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6671\_s3126\_r9364\_p4164

## SM Higgs

mc16\_13TeV.345120.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaul1317.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r9364\_p4164  
mc16\_13TeV.345121.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaulm15hp20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r9364\_p4164

mc16\_13TeV.345122.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaulp15hm20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r9364\_p4164  
 mc16\_13TeV.345123.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautauh30h20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r9364\_p4164  
 mc16\_13TeV.345211.PowhegPy8EG\_NNPDF30\_AZNLO\_WmH125J\_Winc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r9364\_p4164  
 mc16\_13TeV.345212.PowhegPy8EG\_NNPDF30\_AZNLO\_WpH125J\_Winc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r9364\_p4164  
 mc16\_13TeV.345217.PowhegPy8EG\_NNPDF30\_AZNLO\_ZH125J\_Zinc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r9364\_p4164  
 mc16\_13TeV.345324.PowhegPythia8EvtGen\_NNLOPS\_NN30\_ggH125\_WWlvlfv\_EF\_15\_5.deriv.DAOD\_HIGG3D1.e5769\_s3126\_r9364\_p4166  
 mc16\_13TeV.345948.PowhegPy8EG\_NNPDF30\_AZNLOCTEQ6L1\_VBFH125\_WWlvlfv.deriv.DAOD\_HIGG3D1.e7172\_s3126\_r9364\_p4166  
 mc16\_13TeV.346190.PowhegPy8EG\_NNPDF30\_AZNLOCTEQ6L1\_VBFH125\_tautaul1317.deriv.DAOD\_HIGG3D1.e7259\_s3126\_r9364\_p4164

## A.2.2. mc16d

### Gluon-fusion $A$ production

mc16\_13TeV.346025.Pythia8EvtGen\_A14NNPDF23LO\_ggA60\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10201\_p4164  
 mc16\_13TeV.346026.Pythia8EvtGen\_A14NNPDF23LO\_ggA70\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10201\_p4164  
 mc16\_13TeV.346027.Pythia8EvtGen\_A14NNPDF23LO\_ggA80\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10201\_p4164  
 mc16\_13TeV.346028.Pythia8EvtGen\_A14NNPDF23LO\_ggA90\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10201\_p4164  
 mc16\_13TeV.346375.Pythia8EvtGen\_A14NNPDF23LO\_ggA50\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10201\_p4164  
 mc16\_13TeV.451705.Pythia8EvtGen\_A14NNPDF23LO\_ggA40\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e8139\_a875\_r10201\_p4164  
 mc16\_13TeV.801162.Py8EG\_A14NNPDF23LO\_ggA20\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10201\_p4164  
 mc16\_13TeV.801163.Py8EG\_A14NNPDF23LO\_ggA30\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10201\_p4164  
 mc16\_13TeV.801827.Py8EG\_A14NNPDF23LO\_ggA100\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10201\_p4164  
 mc16\_13TeV.801828.Py8EG\_A14NNPDF23LO\_ggA110\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10201\_p4164

$$Z/\gamma^* \rightarrow \tau\tau + \text{jets}$$

mc16\_13TeV.364128.Sherpa\_221\_NNPDF30NNLO\_Zt\_0\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364129.Sherpa\_221\_NNPDF30NNLO\_Zt\_0\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364130.Sherpa\_221\_NNPDF30NNLO\_Zt\_0\_BFilter.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364131.Sherpa\_221\_NNPDF30NNLO\_Zt\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364132.Sherpa\_221\_NNPDF30NNLO\_Zt\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364133.Sherpa\_221\_NNPDF30NNLO\_Zt\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364134.Sherpa\_221\_NNPDF30NNLO\_Zt\_140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364135.Sherpa\_221\_NNPDF30NNLO\_Zt\_140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364136.Sherpa\_221\_NNPDF30NNLO\_Zt\_140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364137.Sherpa\_221\_NNPDF30NNLO\_Zt\_280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364138.Sherpa\_221\_NNPDF30NNLO\_Zt\_280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5313\_s3126\_r10201\_p4164  
 mc16\_13TeV.364139.Sherpa\_221\_NNPDF30NNLO\_Zt\_280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5313\_s3126\_r10201\_p4164  
 mc16\_13TeV.364140.Sherpa\_221\_NNPDF30NNLO\_Zt\_500\_1000.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364141.Sherpa\_221\_NNPDF30NNLO\_Zt\_1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10201\_p4164  
 mc16\_13TeV.364210.Sherpa\_221\_NN30NNLO\_Ztt\_M110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364211.Sherpa\_221\_NN30NNLO\_Ztt\_M110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364212.Sherpa\_221\_NN30NNLO\_Ztt\_M110\_40\_MAXHTPTV0\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364213.Sherpa\_221\_NN30NNLO\_Ztt\_M110\_40\_MAXHTPTV0\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364214.Sherpa\_221\_NN30NNLO\_Ztt\_M110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364215.Sherpa\_221\_NN30NNLO\_Ztt\_M110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364282.Sherpa\_221\_NNPDF30NNLO\_Zt\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r10201\_p4164  
 mc16\_13TeV.364360.Sherpa\_221\_NNPDF30NNLO\_Zt\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10201\_p4164  
 mc16\_13TeV.364363.Sherpa\_221\_NNPDF30NNLO\_Zt\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10201\_p4164

## Diboson

mc16\_13TeV.363355.Sherpa\_221\_NNPDF30NNLO\_ZqqZvv.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10201\_p4164  
 mc16\_13TeV.363356.Sherpa\_221\_NNPDF30NNLO\_ZqqZ11.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10201\_p4164  
 mc16\_13TeV.363357.Sherpa\_221\_NNPDF30NNLO\_WqqZvv.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10201\_p4164  
 mc16\_13TeV.363358.Sherpa\_221\_NNPDF30NNLO\_WqqZ11.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10201\_p4164  
 mc16\_13TeV.363359.Sherpa\_221\_NNPDF30NNLO\_WpqqWmlv.deriv.DAOD\_HIGG3D1.e5583\_s3126\_r10201\_p4164  
 mc16\_13TeV.363360.Sherpa\_221\_NNPDF30NNLO\_WplvWmqq.deriv.DAOD\_HIGG3D1.e5983\_s3126\_r10201\_p4164  
 mc16\_13TeV.363489.Sherpa\_221\_NNPDF30NNLO\_WlvZqq.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10201\_p4164  
 mc16\_13TeV.364250.Sherpa\_222\_NNPDF30NNLO\_l111.deriv.DAOD\_HIGG3D1.e5894\_s3126\_r10201\_p4164  
 mc16\_13TeV.364253.Sherpa\_222\_NNPDF30NNLO\_l11v.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r10201\_p4164  
 mc16\_13TeV.364254.Sherpa\_222\_NNPDF30NNLO\_l1vv.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r10201\_p4166  
 mc16\_13TeV.364255.Sherpa\_222\_NNPDF30NNLO\_lvvv.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r10201\_p4164  
 mc16\_13TeV.364286.Sherpa\_222\_NNPDF30NNLO\_l1vvjj\_ss\_EW4.deriv.DAOD\_HIGG3D1.e6055\_s3126\_r10201\_p4164  
 mc16\_13TeV.364288.Sherpa\_222\_NNPDF30NNLO\_l111\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6096\_s3126\_r10201\_p4164  
 mc16\_13TeV.364289.Sherpa\_222\_NNPDF30NNLO\_l11v\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6133\_s3126\_r10201\_p4164  
 mc16\_13TeV.364290.Sherpa\_222\_NNPDF30NNLO\_l1vv\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6096\_s3126\_r10201\_p4164

## $W(\rightarrow l\nu) + \text{jets}$

mc16\_13TeV.364156.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364157.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364158.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364159.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364160.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364161.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364162.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364163.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364164.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364165.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364166.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364167.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364168.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364169.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364170.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364171.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364172.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364173.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364174.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364175.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364176.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364177.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364178.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364179.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364180.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364181.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364182.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364183.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364184.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164

mc16\_13TeV.364185.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364186.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364187.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364188.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364189.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV0\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364190.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364191.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364192.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364193.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364194.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364195.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364196.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164  
 mc16\_13TeV.364197.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10201\_p4164

### $W\gamma^*$

mc16\_13TeV.700015.Sh\_228\_evgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r10201\_p4164  
 mc16\_13TeV.700016.Sh\_228\_mvgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r10201\_p4569  
 mc16\_13TeV.700017.Sh\_228\_tvgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r10201\_p4164

### $Z/\gamma^*(\rightarrow \mu\mu) + \text{jets}$

mc16\_13TeV.364100.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364101.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364102.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364103.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364104.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364105.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364106.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364107.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364108.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364109.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364110.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364111.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364112.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364113.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10201\_p4164  
 mc16\_13TeV.364198.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364199.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364200.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364201.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364202.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364203.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164  
 mc16\_13TeV.364281.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r10201\_p4164  
 mc16\_13TeV.364359.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10201\_p4164  
 mc16\_13TeV.364362.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10201\_p4164

### $Z/\gamma^*(\rightarrow ee) + \text{jets}$

mc16\_13TeV.364114.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164  
 mc16\_13TeV.364115.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164  
 mc16\_13TeV.364116.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364117.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364118.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364119.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364120.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364121.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364122.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364123.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364124.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364125.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364126.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364127.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10201\_p4164

mc16\_13TeV.364204.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164

mc16\_13TeV.364205.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164

mc16\_13TeV.364206.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV70\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164

mc16\_13TeV.364207.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV70\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164

mc16\_13TeV.364208.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164

mc16\_13TeV.364209.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10201\_p4164

mc16\_13TeV.364280.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r10201\_p4164

mc16\_13TeV.364358.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10201\_p4164

mc16\_13TeV.364361.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10201\_p4164

## $t\bar{t}$ , single top

mc16\_13TeV.410470.Phy8EG\_A14\_ttbar\_hdamp258p75\_nonallhad.deriv.DAOD\_HIGG3D1.e6337\_s3126\_r10201\_p4164

mc16\_13TeV.410644.PowhegPythia8EvtGen\_A14\_singletop\_schan\_lept\_top.deriv.DAOD\_HIGG3D1.e6527\_s3126\_r10201\_p4164

mc16\_13TeV.410645.PowhegPythia8EvtGen\_A14\_singletop\_schan\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6527\_s3126\_r10201\_p4164

mc16\_13TeV.410648.PowhegPythia8EvtGen\_A14\_Wt\_DR\_dilepton\_top.deriv.DAOD\_HIGG3D1.e6615\_s3126\_r10201\_p4164

mc16\_13TeV.410649.PowhegPythia8EvtGen\_A14\_Wt\_DR\_dilepton\_antitop.deriv.DAOD\_HIGG3D1.e6615\_s3126\_r10201\_p4164

mc16\_13TeV.410658.Phy8EG\_A14\_tchan\_BW50\_lept\_top.deriv.DAOD\_HIGG3D1.e6671\_s3126\_r10201\_p4164

mc16\_13TeV.410659.Phy8EG\_A14\_tchan\_BW50\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6671\_s3126\_r10201\_p4164

## SM Higgs

mc16\_13TeV.345120.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaul1317.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10201\_p4164

mc16\_13TeV.345121.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaul15hp20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10201\_p4164

mc16\_13TeV.345122.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaulp15hm20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10201\_p4164

mc16\_13TeV.345123.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautauh30h20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10201\_p4164

mc16\_13TeV.345211.PowhegPy8EG\_NNPDF30\_AZNLO\_WmH125J\_Winc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r10201\_p4164

mc16\_13TeV.345212.PowhegPy8EG\_NNPDF30\_AZNLO\_WpH125J\_Winc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r10201\_p4164

mc16\_13TeV.345217.PowhegPy8EG\_NNPDF30\_AZNLO\_ZH125J\_Zinc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r10201\_p4164

mc16\_13TeV.345324.PowhegPythia8EvtGen\_NNLOPS\_NN30\_ggH125\_WW1v1v\_EF\_15\_5.deriv.DAOD\_HIGG3D1.e5769\_s3126\_r10201\_p4166

mc16\_13TeV.345948.PowhegPy8EG\_NNPDF30\_AZNLOCTEQ6L1\_VBFH125\_WW1v1v.deriv.DAOD\_HIGG3D1.e7172\_s3126\_r10201\_p4166

mc16\_13TeV.346190.PowhegPy8EG\_NNPDF30\_AZNLOCTEQ6L1\_VBFH125\_tautaul1317.deriv.DAOD\_HIGG3D1.e7259\_s3126\_r10201\_p4164

## A.2.3. mc16e

### Gluon-fusion $A$ production

mc16\_13TeV.346025.Pythia8EvtGen\_A14NNPDF23LO\_ggA60\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10724\_p4164  
 mc16\_13TeV.346026.Pythia8EvtGen\_A14NNPDF23LO\_ggA70\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10724\_p4164  
 mc16\_13TeV.346027.Pythia8EvtGen\_A14NNPDF23LO\_ggA80\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10724\_p4164  
 mc16\_13TeV.346028.Pythia8EvtGen\_A14NNPDF23LO\_ggA90\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10724\_p4164  
 mc16\_13TeV.346375.Pythia8EvtGen\_A14NNPDF23LO\_ggA50\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e7122\_a875\_r10724\_p4164  
 mc16\_13TeV.451705.Pythia8EvtGen\_A14NNPDF23LO\_ggA40\_tautau\_leplep.deriv.DAOD\_HIGG3D1.e8139\_a875\_r10724\_p4164  
 mc16\_13TeV.801162.Py8EG\_A14NNPDF23LO\_ggA20\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10724\_p4164  
 mc16\_13TeV.801163.Py8EG\_A14NNPDF23LO\_ggA30\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10724\_p4164  
 mc16\_13TeV.801827.Py8EG\_A14NNPDF23LO\_ggA100\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10724\_p4164  
 mc16\_13TeV.801828.Py8EG\_A14NNPDF23LO\_ggA110\_tautau\_l1314.deriv.DAOD\_HIGG3D1.e8307\_a875\_r10724\_p4164

### $Z/\gamma^* \rightarrow \tau\tau + \text{jets}$

mc16\_13TeV.364128.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364129.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364130.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364131.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV0\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364132.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV0\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364133.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV0\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364134.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364135.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364136.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364137.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364138.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5313\_s3126\_r10724\_p4164  
 mc16\_13TeV.364139.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5313\_s3126\_r10724\_p4164  
 mc16\_13TeV.364140.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364141.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5307\_s3126\_r10724\_p4164  
 mc16\_13TeV.364210.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364211.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364212.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV0\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364213.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV0\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364214.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364215.Sherpa\_221\_NN30NNLO\_Ztt\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364282.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r10724\_p4164  
 mc16\_13TeV.364360.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10724\_p4164  
 mc16\_13TeV.364363.Sherpa\_221\_NNPDF30NNLO\_Ztautau\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10724\_p4164

### Diboson

mc16\_13TeV.363355.Sherpa\_221\_NNPDF30NNLO\_ZqqZvv.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10724\_p4164  
 mc16\_13TeV.363356.Sherpa\_221\_NNPDF30NNLO\_ZqqZll.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10724\_p4164  
 mc16\_13TeV.363357.Sherpa\_221\_NNPDF30NNLO\_WqqZvv.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10724\_p4164  
 mc16\_13TeV.363358.Sherpa\_221\_NNPDF30NNLO\_WqqZll.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10724\_p4164  
 mc16\_13TeV.363359.Sherpa\_221\_NNPDF30NNLO\_WpqqWmlv.deriv.DAOD\_HIGG3D1.e5583\_s3126\_r10724\_p4164  
 mc16\_13TeV.363360.Sherpa\_221\_NNPDF30NNLO\_WpplvWmqq.deriv.DAOD\_HIGG3D1.e5983\_s3126\_r10724\_p4164  
 mc16\_13TeV.363489.Sherpa\_221\_NNPDF30NNLO\_WlvZqq.deriv.DAOD\_HIGG3D1.e5525\_s3126\_r10724\_p4164  
 mc16\_13TeV.364250.Sherpa\_222\_NNPDF30NNLO\_l111.deriv.DAOD\_HIGG3D1.e5894\_s3126\_r10724\_p4164

mc16\_13TeV.364253.Sherpa\_222\_NNPDF30NNLO\_lllv.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r10724\_p4164  
mc16\_13TeV.364254.Sherpa\_222\_NNPDF30NNLO\_llvv.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r10724\_p4166  
mc16\_13TeV.364255.Sherpa\_222\_NNPDF30NNLO\_lvvv.deriv.DAOD\_HIGG3D1.e5916\_s3126\_r10724\_p4164  
mc16\_13TeV.364286.Sherpa\_222\_NNPDF30NNLO\_llvvjj\_ss\_EW4.deriv.DAOD\_HIGG3D1.e6055\_s3126\_r10724\_p4164  
mc16\_13TeV.364288.Sherpa\_222\_NNPDF30NNLO\_l111\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6096\_s3126\_r10724\_p4164  
mc16\_13TeV.364289.Sherpa\_222\_NNPDF30NNLO\_lllv\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6133\_s3126\_r10724\_p4164  
mc16\_13TeV.364290.Sherpa\_222\_NNPDF30NNLO\_llvv\_lowM1PtComplement.deriv.DAOD\_HIGG3D1.e6096\_s3126\_r10724\_p4164

## $W(\rightarrow l\nu) + \text{jets}$

mc16\_13TeV.364156.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364157.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364158.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364159.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364160.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364161.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364162.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364163.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364164.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364165.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364166.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364167.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364168.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364169.Sherpa\_221\_NNPDF30NNLO\_Wmunu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364170.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364171.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364172.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364173.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364174.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364175.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364176.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364177.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364178.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364179.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364180.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364181.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364182.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364183.Sherpa\_221\_NNPDF30NNLO\_Wenu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364184.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364185.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364186.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364187.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV70\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364188.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV70\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364189.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV70\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364190.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364191.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364192.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364193.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
mc16\_13TeV.364194.Sherpa\_221\_NNPDF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164

mc16\_13TeV.364195.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
 mc16\_13TeV.364196.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164  
 mc16\_13TeV.364197.Sherpa\_221\_NNPDPF30NNLO\_Wtaunu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5340\_s3126\_r10724\_p4164

### $W\gamma^*$

mc16\_13TeV.700015.Sh\_228\_evgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r10724\_p4164  
 mc16\_13TeV.700016.Sh\_228\_mvgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r10724\_p4569  
 mc16\_13TeV.700017.Sh\_228\_tvgamma\_pty7\_EnhMaxpTVpTy.deriv.DAOD\_HIGG3D1.e7947\_s3126\_r10724\_p4164

### $Z/\gamma^*(\rightarrow \mu\mu) + \text{jets}$

mc16\_13TeV.364100.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364101.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364102.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364103.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364104.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364105.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV0\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364106.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364107.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364108.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364109.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364110.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364111.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364112.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364113.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5271\_s3126\_r10724\_p4164  
 mc16\_13TeV.364198.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364199.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364200.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364201.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV0\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364202.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364203.Sherpa\_221\_NN30NNLO\_Zmm\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364281.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r10724\_p4164  
 mc16\_13TeV.364359.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10724\_p4164  
 mc16\_13TeV.364362.Sherpa\_221\_NNPDPF30NNLO\_Zmumu\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10724\_p4164

### $Z/\gamma^*(\rightarrow ee) + \text{jets}$

mc16\_13TeV.364114.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_70\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364115.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_70\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364116.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364117.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_140\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364118.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_140\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364119.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV0\_140\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364120.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV140\_280\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364121.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV140\_280\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364122.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV140\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364123.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV280\_500\_CVetoBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364124.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV280\_500\_CFilterBVeto.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364125.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV280\_500\_BFilter.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364126.Sherpa\_221\_NNPDPF30NNLO\_Zee\_MAXHTPTV500\_1000.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164



mc16\_13TeV.364127.Sherpa\_221\_NNPDF30NNLO\_Zee\_MAXHTPTV1000\_E\_CMS.deriv.DAOD\_HIGG3D1.e5299\_s3126\_r10724\_p4164  
 mc16\_13TeV.364204.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV0\_70\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364205.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV0\_70\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364206.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV70\_280\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364207.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV70\_280\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364208.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BVeto.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364209.Sherpa\_221\_NN30NNLO\_Zee\_M1110\_40\_MAXHTPTV280\_E\_CMS\_BFilter.deriv.DAOD\_HIGG3D1.e5421\_s3126\_r10724\_p4164  
 mc16\_13TeV.364280.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV280\_E\_CMS.deriv.DAOD\_HIGG3D1.e6037\_s3126\_r10724\_p4164  
 mc16\_13TeV.364358.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV70\_140.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10724\_p4164  
 mc16\_13TeV.364361.Sherpa\_221\_NNPDF30NNLO\_Zee\_M112M1\_MAXHTPTV140\_280.deriv.DAOD\_HIGG3D1.e6544\_a875\_r10724\_p4164

## $t\bar{t}$ , single top

mc16\_13TeV.410470.Phy8EG\_A14\_ttbar\_hdamp258p75\_nonallhad.deriv.DAOD\_HIGG3D1.e6337\_s3126\_r10724\_p4164  
 mc16\_13TeV.410644.PowhegPythia8EvtGen\_A14\_singletop\_schan\_lept\_top.deriv.DAOD\_HIGG3D1.e6527\_s3126\_r10724\_p4164  
 mc16\_13TeV.410645.PowhegPythia8EvtGen\_A14\_singletop\_schan\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6527\_s3126\_r10724\_p4164  
 mc16\_13TeV.410645.PowhegPythia8EvtGen\_A14\_singletop\_schan\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6527\_s3126\_r9364\_p4164  
 mc16\_13TeV.410648.PowhegPythia8EvtGen\_A14\_Wt\_DR\_dilepton\_top.deriv.DAOD\_HIGG3D1.e6615\_s3126\_r10724\_p4164  
 mc16\_13TeV.410649.PowhegPythia8EvtGen\_A14\_Wt\_DR\_dilepton\_antitop.deriv.DAOD\_HIGG3D1.e6615\_s3126\_r10724\_p4164  
 mc16\_13TeV.410658.Phy8EG\_A14\_tchan\_BW50\_lept\_top.deriv.DAOD\_HIGG3D1.e6671\_s3126\_r10724\_p4164  
 mc16\_13TeV.410659.Phy8EG\_A14\_tchan\_BW50\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6671\_s3126\_r10724\_p4164  
 mc16\_13TeV.410659.Phy8EG\_A14\_tchan\_BW50\_lept\_antitop.deriv.DAOD\_HIGG3D1.e6671\_s3126\_r9364\_p4164

## SM Higgs

mc16\_13TeV.345120.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaul1317.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10724\_p4164  
 mc16\_13TeV.345121.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaulm15hp20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10724\_p4164  
 mc16\_13TeV.345122.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautaulp15hm20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10724\_p4164  
 mc16\_13TeV.345123.PowhegPy8EG\_NNLOPS\_nnlo\_30\_ggH125\_tautauh30h20.deriv.DAOD\_HIGG3D1.e5814\_s3126\_r10724\_p4164  
 mc16\_13TeV.345211.PowhegPy8EG\_NNPDF30\_AZNLO\_WmH125J\_Winc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r10724\_p4164  
 mc16\_13TeV.345212.PowhegPy8EG\_NNPDF30\_AZNLO\_WpH125J\_Winc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r10724\_p4164  
 mc16\_13TeV.345217.PowhegPy8EG\_NNPDF30\_AZNLO\_ZH125J\_Zinc\_MINLO\_tautau.deriv.DAOD\_HIGG3D1.e5808\_s3126\_r10724\_p4164  
 mc16\_13TeV.345324.PowhegPythia8EvtGen\_NNLOPS\_NN30\_ggH125\_WW1v1v\_EF\_15\_5.deriv.DAOD\_HIGG3D1.e5769\_s3126\_r10724\_p4164  
 mc16\_13TeV.345948.PowhegPy8EG\_NNPDF30\_AZNLOCTEQ6L1\_VBFH125\_WW1v1v.deriv.DAOD\_HIGG3D1.e7172\_s3126\_r10724\_p4166  
 mc16\_13TeV.346190.PowhegPy8EG\_NNPDF30\_AZNLOCTEQ6L1\_VBFH125\_tautaul1317.deriv.DAOD\_HIGG3D1.e7259\_s3126\_r10724\_p4164

### A.3. Full List of Triggers

Subsection 4.3.1 gives an overview on the different HLT triggers used for this thesis. Here, a complete list, with the full names of the triggers for the different data-taking periods, is given in Table A.1. All of them are combined electron-muon triggers.

Table A.1.: Triggers for the different data-taking years, used in this thesis with their full names.

Year	Short name	Full ATLAS trigger name
2015	e7mu24	HLT_e7_lhmedium_mu24
	e17mu14	HLT_e17_lhloose_mu14
	e26mu8	HLT_e24_lhmedium_nod0_L1EM20VHI_mu8noL1
2016	e7mu24	HLT_e7_lhmedium_nod0_mu24
	e17mu14	HLT_e17_lhloose_nod0_mu14
	e26mu8	HLT_e26_lhmedium_nod0_L1EM22VHI_mu8noL1
2017-18	e7mu24	HLT_e7_lhmedium_nod0_mu24
	e17mu14	HLT_e17_lhloose_nod0_mu14
	e26mu8	HLT_e26_lhmedium_nod0_mu8noL1

## B. Dependence of MM Efficiencies on the $Z \rightarrow \tau\tau$ Reweighting

This study looks into the influence of the  $Z \rightarrow \tau\tau$  reweighting, discussed in Section 4.5, on the efficiencies needed for the Matrix Method (see Section 4.4). All efficiencies were calculated once without applying the reweighting weights and once with. Figure B.1 shows the direct comparison of a selection of the different efficiencies in the two cases.

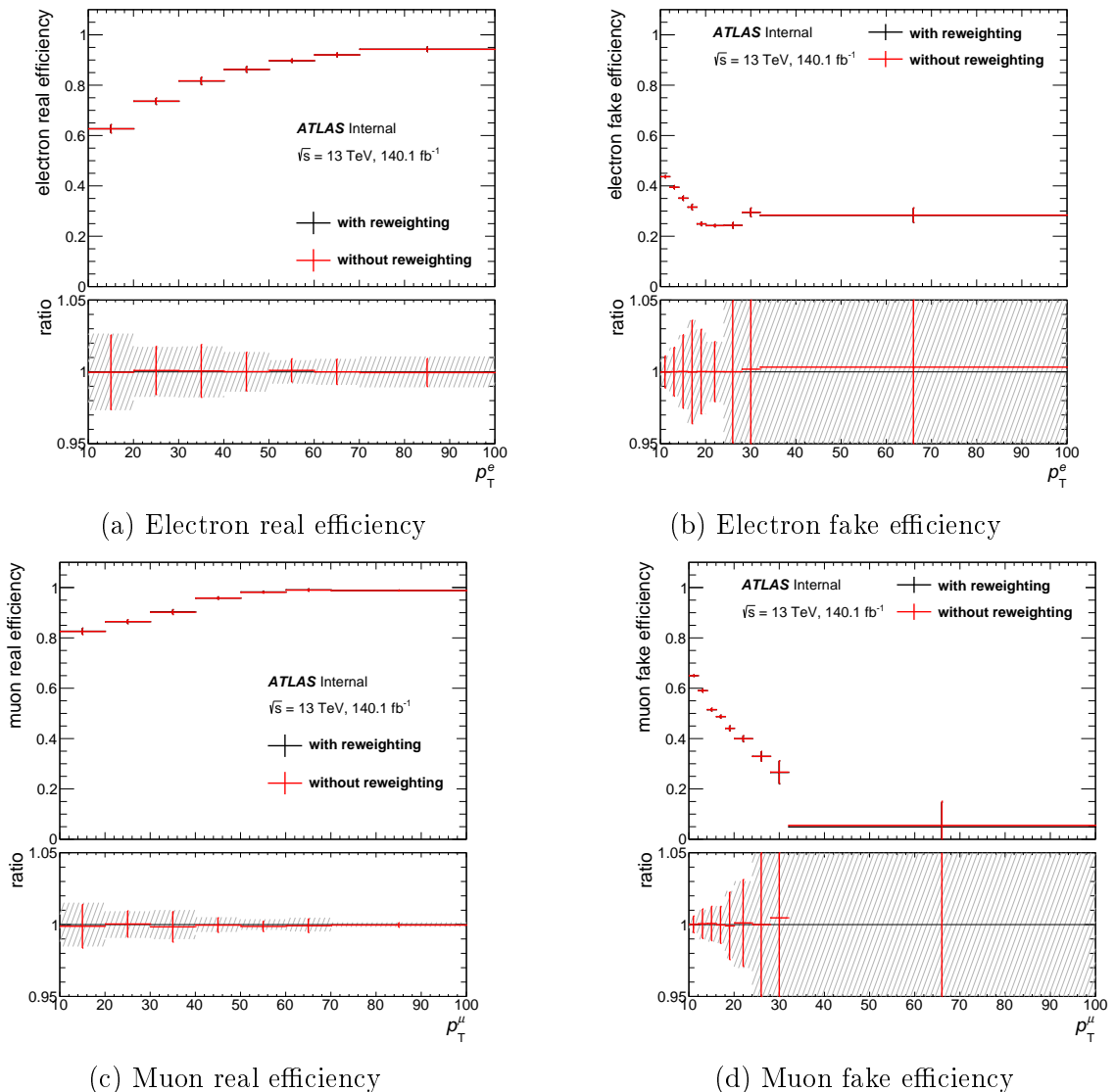


Figure B.1.: Comparing efficiencies calculated with (black) and without (red) the  $Z \rightarrow \tau\tau$  reweighting applied. Selection of efficiencies calculated for the HMSR with the other lepton being tight.

As expected, the differences are very small, for the majority of efficiencies they are below 1% in all bins. Only for muon fake efficiencies the relative differences of the last bin goes up to 10% because the efficiencies are particularly small there.

In all cases the efficiencies calculated with and without reweighting are statistically compatible. All in all, calculating the efficiencies after the reweighting seems to lead to only minor changes. Therefore it is sufficient to use the ones calculated before reweighting.

# C. Mismodelling and Reweighting Studies on the $Z \rightarrow \tau\tau$ -background

This appendix complements Section 4.5. At first, more variable distributions in the ZVR without any reweighting applied are shown in Figure C.1. Here one can see, that not all variables are affected by it in the same way. Additionally, there is the  $E_T^{\text{miss}}$  and  $n_{\text{jets}}$  distributions in the Top VR shown in Figure C.2. These two distributions do not show this strong mismodelling, which comes from overestimating the background processes.

There were several studies performed to investigate the mismodelling observed in the ZVR. Here there are two examples of these studies.

The first one investigated the dependence of the mismodelling on the data-taking year/ the MC campaigns. The distribution of  $E_T^{\text{miss}}$ , with data only recorded in individual years and background from the respective MC campaign, is shown in Figure C.3. It is visible that the distribution from 2015/2016 shows the least mismodelled behaviour. Meanwhile, the distribution from 2018 seems to be the worst of the three. One has to keep in mind, however, that this is the data-taking year with the best statistics and therefore the most meaningful results. All in all, it cannot be concluded from this study that it would be beneficial to exclude data and background from a specific year, especially as this will impair the statistical uncertainties drastically.

Another study delves into the dependence of the mismodelling with the triggers used. Figure C.4 shows the  $E_T^{\text{miss}}$  distribution of events that were selected by the three different electron-muon triggers. All three distributions show a mismodelling to some degree. While one could argue that the distribution triggered by e7mu24 looks the least mismodelled, there is still some disagreement between data and background especially between 30 GeV and 70 GeV. This study shows that there is no strong connection between the three triggers and the mismodelling. It may be interesting for further studies to look if this dependence changes if one uses single lepton triggers instead of the three combined triggers compared here.

Lastly, it was checked how the calculated limits change, if the reweighting including its systematic uncertainties are applied or not applied. It is important to add that the goal of the reweighting was to improve the observed mismodelling and not the final limits. Nevertheless, it is interesting to investigate the influence, especially of the large additional uncertainties. This can be seen in Figure C.5. It is clear that for the two lowest mass hypotheses of 20 GeV and 30 GeV the limit is worsened up to 15% by introducing the reweighting and associated uncertainties. For higher mass hypotheses a slight improvement of the limit can be observed, which can be explained by reducing the amount of background events. The best improvement of around 10% is visible at 70 GeV.

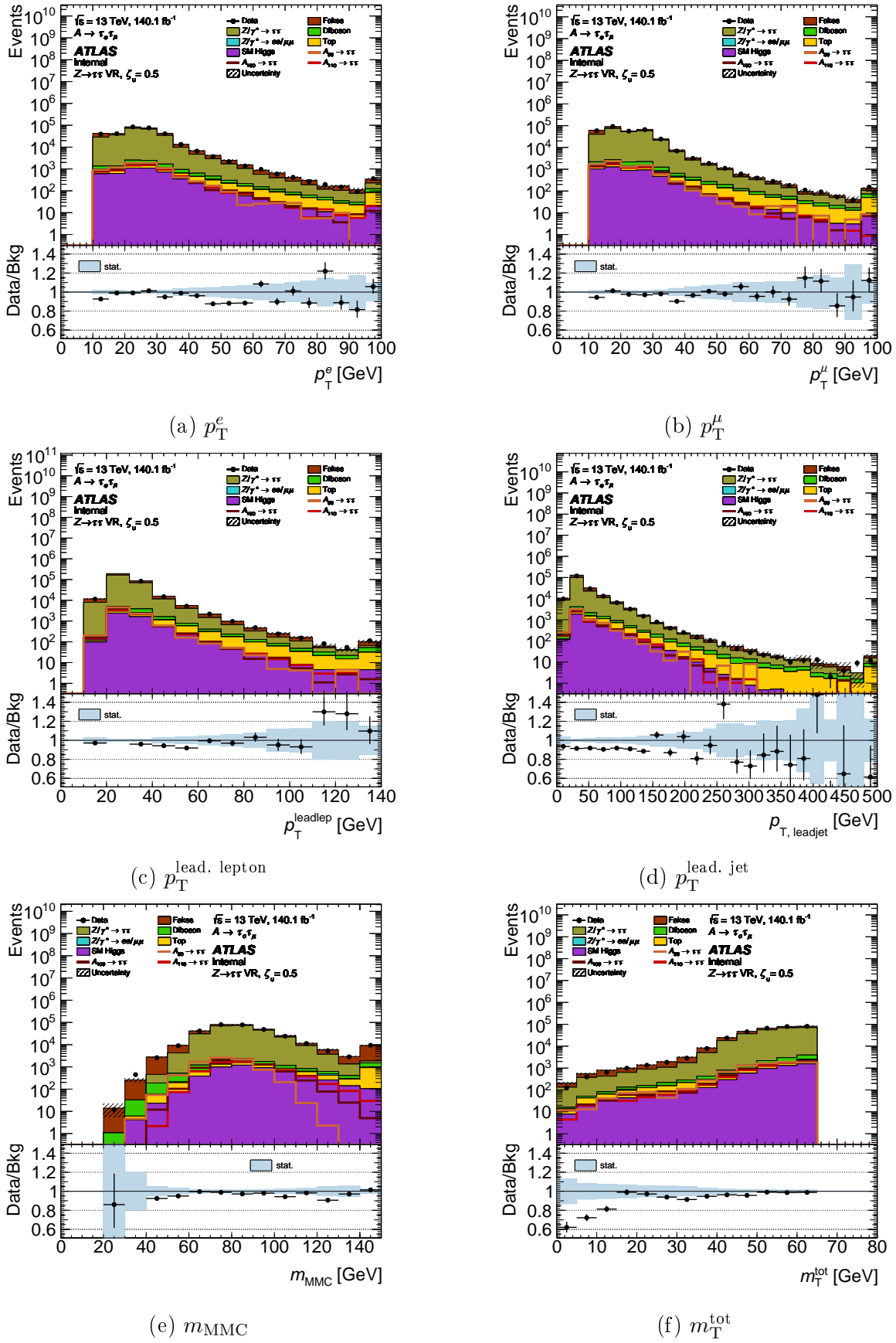


Figure C.1.: Distribution of different variables in the ZVR without applying the  $Z \rightarrow \tau\tau$  reweighting.

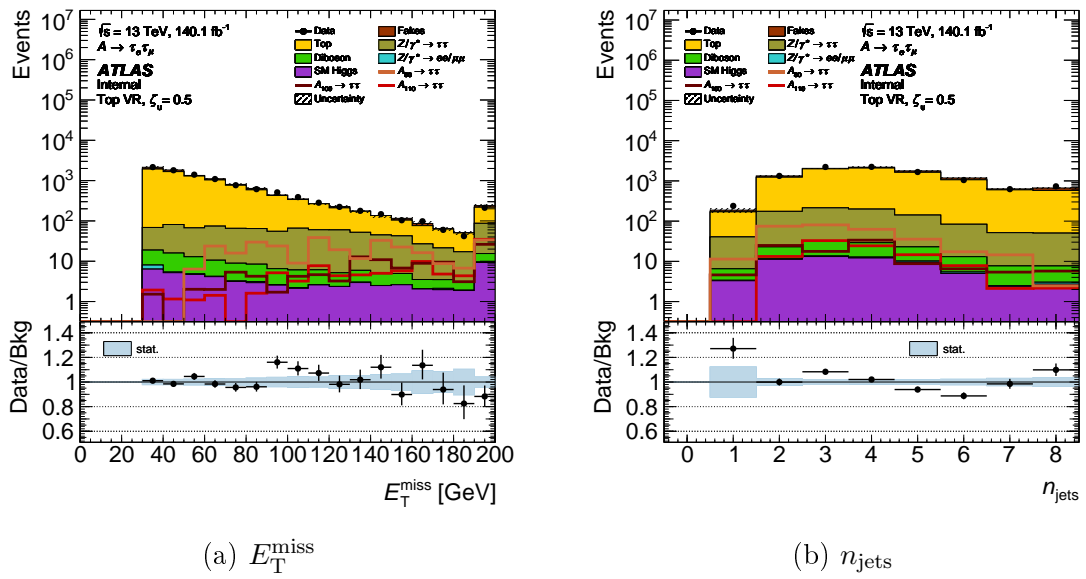
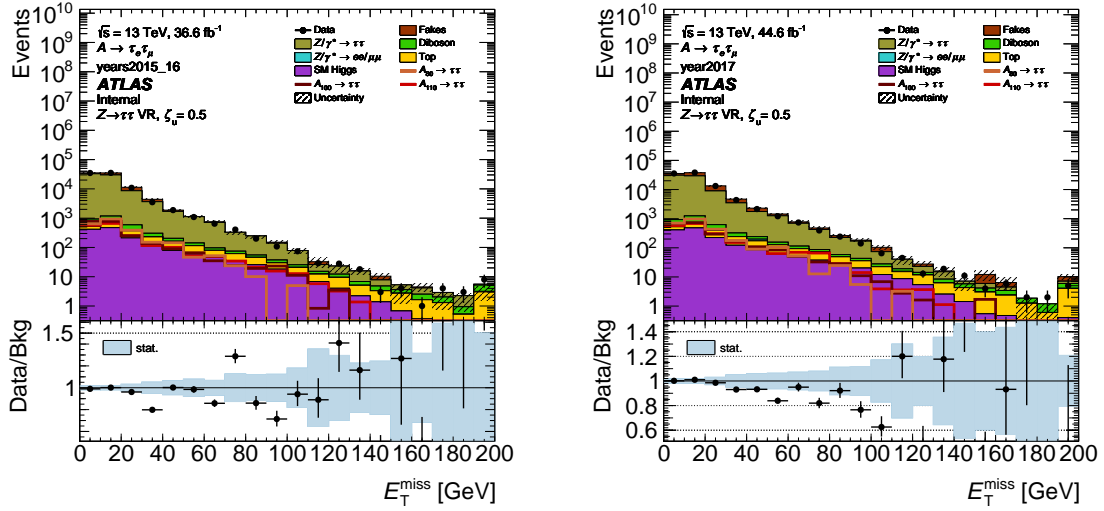
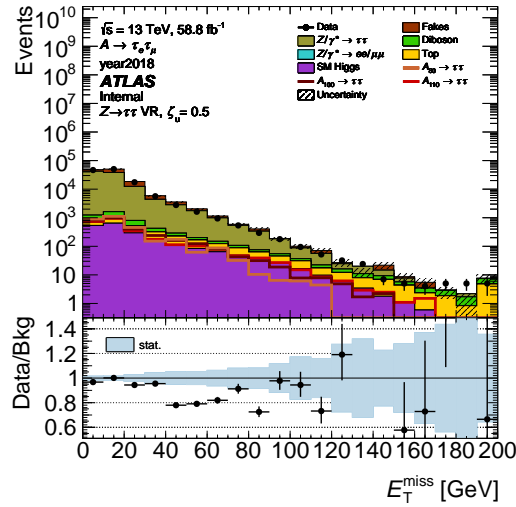


Figure C.2.: Distributions of variables in the TVR, which show a mismodelling in the ZVR.



(a) 2015/2016

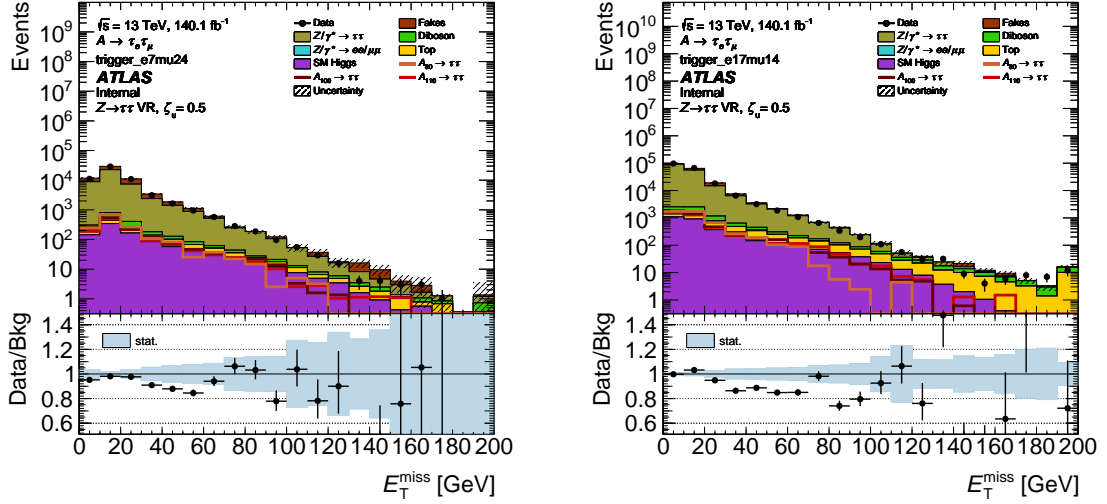
(b) 2017



(c) 2018

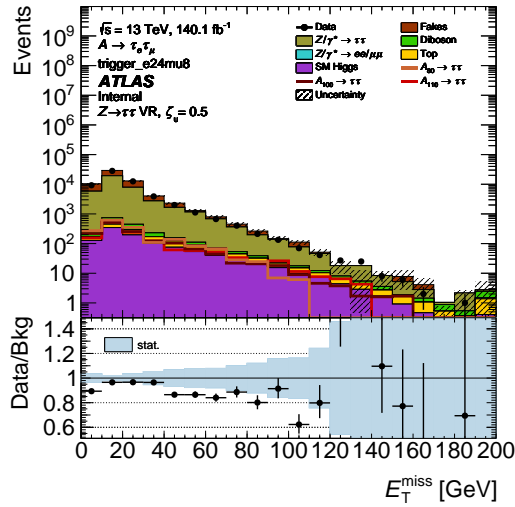
Figure C.3.: Mismodelling of  $E_T^{\text{miss}}$  in the ZVR for different data-taking periods and MC campaigns.





(a) e7mu24

(b) e17mu14



(c) e24mu8

Figure C.4.: Mismodelling of  $E_T^{\text{miss}}$  in the ZVR for different combined electron-muon triggers.

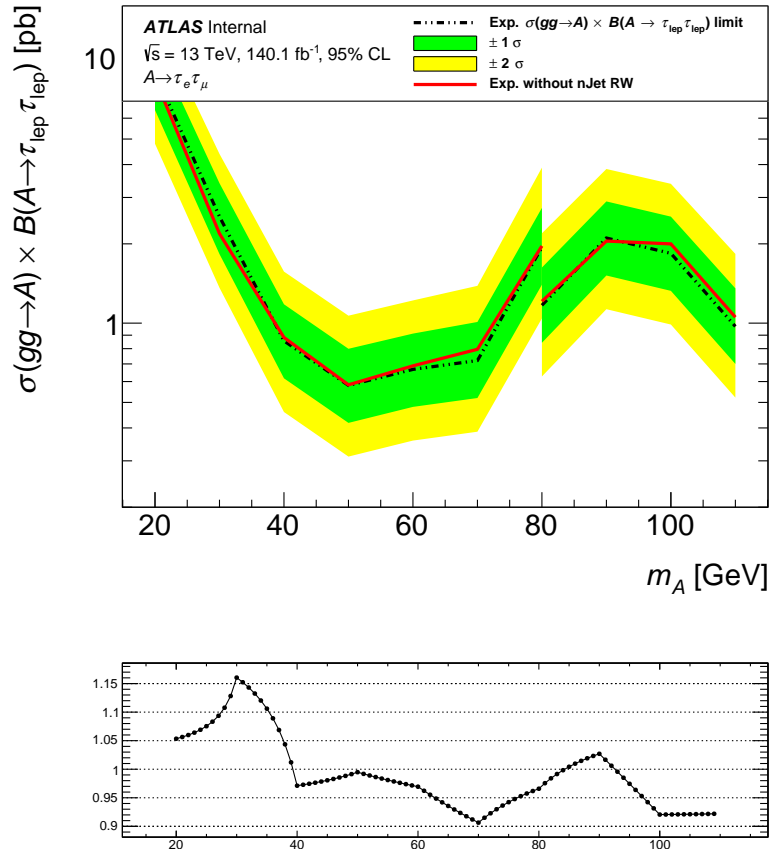


Figure C.5.: Influence of the reweighting of the  $Z \rightarrow \tau\tau$  samples, including the additional systematic uncertainties, on the expected cross section times branching ratio limit. The dashed line describes the limit **with** reweighting applied and the red line shows it **without**. The plot at the bottom shows the ratio of the two limits.

# D. Neural Network Extension to the Analysis

## D.1. Example Configuration of the Hyperparameter Optimisation Framework

```
import os
import numpy as np
from OPTIMA import KerasTools
from OPTIMA import FiguresOfMerit

# input data and splittings
use_testing_dataset = True
fixed_testing_dataset = False
use_custom_splitting = True
custom_splitting_val = 0
custom_splitting_test = 1
custom_splitting_N = 5
validation_fraction = 0.2
test_fraction = 0.2
max_num_events = np.inf

# Initial input variables
input_vars = [
    "met",
    "ele_pt1",
    "ele_eta0",
    "ele_phi0",
    "muon_pt1",
    "muon_eta0",
    "muon_phi0",
    "mTtot",
    "MMC_mass",
    "higgs_pt",
    "m_lep_lep",
    "dR_lep_lep",
    "dPhi_leadlep_met",
    "dPhi_subleadlep_met",
    "jet_0_pt",
    "mass_A",
    "d0sig_ele1",
```

```
"d0sig_muon1",
"delta_z0_sinTheta_ele1",
"delta_z0_sinTheta_muon1",
"jet_eta0",
"jet_phi0",
"n_jets",
"m_jet_jet",
"dPhi_lep_lep",
"dPhi_leadjet_met"
]

# input scaling inputs_scaling_dict = {
    "met": (lambda x: x, "met"),
    "ele_pt1": (lambda x: x, "ele_pt1"),
    "ele_eta0": (lambda x: x, "ele_eta0"),
    "ele_phi0": (lambda x: x, "ele_phi0"),
    "muon_pt1": (lambda x: x, "muon_pt1"),
    "muon_eta0": (lambda x: x, "muon_eta0"),
    "muon_phi0": (lambda x: x, "muon_phi0"),
    "mTtot": (lambda x: x, "mTtot"),
    "MMC_mass": (lambda x: x, "MMC_mass"),
    "higgs_pt": (lambda x: x, "higgs_pt"),
    "m_lep_lep": (lambda x: x, "m_lep_lep"),
    "dR_lep_lep": (lambda x: x, "dR_lep_lep"),
    "dPhi_leadlep_met": (lambda x: x, "dPhi_leadlep_met"),
    "dPhi_subleadlep_met": (lambda x: x, "dPhi_subleadlep_met"),
    "jet_0_pt": (lambda x: x, "jet_0_pt"),
    "mass_A": (lambda x: x, "mass_A"),
    "d0sig_ele1": (lambda x: x, "d0sig_ele1"),
    "d0sig_muon1": (lambda x: x, "d0sig_muon1"),
    "delta_z0_sinTheta_ele1": (lambda x: x, "delta_z0_sinTheta_ele1"),
    "delta_z0_sinTheta_muon1": (lambda x: x, "delta_z0_sinTheta_muon1"),
    "jet_eta0": (lambda x: x, "jet_eta0"),
    "jet_phi0": (lambda x: x, "jet_phi0"),
    "n_jets": (lambda x: x, "n_jets"),
    "m_jet_jet": (lambda x: x, "m_jet_jet"),
    "dPhi_lep_lep": (lambda x: x, "dPhi_lep_lep"),
    "dPhi_leadjet_met": (lambda x: x, "dPhi_leadjet_met")
}

# general settings for the optimization
monitor_name = 'val_bce_loss'
monitor_op = 'min'
optimize_on_best_value = False
restore_on_best_checkpoint = False
max_epochs = 1000
early_stopping_patience = 20
overtraining_patience = 6
```

---

---

```
# settings for the evaluation
fit_min_R_squared = 0.95
check_overtraining_with_fit = True
use_early_stopping_for_crossvalidation = False
use_OT_conditions_for_crossvalidation = False

# settings for the initial optimization phase and variable optimization
perform_variable_opt = True
use_fit_results_for_varOpt = False
variable_optimization_mode = 'shuffle'
num_samples_variableOpt = 400
var_metric = 'bce_loss'
var_metric_op = 'min'
max_rel_change = 0.02
num_shuffles = 1
use_median_for_averages = True
choose_best_var = True
var_opt_patience = 3

# settings for Optuna and the Asynchronous HyperBand scheduler (ASHA)
use_optuna_opt = True
use_TPESampler = True
use_ASHAScheduler = True
num_samples_optuna = 1000
ASHA_grace_period = 10
ASHA_max_t = 75
ASHA_reduction_factor = 2
use_multivariate_TPE = False

# PBT specific settings
use_population_based_training = False
use_fit_results_for_PBT = False
num_samples_PBT = 400
perturbation_interval = 4
burn_in_period = 6

# hyperparameter search space
search_space = {
    "num_layers": ('uniform', 2, 3, 1),
    "units_1": ('log', 4, 128, 1),
    "units_2": ('log', 4, 128, 1),
    "units_3": ('log', 4, 128, 1),
    "activation": 'sigmoid',
    "kernel_initializer": 'auto',
    "bias_initializer": 'auto',
    "l1_lambda": 0,
    "l2_lambda": 0,
```

---

```
"dropout": 0,
"batch_size": ('log', 4, 256),
"learning_rate": ('log', 1e-5, 1e-2),
"Adam_beta_1": 0.9,
"one_minus_Adam_beta_2": 0.9999,
"Adam_epsilon": 1e-7,
"loss_function": 'BinaryCrossentropy',
"loss_signal_weight": 1.
}
optimize_units_per_layer = True

# metrics custom_metrics = [
    ('bce_loss',
     KerasTools.WeightedBinaryCrossentropy(signal_weight=1.0,
                                             only_numpy=True).calc_loss)
]
custom_Keras_metrics = []
custom_weighted_Keras_metrics = []
composite_metrics = []

# overtraining conditions
overtraining_conditions = []

# custom objects
custom_objects = {}
```

---

---

## D.2. Distribution of All Input Variables to the NN

This section provides distributions of all input variables to the NN in different regions. This includes the mvaVR, which is used to extract training data, but also the ZVR to check the modelling of input variables.

The distributions of input variables to the NN can be expected to differ to the ones, that can be found in this section. The plots shown here, utilise the event weights that are used in the base analysis, the weights used during the training of the NN are slightly different. A description of the event weights used during training can be found in Section 6.1. This should only lead to smaller differences in the distributions, as the main difference between the weighting strategies are scale factors that are applied in the base analysis, but are close to one for all events.

Also the standard Matrix Method is used here for the estimation of fake background, this also implies adding a cut on the truth information of all MC generated backgrounds. This cut is not used for the input to the NN. Also the fake events are represented by data from a not-Tight mvaVR, with their weights simply matched to the total normal background contribution. This procedure is described in more detail in Section 6.1. Unfortunately it was not possible to implement this fake estimation into the base analysis that can produce plots of the distributions.

### D.2.1. Distributions in the mvaVR

This section shows the distributions of all variables that are used as input for the NN. Here the distributions in the mvaVR, which is the region that is used for the training of the NN, are displayed. They are generated without using the  $Z \rightarrow \tau\tau$  reweighting that was introduced in Section 4.5, as it is not yet implemented for the calculation of weights for the training.

---

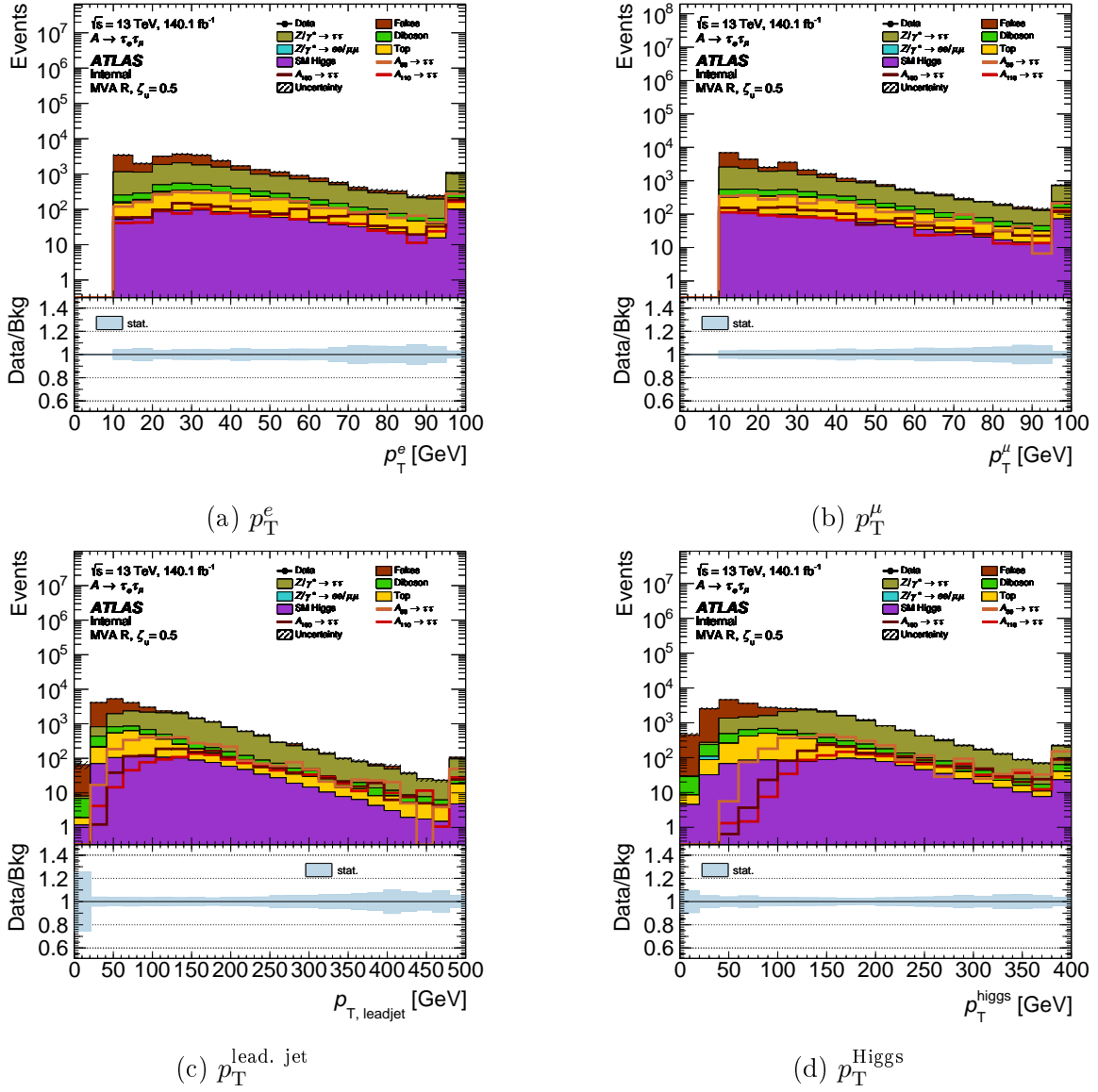


Figure D.1.: Distributions of different variables in the mvaVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.



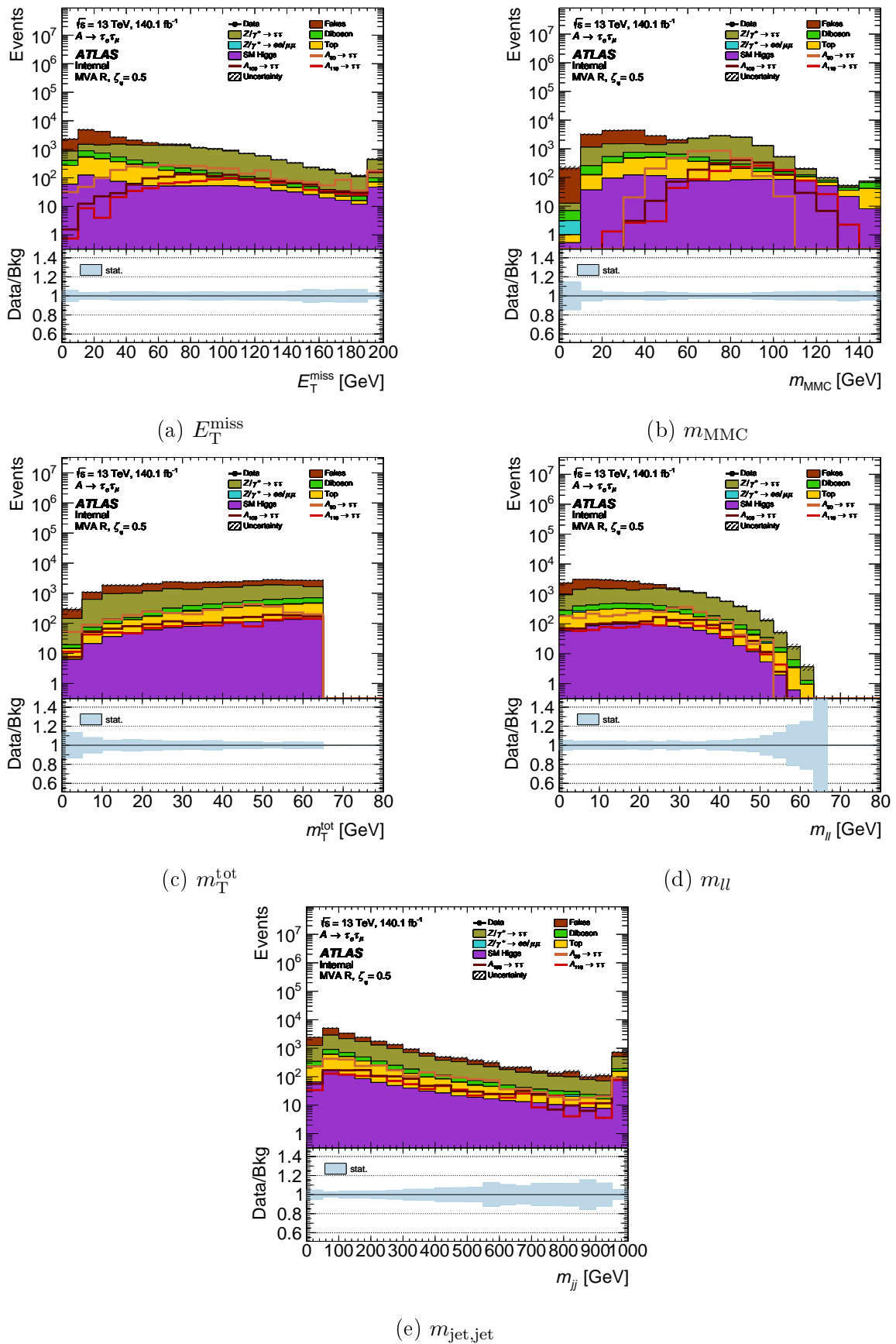


Figure D.2.: More distributions of different variables in the mvaVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.

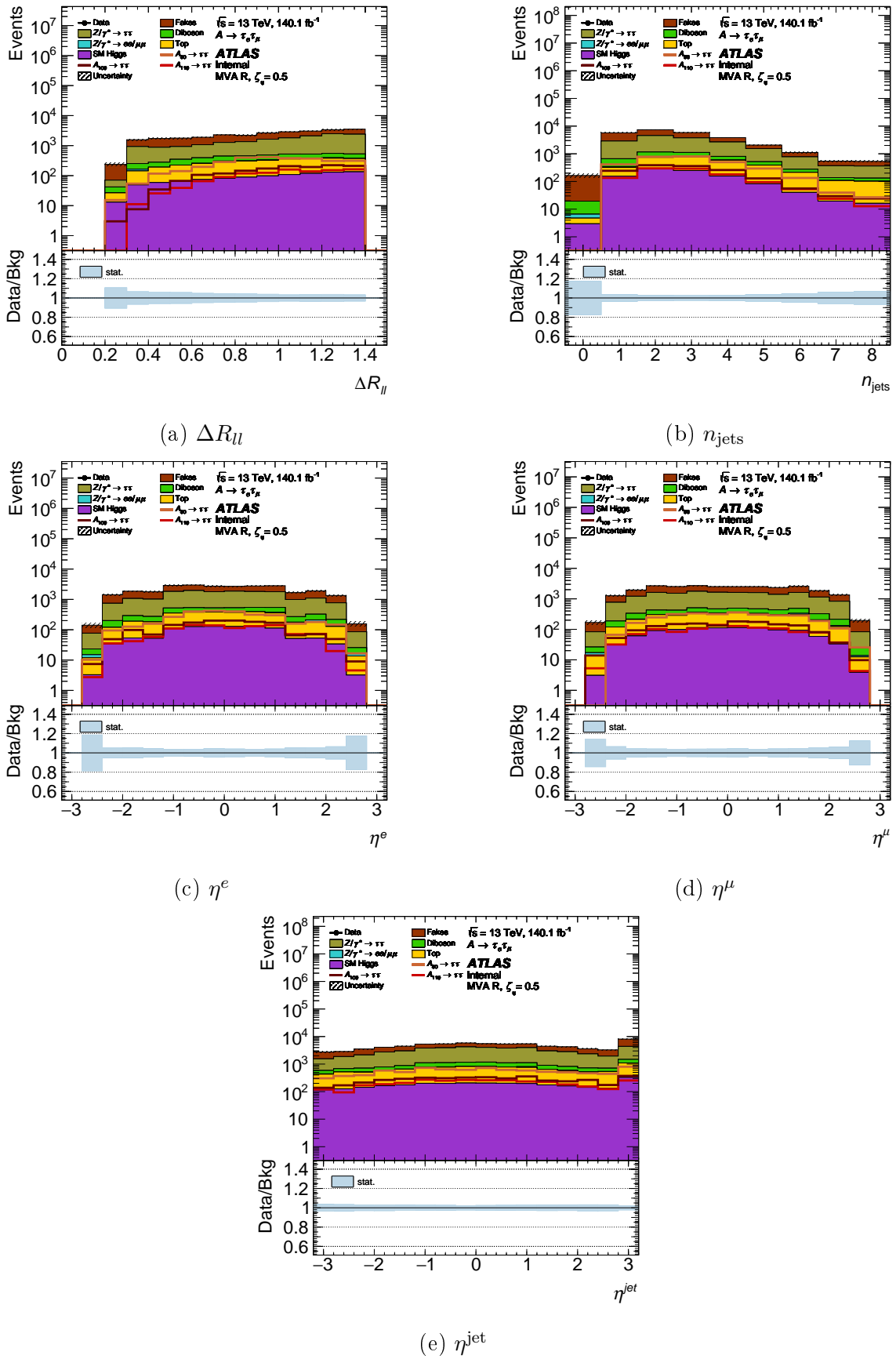


Figure D.3.: More distributions of different variables in the mvaVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.

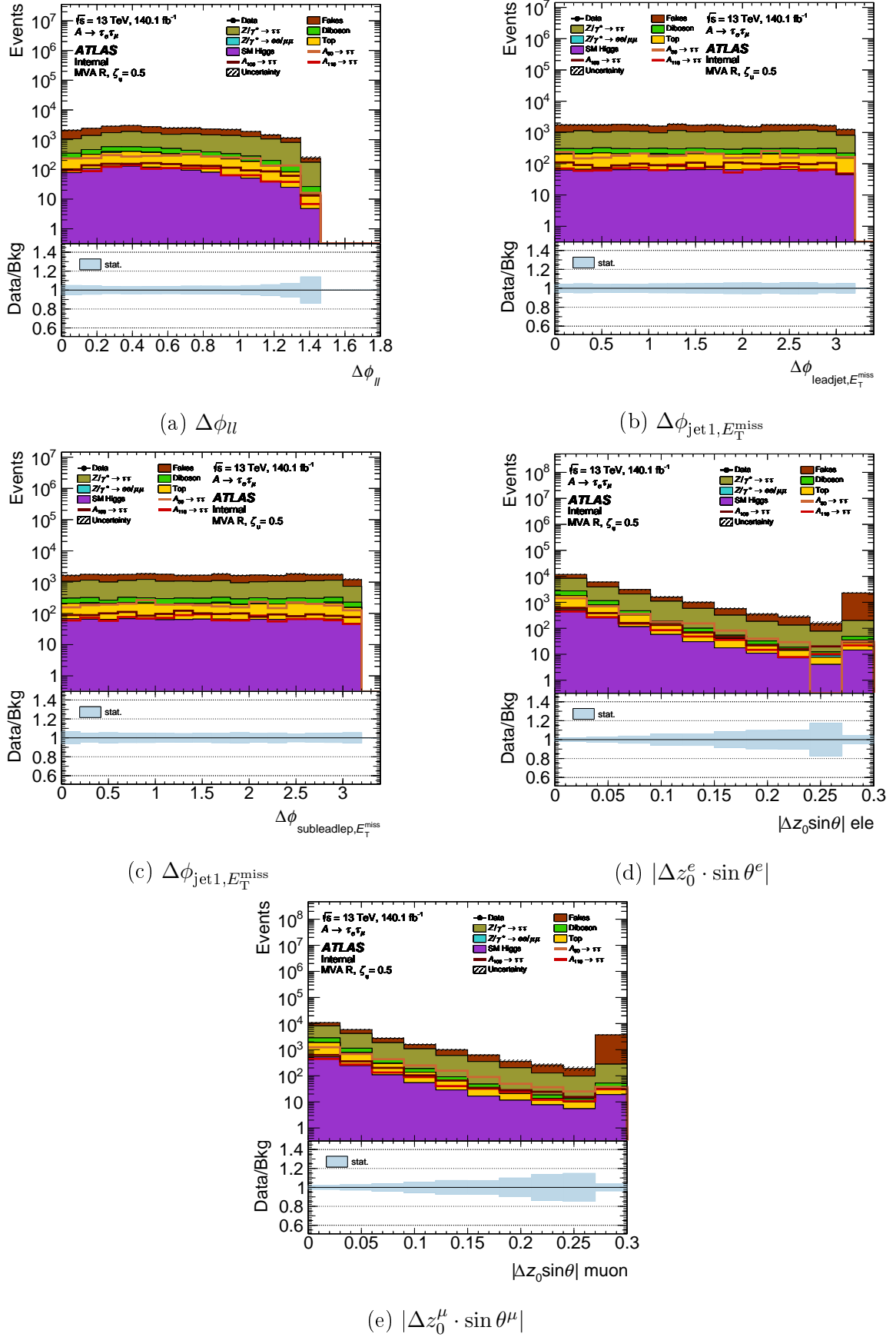


Figure D.4.: More distributions of different variables in the mvaVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.

### D.2.2. Distributions in the ZVR

This section shows the distributions of all variables that are used as input for the NN. Here the distributions in the ZVR are displayed, to investigate their modelling. They are generated without using the  $Z \rightarrow \tau\tau$  reweighting that was introduced in Section 4.5, as it is not yet implemented for the calculation of weights for the training.

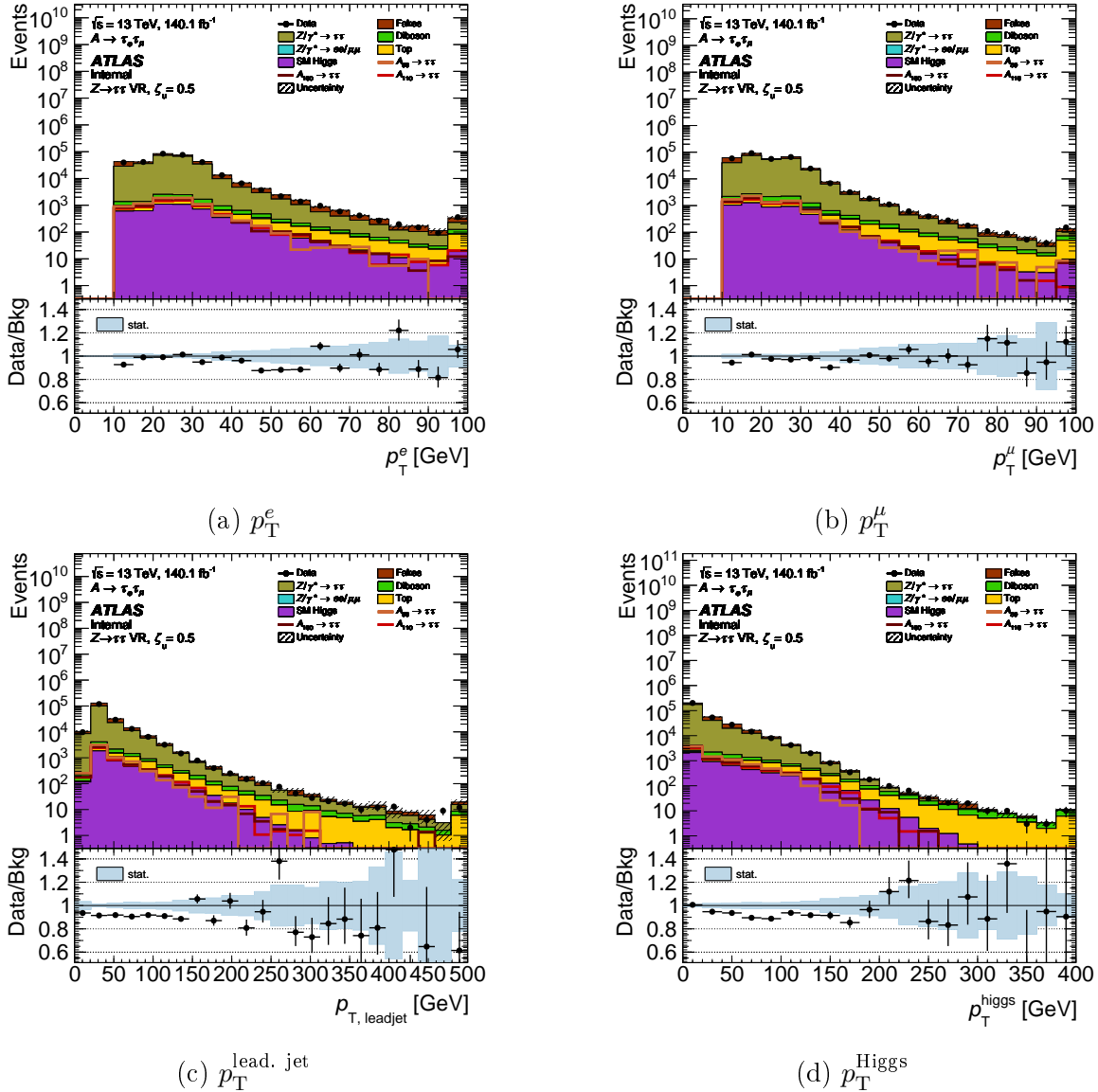


Figure D.5.: Distributions of different variables in the ZVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.

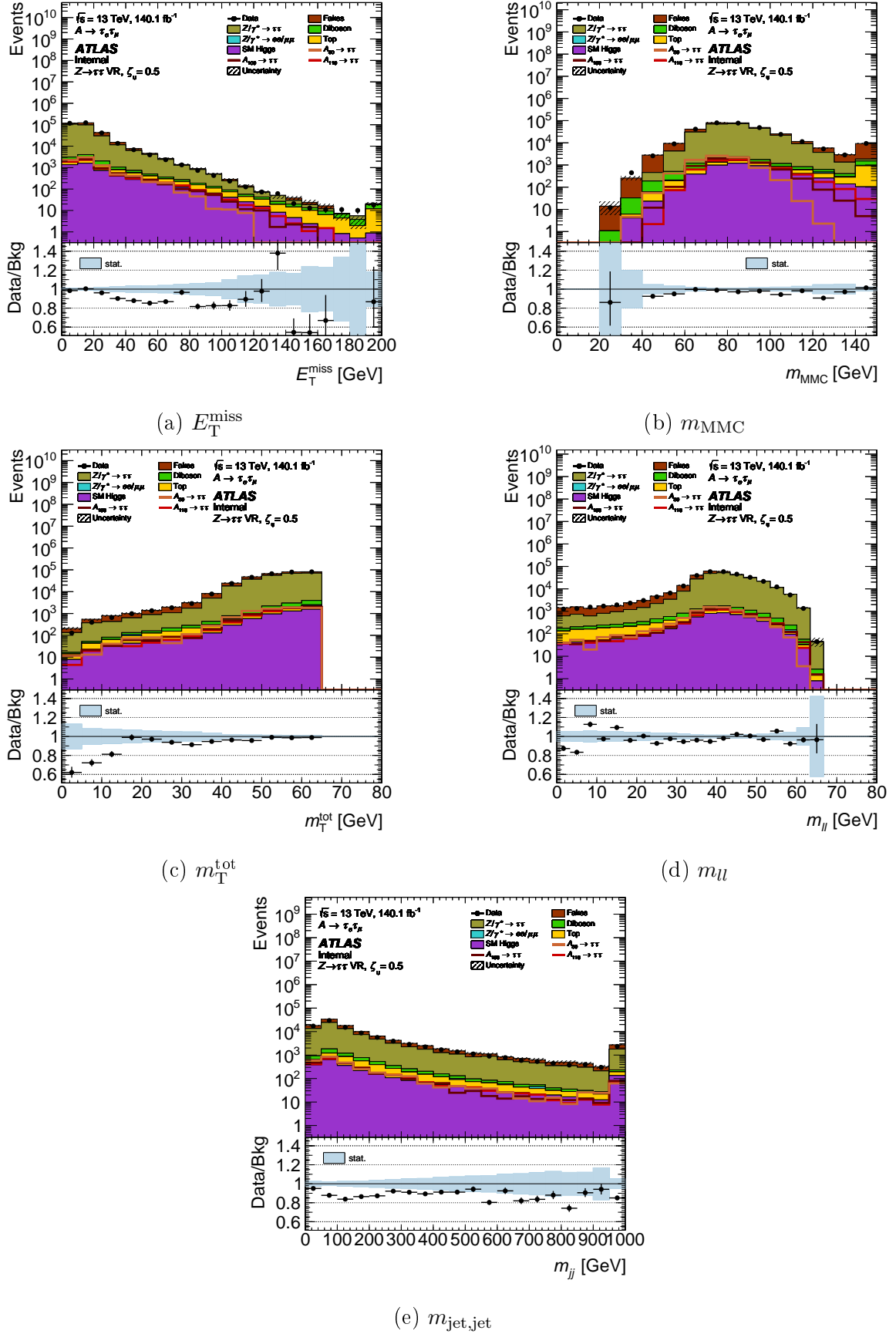


Figure D.6.: More distributions of different variables in the ZVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.

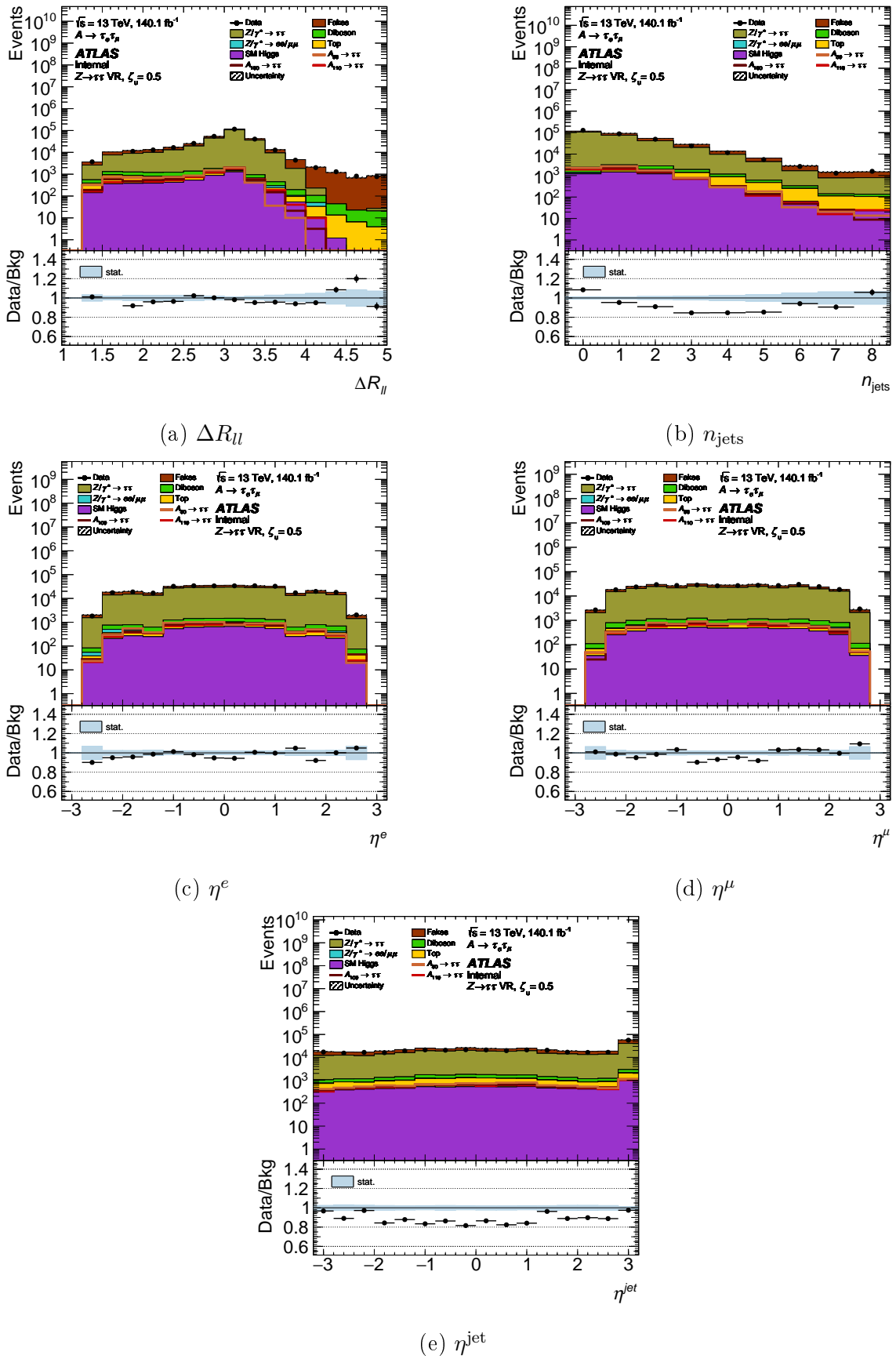


Figure D.7.: More distributions of different variables in the ZVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.

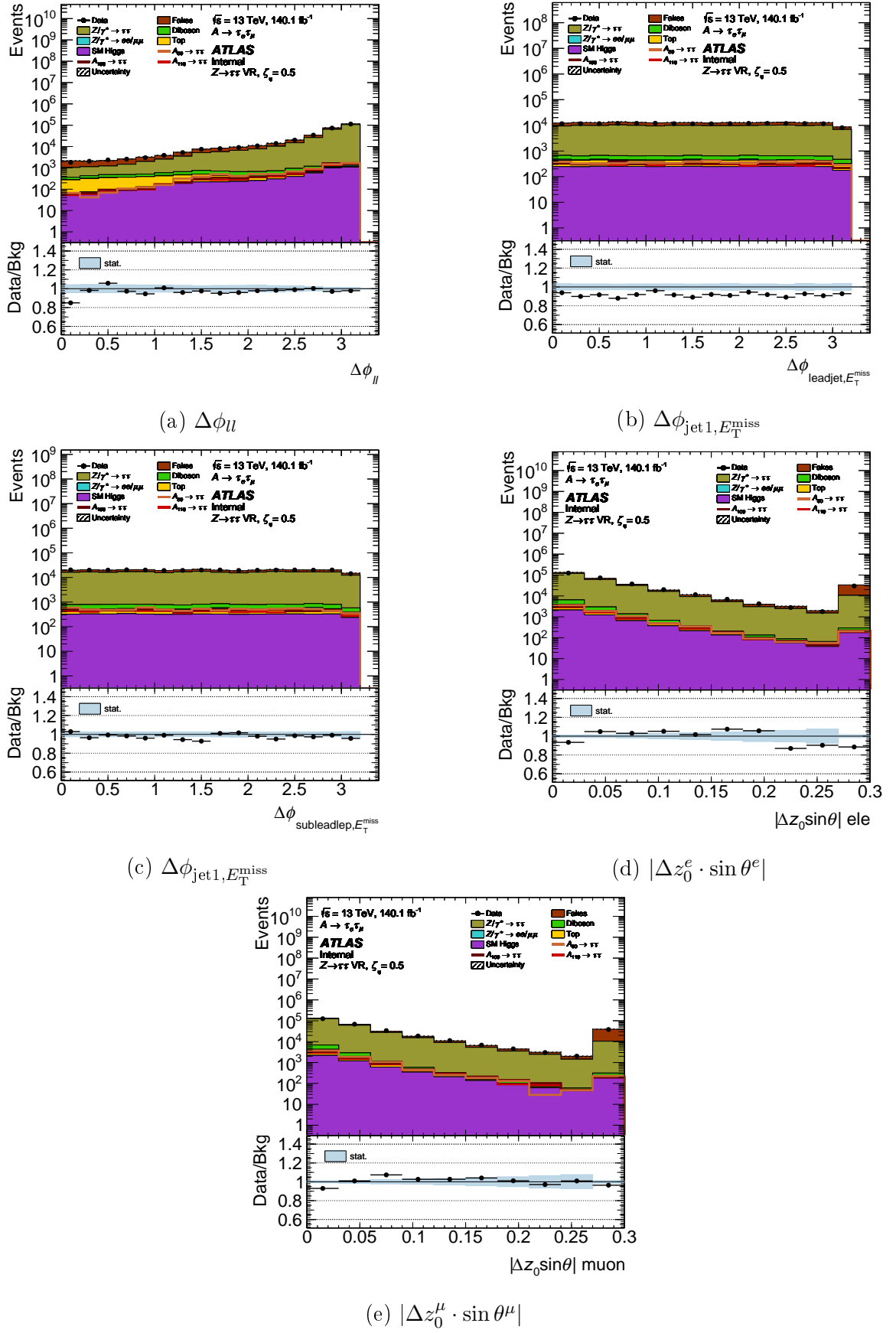


Figure D.8.: More distributions of different variables in the ZVR that are used as input to the neural network. The shown distributions do not use the  $Z \rightarrow \tau\tau$  reweighting as it was not implemented for the input of the NN yet.

### D.3. Overfitting in a Longer Training

The training strategy applied in this thesis does not contain any early stopping methods. This means overfitting must be prevented by hand, by choosing the number of epochs for a training. Figure D.9 shows the development of the training and validation loss during the training of a NN, that uses exactly the same configuration as described in Chapter 6. The only difference is a higher number of 100 epochs.

It is apparent in Figure D.9 that there is overfitting happening in later epochs of the training. The training loss keeps decreasing, while the validation loss becomes constant. Starting at around epoch 50 this behaviour is so strong that even the fluctuations from the different k-folds (transparent red and blue area in the plot) are not compatible any longer. For this reason a final number of 50 epochs was chosen for the NN discussed in Chapter 6.

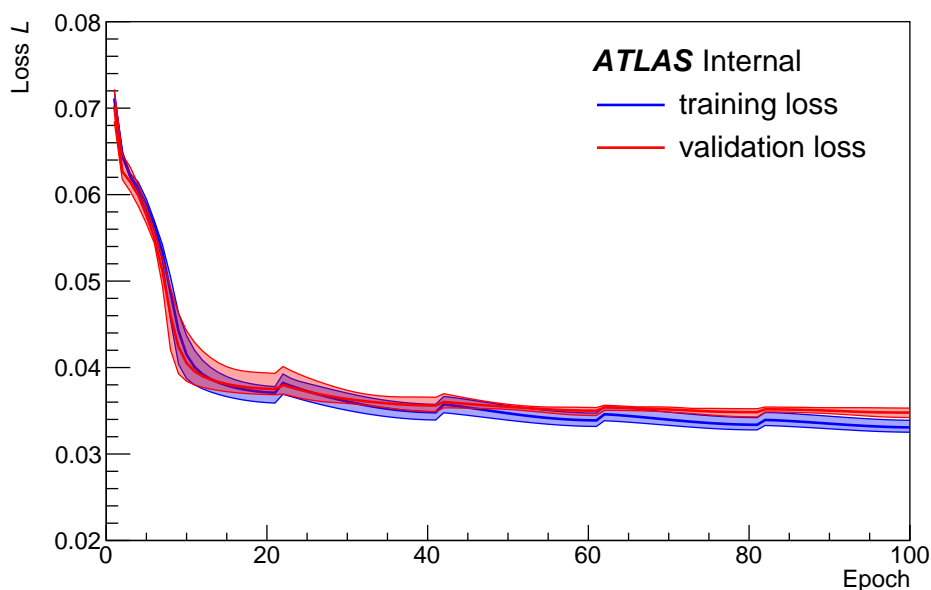


Figure D.9.: Diagram showing the development of training (blue) and validation (red) loss over the course of a longer training of 100 epochs. Weighted binary cross entropy is chosen as loss function. The transparent areas denote the full range of loss values over the different NNs from the k-fold technique and the line shows their mean value for each epoch.



---

## D.4. Distribution of All Score Variables Calculated by the NN

The distributions of score variables in different regions can be found here. This includes diagrams in the *mvaVR*, where the NN was trained, as well as distributions in the *TVR* and *ZVR* to check the modelling of the final scores.

### D.4.1. Score Distributions in the *mvaVR*

This section shows the distributions of score variables that are calculated by the NN, that is described throughout Chapter 6, in the *mvaVR*, where the NN was trained. The histograms shown here, are not binned in the way that is described in Section 6.7 for the calculation of final limits. Instead a simple uniform binning is applied. The fake background is estimated using the standard Matrix Method (see Section 4.4) and the  $Z \rightarrow \tau\tau$  reweighting is applied.

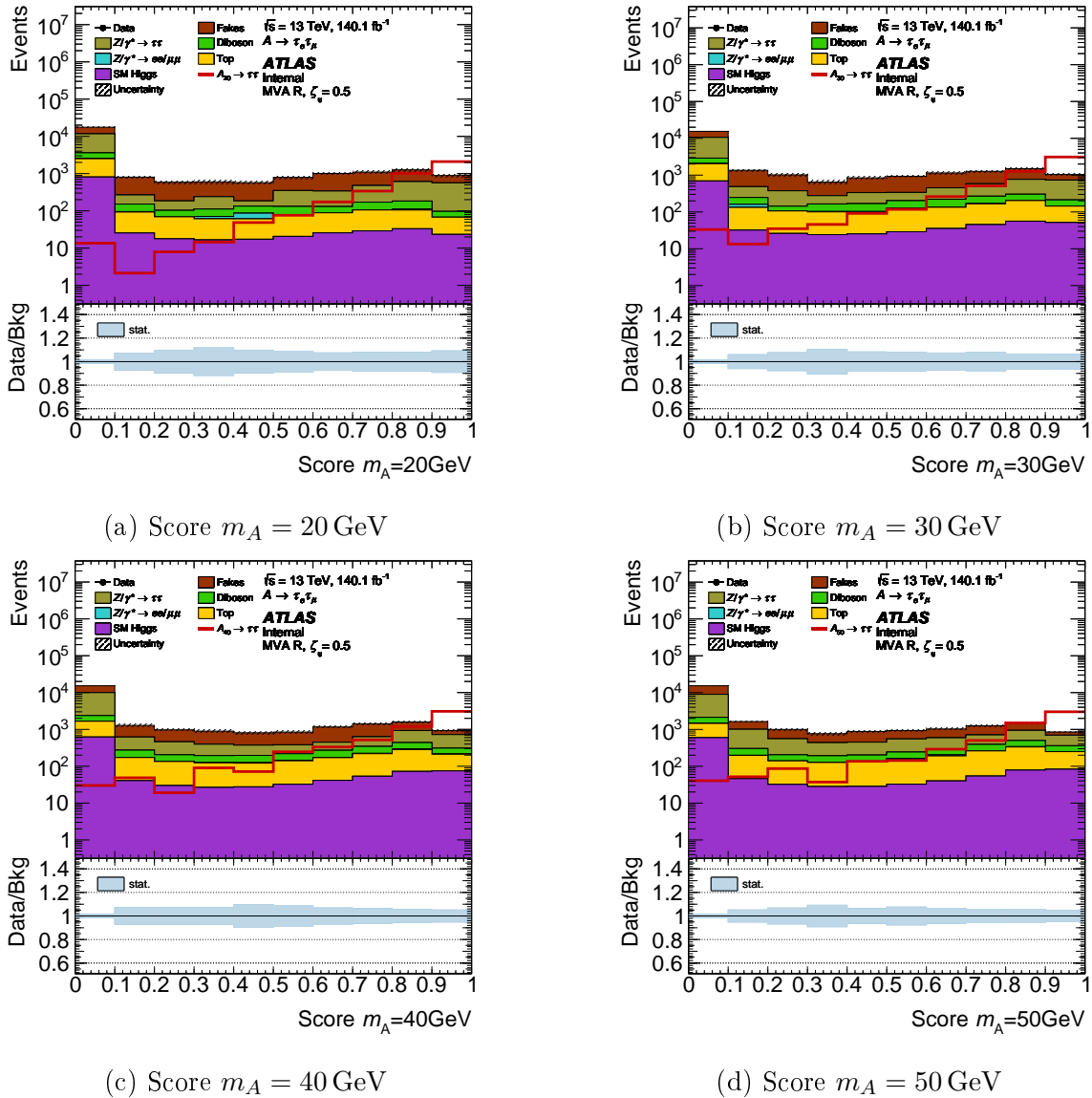


Figure D.10.: Distributions of score variables for  $m_A$  from 20 GeV to 50 GeV in the mvaVR that were calculated by the neural network, described in Chapter 6. The fake background is estimated using the standard Matrix Method (see Section 4.4) and the  $Z \rightarrow \tau\tau$  reweighting is applied.

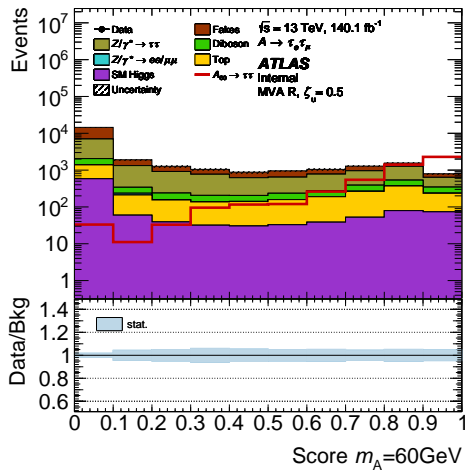
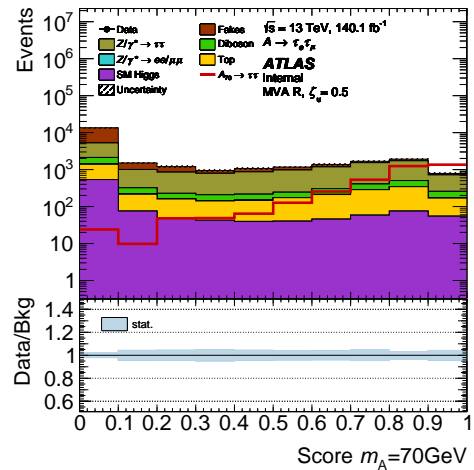
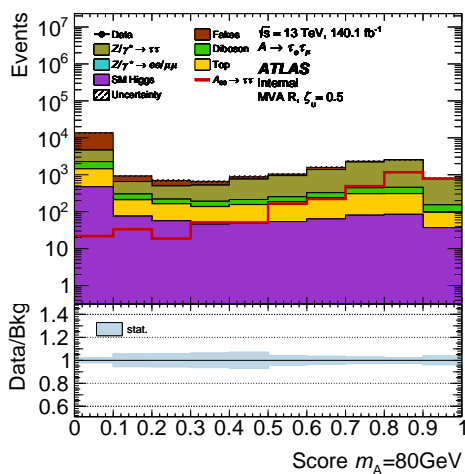
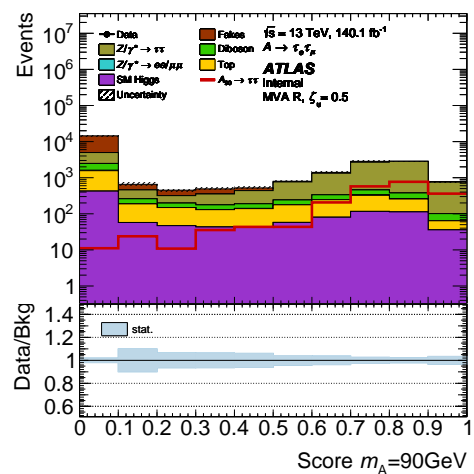
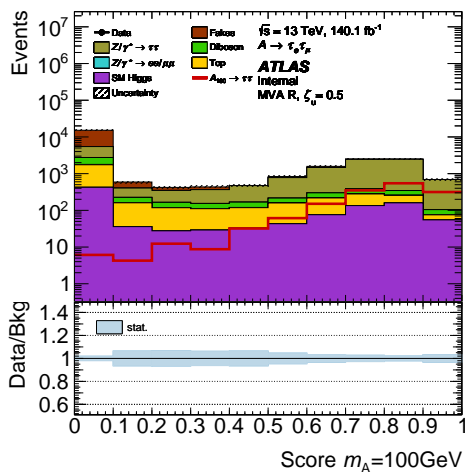
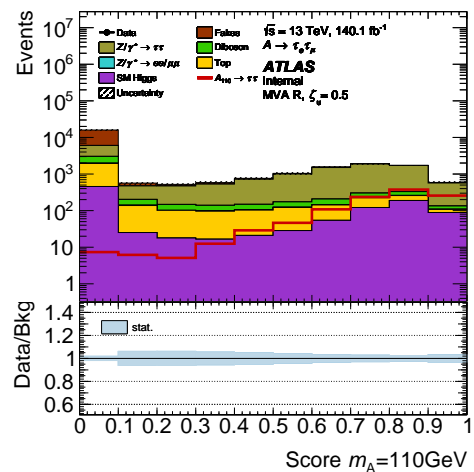
(a) Score  $m_A = 60$  GeV(b) Score  $m_A = 70$  GeV(c) Score  $m_A = 80$  GeV(d) Score  $m_A = 90$  GeV(e) Score  $m_A = 100$  GeV(f) Score  $m_A = 110$  GeV

Figure D.11.: Distributions of score variables for  $m_A$  from 60 GeV to 110 GeV in the mvaVR that were calculated by the neural network, described in Chapter 6. The fake background is estimated using the standard MM and the  $Z \rightarrow \tau\tau$  reweighting is applied.

### D.4.2. Score Distributions in the TVR

This section shows the distributions of score variables that are calculated by the NN in the TVR to investigate their modelling. The histograms shown here, are not binned in the way that is described in Section 6.7 for the calculation of final limits. Instead a simple uniform binning is applied.

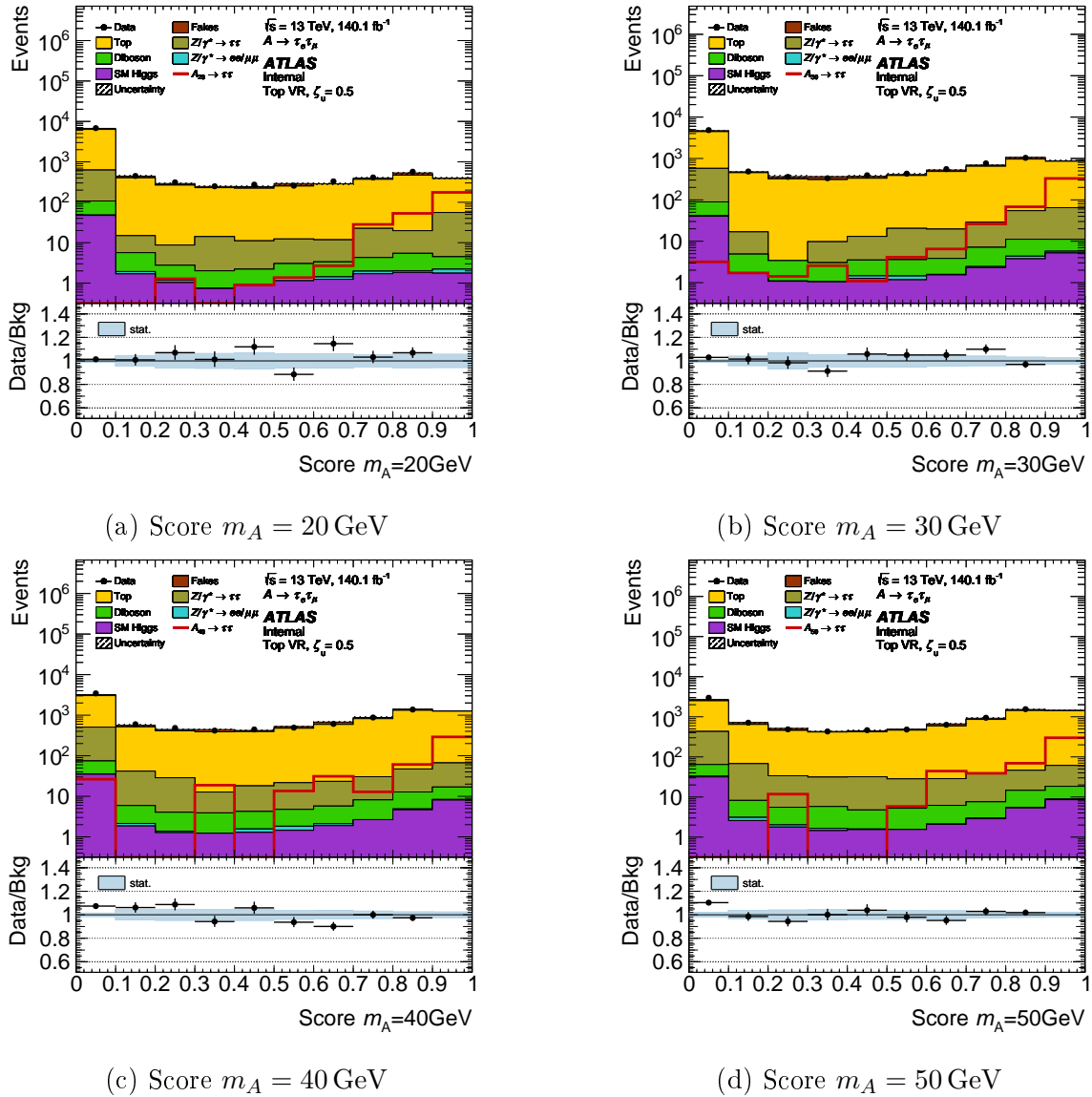


Figure D.12.: Distributions of score variables for  $m_A$  from 20 GeV to 50 GeV in the TVR that were calculated by the neural network, described in Chapter 6. The fake background is estimated using the standard Matrix Method (see Section 4.4) and the  $Z \rightarrow \tau\tau$  reweighting is applied.

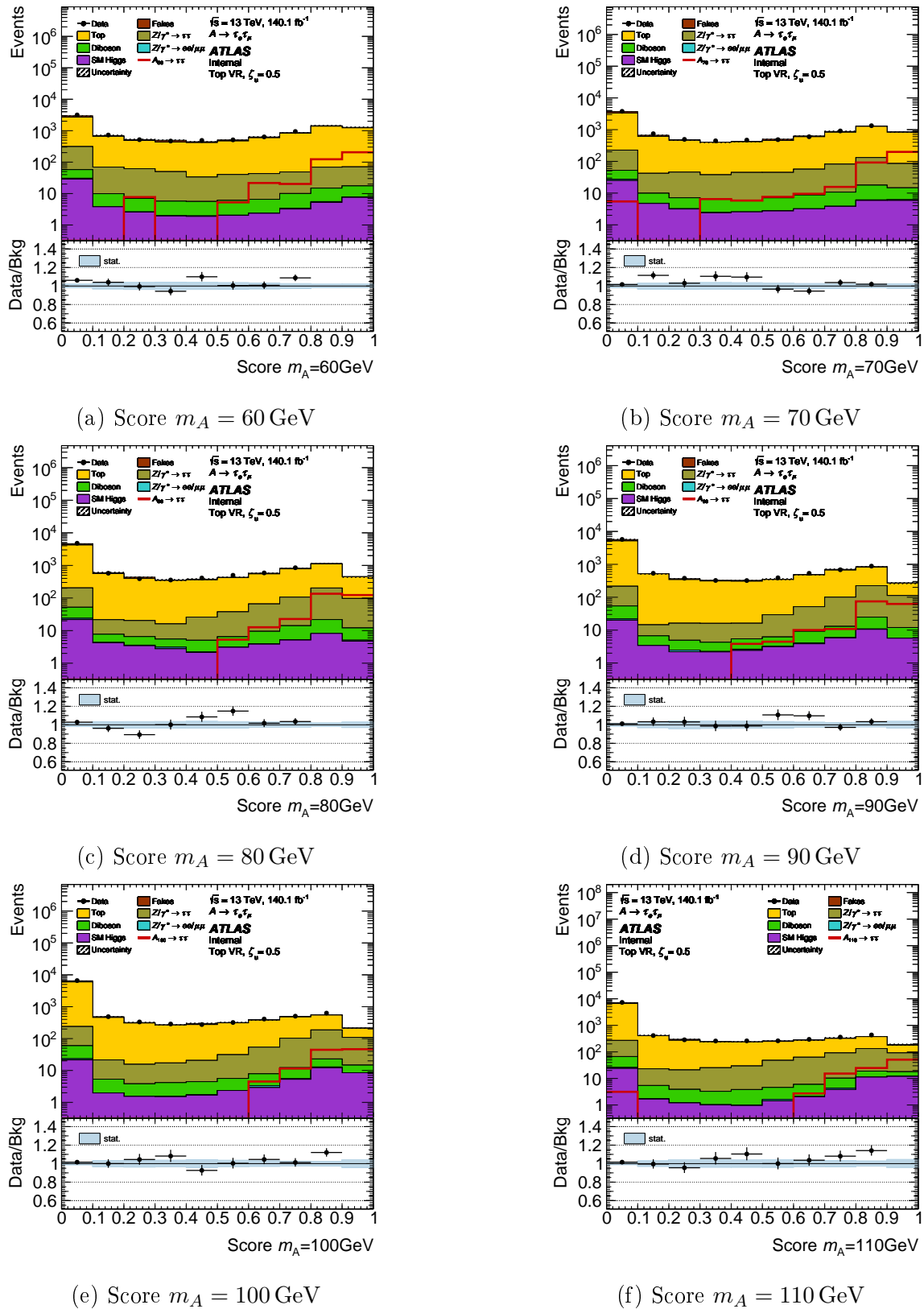


Figure D.13.: Distributions of score variables for  $m_A$  from 60 GeV to 110 GeV in the TVR that were calculated by the neural network, described in Chapter 6. The fake background is estimated using the standard MM and the  $Z \rightarrow \tau\tau$  reweighting is applied.

### D.4.3. Score Distributions in the ZVR

This section shows the distributions of score variables that are calculated by the NN in the ZVR to investigate their modelling. The histograms shown here, are not binned in the way that is described in Section 6.7 for the calculation of final limits. Instead a simple uniform binning is applied.

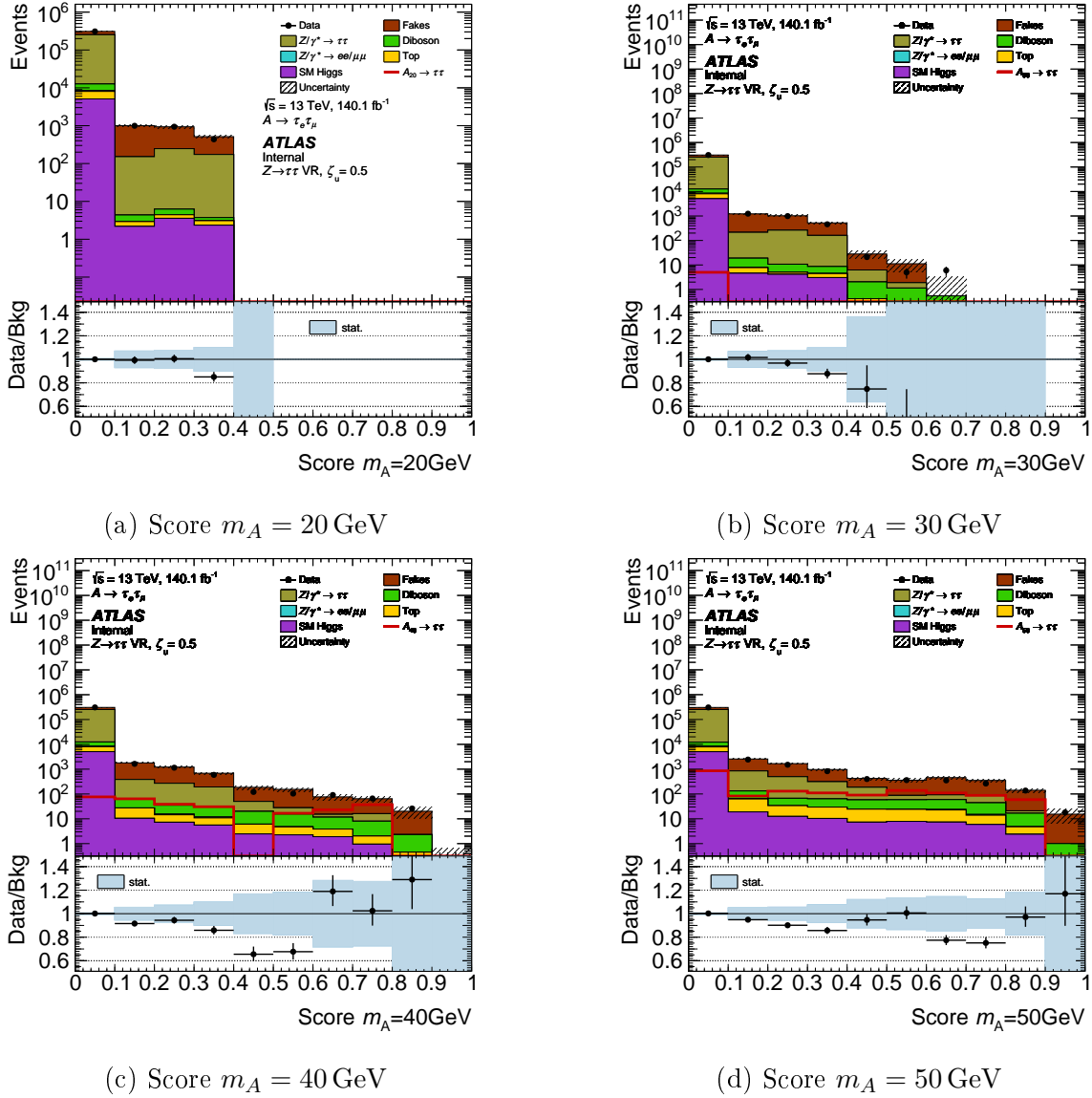


Figure D.14.: Distributions of score variables for  $m_A$  from 20 GeV to 50 GeV in the ZVR that were calculated by the neural network, described in Chapter 6. The fake background is estimated using the standard Matrix Method (see Section 4.4) and the  $Z \rightarrow \tau\tau$  reweighting is applied.

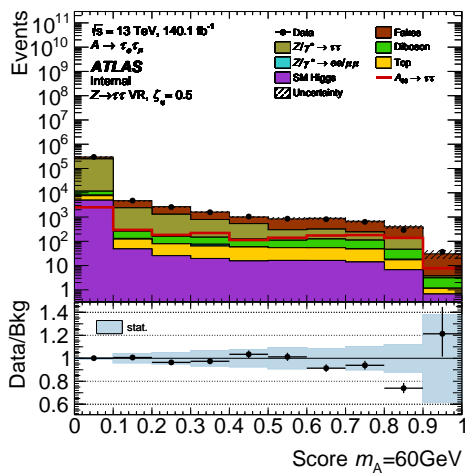
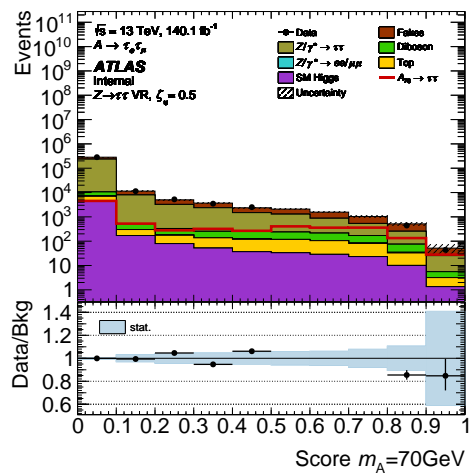
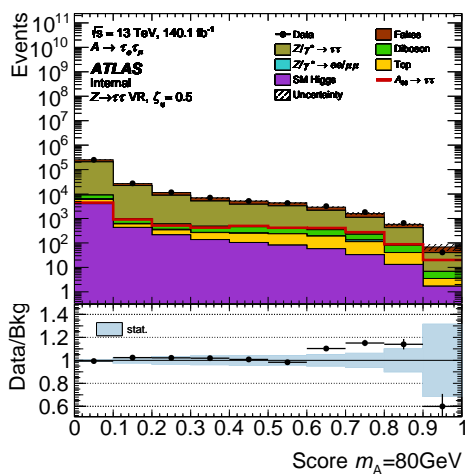
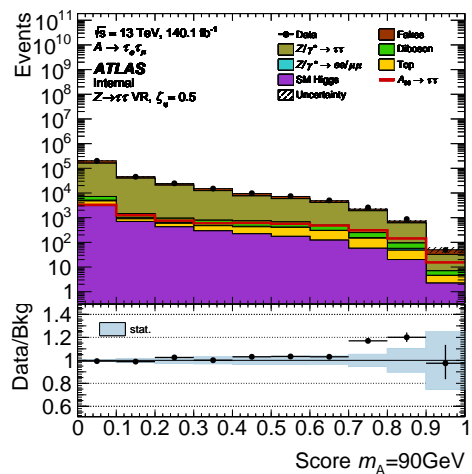
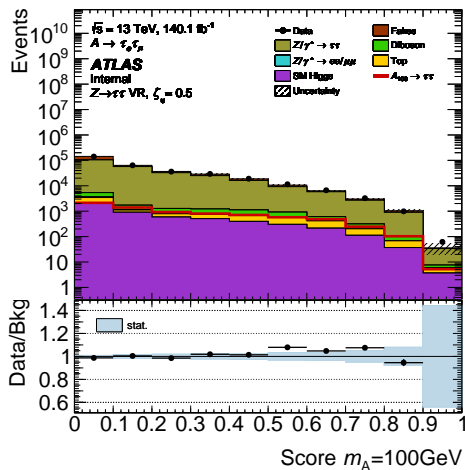
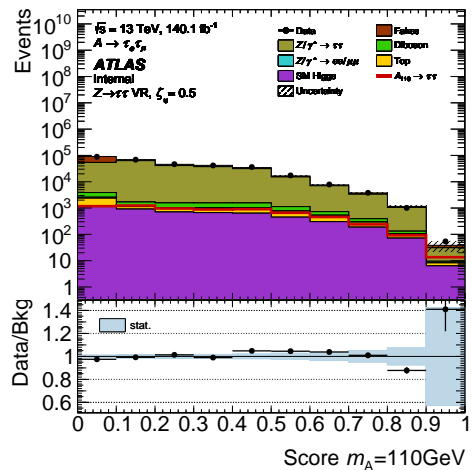
(a) Score  $m_A = 60$  GeV(b) Score  $m_A = 70$  GeV(c) Score  $m_A = 80$  GeV(d) Score  $m_A = 90$  GeV(e) Score  $m_A = 100$  GeV(f) Score  $m_A = 110$  GeV

Figure D.15.: Distributions of score variables for  $m_A$  from 60 GeV to 110 GeV in the ZVR that were calculated by the neural network, described in Chapter 6. The fake background is estimated using the standard MM and the  $Z \rightarrow \tau\tau$  reweighting is applied.

# E. Limit Studies

## E.1. Influence of the Input-Binning on Limits

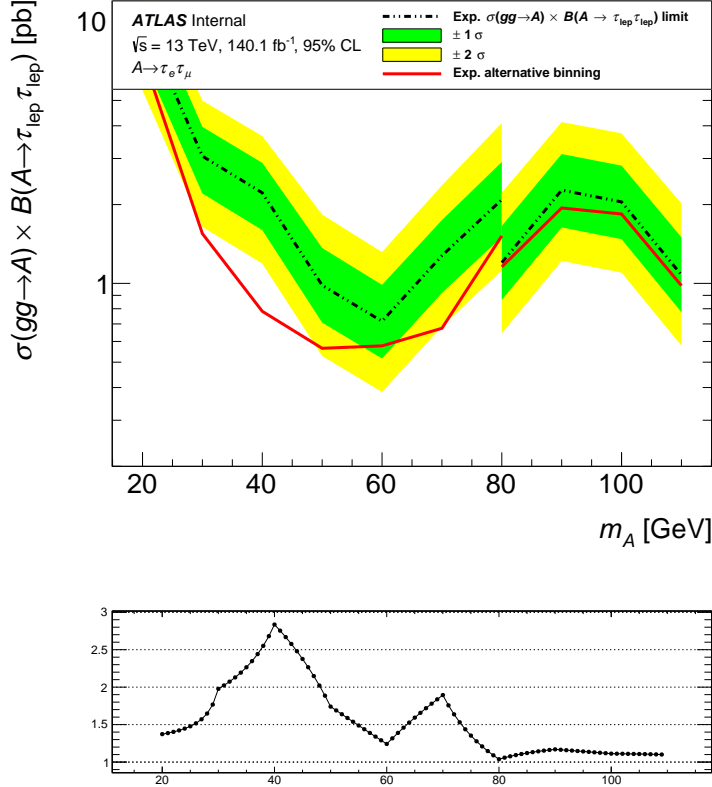


Figure E.1.: Expected limit on the production cross section for  $gg \rightarrow A$  times the branching ratio for  $A$  decaying into two leptonic  $\tau$ -leptons. The results are calculated based on the same  $m_{\text{MMC}}$  distributions but binned differently. The red curve is calculated with the standard binning as described in Section 4.6 and the dashed one is based on the alternative binning described here. The limit is calculated for  $A$  boson masses from 20 to 110 GeV and includes systematic uncertainties in both cases. The green area denotes the  $1\sigma$  uncertainty band for the results using the NN and yellow marks the  $2\sigma$  area.

It was observed that the expected limits depend strongly on the chosen binning of the input distributions. To investigate this influence, the final results of the base analysis using the standard binning method is compared with another exemplary limit calculated from the exact same input distributions just binned differently.

As a reminder the standard binning technique requires a minimum of ten events in



---

each bin, a bin size of at least 2 GeV and a maximum relative statistical uncertainty of 10%. The alternative binning strategy chosen here for this comparison also uses at least ten events in each bin, but drops the minimal bin size and allows for a maximum of 5% relative statistical uncertainty. Lowering the maximal statistical uncertainty means that each bin will generally need to contain more events than in the standard procedure, which will lead to broader bins.

Comparing the two limits in Figure E.1 it is very clear that the binning of the input distributions has a very big influence. The limit calculated with the (smaller) standard bins performs way better, than the one calculated with the broader alternative binning. In each mass point the limit is at least 10% worse and reaches the biggest deviation of roughly 280% at  $m_A = 40$  GeV. This confirms a general observation that the smaller the input bins, the better is the final limit.

This strong dependence on the input binning is not a big problem if one compares for example the influence of an additional systematic uncertainty on the final limit. In this case simply the same input distributions with the same binning can be used. This is, however, a problem if one tries to compare the limits calculated based on different variable distributions. In this case it is generally not possible to use exactly the same strategy of determining the bin edges. Especially the requirement of a minimum bin size is hard to translate to other variables, but is important at the same time to suppress the influence of random bin migration.

---

## E.2. Neural Network Limits Calculated in the Signal Regions

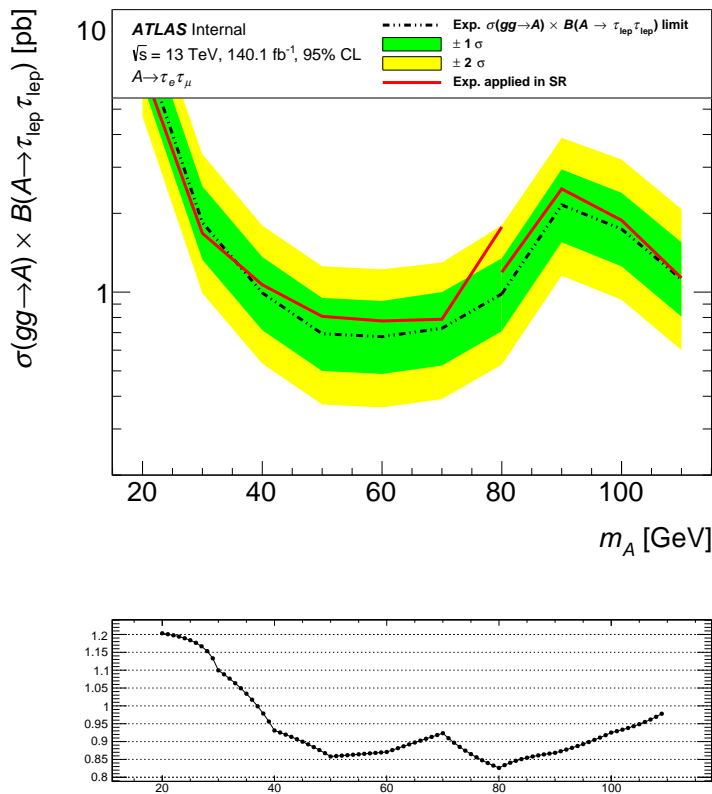


Figure E.2.: Expected limit on the production cross section for  $gg \rightarrow A$  times the branching ratio for  $A$  decaying into two leptonic  $\tau$ -leptons. The results are calculated based on the score distribution of the corresponding mass hypotheses in the mvaVR (dashed line) or the high-mass and low-mass SR (red). The limit is calculated for  $A$  boson masses from 20 to 110 GeV and includes systematic uncertainties in both cases. The green area denotes the  $1\sigma$  uncertainty band for the results using the mvaVR and yellow marks the  $2\sigma$  area. The same binning is used in both cases.

The standard version of the NN based analysis uses the score distribution in the mvaVR as input for the statistical analysis and the calculation of the expected production cross section times branching ratio limits. This study looks at the change in the final limit if the score distributions in the two signal regions are used instead. A comparison is shown in Figure E.2.

For this procedure the exact same NN was used and only the selection of events changed. The signal region distributions have lower statistics but a better overall signal-to-background ratio. It seems like this leads to a generally slightly worse limit. An exception are the two lowest mass points, where an improvement can be observed. Overall, the better statistics in the mvaVR and already quite good separation between signal and background seems to be superior. For this reason the final limit is based on inputs from the mvaVR.

### E.3. Statistics-Only Final Limit

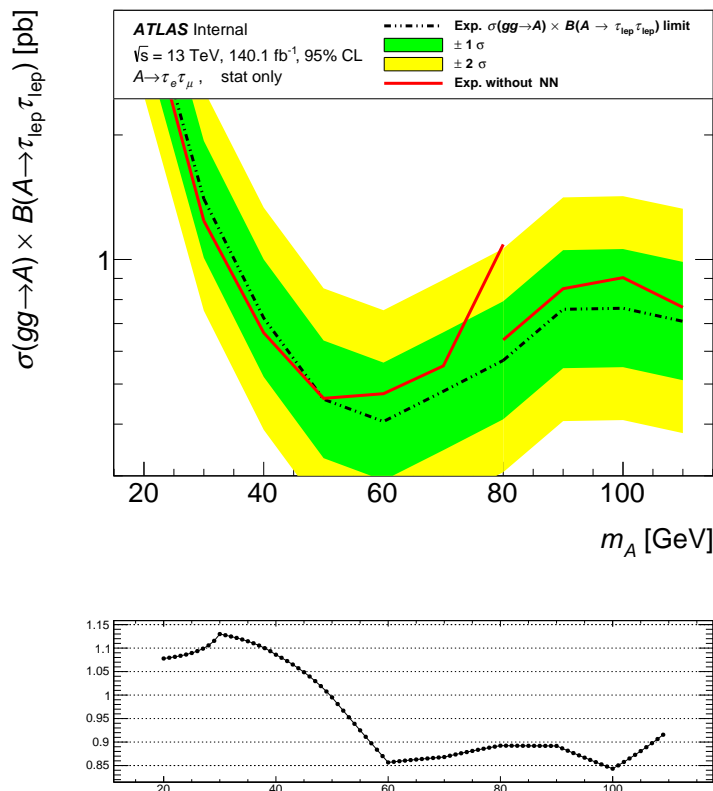


Figure E.3.: Expected limit on the production cross section for  $gg \rightarrow A$  times the branching ratio for  $A$  decaying into two leptonic  $\tau$ -leptons. The results from the base analysis (red) can be compared to the one based on the score distributions (dashed). The limit is calculated for  $A$  boson masses from 20 to 110 GeV and includes only statistical uncertainties in both cases. The green area denotes the  $1\sigma$  uncertainty band for the results using the NN and yellow marks the  $2\sigma$  area.

The final expected production cross section times branching ratio limit shown in Figure 6.5 takes into account both statistical and systematic uncertainties on the input distributions. This study compares the final limit of the base and NN analyses, when only statistical uncertainties are considered. The result can be seen in Figure E.3.

As expected both limits are better if only statistical uncertainties are taken into account. All considerations in Section 6.7 regarding the influence of the different binning strategies are still valid. In this comparison the NN limit seems to perform better than in the limit using also systematic uncertainties. Especially for higher masses above 50 GeV it looks like the NN limit is better than the nominal limit. In the lower mass ranges the nominal limit still outperforms the NN, but they are more similar than with systematics.

Overall, the NN limits seems to be dominated more by systematic uncertainties than the nominal one. This makes sense, as the SRs used in the base analysis contain less events than the mvaVR. Therefore the nominal limit is naturally more dominated by statistical uncertainties than the NN limit.



## Danksagung

Zuletzt möchte ich die Gelegenheit nutzen mich bei all den Personen zu bedanken, die mich im Verlauf des letzten Jahres maßgeblich unterstützt haben. Zu aller erst möchte ich meinem Betreuer Prof. Arno Straessner danken, für die Gelegenheit bei ihm in der Arbeitsgruppe meine Masterarbeit schreiben zu können, aber vor allem für die hilfreichen Anmerkungen und Hinweise über die Zeit hinweg.

Ein großer Dank geht auch an Manuel und Tom, die mich am Anfang der Masterarbeit in die Analyse eingeführt haben und die mir das ganze Jahr über immer mit Rat zur Seite standen. Danke auch an alle anderen in der Arbeitsgruppe, also an Asma, Christian und Valentin. Ich habe das nette Miteinander im Büro und im ganzen Institut allgemein immer sehr gewertschätzt und werde die Mittagessen mit der erweiterten ATLAS Gruppe sehr vermissen. Ich weiß es sehr zu schätzen, dass sich Manuel, Tom, Christian und Vincent die Zeit genommen haben, diese Arbeit zu lesen und mir wertvolle Hinweise zu ihrer Verbesserung zu geben.

Vielen Dank an meine Freunde, bei denen ich mich über aufgetretene Probleme auslassen konnte, aber auch mal über andere Sachen als Physik reden konnte. Zum Schluss möchte ich mich bei meiner ganzen Familie bedanken, die mir nicht nur während der Zeit der Masterarbeit immer motivierend zur Seite gestanden haben, sondern mein gesamtes Studium in dieser Form überhaupt erst möglich gemacht haben.

---



## Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit im Rahmen der Betreuung am Institut für Kern- und Teilchenphysik ohne unzulässige Hilfe Dritter verfasst habe und alle verwendeten Quellen als solche gekennzeichnet habe.

Hannah Jacobi  
Dresden, November 2023

---